Development and Extrapolation of a General Light Use Efficiency Model for the Gross Primary Production

Dissertation der Fakulät für Geowissenschaften der Ludwig–Maximilians–Universität München
zur Erlangung des akademischen Grades
"doctor rerum naturalium" (Dr. rer. nat.)

vorgelegt von
Judith E. Horn
aus Offenburg

München, den 14.2.2011

Erstgutachter: Prof. Dr. Karsten Schulz

Zweitgutachter: Prof. Dr. Andreas Huth

Tag der mündlichen Prüfung: 13.7.2011

"That's the whole problem with science. You've got a bunch of empiricists trying to describe things of unimaginable wonder."

Bill Watterson in "Calvin And Hobbes"

Acknowledgments

I am very grateful to Vanessa Stauch. In her cheerful and warm manner she introduced me to Leipzig and the working group. She made it easy for me to become familiar with the data and required methods and the philosophy behind the project.

I would like to thank Prof. Andreas Huth for spontaneously accepting the second review of this thesis.

I wish to express my warmest gratitude to my advisor Prof. Karsten Schulz for supporting and encouraging me over the past years. He gave new impulses when I was stuck and shared his thoughts, experiences and inspiring ideas with me. And he taught me: Science can be fun!

My thanks also go to my brothers Pirmin, Dominik and Simon for being the reliable companions they are; I am very glad I have these three brothers and I know I can count on them when it matters. I always enjoy the refreshing and inspiring talks with them on the meaningful senselessness and the important things in life. Thanks to Virginia for her valuable assistance in the field of graphic design.

I am indebted to my parents and thank them sincerely for their unconditional love, support and faith in me. It is them I want to dedicate this work to.

And last but not least, I thank you with all my heart, Peter, for all the love you give. Thank you for your unlimited support, for your strong confidence in me, for your infinite optimisms, for sharing your worries, happiness and dreams with me. Thank you for being there.

Summary

The global carbon cycle is substantially influenced by the terrestrial biosphere. Against the background of accelerating global change and in particular of the rising CO_2-concentration in the atmosphere and associated climate change, the scientific community is highly interested in analyzing and understanding the dynamics of the global carbon cycle. The highly complex processes determining the gross primary production – the carbon uptake and assimilation by photosynthetically active plants at ecosystem level – are of particular relevance in this context and form the focus of this thesis. Only a comprehensive observation network and thorough analyses of those interdependencies enable scientists to achieve the objective of developing future scenarios and, finally, establishing adaptation and mitigation strategies to combat the consequences of climate change.

To meet this challenge, measurement networks such as FLUXNET have been initiated. Around the world, FLUXNET participants collect and provide data on the exchange of energy and matter between vegetation stands and the atmosphere with the so-called eddy-covariance (EC) technique. These EC measurements support the identification and understanding of ecosystem processes and are indispensable for purposes of developing, calibrating and validating simulation models. However, if these soil-vegetation-atmosphere-transfer (SVAT) models are to be applied on larger scales, spatially continuous input data are needed. Remote sensing as only source of large-scale information on the land surface can serve this purpose. MODIS, a satellite sensor specifically designed for this task, has proven to be a key sensor for observing ecosystem states and tracking changes. Modeling, EC measurements and remote sensing complement each other synergistically.

Existing SVAT models have shown to suffer from limitations: A mismatch between the small scale and high complexity of the process descriptions implemented in the model and the larger model application scale with scarce data sources has been detected. This mismatch can render parameter calibrations difficult and predictions uncertain. Other models specifically established for large-scale applications are, on the contrary, typically formulated with parsimonious model structures. They are parametrized with fixed or coarsely grouped values as well as predefined relationships between ecosystem processes and environmental drivers that often do not hold when compared with EC measurement data.

The overarching objective of this thesis is to exploit the powerful combination of modeling, EC measurements and remote sensing in order to develop a simple but robust model for the gross primary production of vegetation stands. Furthermore, an extrapolation

scheme is sought with which the model parameters calibrated at FLUXNET sites can be regionalized. To achieve this aim, the well-established concept of light use efficiency is chosen as modeling framework. The basic equation is successively refined until the model complexity commensurates with the observations made at daily time steps. To serve this purpose, non-linear data analysis tools are applied to exploit the information content of EC measurement data. The functional forms describing the relationships between the photosynthetic gross CO_2 flux and environmental drivers are directly identified from the measurement data without making strong *a priori* assumptions. In this way, physical process understanding is coupled with the information on ecosystem processes as reflected in the measurement data. To apply the MODIS data products in a best possible way and avoid usage pitfalls, a thorough analysis of a core MODIS product used throughout this thesis is carried out.

The derived model demands not more than three environmental variables, namely the absorbed photosynthetically active radiation, temperature and a water availability measure. Despite its simplicity, the model captures a great proportion of the day-to-day variations of the gross primary production as measured at the study sites. The unique model structure accounts for variable influences of temperature and droughts on the photosynthetic gross CO_2 flux at different sites. This inherent model characteristic renders the model widely applicable and enables its usage for differing vegetation types and environmental conditions ranging from boreal needleleaf forests to semi-arid grasslands. The optimized set of model parameters is well defined and the model uncertainty due to the parameter calibration is generally found to be low.

To allow for model applications at larger scales, the calibrated model parameters are related to climatic and biophysical site characteristics by means of support vector regression, a powerful machine learning technique. A novel framework is set up which automatically and objectively selects the explanatory features for each model parameter out of a large set of site characteristics. In a cross-validation, the time series of the photosynthetic gross CO_2 flux modeled with the extrapolated parameters correlates very well with the measured dynamics. Likewise, the cumulative sums of the gross primary production as stringent performance criterion compare satisfyingly with the measured sums. Overall, the performed cross-validation proved the proposed scheme to be highly suitable for the extrapolation of model parameters and thus allows the model application at larger scales.

The modeling and extrapolation framework presented in this thesis contributes substantially to the efforts of the scientific community to predict the gross primary production under future environmental conditions. The data-driven approach followed in this thesis likewise appears to be appropriate for modeling other ecosystem fluxes such as the evapotranspiration. This crucial component of the water cycle, which is intrinsically linked to the carbon cycle, is also of major importance in a changing environment.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

All units of the variable and parameter are given for the case of application in daily time step models.

α	Lag parameter of the light use efficiency model [-]
APAR	Absorbed photosynthetically radiation [MJ m^{-2} d^{-1}]
API	Antecedent precipitation index [mm]
b	Function offset
C	Constant support vector regression parameter: weighting factor for the training error
c	Regression parameter
w'	Fluctuations of the mixing ratio [kg/kg]
$C_1 - C_4$	EVI-coefficients
CO_2	Carbon dioxide
c_p	Specific heat of air [J kg^{-1} K^{-1}]
D	Displacement height [m]
d	Time step in the time window Z before the current time step
D^2	Second derivate of a function
DDS	Dynamically dimensioned search
Δ	Rate of change of saturated specific humidity with temperature [kg kg^{-1} K^{-1}]
δq	Specific humidity deficit [kg kg^{-1}]
DLR	Dynamic linear regression
DOY	Day of year
EC	Eddy covariance
EF	Evaporative fraction [-]
EF_I	Constant light use efficiency model parameter determining the inflection point of the subfunction f_W in case of using EF [-]
ϵ	Light use efficiency [gC MJ^{-1}]
ϵ_J	Light use efficiency as used in Jarvis *et al.* (2004) [gC MJ^{-1}]
ϵ_{max}	Maximum light use efficiency [gC MJ^{-1}]
$\epsilon_{max,J}$	Maximum light use efficiency as used in Jarvis *et al.* (2004) [gC MJ^{-1}]
ϵ_{SVR}	Constant support vector regression parameter: width of the band insensitive to training errors
η	Noise sequence
EVI	Enhanced Vegetation Index

F_G	Flux of carbon uptake by a canopy, gross primary production [gC m^{-2} d^{-1}
F_N	Net flux of carbon uptake by a canopy [gC m^{-2} d^{-1}]
FPAR	Fraction of absorbed photosynthetically radiation [-]
F_R	Respiration flux by a canopy [gC m^{-2} d^{-1}]
f_T	Subfunction of the light use efficiency model for the temperature
f_W	Subfunction of the light use efficiency model for the water availability
γ	Parameter of the radial basis function
γ_P	Psychrometric constant [kg kg^{-1} K^{-1}]
H	Sensible heat flux [MJ m^{-2} d^{-1}]
H_U	Vegetation height of the understorey [m]
H_V	Vegetation height [m]
i	Context-specific index
κ	Recession constant for calculation of the API
k_T	Constant light use efficiency model parameter determining the rate of change of the subfunction f_T [°C $^{-1}$]
$k_{T,J}$	Constant light use efficiency model parameter determining the rate of increase of the temperature subfunction in Jarvis *et al.* (2004) [°C $^{-1}$]
k_W	Constant light use efficiency model parameter determining the rate of increase of the subfunction f_W
LAI	Leaf area index [m^2 m^{-2}]
λE	Latent heat flux [MJ m^{-2} d^{-1}]
N	Context-specific total number
NDVI	Normalized Difference Vegetation Index
nRMSE	Normalized root mean squared error
NS	Nash-Sutcliffe efficiency criterion
NVR	Noise-variance ratio
ω	Weighting function
P	Precipitation [mm]
p	Constant light use efficiency model parameter assigning the contribution of the two subfunctions [-]
PAR	Photosynthetically radiation [MJ m^{-2} d^{-1}]
Φ	Objective function within the support vector regression framework
r_a	Aerodynamic resistance [s m^{-1}]
RBF	radial basis function
RE$_M$	Relative predictive error regarding the means
RE$_S$	Relative predictive error regarding the cumulative sums
ρ_a	Air density [kg m^{-3}]
ρ_{NIR}	NIR infrared reflectance band
ρ_{blue}	Radiation reflectance band in the blue
ρ_{red}	Radiation reflectance band in the red

R_L	Roughness length [m]
RMSE	Root mean squared error
R_N	Net radiation [MJ m^{-2} d^{-1}]
r_s	Surface resistance [s m^{-1}]
r^2	Coefficient of determination
r_w^2	Weighted coefficient of determination
S_0	Incident solar radiation [MJ m^{-2} d^{-1}]
SDP	State dependent parameter estimation
SSE	Sum of squared errors
SSE$_W$	Weighted sum of squared errors
SVAT	Soil-vegetation-atmosphere-transfer
SVM	Support vector machines
SVR	Support vector regression
SWC	Soil water volumetric content [%]
t	time
T_F	Time-lagged soil temperature in the model of Jarvis *et al.* (2004) [°C]
T_I	Constant light use efficiency model parameter determining the inflection point of the temperature subfunction in Jarvis *et al.* (2004) [°C]
T_{opt}	Constant light use efficiency model parameter determining the maximum of the subfunction f_T [°C]
T_S	Soil temperature [°C]
u	System state in a state-dependent regression model
VPD	Vapor pressure deficit [kPa]
W	Water availability measure such as EF, SWC or API
w	Norm vector
w'	Fluctuations of the vertical air movement [m/s]
W_I	Constant light use efficiency model parameter determining the inflection point of the subfunction f_W
x	Regressor in a time- or state-dependent regression model
ξ^-, ξ^+	Slack variables within the support vector regression framework
y	Context-specific dependent variable, function value
Z	Time window
ζ	Regression model error series

1

Introduction

1.1 Thematic Background and Motivation

1.1.1 The Terrestrial Biosphere and the Carbon Cycle

The terrestrial biosphere and the atmosphere are tightly coupled through the exchange of energy and matter (Monteith & Unsworth, 2008). A central component of this interdependency is the assimilation of CO_2 by photosynthesis of plants. Through this physiological process – at the ecosystem level also referred to as gross primary production – terrestrial ecosystems assimilate approximately 120 Pg carbon per year (Beer *et al.*, 2010) – compared to 5 Pg carbon released from fossil fuel burning (Janzen, 2004). Thus this assimilation process by plants modulates substantially the global carbon cycle (Schimel *et al.*, 2001), especially in the Northern hemisphere (Fan *et al.*, 1998).

Motivated by the substantial anthropogenic increase in atmospheric CO_2 and associated climate change scenarios (Cox *et al.*, 2000), the scientific community has payed increased attention to the respective role of the terrestrial biosphere within this complex system (Running, 2008; Thornton *et al.*, 2009; Mahecha *et al.*, 2010). Addressing this scientific issue is complicated by the fact that the rising atmospheric CO_2 concentration and the associated changing climate factors have implications on the functioning of ecosystems. Vice versa, small alterations in the terrestrial carbon balance are likely to pose a significant impact on CO_2 concentrations in the atmosphere. Hence, feedback mechanisms are provoked in a complex and non-linear manner rendering predictions difficult and uncertain (Cramer *et al.*, 2001; Heimann & Reichstein, 2008; Piao *et al.*, 2009; Thornton *et al.*, 2009; Arneth *et al.*, 2010; Zaehle *et al.*, 2010). The question if terrestrial biosphere-atmosphere interactions mitigate or amplify anthropogenic climate change under future climate scenarios is still an active field of research (Frank *et al.*, 2010; Friedlingstein & Prentice, 2010; Zhao & Running, 2010; Betts *et al.*, 2011) subject to immense public interest and being of politically explosive nature (Klein *et al.*, 2005; Parry, 2007; Lahsen, 2009).

Quantifying the global carbon balance under current conditions is a prerequisite for developing future scenarios; the identification of the main drivers of photosynthetic CO_2 uptake as largest terrestrial carbon flux component is of central importance in this respect (Running, 2008; Beer *et al.*, 2010; Yi *et al.*, 2010). This dissertation approaches this issue and focuses on developing a scheme to estimate the gross CO_2 uptake by plants for a wide range of biomes. To meet this challenge, the combination of (i) field measurements and experiments, (ii) remote sensing from airborne or satellite platforms, and (iii) simulation models offer the best promise (Roughgarden *et al.*, 1991; Knorr & Heimann, 2001b; Lefsky *et al.*, 2002; Turner *et al.*, 2004b; Gamon *et al.*, 2006). None of these three approaches alone is feasible to serve the purpose of understanding, monitoring and prognosticating the carbon exchange by the terrestrial biosphere, but together they frame a powerful, self-amplifying triangle and provide a promising pathway for the collaboration of ecologists, biologists and chemists, information scientists and physicists (Baldocchi *et al.*, 1996; Running *et al.*, 1999; Ustin & Gamon, 2010).

1.1.2   Micrometeorological Measurements

In the last two decades, large efforts in gathering data around the globe have been undertaken to monitor exchanges of CO_2, water vapor and energy between vegetation and the atmosphere in field-measurement campaigns. FLUXNET (Baldocchi *et al.*, 1988, 2001), a highly recognized, world-wide measurement network using eddy-covariance (EC) techniques, arose from these endeavors (Friend *et al.*, 2007; Baldocchi, 2008) and led to concerted research projects aiming at the quantification and characterization of the terrestrial carbon exchange (Falge *et al.*, 2002; Luyssaert *et al.*, 2007; Yi *et al.*, 2010). FLUXNET is a global research network of currently over 500 EC towers, which measure the exchange of energy, water vapor and CO_2 along with further important meteorological variables. These long-term measurement efforts provide valuable insights into the functioning of ecosystems (Friend *et al.*, 2007). FLUXNET merges these valuable data collected around the world in a comprehensive database and provides the scientific community a unique opportunity to establish understanding of ecosystem exchange processes and to set up, calibrate and validate models to predict ecosystem behavior in a changing environment (Wang *et al.*, 2007b; Stoeckli *et al.*, 2008; Williams *et al.*, 2009).

1.1.3   Remote Sensing

Continuous EC measurements by micrometeorological flux towers have become indispensable in monitoring the exchange of energy, carbon and water fluxes between the terrestrial biosphere and the atmosphere. However, flux towers can only catch a glimpse on ecosystem processes given their quasi-point measurement nature, which stands in contrast to the usually more extensive, spatially continuous model application and map-

ping scale. Satellite remote sensing, instead, offers spatially continuous information. It can consequently help to dissolve the dilemma of having scattered point-like measurements on the one hand and the need to determine and predict exchange processes at a local, regional or even global scale on the other hand by bridging the gap between these scales (Running *et al.*, 1999; Turner *et al.*, 2004a; Coops *et al.*, 2007; Schaepman *et al.*, 2009). Furthermore, remote sensing can provide ecosystem characteristics that are time-consuming or difficult to gather in the field (Hanan & Begue, 1995; Juarez *et al.*, 2009; Zheng & Moskal, 2009). Vice versa, the validation of remote sensing products relies on continuous measurements such as done at the FLUXNET sites (Cohen *et al.*, 2003; Baret *et al.*, 2006).

As recently as in the year 1991, it was noted: "There has never been a remote sensing instrument designed for ecological research" (Roughgarden *et al.*, 1991, p. 1921). Fortunately, this situation has improved: Sensors such as the Advanced Very High Resolution Radiometer (AVHRR) onboard the Polar Orbiting Environmental Satellites (POES), SPOT sensors (Satellite Pour l'Observation de la Terre) of the French Space Agency, the CHRIS (Compact High Resolution Imaging Spectrometer) sensor on the European Space Agency's PROBA (PRoject for On-Board Autonomy) satellite or the NASA's (National Aeronautics and Space Administration) MODIS (Moderate Resolution Imaging Spectroradiometer) were specifically designed to serve the purposes and demands of environmental research.

Amongst them, the MODIS sensor (King *et al.*, 1992; Running *et al.*, 1994; Justice *et al.*, 1998; Xiong *et al.*, 2009) has proven itself a key sensor in remote sensing of ecosystem dynamics and land surface processes. The sensor is mounted onboard the satellites Terra and Aqua, which were launched within the framework of NASA's EOS program (Earth Observing System). MODIS data have been widely applied in studies to detect the vegetation physiological status (e.g. Huete *et al.*, 2002; Doraiswamy *et al.*, 2005; Houborg *et al.*, 2007; Roehrig & Laudien, 2009), to track the course of phenology (Zhang *et al.*, 2003; Soudani *et al.*, 2008; Julien & Sobrino, 2009), detect biodiversity changes (Duro *et al.*, 2007; Huang *et al.*, 2009; He *et al.*, 2009), dust storms (Hao & Qu, 2007) and burned areas (Giglio *et al.*, 2006; Chuvieco *et al.*, 2008; Al-Saadi *et al.*, 2008; Soja *et al.*, 2009), to document vegetation responses to global climate patterns (Potter *et al.*, 2008), to assist in protected area management (Crabtree *et al.*, 2009; Nemani *et al.*, 2009) and drought monitoring (Wang *et al.*, 2009), or to study explicatively the global carbon cycle (Jung *et al.*, 2008; Turner *et al.*, 2006; Zhao & Qualls, 2006).

The ORNL DAAC (Oak Ridge National Laboratory Distributed Active Archive Center) supports the synergetic interaction between micrometeorological tower measurements, modeling and remote sensing: The institute makes resampled key remote sensing products from MODIS as so called "MODIS Land Product Subsets" freely available. These files are provided in an easy to handle ASCII-format and comprise selected MODIS land products, which are resampled specifically for FLUXNET and other long-term measure-

ment sites. The subsets contain a grid with 7x7 1-km^2 pixels, whereas the measurement tower is located in the center pixel. One of the subset products retrieved from MODIS Terra and Aqua data is MOD15A2 or MYD15A2, respectively, which provides 8-day estimates of the leaf area index (LAI) accompanied by the closely related FPAR (Fraction of Absorbed Photosynthetically Active Radiation Myneni *et al.* (2002)). At least one of these two variables is a fundamental characteristic in almost every model concerning the exchange of energy, water and carbon of ecosystems. In this context the MODIS LAI/FPAR product has been used frequently for gross primary production models (e.g. Yang *et al.*, 2007; Xiao *et al.*, 2008a; Zhao & Running, 2010) and is employed in this thesis, too.

However, the time series provided by the MODIS LAI/FPAR product has shown to be quite unstable with respect to the smooth LAI/FPAR time series that can be expected from a vegetation stand that had been not exposed to extreme events (Weiss *et al.*, 2004; Lu *et al.*, 2007; Xiao *et al.*, 2009). Furthermore, the user is confronted with a variety of options regarding quality filtering, spatial aggregation and sensor choice, for which no standardized application is common amongst the users. The necessity of spatial and temporal smoothing and interpolating of the noisy and incomplete time series has often been formulated and respective methods have been proposed (Gu *et al.*, 2006; Fang *et al.*, 2008; Gao *et al.*, 2008; Borak & Jasinski, 2009). However, the effects of all the mentioned options (spatial aggregations, quality filtering, sensor choice) on the resulting time series and the consequences for model applications have not been studied explicitly yet. This dissertation takes a critical look at these issues.

1.1.4   Modeling of Gross Primary Production

Prior to the availability of comprehensive data sets from observation networks, research had rather focused on studying biochemical and biophysical processes of photosynthesis under laboratory conditions. These processes are well understood on cell, leaf and plant scales (e.g. Farquhar *et al.*, 1980; Stitt, 2006). Sophisticated process models simulating carbon fluxes have been formulated at these scales and incorporated into soil-vegetation-atmosphere transfer model schemes (Collatz *et al.*, 1991), which have more and more often found their way into global circulation models (Sellers *et al.*, 1997; Cox *et al.*, 1998; Cowling *et al.*, 2009).

However, the up-scaling in time and space from a cell, leaf and plant scale to the stand-level, landscapes or even regional dimensions is not straightforward and incorporates non-linear relationships (Leuning *et al.*, 1995). Furthermore, the complex photosynthesis models need many detailed input parameters, which are often not measurable or not available at canopy or regional scales. This situation can lead to the problem of parameter equifinality referring to the non-uniqueness of optimized parameter sets (Franks *et al.*, 1997; Wang *et al.*, 2001). This means that a more or less equally good model performance

can be achieved by various parameter sets, a fact which consequently renders model predictions uncertain (Beven & Freer, 2001; Schulz *et al.*, 2001).

An alternative strategy to circumvent these difficulties is the development of plant-physiological process models specifically for regional scale and inter-seasonal or inter-annual purposes. Biosphere models like Biome-BGC (Running & Hunt, 1993) or BETHY (Knorr, 2000) with daily time steps were designed with the compromise between mechanistic details on the one hand and simplified process description and integrated consideration of biology and geochemistry on the other hand. 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, which enhances the forecast uncertainty significantly (Knorr & 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 & Kattge, 2005; Rayner *et al.*, 2005; Williams *et al.*, 2004). This tactic has been made feasible by the 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: They chose parsimonious model structures without the implementation of explicit physiological processes occurring at cell and leaf scale. Such diagnostic models treat canopies as functional units by aggregating and averaging processes over space and time. A very popular approach amongst those models integrating system dynamics uses the concept of light use efficiency, which represents the ratio of carbon biomass production per unit of absorbed photosynthetically active radiation (Watson, 1947; Monteith, 1972; Monteith & Unsworth, 2008):

$$F_G = \epsilon \cdot APAR \tag{1.1}$$

with F_G being the gross carbon uptake, ϵ being the light use efficiency, and $APAR$ the photosynthetically active radiation being absorbed. The latter is mainly a function of the amount of incident radiation, of the available leaf surface area and chlorophyll content (Sellers, 1985; Dawson *et al.*, 2003) as well as a function of the leaves geometrical position in relation to the incoming radiation (Chen & Black, 1992). The light use efficiency concept is based on the "functional convergence hypothesis" stating that canopies act as resource constrained units that optimize the capture of light to maximize carbon gain as a result of evolutionary processes (Field, 1991). Convergence of form and functions are a consequence of resource limitations and evolutionary adaption processes to environmental constraints (Funk & Vitousek, 2007; Ustin & Gamon, 2010). The actually reached efficiency with which the absorbed radiation as primary resource of photosynthesis is used, is thus

assumed to be the result of a maximal possible efficiency limited by other resources such as water, and environmental conditions such as the ambient temperature (Runyon *et al.*, 1994). Various studies have proven the light use efficiency to be quite constant over the day, a fact which makes the light use efficiency concept particularly suitable for daily time step models (Ruimy *et al.*, 1995; Rosati & Dejong, 2003; Sims *et al.*, 2005).

The light use efficiency approach has been used as a stand-alone application (Yuan *et al.*, 2007; Mäkelä *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 & Waring, 1994; Prince *et al.*, 1995; Zhao & Running, 2008). The NASA's EOS program makes a global estimation of the gross primary production available (Running *et al.*, 1999; Zhao *et al.*, 2005); the product is based on the light use efficiency approach, too. The light use efficiency parameter has been implemented in ecosystem models as a constant (Landsberg & Waring, 1997; Veroustraete *et al.*, 2002) or modified by restricting environmental factors such as temperature and vapor pressure deficit with predefined functions (McMurtrie *et al.*, 1994; Prince *et al.*, 1995; Veroustraete *et al.*, 2002; Xiao *et al.*, 2004a; Yuan *et al.*, 2007; Mäkelä *et al.*, 2008). To do so, the maximum light use efficiency is usually multiplied with scalars defined by *a priori* specified functions of the limiting conditions (Potter *et al.*, 1993).

Overall, numerous gross primary production models have been proposed and cross-validated with the help of FLUXNET data, but they have mostly been limited to specific vegetation types and regions (e.g. Xiao *et al.*, 2004a,b, 2005; Mäkelä *et al.*, 2008; Yan *et al.*, 2009). Furthermore, they are either too complex for the application at stand-level or larger scales demanding a data information content that can not be satisfied, or they make too strong assumptions that can not hold for a variety of vegetation stands. Many questions have remained unanswered in the field of light use efficiency modeling, that is still "an active area of research with issues remaining to be solved on the leaf, stand, and landscape level [...] targeting issues of upscaling from site observations to ecoregion, biome, and global level [...] with major challenges evident in the areas of modelling" (Hilker *et al.*, 2008, p. 418). There is still little clear evidence for relationships between the variability of the light use efficiency parameter for differing vegetation types and climatic drivers (Turner *et al.*, 2003b; Still *et al.*, 2004). Furthermore, as stated by Garbulsky *et al.* (2010), many of the relationships applied in typical models – such as used as function limiting the maximum light use efficiency – were derived from other models rather than measurements matching the spatial and temporal model application scales. This thesis is a response to the discussed problems and open questions of gross primary production models.

1.1.5 Extrapolation

Estimating gross primary production apart from measurement sites requires beside a profound understanding of the ecosystem dynamics an extensive, scale-corresponding data basis. At the latest when large-scale patterns are to be captured and spatially continuous maps of gross primary production are to be produced, remote sensing has to complement modeling approaches to allow for this task. As the MODIS gross primary production algorithm mentioned above demonstrates, there have been attempts to produce spatially continuous maps of gross primary production on regional or even global scale by means of light use efficiency models, which have been tested at field measurement sites and applied with the help of remote sensing.

In these cases, the light use efficiency models are driven by remotely sensed vegetation properties such as vegetation indices, which reflect the physiological status of plants, and by spatial fields of meteorological variables to detect conditions limiting photosynthesis. The latter are usually either produced by the interpolation of measurements made by meteorological networks such as provided by the NASA's Data Assimilation Office (DAO) or by retrieving the meteorological variables or surrogates of them, respectively, with remote sensing. Typically, the maximum light use efficiency (ϵ_{max}) – to name the most important parameter of a light use efficiency model – is either set to a fixed value or to a land class dependent constant stored in look-up tables (Running *et al.*, 2004; Turner *et al.*, 2003a). The well-known MODIS gross primary production algorithm uses a defined ϵ_{max} for the six considered vegetation classes (Zhao & Running, 2010). This assignment of ϵ_{max}-values to a few, broad vegetation classes is assumed to be the major cause that the MODIS gross primary production product reveals significant problems when compared to EC measurement data (Heinsch *et al.*, 2006; Yang *et al.*, 2007; Zhang *et al.*, 2008; Wu *et al.*, 2010a). Generally, the assumption of a constant maximum light use efficiency "for different sites within a given biome is far from optimum and is the possible cause of the low performance of the photosynthetic uptake models" (Garbulsky *et al.*, 2010, p. 255). Stoy *et al.* (2009) noted, too, that static ecosystem assumptions may not be valid over longer time and larger spatial scales with probable changes in ecosystem structure and functioning being involved. In recent studies, Groenendijk *et al.* (2011) and Mäkelä *et al.* (2008) confirmed that there is no general set of parameters for gross primary production models; even more, the parameters vary strongly within broad vegetation classes and even within simple plant functional types specified by coarse vegetation and climate classifications. Another way some studies followed to tackle the problem of deriving spatially continuous gross primary production estimates is the use of model-independent, purely empirical machine learning algorithms trained with meteorological data, remotely sensed vegetation properties and vegetation type in a "black-box"-fashion (Papale & Valentini, 2003; Yang *et al.*, 2007).

Overall, "current remote sensing methodologies for estimating gross primary productivity are not satisfactory" (Garbulsky *et al.*, 2010, p.2860) despite numerous attempts made. A further alternative strategy could be constituted by relating calibrated light use efficiency model parameters to site specific information and utilize these relationships to regionalize the parameters. This approach would lead to more flexible model parameterizations facilitating the model application to a broad range of vegetation types and climatological conditions. This dissertation follows this approach, since – unlike in the field of hydrology – explicit regionalization strategies for calibrated model parameters across biomes have not been pursued with the necessary emphasis until recently. Respective efforts – sparked by unprecedented, harmonized data assemblies – have recently got new impetus (Jung *et al.*, 2009; Garbulsky *et al.*, 2010; Groenendijk *et al.*, 2011).

1.2   Research Objectives

1.2.1   Outline

The comprehensive FLUXNET and MODIS data sets represent a substantial financial investment into understanding the role of the terrestrial biosphere in the Earth system. This work is dedicated to utilize these data collections and exploit relevant information content extractable from FLUXNET data sets with non-stationary and non-linear dynamic time series analysis tools in order to derive a robust light use efficiency model with daily time steps to account for seasonal dynamics. Measurement data are given more weight in the model building process than usual. In this way problems with highly uncertain model parameter calibrations due to lacking information content of the available data shall be circumvented. At the same time, the structure of the light use efficiency model is sought to be so flexible that it is applicable to various vegetation types under a wide range of climatic conditions. The site-specific calibration of the model parameters such as the maximum light use efficiency allows for parameter sets whose dependence on site conditions is not restricted to broad vegetation classes. The development of an extrapolation scheme is a first attempt towards a robust temporal and spatial extrapolation of flux measurements via model parameter regionalization. The MODIS LAI/FPAR product is applied in both steps – model building and extrapolation – and thus an important product within this study. Therefore, a thorough analysis of possible usage options is carried out to deduce optimal application strategies.

This thesis is structured in three main chapters, which build upon each other and present the research on the three discussed topics carried out in this thesis: analysis of the usage options of the MODIS LAI/FPAR product (chapter 3), formulation of a light use efficiency model (chapter 4) and, finally, exploration of a proposed extrapolation scheme (chapter

5). A preceding chapter (chapter 2) presents the data used throughout all study parts of this thesis as well as general data processing methods. Chapter 6, finally, completes this thesis with a synthesis and an overall discussion of the three study parts and makes suggestions for further research activities. The appendix provides additional tables on the study sites and results. In the following, a short introduction is given to the three core chapters of this thesis.

1.2.2 Investigating Usage Options of MODIS LAI/FPAR Data

Chapter 3 deals with several usage options of MODIS LAI/FPAR data, since this remote sensing product provides key input data for ecosystem models and in this function it will be extensively used in this thesis, too: FPAR is an input variable for the derived light use efficiency model, and the LAI serves as important characteristic for the extrapolation scheme. But as a matter of fact, there is still no consensus on the product's mode of usage and its combination with FLUXNET data concerning (i) spatial aggregating, (ii) sensor choice (Terra, Aqua, or both) and (iii) the employment of the additionally provided quality criteria (QC), despite the frequent application of this product in ecosystem models. With the final aim of finding a way for an optimal application of the MODIS LAI/FPAR subsets, various usage options are explored. Using the example of the LAI values provided by this MODIS product, the effects of applying various QC filters on magnitude and temporal dynamics of MODIS LAI time series are analyzed at six exemplary FLUXNET sites. Furthermore, the advantage of a spatial aggregation of LAI values of the pixels available in the MODIS subset around the pixel containing the measurement tower is explored. The benefits of taking Terra, Aqua or both sensor products into account is studied. Finally, the sensitivity of a simple soil-vegetation-atmosphere-transfer (SVAT) model on differently post-processed LAI times series is assessed. The results of this analysis enable a best possible utilization of the MODIS LAI/FPAR product in the subsequent modeling and extrapolation studies.

1.2.3 Deriving a Light Use Efficiency Model

A novel model for the gross primary production on stand-level is formulated in chapter 4 to circumvent some of the typical problems of ecosystem models elaborated above (see section 1.1.4). This study distinguishes itself from existing ones by (i) its data-driven model building approach which makes a minimum of prior functional assumptions, by (ii) avoiding the assignment of fixed parameter values for broad vegetation classes without (iii) making the model so complex that difficulties in the context of parameter calibration arise, by (iv) allowing for an explicit extrapolation to larger scales with remote sensing data, and by (v) the resulting flexible model structure which makes it possible to apply the model to a wide variety of climate conditions.

As mentioned above, relationships like the basic light use efficiency equation are well suited as a starting point for model identification procedures in a top-down fashion. The model building process in this study starts therefore with an "oversimplified" model structure based on the light use efficiency concept. Micrometeorological measurements from a broad range of FLUXNET sites are then analyzed to further refine the modeling scheme. This approach enables the formulation of a flexible but robust hybrid stochastic-mechanistic model with not more complexity than can be supported by the observation data information content.

Suitable tools for this methodology are non-parametric state-dependent parameter estimation (SDP) and dynamic linear regression (DLR) based on filtering and smoothing techniques. They allow for the time and state dependent evolution of model parameters to be estimated directly from measured time series (Young & Pedegral, 1999; Young et al., 2001) and have shown to be capable of capturing seasonal behavior of ecosystems (Young, 1998; Schulz & Jarvis, 2004; Jarvis et al., 2004; Gamier, 2006; Taylor et al., 2007). Uncertainties inevitably arise when constraining a model with observations, but using the mentioned analysis tools in conjunction with the selection of an appropriate noise variance ratio (see below) when estimating the time or state dependencies, the uncertainties remain traceable.

In this study, SDP and DLR are employed to derive a simple but – in contrast to many models published earlier – broadly applicable model structure for the estimation of gross primary production by extracting state-dependencies of the time-varying light use efficiency parameter ϵ. Based on these results, functional forms describing ϵ are incorporated into the model structure. The model parameters are calibrated site specifically following the assumption that there is no single set of parameters that describes the behavior of sites across climate classes and vegetation types. Finally, the optimized model parameters are qualitatively analyzed for patterns which relate themselves to site specific characteristics serving the final aim of regionalizing the model parameters.

1.2.4   Finding an Extrapolation Scheme

In chapter 5, finally, the model parameters calibrated at FLUXNET sites are related to a variety of site characteristics in oder to spatially extrapolate them to unobserved sites as a first step towards regionalization schemes for spatially continuous gross primary production maps. To do so, site characteristics are extracted from the available FLUXNET and MODIS data: General vegetations and climate classes are selected, climate characteristics such as the average annual temperature amplitude are taken into account, as well as physiological and phenological characteristics. Support vector regression (SVR), a powerful machine learning technique, is applied to relate the model parameters to these characteristics. SVR has a high generalization capacity and offers the advantage over many similar methods in avoiding over-fitting of training data. It is suitable for capturing

highly non-linear relationships and, moreover, is capable of exploiting information from relatively small training data sets.

Apparently, the described spatial extrapolation approach and even more the data-based derivation of a light use efficiency model sketched above rely on the existence of suitable data sets. Indeed, the growing number of micrometeorological and flux measurements within the FLUXNET framework makes such an approach as proposed in this thesis more feasible than ever.

2

Data and
Data Processing Methods

2.1 Overview

Throughout this thesis, d**a**ta from two sources are used: micrometeorological measurements provided by FLUXNET and products from the satellite sensor MODIS. FLUXNET data as well as MODIS products are coordinated by the ORNL DAAC (Oak Ridge National Laboratory Distributed Active Archive Center), one of the large NASA's (National Aeronautics and Space Administration) Earth Observing System Data and Information System (EOSDIS) data centers. In this chapter, the FLUXNET data are characterized, data processing measures are explained and the applied MODIS products are described. Finally, general model performance measures and criteria for the comparison of time series are introduced.

2.2 Micrometeorological Data

2.2.1 FLUXNET Study Sites

FLUXNET – organized under the roof of NASA's ORNL DAAC – is a global research network of currently over 500 measurement towers, which continuously gather standardized data on the exchange of energy, water and CO_2 at the canopy level (Baldocchi *et al.*, 2001); additionally, meteorological information and relevant site characteristics are collected. FLUXNET itself is an umbrella organization for regional measurement networks such as AmeriFlux and CarboEurope, naming the largest ones. The distribution of the vegetation types, in which the towers are located in, mirrors roughly the fraction the respective
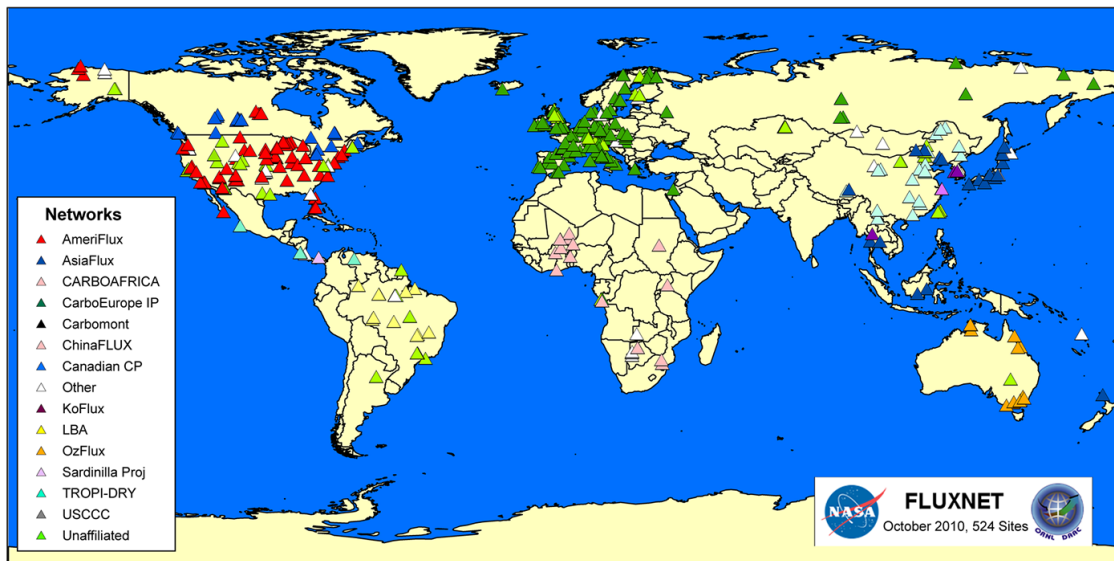
Figure 2.1: Geographical location of the FLUXNET measurement towers and their organizational affiliation. (FLUXNET)

vegetation types cover globally. Figure 2.1 showing the flux towers distributed throughout the world underlines the great scientific potential of this network to elucidate temporal and spatial dynamics and variability of carbon, water and energy fluxes.

44 forest and grassland FLUXNET sites in climate zones reaching from boreal to semi-arid were chosen as data base for this study. The selection criterion was the existence of at least three measurement years at the beginning of this study and no measurement gaps exceeding 14 consecutive missing days. The regional networks outside North America and Europe are growing fast but at the time starting this project, longer time series were only available from AmeriFlux and CarboEurope. The selected sites are therefore located in North America and Europe (Figure 2.2) and comprise 18 coniferous forest sites, 12 deciduous, 5 mixed, 2 evergreen forests as well as 7 grasslands. The majority of them is located in the continental climate type according to the Köppen-Geiger classification (21 sites), but the temperate climate is (with 20 sites) nearly as strongly represented. Only three sites are characterized by a dry climate. Table A.1 summarizes the main characteristics of the selected sites. Meanwhile, a standardized synthesis data set ("La Thuile") is available for FLUXNET data providers and selected project teams (Agarwal *et al.*, 2008). Since our team did not got access before the final stage of this thesis, the La Thuile data set is only used for comparison with the flux data processing method applied in this thesis (see section 2.2.3).
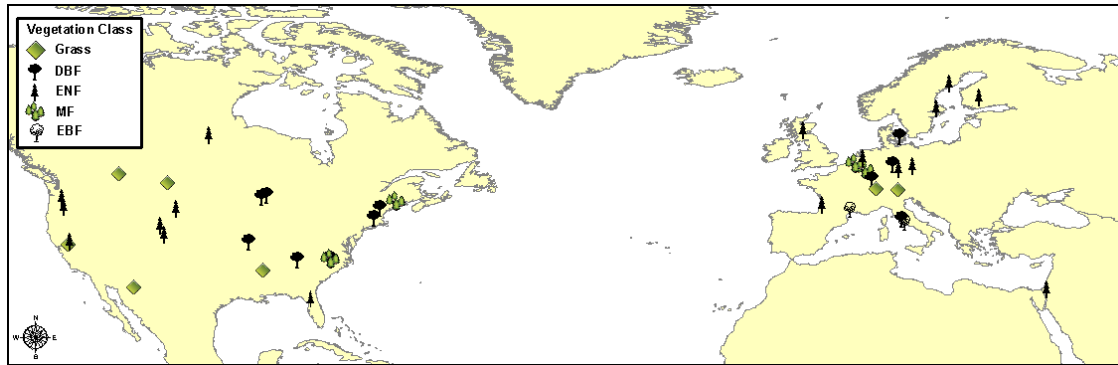
Figure 2.2: Geographical location of the 44 study sites and their dominant vegetation type. DBF: deciduous broadleaf forest, ENF: evergreen needleleaf forest, MF: mixed forest, EBF: evergreen broadleaf forest. The center of the markers indicates the location of the towers.

2.2.2 Eddy-Covariance Measurements

To measure the exchange of energy, water and CO_2 the eddy-covariance (EC) measurement technique is the method of choice. FLUXNET participants use the sophisticated EC technique, too, and that in a standardized manner regarding instrumentation and data processing to facilitate cross-site comparisons (Papale *et al.*, 2006). The underlying idea behind the EC technique is the assumption that the vertical air flow above a canopy – and with it the transportation of energy and matter – consists of air parcels in turbulent motion and this turbulence is the main driver of transfers of energy and matter; these air parcels are called eddies. The EC technique quantifies this motion along with concentration measurements of the desired entity. These two turbulent variables are decomposed into mean and fluctuating components (Arya, 2001). After this so-called Reynolds decomposition and averaging, the resulting flux density is finally calculated as the covariance between the fluctuations of the velocity of the vertical air movements, w', and the mixing ratio of interest (Baldocchi, 2003). In case of the net CO_2 flux, F_N, this is:

$$F_N = \rho_a \cdot \overline{w' \cdot c'} \tag{2.1}$$

with ρ_a being the air density and c' the fluctuation of the CO_2 mixing ratio. This equation is valid under the assumption of negligible density fluctuations and mean vertical flow over horizontal, homogeneous terrain. The measurements are representative of a specific up-wind area, called the "footprint" or "fetch" (Schmid *et al.*, 2000). This area depends mainly on the height of the measurements made, the surface roughness length, wind speed and direction and atmospheric stability (Chen *et al.*, 2009).

The technique obviously requires measurements being done above the canopy but still within the boundary layer of interest, so the instruments are mounted on a tower; its height depends on the height of the considered canopy (Figure 2.3a and b). The main instruments are a sonic anemometer and an open- or closed-path infrared gas analyzer (Figures 2.3c). Turbulent fluctuations happen fast, so their recording demands high-end

instrumentations with an ideal temporal resolution of 10 to 50 Hz (Bosveld & Beljaars, 2001). Usually, 30 or 60 minutes are chosen as averaging periods for the flux density calculations.
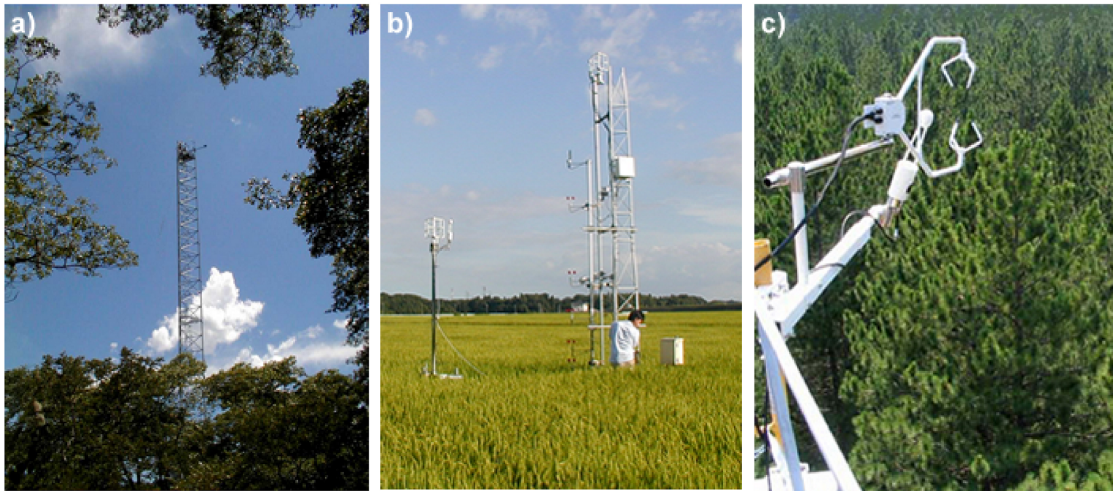
Figure 2.3: Eddy-covariance measurement towers above a forest (a) and a corn filed (b). The basic sensors required for this technique are shown in c: Sonic anemometer and (open-path) CO_2 and H_2O analyzer.

However, instrument limitations and assumptions that often cannot be met in reality require sophisticated corrections of the flux measurements. Spectral corrections are to be done as well as adjustments for drainage or advection, for limitations in measuring low frequency contributions, diffusive exchange, the coordinate system, for air density deviations and for stable stratifications often occurring at night (Massman & Lee, 2002). However, even post-processing measures cannot prevent errors being made and EC data are therefore afflicted with noise (Richardson *et al.*, 2008; Stauch *et al.*, 2008) what has to be taken into account when using the measurement data as basis for modeling purposes.

2.2.3 Data Processing Methods

The data were downloaded from the web gateways of the regional FLUXNET sub-networks AmeriFlux (AmeriFlux, 2009) and CarboEurope (CarboEuropeIP, 2009) as so-called "level 2 data". At this processing level, the raw measurement data have been processed and corrections schemes applied by the individual site measurement teams but no other processing steps have been carried out. The downloaded data including energy and carbon fluxes along with meteorological variables have measurement gaps. These are filled in the following way: Short gaps up to three hours in meteorological and soil moisture variables are 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.*,

2001a) serve to fill gaps of medium length up to 4 days. Even larger gaps are replaced with the respective values averaged over the whole time series available.

State dependent parameter modeling (SDP) is used to fill gaps in the latent and sensible heat fluxes. SDP employs linear models in which the parameters (c_i) vary as functions of system states (u_i):

$$y(t) = \sum_{i=1}^{N} c_i(u_i(t)) \cdot x_i(t) + \zeta(t) \tag{2.2}$$

with y being the function values, x_i the depending variable, N the number of considered dependencies and ζ being the model error series; all variables are varying with time, t. A more detailed explanation of SDP is given in section 4.3, in which this tool is applied to find new model structures by analyzing measured time series. For the latent heat flux, λE, following model realization is chosen:

$$\lambda E(t) = c_1(T_A(t)) \cdot R_N(t) + c_2(VPD(t)) + \zeta(t) \tag{2.3}$$

where R_N is the net radiation, T_A the air temperature and VPD is the vapor pressure deficit. Thus, λE is a function of the available energy, the temperature and the moisture in the air. The remaining energy, $R_N - \lambda E$, the soil temperature T_S and, again, T_A serve to model the sensible heat, H, to fill its measurement gaps:

$$H = c_1(T_A(t)) \cdot (R_N(t) - \lambda E(t)) + c_2(T_A(t)) + \zeta(t) \tag{2.4}$$

Missing data in the time series of the net CO_2 flux, F_N, are replenished (Figure 2.4a) with the multi-dimensional semi-parametric spline interpolation scheme developed by Stauch & Jarvis (2006). The methodology combines the benefit of using current observations and knowledge about the major driving forces of F_N (light, temperature and time) and assumes that F_N can be described by a three dimensional function, f, of the mentioned variables:

$$F_N = f(S_0, T_A, t) + \zeta(t) \tag{2.5}$$

where S_0 is the incident radiation. The model error series, ζ, includes the stochastic behavior of the system as well as the uncertainty of the EC measurements. This unknown function is obtained by fitting a piecewise cubic spline hyper-surface to the available data of every site and each year. This is done by setting several fixed nodes in the variable space. The intervals between these nodes are interpolated by Hermite cubic splines. The values at the nodes are optimized against the data through nonlinear least squares minimization. Figure 2.5a shows an example of such a hyper-surface, which is for simplicity three-dimensionally plotted against the temperature and the time.

This procedure not only fills missing data but also extracts the deterministic component of the measured time series (Figure 2.4b). A stochastic noise component is then added to this signal component. The added noise is retrieved from the spline model residuals using empirical cumulative probability distributions. A Monte Carlo framework for multiple

realizations of the joint signal-noise model leads to the final flux estimates. In this way, the stochastic nature of EC measurements is explicitly addressed (Stauch *et al.*, 2008). Since the goal of this thesis is to exploit the information on systematic behavior of the CO_2 fluxes as apparent in EC measurements, only the so derived deterministic signal component of F_N is used in the following.
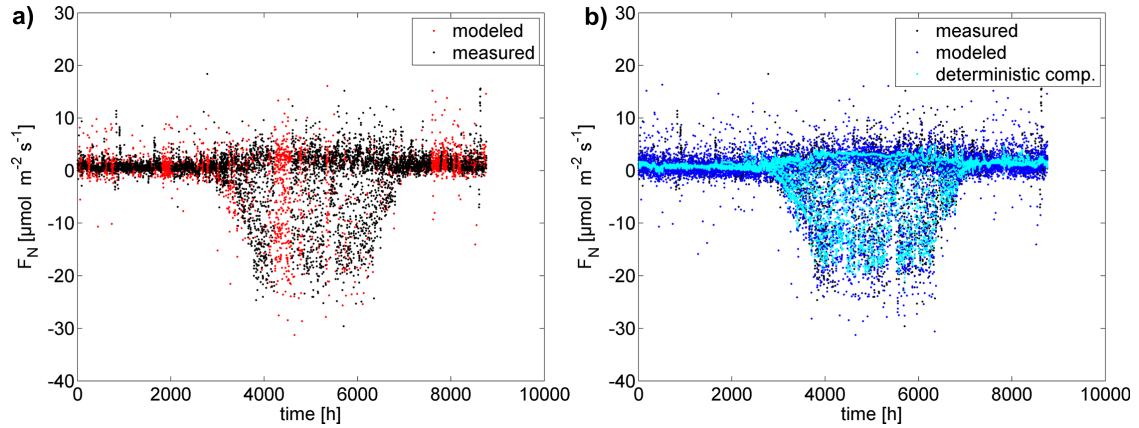
Figure 2.4: a): The measured, incomplete time series of F_N and the gap-filled time series. b) Same fluxes as in the left plot but with the extracted deterministic component (cyan) being highlighted. For gap-filling and determination of the deterministic component the semi-parametric spline method was applied. Data from Harvard Forest, 1992.

The methodology compares very well to other gap-filling techniques for eddy covariance net carbon fluxes. The performance of 15 techniques was tested for artificial gap scenarios based on a set of ten benchmark datasets, and the method applied here performed best together with non-linear regression techniques, a look-up table approach and marginal distribution sampling (Moffat *et al.*, 2007). Figure 2.5b shows an exemplary comparison of the time series resulting from the semi-parametric spline interpolation (deterministic plus stochastic component) with the respective flux from the La Thuile data set.

The so derived net CO_2 flux is finally split up into its components, the respiration and the gross flux of carbon uptake (F_G) by using the semi-parametric model for the case $S_0 = 0$ to quantify the respiration (Figure 2.5a). However, whilst the method performed well in a comprehensive comparison of various partitioning methods used in the FLUXNET community applied to a synthetic data set and European forests sites (Desai *et al.*, 2008b), problems of the method get evident if the respiration flux has to be extrapolated over larger areas in the system state space. This is the case if either measurement gaps are large or if day temperatures exceed those occurring at night considerably. Thus, the hyper-surface for the case $S_0 = 0$ has only information for lower temperatures. Whilst otherwise performing satisfyingly, the Hermite cubic splines proposed in Stauch & Jarvis (2006) can in these cases get somewhat unstable and lead to possibly unrealistic high or low values (Figure 2.6). The method applied for the generation of the La Thuile data set
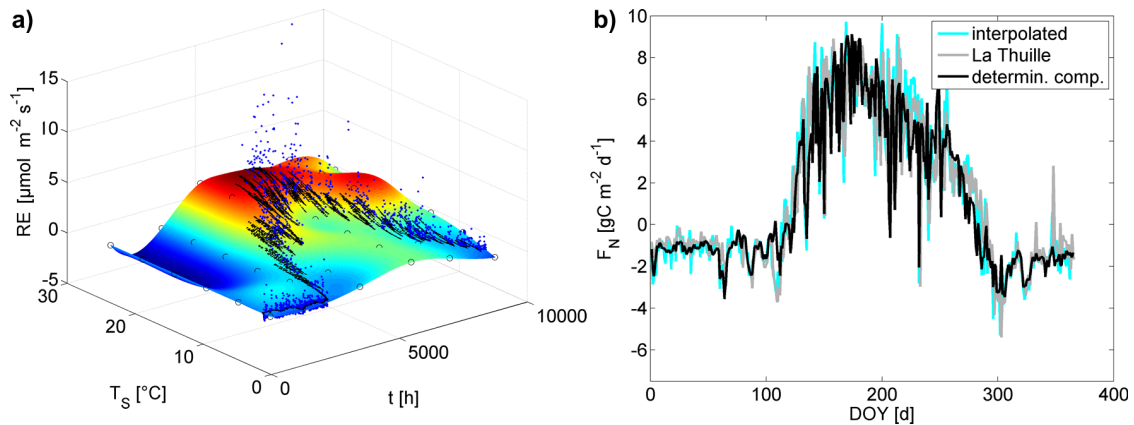
Figure 2.5: a) Example (forest site UMBS, 2001) for a hyper-surface that is spanned through the measured data to fill measurement gaps and extract the deterministic component. The fourth dimension, the incident radiation, is omitted for clearness and the model is plotted for $S_0 = 0$, thus the night fluxes are shown. The blue points are the measurement data, the black points show the realized data points upon the hyper-surface representing the deterministic component of the noisy measurement data. b) The gap-filled time series compared with the respective time series in the La Thuile data set. Exemplarily shown for the forest site MMSF, 2004.

also shows this "overshooting" behavior in some cases; Figure 2.7 is an extreme example for this.

Therefore, a more robust method is used for the extrapolation of the night fluxes to the daytime respiration fluxes occurring in the daytime, namely the thin-plate smoothing spline method (Wahba, 1979) as implemented in the MATLAB® Curve Fitting Toolbox 3.0. Thin-plate smoothing splines are especially helpful when noisy multi-dimensional data have to be inter- and extrapolated and have been used before within a meteorological context (Boer *et al.*, 2001; Tait *et al.*, 2006). Figure 2.7a shows an example of such a thin-plate smoothing spline surface. Obviously, the robustness and more conservative behavior is somewhat at the expense of short-term variability. But overall, this methods leads to more reliable results. Figure 2.8 shows two examples of disaggregated F_N time series from the La Thuile data set and those resulting from the method applied in this thesis.

Having determined the respiration flux, the gross flux of carbon uptake (F_G) is afterwards simply calculated as difference of net flux and respiration. The so-derived flux of carbon uptake, the gross primary production, is used in the further considerations within this thesis. All values – meteorological and ecosystem fluxes – are aggregated to daily values: climatological variables are averaged, fluxed are summed up. These are used throughout this thesis with one exception when a SVAT model is run with hourly data to analyze the effects of different LAI time series on the results of evapotranspiration calculations. At all other stages of this thesis – for finding new model structures, for calibrating, validating and spatially extrapolating the light use efficiency model – daily data are used. On this
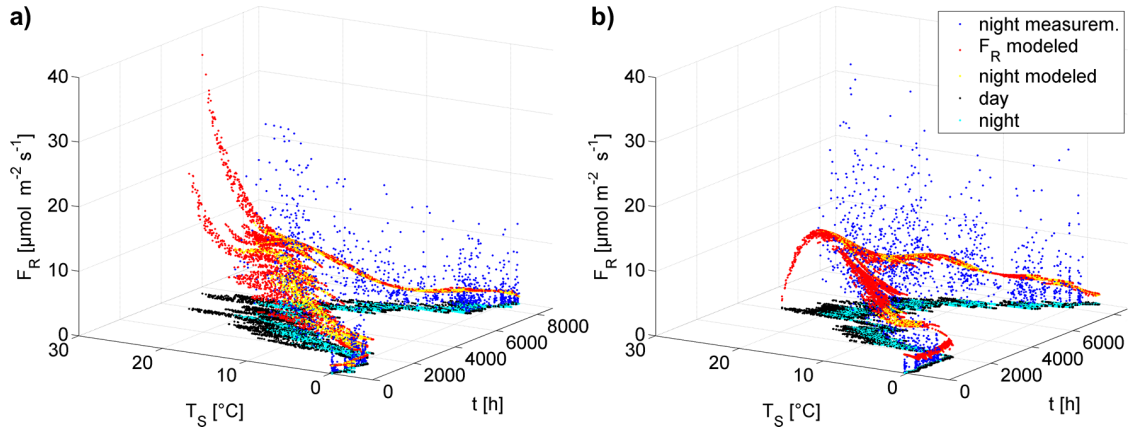
Figure 2.6: Example for a spline hyper-surface within the temperature-time-space (T_S, t) used for modeling the respiration flux, F_R, by extrapolation of the measured night fluxes ($S_0 = 0$). The blue dots are the night fluxes, the modeled values (deterministic component) are dotted in red. The yellow dots represent the model values within the range of variable input values occurring at night. For a better indication of the extrapolation to be done, the day (black) and night (cyan) fluxes are projected to the T_S-t-plane. Data from the grassland site Neustift, 2003 (a), and 2006 (b).

daily basis, the carbon fluxes are expressed in terms of the units gC m^{-2} d^{-1}, the energy fluxes have the units MJ m^{-2} d^{-1}.

2.3 MODIS Data

2.3.1 The MODIS Sensor

As sensor on-board Terra and Aqua, two sun-synchronous and near-circular satellites, MODIS scans the earth surface every second day, above 40° latitude even every day with a swath width of 2330 km. The 36 spectral channels in the visible and infrared spectrum (0.4 to 14.4 μm) have a spatial resolution of 250 to 1000 m, with the shorter wavelength having the greater resolution (Hyman, 1996). 44 MODIS products are offered, ten of which characterize the land surface. Key products of this collection are prepared for the usage with FLUXNET data. Three of these products – the LAI/FPAR product, the land classification product as well as the product containing vegetation indices – are used in this thesis.

2.3.2 The MODIS LAI/FPAR Product

The MODIS LAI/FPAR product, retrieved from MODIS Terra as well as Aqua data, is referred to as MOD15A2 or MYD15A2, respectively (Myneni *et al.*, 2002). It provides 8-day estimates of the leaf area index (LAI) accompanied by the closely related FPAR
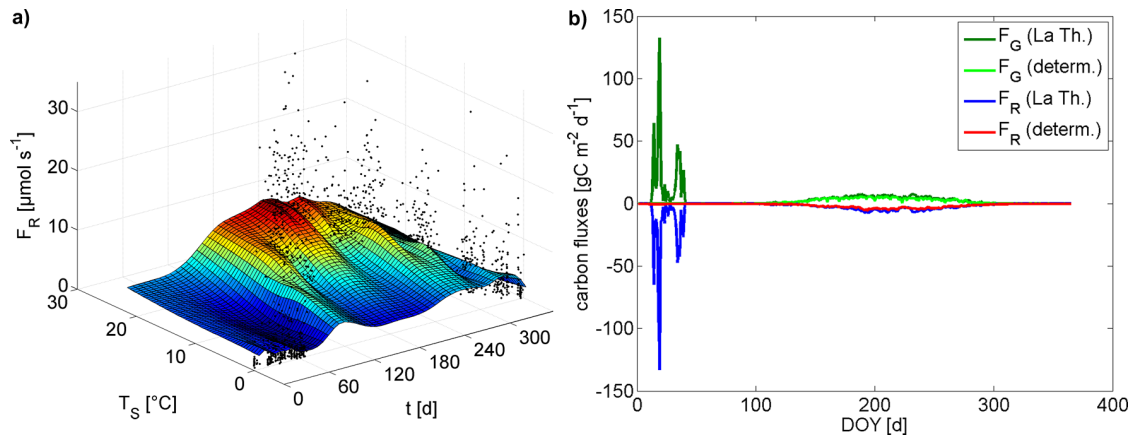
Figure 2.7: a) Hyper-surface for modeling the respiration flux, F_R, with thin plate smoothing splines at the grassland site Neustift, 2006. b) Instable behavior of the disaggregation method used in the La Thuile data set occurring at the forest sites Boreas, 1997, and comparison with the fluxes resulting from the more robust method used in this thesis.
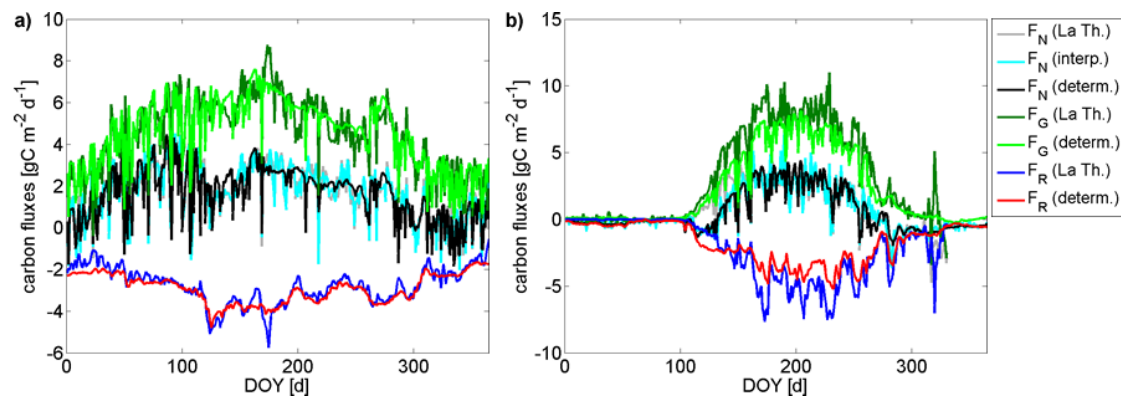
Figure 2.8: Two exemplary comparisons between the net and disaggregated fluxes resulting from the semi-parametric method applied in this thesis and those from the La Thuile data set. a) Mediterranean forest site Castelporziano, 2004, b) continental, deciduous forest site Sylvania Wilderness, 2003.

(Fraction of Absorbed Photosynthetically Active Radiation). To retrieve these products, up to seven spectral bands are utilized to solve an inverse problem with a radiative transfer model. The parameters for this model are stored in a look-up table with specific values for eight biomes. If this main algorithm fails, a back-up algorithm is triggered, which regresses the LAI and FPAR values on the basis of an empirical relationship between these variables and the NDVI (Normalized Difference Vegetation Index). The final product has a resolution of 1 km. The daily values are merged to 8-day composites to reduce the impact of clouds and aerosols as disturbing factors. In order to do so, the LAI and FPAR values of the day with the maximum FPAR value is selected for the respective 8-day period (Knyazikhin *et al.*, 1999) since this value is expected to show the lowest atmospheric influence.

To create the MODIS Land Product Subsets for the specific usage with FLUXNET data, the LAI/FPAR product is resampled in such a way that the center pixel of a 7x7 grid of 1-km^2 pixels contains the FLUXNET tower. The subsets contain six values for every time step: a LAI as well as a FPAR value, their standard deviations, a general and a detailed quality criterion (QC). For the usage in this thesis, MOD15A2 and MYD15A2 are downloaded in ASCII-format from the ORNL DAAC website (ORNL DAAC, 2009). The LAI and FPAR values and the general QC are extracted from the data sets of both sensors.

The general QC provides four quality categories in 8 bits. One quality criterion is called "MODLAND_Q" and allows a first check on the LAI quality. Since the algorithm path is the main factor influencing the LAI quality (Knyazikhin *et al.*, 1999; Yang *et al.*, 2006b), MODLAND_Q informs the user if the main algorithm was used, if the back-up algorithm had to fill in or if the value could not be retrieved at all. Another quality criterion indicates if the detector worked for up to 50 % of the channels or was mainly dead. The cloud state quality bit specifies if the pixel was clear, a significant cloud coverage or mixed clouds were present, or if the cloud status could not have been identified. Finally, additional details about the algorithm path are provided: A retrieval with the main algorithm without saturation is classified as "best possible result", with saturation as "good result". Saturation means, the reflectances do not deliver precise information about the surface and the canopy radiation transfer model is insensitive to a set of possible canopy realizations, thus the probability for different LAI values is of the same magnitude. This situation occurs with increasing frequency for LAI values greater than 4 and affects almost every value greater than 5 since the reflectances at these high LAI values get more and more insensitive to LAI differences (Knyazikhin *et al.*, 1999). It is further stated in the detailed QC bits if the main algorithm failed and the back-up algorithm was used due to geometry problems or for other reasons than bad geometry. A complete failure is the last possible outcome of the algorithm path analysis.

Chapter 3 investigates the usage of these QC meta-data and analyzes the consequences of different modes of application; the same chapter furthermore pursues the questions if aggregating the provided 7x7 pixels around the measurement towers as well as if merging the Terra and Aqua data sets is useful for continuous and meaningful LAI/FPAR time series. Finally, an interpolation scheme is proposed.

2.3.3   MODIS Land Cover Data

Additionally, the MODIS land cover product (MCD12Q1) retrieved with data from Terra and Aqua (Friedl *et al.*, 2002), is downloaded to further characterize the pixels around the tower and to filter pixels of strongly differing land classes. MCD12Q1 subsets are available for the years 2001 to 2005. They provide annual land class estimations around the tower with a spatial resolution of 250 m. Five land classifications are offered. For this study, the third land cover scheme, the so called "MODIS-derived LAI/FPAR scheme" has

been selected. This scheme distinguishes between 8 vegetation classes beside water, non-vegetated and urban classes: 1. Grasses and cereal crops, 2. shrubs, 3. broadleaf crops, 4. savanna, 5. evergreen broadleaf forests, 6. deciduous broadleaf forest, 7. evergreen needleleaf forest, 8. deciduous needleleaf forest. These eight classes are exactly those which are used in the LAI retrieval. They have shown to be somewhat more reliable than the other classifications provided (Heinsch *et al.*, 2006).

It is assumed that there have been no substantial land use changes in the areas around the towers neither in the years 2001 to 2005 nor later on. For every study site, a single land class map for all study years is generated by comparing the corresponding pixels of all available years. That land class is assigned to the pixel that at least occurs three times. After the temporal aggregation, the pixels are spatially aggregated: A mask with its resolution matching that of the LAI grid is generated, for what the four corresponding 250-m pixels of each 1-km pixel are examined. That land class which occurs at least at three of the four sub-pixels is assigned to the 1-km pixel. If no land class appears more than two times, this pixel is treated as unclassified and not considered in the further analysis. Additionally those pixels are excluded from the further study which have a completely different land cover and consequently not comparable LAI values in relation to the tower pixel. Borak & Jasinski (2009) have shown spatial aggregations based on land class selections to be superior. However, pixels with another but similar land class assigned to by the MODIS classification are accepted for the further analysis. Our reason for this procedure is that the MODIS land cover product often seems to have difficulties to distinguish between similar vegetation classes, especially in inhomogeneous areas (Cohen *et al.*, 2003; Heinsch *et al.*, 2006). Misclassifications in relatively similar biomes have moderate consequences on the LAI retrieval (Heinsch *et al.*, 2006). Therefore, the simplification is made that all the forest classes 5 to 8 are considered as similar in the following for the MODIS LAI subsets; the classes 1 to 4 serve as similar classes for the grassland sites. The classification problems and the necessity of this simplification are clearly evident in the downloaded subsets. For example, the concerned MODIS classification cannot detect the loblolly pine plantation within a hardwood forest, but identifies a broadleaf forest in the whole region around the tower at Duke Forest. Additionally, many pixels are variantly classified as shrub, savanna and grass and evergreen and deciduous forest during the five years; the severe ice storm disturbances and the rich understorey certainly play a role in this context (see section 2.2.1). At the grassland sites, the grassland class alternates spatially and temporally with the savanna class and in some cases with the crop class. The center pixel at Vaira, for instance, is even classified as savanna in every year. For these reasons, only those pixels around the tower are taken into account in the following which have a similar land class as the known vegetation class of the FLUXNET site according to the assumptions explained above.

2.3.4 MODIS Vegetation Indices

The MODIS land products MOD13Q1 and MYD13Q1 from Terra and Aqua data (Huete *et al.*, 2002) contain the vegetation indices NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index). For the usage in this study, they are also downloaded as regridded MODIS Land Product Subsets and interpolated with smoothing splines in the same way as the MODIS LAI/FPAR subsets (see chapter 3).

The NDVI is the ratio of the near infrared (ρ_{NIR}) to red reflectance band (ρ_{red}) and standardized to values between -1 and 1:

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \tag{2.6}$$

The calculation of EVI (Rahman *et al.*, 2005) is based on the NDVI but takes additionally canopy background and atmospheric influences into account and incorporates blue band reflectance (ρ_{blue}):

$$EVI = C_1 \cdot \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_2 \cdot \rho_{red} - C_3 \cdot \rho_{blue} + C_4} \tag{2.7}$$

with the coefficients C_1, C_2, C_3, and C_4.

Vegetation indices such as NDVI and EVI represent a measure of both photosynthetic activity and canopy structure in an integrative manner. They are useful in vegetation monitoring, time series analysis as well as change detection studies. Spatial and temporal variability as expressed in NDVI and EVI data arise from several vegetation related properties, including LAI, green biomass, optical leaf properties, canopy structure and architecture, species composition as well as land cover type (Huete *et al.*, 2002). The MODIS NDVI/EVI therefore has been used to serve a multitude of purposes: The product has been applied to detect seasonal patterns of leaf phenology (Hess *et al.*, 2010), to model gross primary production (Xiao *et al.*, 2004a,b; Li *et al.*, 2007) or to monitor rapid vegetation succession (Zhao *et al.*, 2009). The indices were linked to vegetation water content (Cheng *et al.*, 2006) and live fuel moisture content for fire risk assessment (Yebra *et al.*, 2008) or even applied to estimate the root zone soil moisture content (Schnur *et al.*, 2010). In this study, EVI values are used amongst other attributes to relate optimized model parameters to site characteristics.

2.4 Model Performance Measures

Several measures are used throughout this thesis to assess the predictive power of the derived modeling schemes or to evaluate the accuracy with which remote sensing products represent field-measured data. In the following, O denotes an observed variable and P its predicted counterpart, both having N elements.

2.4.1   Coefficient of Determination

The widely used coefficient of determination, r^2, is a measure for the goodness of fit. It explains how much of the variance of observed values is explained by predicted values. A common definition according to Bravais-Pearson is (Krause *et al.*, 2005):

$$r^2 = \left(\frac{\sum_{i=1}^{N}(O_i - \bar{O}) \cdot \sum_{i=1}^{N}(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \bar{O})^2} \cdot \sqrt{\sum_{i=1}^{N}(P_i - \bar{P})^2}} \right)^2 \tag{2.8}$$

with i running from 1 to N, the number of samples. A perfect fit is indicated by a r^2-value of one. A major drawback of its usage is the fact that it is oversensitive to outliers and does not take systematic under- or overprediction of observed values into account since only the dispersion is quantified. This, however, is insensitive to proportional or additive differences (Legates & McCabe Jr, 1999).

2.4.2   Nash-Sutcliffe Efficiency Criterion

Therefore, other quality measures are often additionally used. The Nash-Sutcliffe efficiency criterion (Nash & Sutcliffe, 1970; Krause *et al.*, 2005), in this thesis denoted as NS, is widely used in hydrology to evaluate modeled time series with measurements and is defined as the variance of residuals of predicted and observed values normalized by the variance of the observed values and subtracted from one:

$$NS = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}. \tag{2.9}$$

This quality criterion ranges from a value of one for a perfect fit to $-\infty$ whereas negative EC values indicate that the mean of the observed time series would have been superior to the predictor.

2.4.3   Squared Errors

The sum of squared errors, SSE, is a basic measure that simply sums the squared deviations of the modeled to the observed values:

$$SSE = \sum_{i=1}^{N}(P_i - O_i)^2. \tag{2.10}$$

Similar to SSE, the root mean squared error, RMSE, considers the differences between the observed and modeled values, averages them and takes the root to give a mean of the deviation expressed in units of the considered variables:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2)}{N}}. \tag{2.11}$$

To facilitate an comparison of model performances regarding time series of different magnitudes, the RMSE is normalized with the range of the observed values:

$$nRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{N}}}{O_{max} - O_{min}} \tag{2.12}$$

where O_{max} is the maximum O-value and O_{min} the minimum O-value.

2.4.4 Weighted Performance Indicators

The weighted SSE, SSE_w, takes different weights into account that are assigned to the data points by a weight vector w:

$$SSE_w = \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{w_i} \right)^2. \tag{2.13}$$

A weighted form of the r^2 is also used in this thesis and defined as:

$$r_w^2 = \frac{\sum_{i=1}^{N}(w_i \cdot (O_i - O_w) \cdot (P_i - P_w))}{\sqrt{\sum_{i=1}^{N}(w_i \cdot (O - O_w)^2)} \cdot \sqrt{\sum_{i=1}^{N}(w_i \cdot (P_i - P_w)^2)}} \tag{2.14}$$

with the weighted means (Kulinskaya & Staudte, 2006):

$$O_w = \frac{\sum_{i=1}^{N}(w_i \cdot O_i)}{\sum_{i=1}^{N} w_i}, P_w = \frac{\sum_{i=1}^{N}(w_i \cdot P_i)}{\sum_{i=1}^{N} w_i}. \tag{2.15}$$

2.4.5 Comparison of Means and Sums

The bias reveals systematic differences between observed and modeled values. It is defined in this thesis with respect to their means:

$$bias = \bar{P} - \bar{O}. \tag{2.16}$$

The relative error, RE_S [%], is used in this thesis to measure the difference between the cumulative sums of a measured and modeled time series relative to the measured sums

$$RE_S = \frac{\sum_i^N P_i - \sum_i^N O_i}{\sum_i^N O_i} \cdot 100. \tag{2.17}$$

Comparably, RE_M [%] is referred to the means of the modeled and measured time series:

$$RE_M = \frac{\bar{P}_i - \bar{O}_i}{\bar{O}_i} \cdot 100. \tag{2.18}$$

3

Analysis of the
MODIS Leaf Area Index Product

3.1 Introduction

The leaf area index (LAI) and the closely related fraction of absorbed photosynthetically active radiation (FPAR) are indispensable as biophysical variables for the interpretation of measured carbon, water and energy fluxes and their modeling (Bonan, 1993; Sellers *et al.*, 1997). The dimensionless LAI is defined as one-sided surface area of leaves and needles per ground area (Watson, 1947; Breda, 2003) and thus provides information on the seasonal course of vegetation and characterizes functional and structural attributes of vegetation stands. FPAR indicates the fraction of the incoming solar radiation in the spectral region that can be absorbed by vegetation for photosynthetic processes and depends – amongst others – directly on the leaf area available to absorb radiation (Bonan, 2002). Therefore, LAI and FPAR as crucial biophysical variable are intrinsically tied to the primary productivity of photosynthesis and in this context subject of the following study.

Satellite remote sensing offers a unique opportunity to estimate this basic parameter over large areas in an efficient manner. Global LAI/FPAR products are routinely retrieved from satellite sensors such as MODIS. The widely used MODIS LAI/FPAR product MOD15A2 from Terra and MYD15A2 from Aqua data – introduced in section 2.3.2 – has been provided globally since the year 2000 (Myneni *et al.*, 2002). The product quality varies, mainly due to cloud and snow cover, detector problems and other noise sources as

well as due to retrieval algorithm inadequatenesses under certain boundary conditions (Fang *et al.*, 2008; Xiao *et al.*, 2009).

Nonetheless, the MODIS LAI product is considered as having reached the second validation stage what means that it has been evaluated in several field studies over time periods regarded as sufficiently, and all major biomes have been covered by these validation analysis (Morisette *et al.*, 2002; Yang *et al.*, 2006a). Over all these biomes, the LAI shows an accuracy of 0.66 (RMSE) according to the validation website of the MODIS Land Team (Myneni, 2009). Still, several problems persist even in the actual Collection 5 and the user is confronted with the problem of how to use the product data. This situation is reflected by the different approaches using MODIS LAI data and comparing it with other LAI data sources as well as the ongoing development of smoothing techniques and alternative algorithms (Houborg *et al.*, 2007; Pisek & Chen, 2007; Borak & Jasinski, 2009). Consequently, when combining MODIS LAI subsets with FLUXNET data (see section 2.2.1) the user has to make several decisions regarding the data post-processing: Which quality filter shall be applied? Should neighboring pixels be considered and if yes, how wide shall the data window be drawn in the 7x7 1-km^2 pixel subsets? Shall only Terra or Aqua data be used or shall they be combined? These questions have to be answered before methods are picked to fill missing data and smooth the time series if necessary.

Usually, MOD15A2, hence the more extensively validated Terra product, is downloaded by researchers, but Yang *et al.* (2006b) showed that there are no significant differences between the two sensors at the continental and tile scale and concluded that the combination of them helps to enlarge the number of high quality retrievals; on a pixel by pixel basis, however, they found that large differences can occur. The problem of choosing an appropriate LAI quality filter at FLUXNET sites has been tackled differently in MODIS LAI studies: All variations are used from selecting only those values signed as "best" to no filtering: For example, Wang *et al.* (2005a) preferred data flagged as "best", Kanniah *et al.* (2009) and Verger *et al.* (2008) recognized all "good" data, Borak & Jasinski (2009) screened low quality data not marked at least as "good" and those contaminated with clouds, Zhang *et al.* (2006) filtered out cloud contaminated pixels, and Leuning *et al.* (2005) averaged over all values. Yang *et al.* (2006b) advise to use back-up retrievals with care. Likewise, the question whether surrounding pixels should be taken into account and if yes how many pixels around the tower should be used has been answered differently by scientists: For example, only the central tower pixel is considered by Wang *et al.* (2005a); in other studies, 3x3, 5x5 or 7x7 mean or median averaging windows are drawn around the tower pixel with the purpose of better representing the flux tower footprint and reducing geolocation and pixel-shift errors with the additional advantage of reducing the number of missing data points (Xiao *et al.*, 2008b; Kang *et al.*, 2005; Leuning *et al.*, 2005). The latter problem of spatial and temporal discontinuity is an inherent characteristic of remote sensing data in the visible and near infrared due to the sensitivity of the radiance to cloud cover, snow and increased aerosol loading. Several studies have recently dealt with this

problem and developed spatial-temporal smoothing, filtering and gap-filling techniques (Gu *et al.*, 2006; Fang *et al.*, 2008; Gao *et al.*, 2008; Borak & Jasinski, 2009). These approaches make, together with complex statistical and modeling techniques, again usage of quality flag and land class filtering as well as spatial averaging around a center pixel. However, the analysis of effects of these basic methods and their modification have been somewhat neglected in the literature.

In what follows, the consequences of some fundamental post-processing methods on the magnitude, temporal variability and consistency of MODIS LAI values is explicitly analyzed from an end user perspective. This is done exemplarily at six FLUXNET sites of different vegetation classes. Specifically, the impact of several quality filters is illustrated, the consequences of averaging over different window sizes around the tower pixel are assessed, Aqua and Terra subsets are compared and the combination of them is considered. Finally, the consequences on the output of a simple evapotranspiration model ("BUCKUP") resulting from different LAI input data sets are evaluated. Following QC cases are considered: unfiltered, cloudless, "good" values, "good" values without clouds, "best" values, "best" values without clouds, back-up retrieval and back-up retrieval without clouds. Terra began to provide MODIS data from the beginning of 2000, Aqua followed in April 2002. For this analysis, data from the sensors on both satellites are recognized from the latter date when both MODIS products are available until September 2009.

3.2   Study Sites

Six FLUXNET test sites in North America are selected (Table 3.1): Duke Forest, North Carolina, USA; Howland, Maine, USA; Harvard Forest, Massachusetts, USA; Morgan Monroe State Forest (MMSF), Indiana, USA; Vaira Ranch, California, USA; and Lethbridge, Alberta, Canada. These sites have been arbitrarily chosen between those featuring comprehensive multi-annual time-series with a good temporal coverage and being often used in the FLUXNET community (e.g. Falge *et al.* (2001b); Yuan *et al.* (2007)).

Duke Forest is an evergreen needleleaf forest plantation of about 25 years in a humid climate with mild winters and hot summers. Its overstorey almost solely consists of loblolly pine (Pinus taeda L.) with a mean height of about 20 m; 26 different hardwood species form the rich understorey (McCarthy *et al.*, 2007). In December of 2002 an ice storm damaged one third of the trees (McCarthy *et al.*, 2006). In the temperate continental Howland Forest, hardwoods such as red maple and paper birch (Betula papyrifera) occur among the still dominating coniferous species as red spruce (Picea rubens), eastern hemlock (Tsuga canadensis), balsam fir (Abies balsamea), and white pine (Pinus strobus). The median stand-age of this boreal-northern hardwood transitional forest is about 100 years. The canopy height is estimated as 20 m (Hollinger *et al.*, 1999). Harvard Forest

Table 3.1: Characteristics of the study sites according to FLUXNET and LocClim, the FAO Local Climate Estimator (Grieser *et al.*, 2006). Geographic location: latitude (lat), west longitude (long). h: elevation. Vegetation types (VT): evergreen needleleaf forest (ENF), deciduous broadleaf forest (DBF), grass (G). Koeppen-Geiger climate classes: C: temperate, D: continental, f: fully humid, s: summer dry, a: hot summer, b: warm summer. Temperature (T) and precipitation (P) are annual mean averages.

site	lat, long [°]	h [m]	VT	climate	T [°C]	P [mm]
Duke	35.98, 79.09	163	ENF	Cfa	14.4	1169
Howland	45.20, 68.74	60	MF/ENF	Dfb	5.3	1070
Harvard	42.53, 72.17	340	MF/DBF	Dfb	6.6	1071
MMSF	39.32, 86.41	275	DBF	Dfa	10.9	1032
Vaira	38.41, 120.95	129	G	Csa	15.9	544
Lethbridge	49.71, 112.94	960	G	Dfb	5.4	398

is mainly composed by deciduous broadleaf trees with a stand-age of about 80 years; however, more and more needleleaf species are present, so it is meanwhile often referred to as mixed forest. The main species are red oak (Quercus rubra), red maple (Acer rubrum), black birch (Betula lenta), white pine (Pinus strobus), hemlock (Tsuga canadensis), white oak (Quercus alba), black oak (Quercus velutina), and hickory (Carya ovata) with a mean height of 23 m. The climate is characterized by cold winters and warm summers (Goulden *et al.*, 2006). MMSF is classified as deciduous broadleaf forest dominantly composed by sugar maple (Acer saccharum), tulip poplar (Liriodendron tulipifera), sassafras (Sassafras albidium), white oak (Quercus alba), black oak (Quercus nigra) with a mean age of almost 80 years. Cold winters and hot summers form a temperate continental climate (Schmid *et al.*, 2000).

The two selected grassland sites represent contrasting ecosystems: Vaira Ranch is located in the lower foothills of the Sierra Nevada Mountains on managed farmland. The C3 annual grasses grow in a Mediterranean climate with a distinct drought season in summer (Xu & Baldocchi, 2004). Lethbridge, however, east of the Canadian Rocky Mountains, is characterized by a humid climate with cold winters and warm summers. The short prairie consists of C3 and C4 species (Flanagan *et al.*, 2002).

For all sites but Lethbridge, field measured LAI data were available via the website of AmeriFlux, the regional subdivision of FLUXNET. An overview on the time range with available measurements from 2002 to 2009, the number of measurements as well as remarks to the measurements can be found in Table 3.1.

Table 3.2: Available field measurements specified by the covered time range and the total number of sample dates within the whole time period along with information on the measurements.

site	time range	number	measurement
Duke	2002-2005	8	Multiple techniques assimilated
Howland	2006	7	LAI-2000; 200 m transect every 10 m
MMSF	2002-2006	95	LAI-2000
Harvard	2005-2008	41	33 samples per date
Vaira Ranch	2002-2006	48	Li-Cor 3100, 4 samples per date

3.3 Analysis

3.3.1 Overview

In what follows, the frequency of occurrence of the various quality criteria and their effect upon the seasonal evolution and statistical characteristics of the analyzed LAI time series are investigated. The various data sets resulting from the application of quality criteria and the spatial aggregation from 3x3 to 7x7 pixels or mere usage of the tower pixel, respectively, are thereafter analyzed and presented. Subsequently, the Terra and Aqua data sets are directly compared and combined to a single time series. Finally, it is shown of which magnitude the differences of a simple evapotranspiration model are when using various MODIS LAI data sets as model input.

3.3.2 Quality Criteria

Occurrences of QC Classes

To characterize the quality sets, the proportion of the various quality classes (unfiltered, "good", "best" and back-up) with and without clouds as percentage of all retrievals is quantified as a first step. Figure 3.1 uncovers that significant more LAI values were retrieved under cloud-free than under cloudy conditions according to the QC bits (59 - 94 %). Aqua data tend to have a little more cloudy pixels than Terra data (up to 6%). Cloudy values have a higher percentage of back-up values. 68 to 100 % of all retrievals resulted in "good" results via the main algorithm with the maximum number at Duke Forest and Vaira; at the latter site, the back-up algorithm has barely had to be triggered. Most "good" values are associated with a clear sky (76 - 99 % of all "good" values). Vice versa, the percentage of back-up values increases for cloudy pixels up to 53 % The "best" retrievals make up 62 to 82 % of the "good" set at the forest sites and almost all "good" retrievals are "best" at grassland sites, so barely no saturation occurred. Overall, about 50 to 64 % of all retrievals achieved a "best" result, with the exception of Vaira, where almost all LAI values are rated as "best". Most "good" values with clouds belong to the "best"

quality set with up to 98 % at forest sites. In contrast, the percentage of "best" retrievals amount to 62 - 82 % of the cloud-free "good" retrievals.

Figure 3.1: Quality sets of MODIS LAI data used in this study. The number are percentages of the respective superordinate quality set for Terra, Aqua and the combined time series explained below (listed in this order in %). Cloudy and cloudless values are subsets of the whole quantity of LAI values. These sets characterized by the cloud state have subsets of back-up and "good" quality sets. The sets of "good" retrievals have a subset of "best" values. The percentages refer to the directly superordinate set. This means i.e. for the LAI Terra data at Duke Forest, there are 6 % cloudy and 94 % cloudless retrievals. 95 % of the cloudless data are classified as "good", only 5 % have been retrieved by the back-up algorithm; 62 % of the "good" main algorithm retrievals achieve even a "best" result.

Seasonal Evolution

As a second step to assess the impact of quality criteria on LAI time series, a visual inspection is carried out. For this purpose, Figure 3.2 plots the LAI-time series of all pixels for the various quality sets in following order of increasing filtering: all retrievals, "good" retrievals, "good" retrievals without clouds, "best" retrievals, and "best" retrievals without clouds. Earlier plotted values belonging to the "lower", less stringent quality class are therefore covered by the dots of higher classes; values which are filtered out remain visible. What catches the eye first is a division of LAI values into two domains: "best" values make up the LAI values from 0 to about 4 m^2 m^{-2} and "good" values without the "best" are located above up to the maximum LAI. Even in summer, the LAI values are scattered almost throughout the whole range of LAI values. Unfiltered values which are not overlapped by the higher quality values, hence back-up values, appear in the whole LAI range, but tend to agglomerate in winter. Often they appear as a vertical sequence of dots. These can reach values that are lower or higher than the average at the considered

time step. But beside the very low outliers in winter they form a time series that seems to be not different from higher quality data. In summer the back-up retrievals at high LAIs even tend to be not as noisy as the values flagged as "good" and adopt values around the mean of "good" values. Values from cloudy pixels, back-up retrievals and values in the "best" set accumulate frequently in clusters in the lower LAI range.

The phenomenon of segmentation of the "good" values in the saturation domain above 4 m^2 m^{-2} and "best" values at forest sites beneath it becomes even more apparent in the mean time series averaged over all years (Figure 3.4): Whereas they are similar in winter, they diverge strongly in summer. The time series build from the "best" values reach only about 60 % of the summer level of the time series resulting from the "good" values. Since the set of "good" comprises the "best", the multi-annual mean of the "good" values without the "best" retrievals in summer (averaged at the particular days of year) is even higher. This time series of the "good" without the "best" in summer is plotted in Figure 3.3 and exhibits significant better seasonal dynamics than the unfiltered time series or the "best" time series as shown in Figure 3.2 in blue color. It is generated in the following way: As soon as the median of the 7x7 LAI values gets greater than 4 in spring, all values are deleted which do not belong to the saturation domain ("best"). This approach is applied for the whole summer until the median of the LAI gets smaller than 4 again. This value is chosen as a limit because the non-saturation frequency increases drastically according to the product's theoretical basis document (Knyazikhin *et al.*, 1999). And indeed, less than 2 %, in most cases even less than 1% of the "best" values are greater than 4.

Statistical Characteristics

In statistical terms, the single quality sets show differences, too. The means of the quality sets can be assigned to three groups of similar set means: Unfiltered values, retrievals without clouds, "good" and "good" values without clouds form a group of means, "best" retrievals and those with clouds build another group with lower means, and back-up values, finally, present the third group with higher means with respect to the other groups. Since there are almost no "good" values with saturation at grassland sites, hence "good" values are also rated as "best", only two groups can be identified: unfiltered and main algorithm values as well as back-up values. At the grasslands there is no tendency regarding the difference between the group means. Within all groups, retrievals from cloud-free pixels have a higher mean than their unfiltered counterparts, and a even higher mean than the cloud affected retrievals. The differences between the quality set means reach values in the range of the LAI magnitude of the unfiltered mean; at MMSF the difference between the back-up values and the "best" amounts to nearly 4 LAI units. The medians of the quality sets are lower than the means, only the back-up medians are higher in a few cases. The medians have the same characteristics as the means, but show with up to 5 LAI units even higher differences between the quality sets. The analysis of the
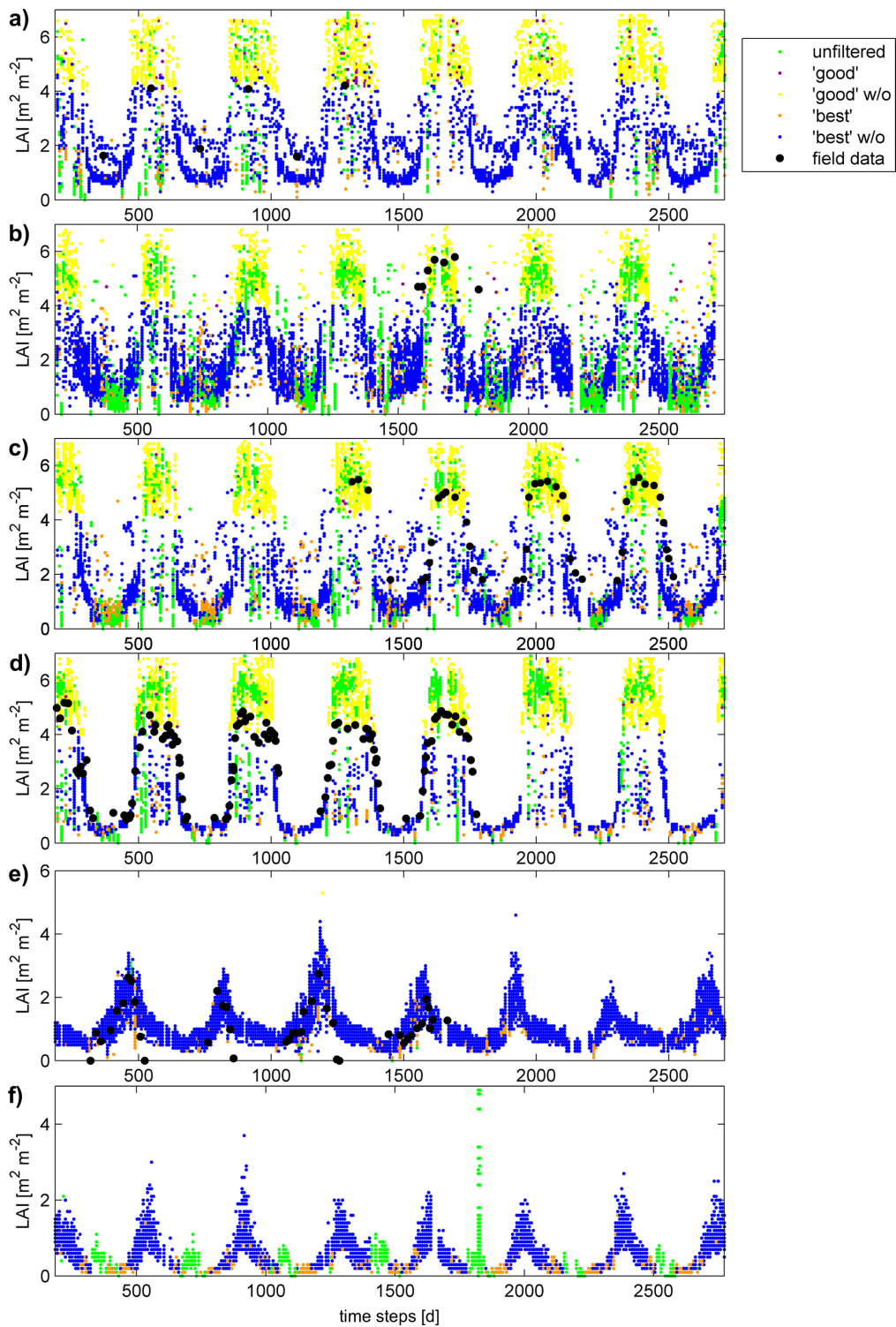
Figure 3.2: LAI time series of the Aqua data sets at Duke (a), Howland (b), Harvard (c), MMSF (d), Vaira Ranch (e) and Lethbridge (f) showing the various quality criteria classes the LAI values belong to. These have been plotted in the order unfiltered, "good", "good" without clouds, "best", "best" without clouds. Dots plotted later cover previous plotted dots. LAI field data are plotted as a rough reference.

quality sets' variances draws a similar picture: lower variances of "best" and cloudy values and higher for back-up retrievals compared to the unfiltered and "good" retrievals. The average variance of time series without clouds is lower than that of the unfiltered time series. The differences between the temporal variance of each pixel of the 7x7 pixels when compared to each other are significant larger for time steps with clouds than without cloud coverage, so the temporal variances are much more unsteady.
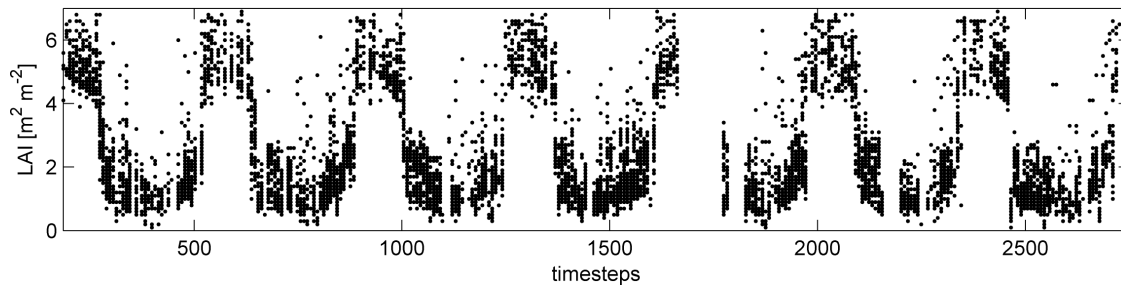
Figure 3.3: The set of "good" values without the "best" values in summer. This time series of the Terra sensor at Howland is typical for the analyzed forest sites.

The LAI frequency distributions of the quality sets provide an explanation for the differing means and medians as Figure 3.5 demonstrates: The unfiltered values show a frequency distribution with two peaks around 1 as well as 5 m^2 m^{-2}. The "good" LAI values form a similar bimodal frequency distribution pattern with somewhat more distinct maxima at the most sites. Likewise, the frequency distribution of the back-up retrievals has two peaks, but these are shifted towards the upper and lower end of the LAI range. The "best" values, however, exhibit an uni-modal distribution with a peak between 0 and 1 m^2 m^{-2}. Most quality sets free of cloud contamination show a smaller occurrence of lower values and a higher occurrence of higher values; this shift to higher values is most distinctive for the back-up sets.

With the non-parametric Kolmogorov-Smirnov-test all quality sets are compared mutually with each other with respect to their frequency distributions. The results indicate that almost all quality sets at the forest sites can be assumed to have been drawn from different distributions with two exceptions: first, the different cloudy data sets of the main algorithm – which is simply a consequence of the fact that the most cloudy data from the "good" set are classified as "best" since they are in the lower LAI domain and hence not subject to saturation – and second, the "good" values without clouds and the data set without non-saturation values in summer. At Duke Forest, the "good" and the unfiltered set without clouds additionally comply with the null hypothesis that they are from the same distribution and the same is true for the "best" sets with and without clouds. At the grassland sites, unfiltered quality sets and those of the main algorithm are found to be similar when they have the same cloud status; this similarity is not surprising since the classes share almost the same elements. The additionally executed Kruskal-Wallis hypothesis test, a one-way analysis of variance by ranks, draws a similar picture. In
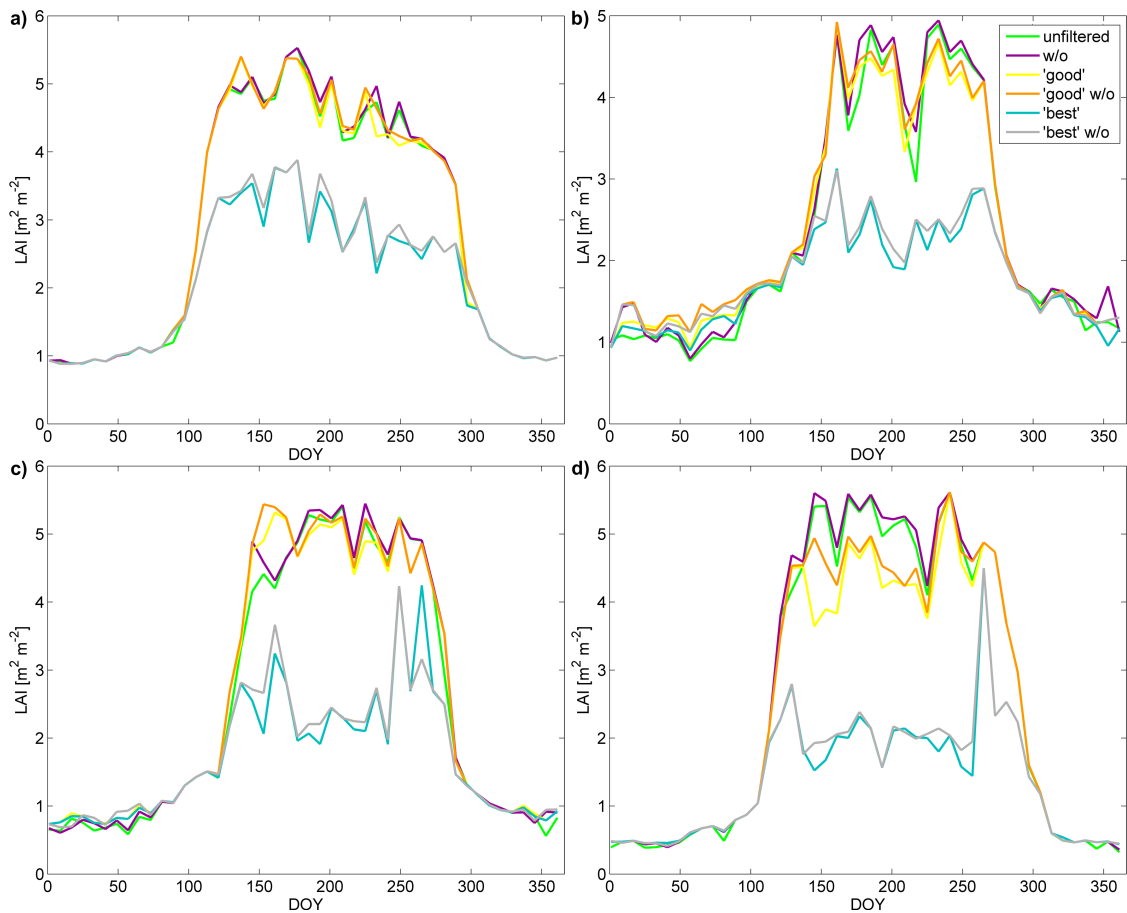
Figure 3.4: Multi-year average of the various quality classes in the Terra data sets at Duke (a), Howland (b), Harvard (c), MMSF (d).
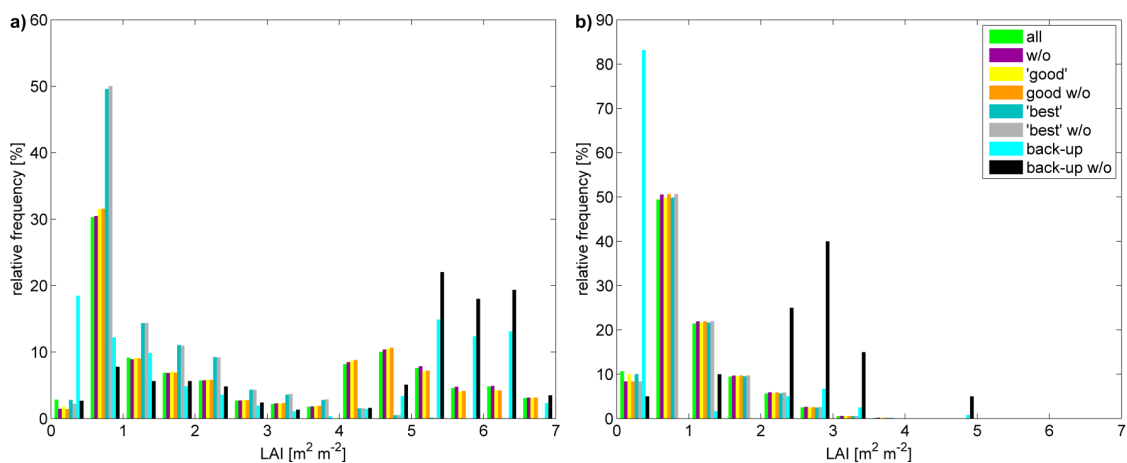
Figure 3.5: Relative frequency distributions of the quality sets using the example of Terra data at Duke Forest and Vaira Ranch. Frequencies of the sets are relative to the number of all retrievals in the respective sets.

summary, the two applied hypothesis tests show the tendency of the "best" sets being different from the other sets at the forest sites and the cloud state playing a significant role when comparing data sets.

Comparison with Field Data

A comparison with the available LAI field data is done as next step in comparing the quality sets. Since the LAI is not measured with the same methods at the test sites and the field measurement and MODIS scales are different the comparison is executed to show trends and not to validate the MODIS LAI data explicitly. The comparison shall serve as an additional piece of the puzzle of hints regarding the best usage of MODIS LAI data. A visual comparison in Figure 3.2 reveals that the MODIS LAI follows quite good the seasonal LAI dynamics. In the critical summer season, the field measured LAI values from Harvard and Howland lie in the range spanned by the "good" MODIS values, at Duke Forest and MMSF, however, the field data are located at the lower end of range of "good" values. In winter, however, the MODIS LAI values rather underestimate the measured data at the needle leaf Duke Forest and tend to overestimate it at the broadleaf MMSF – a phenomenon which has often been reported before (Chen *et al.*, 2005).

A quantitative comparison at the three sites with the longest field measured time series (Harvard, MMSF, Vaira) is carried out by applying the RMSE, the bias (as difference between the means) as well as the Nash-Sutcliffe efficiency (NS). NS, RMSE and bias show a comparable behavior of the unfiltered and "good" values with respect to the field data at the forest sites (Figure 3.6a). This is true for all spatial configurations from the exclusive use of the tower pixel to the exploitation of all 7x7 pixels with a similar land class available in the subset whereas it appears that the more pixels are taken into account the better is the match of field and remote sensing data. If the "good" values subject to saturation are removed in summer comparison statistics improve a little bit. At MMSF and Duke, however, the "good" values in the saturation domain are too high compared to the field data; the consequence is that the unfiltered LAI data compare best to the field data, since the too low "best" values and the too high "good" values in the saturation domain compensate each other. Taking only the "best" values into account, however, the goodness of fit between field and MODIS LAI data deteriorates and results in a greater bias and RMSE as well as a lower NS at the forest sites. The retrievals without clouds tend to be somewhat more concordant with the field measurements at all sites; this is especially apparent at Vaira Ranch (Figure 3.6b).

3.3.3 Spatial Aggregation

Figure 3.6 discussed above has shown that the spatial aggregation by taking the mean of the considered pixel window has an influence on the comparability with the field
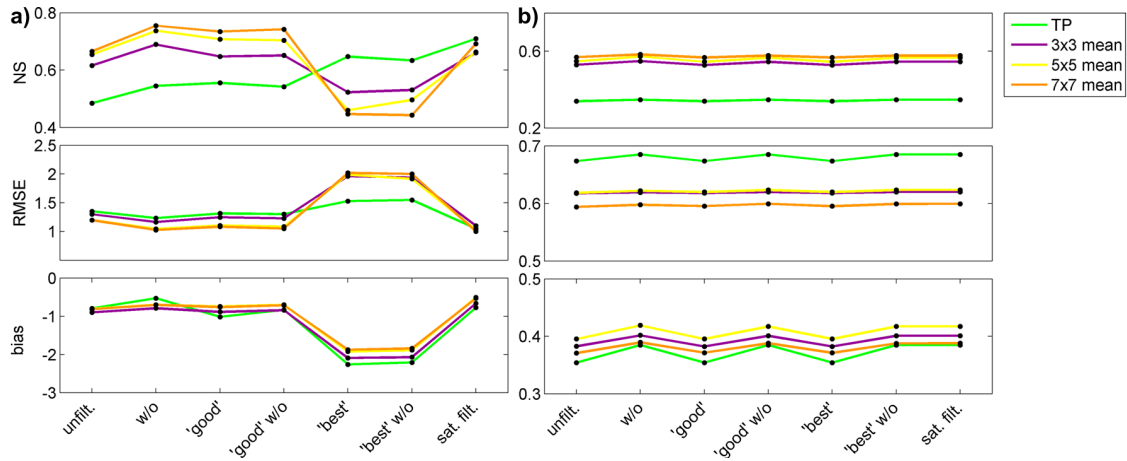
Figure 3.6: Comparison of field measured data with the MODIS LAI values for the various quality sets at Harvard (a) and Vaira Ranch (b). The tower pixel (TP) alone as well as the 3x3, 5x5, 7x7 pixels around the tower pixel are compared with the field measurements.

measurements: The more pixels around the tower are used, the better the field and MODIS data are comparable tendentially. Do the spatial aggregations also have an positive effect on the temporal consistency of the resulting time series? The temporal evolution during a typical year without extreme events as storms, drought or severe fires is supposed to be relatively smooth; MODIS LAI products, however, tend to show a higher temporal noise than expected in a phenologically sound sense (Verger *et al.*, 2008; Tan *et al.*, 2005; Pisek & Chen, 2007).

A comparison of the variances σ for the tower pixel and spatially aggregated time series shows indeed a decrease with an increasing pixel window size in the most cases, whereas at the same time, the means remain very similar for the aggregations, except for Howland Forest. An inspection of the Howland time series reveals lower values in winter for the more extensive aggregations, which is a consequence of a changing tree composition in favor of deciduous species; this assumption is supported by the MODIS land class map.

But to evaluate the smoothness of the time series, the magnitude of the up-and-downs at consecutive time steps is even more important, for what reason the smoothness measure δ is consulted. This method has already been used by Verger *et al.* (2008) to compare the temporal consistency of MODIS and CYCLOPES LAI time series. δ calculates the difference for each time series element to the linearly interpolating line between its two adjacent neighbors:

$$\delta = (0.5 \cdot (LAI(t + \Delta t) + LAI(t - \Delta t))) - LAI(t) \tag{3.1}$$

A single drop of a time series element would for example lead to a high δ value at this time point, a sudden jump to a high negative value. Histograms of the relative frequencies of all occurring δ-values show indeed a change of the δ-distributions for the different spatial aggregations towards the low δ-bins for the aggregations compared to the tower

pixels' δ-values: The greater the aggregation, the steeper is the decrease and the earlier the frequencies approach zero occurrences towards larger δ-bins (Figure 3.7a).

As second method to measure the smoothness and noise of the LAI time series, frequency analysis is applied and realized with the Matlab® function "spa", which calculates the frequency response and power spectrum of a signal using Fourier analysis. A smooth, periodic time series would result in a power spectrum with a sharp peak at low frequencies and low values at higher frequencies. A highly noisy time series would result in a large horizontal part in the power spectrum. The application of this method to the various aggregation show a faster drop of the power spectrum towards high frequencies, the larger the aggregation window around the tower pixel is. Figure 3.7b shows this behavior for Lethbridge. The aggregation not surprisingly tends to blur the difference between the quality sets (Figure 3.7b); but still, the cloud filtered time series has a lower power spectrum at high frequencies and the main algorithm time series without clouds tends to show even more the characteristics of a slightly smoother time series. Using only the tower pixel makes it more important to understand and apply the quality flags because the difference between them gets larger (Figure 3.7d). The frequency analysis also shows that the empirical back-up values are not necessarily more noisy than the main algorithm values.

3.3.4   Comparison of Terra and Aqua Data

LAI Retrievals and Quality Classes

As a first step of the Terra-Aqua-analysis, the questions is explored of whether it is advantageous to combine MOD15A2 with MYD15A2. The number of overall retrievals would clearly profit from the combination of Terra and Aqua. In 6 to 7.4 % of all pixels and retrieval dates one value lacks in the Terra or Aqua data set but the other is available. In all but a few cases Aqua delivers 2 to 4 % more values than Terra, which is mainly a consequence of a longer gap in the Terra time series in 2006. The proportions of the quality classes are similar in the two data sets. The greatest differences occur at Duke and Lethbridge with the Aqua data sets having 3 % or almost 7 %, respectively, less main algorithm retrievals without clouds than Terra. In at least 4 % of all retrieval dates, the pixel of one data set is cloudy, the other is not. The Canadian grassland Lethbridge experiences the greatest cloud coverage of about 40 %; the combination of the two sensors could reduce this amount to 26 %.

Statistical Characteristics

But are the two data sets really comparable and hence combinable? Analyzing the statistical characteristics of the two data sets, it becomes evident that the overall means of
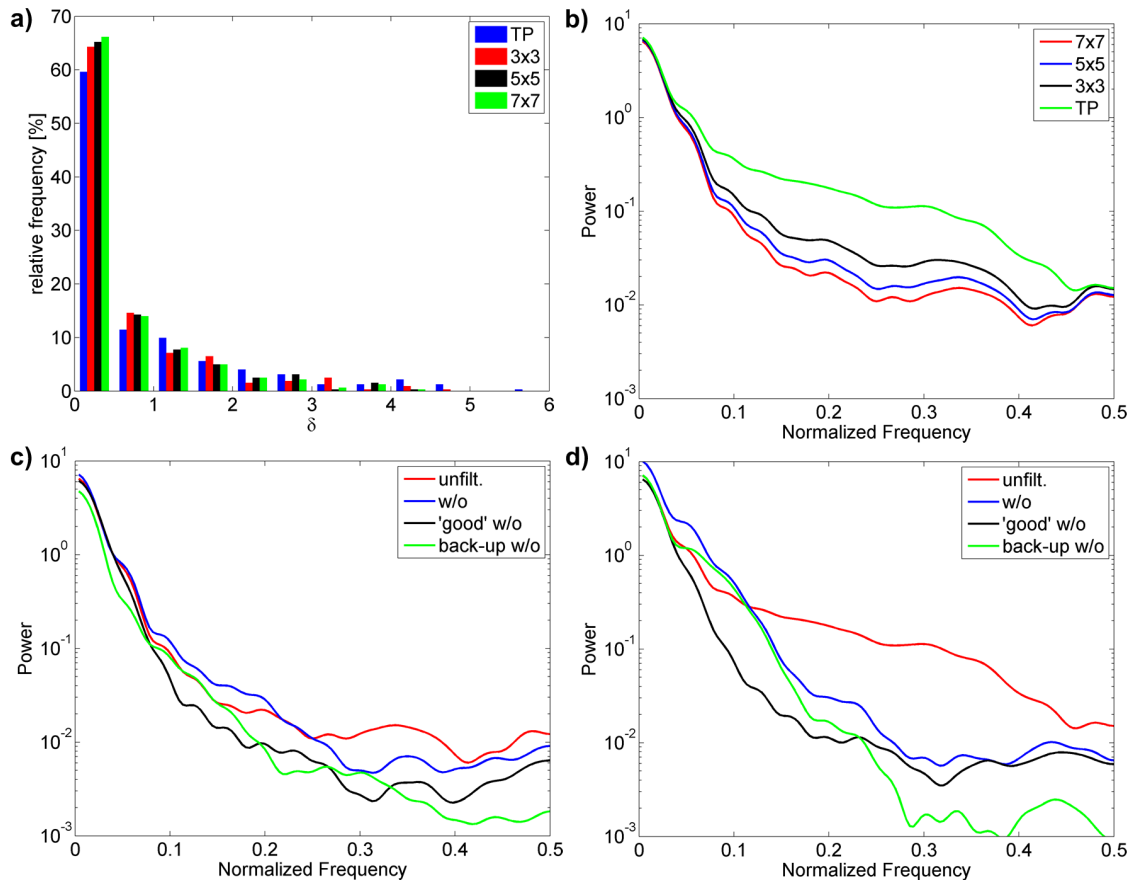
Figure 3.7: Histogram of relative frequencies of δ for the spatial aggregations, namely for the time series of the tower pixel (TP) up to the averaged (median) 7x7 pixel window, exemplarily shown for the unfiltered MMSF Aqua data set (a). Power spectra for the aggregated Lethbridge data sets (b) and for the various quality classes for the 7x7 aggregation (c) as well as the tower pixel (d).

Terra are slightly higher than those of the Aqua data sets (0.01 to 0.15 m^2 m^{-2}), only at Duke, the bias constitutes even nearly 0.3. The medians are equal in all cases, but again at Duke, the median differs even by 0.4. These differences get evident in the multi-year averaged LAI-dynamics (Figure 3.8) as well as in the frequency distributions (Figure 3.9): Average Terra values in summer are noticeably higher than the Aqua multi-year means, and the relative frequencies of Terra retrievals in the second half of the LAI range exceed the relative frequencies of Aqua retrievals. Vice versa, the Aqua data sets have relatively more LAI values in the lower LAI range. This is true for both forests and grasslands, even though the phenomenon is less distinctive at grassland sites.

The residuals of the two sets calculated as Terra minus Aqua values range up to 6 m^2 m^{-2} in the forest data sets and up to 3 at the grassland data sets. Their frequency distributions diverge from the normal distribution with a positive kurtosis and skewness rightwards, so the deviations do not seem to be random but systematic. The Kruskal-Wallis-test denies the null hypothesis that the two data sets are statistically similar and suggests that they are drawn from different populations. For higher quality data, the null hypothesis
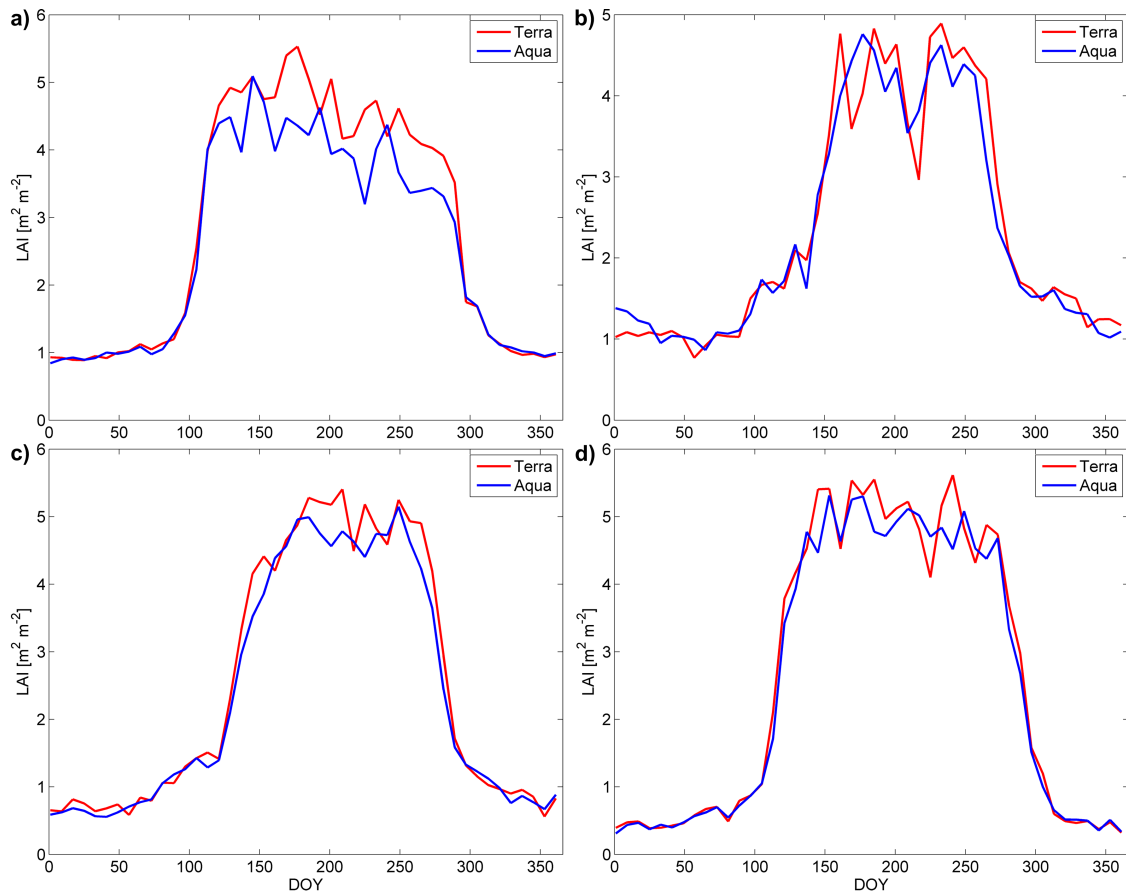
Figure 3.8: Multi-year average LAI at a) Duke, b) Howland, c) Harvard, d) MMSF showing the long-term difference between the MODIS sensors onboard Terra and Aqua.

gets somewhat more probable. According to the RMSE, the differences range from values of 0.15 (Lethbridge) to 1.35 (Duke Forest).
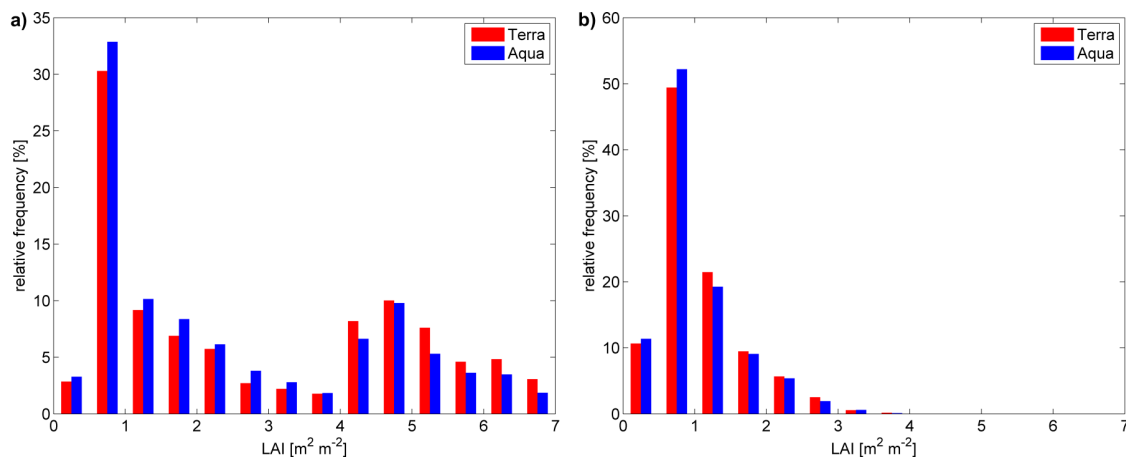
Figure 3.9: Relative frequency of LAI values in bins of 0.5 LAI units as percentage of all retrievals in the subsets of a) Duke and b) Vaira. Relative frequency distributions at the other study sites are similar.

Plotting the LAI values from the two data sets directly against each other (Figure 3.10) and counting the occurrences in 0.1 m^2 m^{-2} wide bins reveals different patterns at the forest and the grassland sites: In both biomes, the highest frequencies of data points occur around the 1:1 line of a perfect match between the two data sets. At grasslands, the data points in the sensor space form an elliptic-like shape with some few scattered points around it. At the four forest sites, however, the scatter pattern clusters in four parts: One part is similar to an ellipse around the 1:1 line at LAI-values up to 4 and contains the most values. After a gap with no or very few values, this part is continued at along the 1:1 line at higher LAI-values by a second, wider cluster. Two additional clusters are located in the left upper and right lower quadrant of the scatter plot, namely at low Terra values and high Aqua values and vice versa. In summary, the data of Terra and Aqua are either very similar or very different, whereas the first possibility has a significantly higher probability, especially at low LAI values. Taking only values flagged as "good" and without clouds into account, the clusters somewhat narrow, especially those two clusters indicating high differences get smaller and less dense.
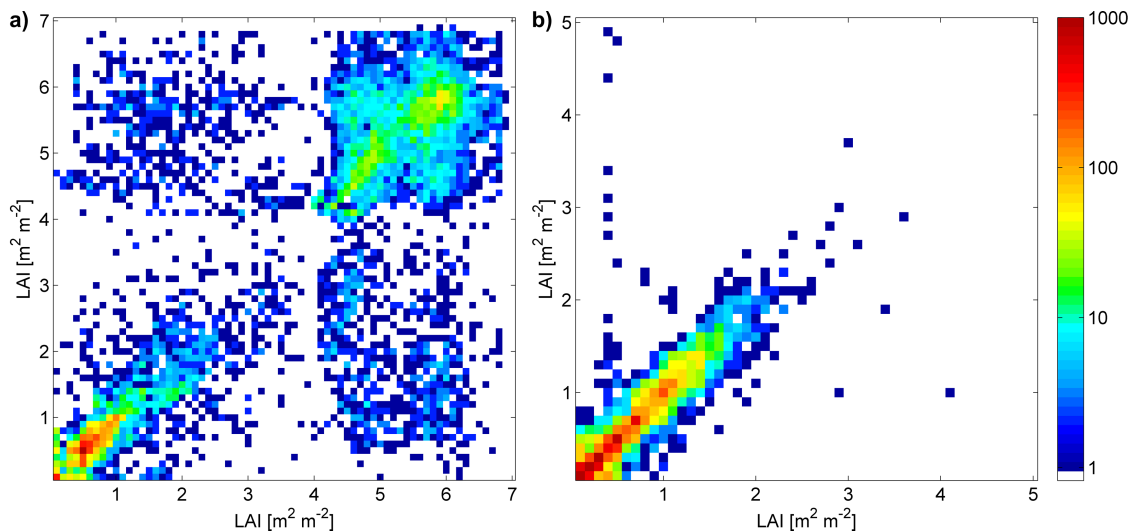
Figure 3.10: Bivariate histograms of Terra (x-axes) and Aqua (y-axes) LAI values, exemplarily for a) MMSF and b) Lethbridge. White presents no occurrence, dark blue shows frequencies of 1, dark red frequencies of 1000 cases and more.

Combination

Despite the clues for systematic but compared to other criteria rather small differences between the Terra and Aqua data sets at these test sites, a composition scheme is applied in what follows. To generate the combined LAI values this study follows the "official" product combination approach (Yang *et al.*, 2006b) but enhances it with the lessons learned in this study regarding the influence of the cloud state. In doing so, the general quality criterion are compared first: If one was retrieved by the main algorithm and the other

not, the "good" value is preferred; if the algorithm path is equal the cloud free retrieval is chosen; if the cloud state is also the same for both sensors, the LAI value with the higher corresponding FPAR value is taken since this value is assumed to be less atmospherically affected.

The advantage of the combined product is clearly the higher data availability as already mentioned above. Due to the preference of "good" retrievals without clouds in the combination procedure, the number of these retrievals is enhanced by up to almost 20 % and the number of time steps, at which at least 45 pixels "good" retrievals without clouds are available, rises by up to 40 % (Figure 3.11). The temporal consistency as described by δ and frequency analysis is merely promoted by the combination procedure. The average temporal variance shows no decline, too. The spatial variance at each time step, however, tends to be moderated by the combination of the two sensors.
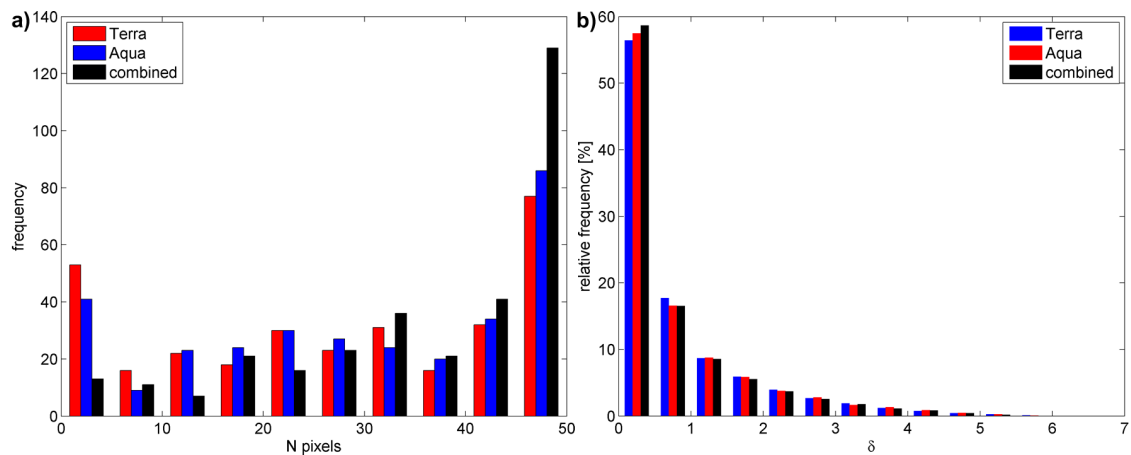
Figure 3.11: The combined product shows an increased frequency of "good" pixels without clouds for the combined product at every time step (a). The temporal consistency analyzed with δ , instead, shows only a small improvement for the combined time series (b). Plots shown for Howland.

3.3.5   Model Sensitivity to LAI

The leaf surfaces represent the exchange medium between the plants and the atmosphere. The LAI as quantity for this exchange surface is consequently a critical variable for all biophysical models describing fluxes of energy, momentum and matter such as the exchange of carbon and water (Bonan, 2002). The index is of particular importance to modelers who want to upscale fluxes measured by eddy-covariance towers (Cleugh *et al.*, 2007) or apply biomass models on a global basis or for agricultural purposes (Demarty *et al.*, 2007). Since biosphere-atmosphere interactions are more and more implemented in global climate and hydrological models, the LAI is meanwhile also a key input variable for these model types (Chase *et al.*, 1996; Andersen *et al.*, 2002).

The Penman-Monteith equation is broadly applied in SVAT (soil-vegetation-atmosphere-transfer) models involving an estimation of the evapotranspiration (Monteith, 1981) and is often executed depending on LAI (Lindroth, 1993; Leuning *et al.*, 2008). As final part of this analysis, the potential consequences of using various LAI time series for the calculation of latent heat fluxes are tested based on various quality criteria . To do so, the simple bucket type SVAT-scheme "BUCKUP" (Franks *et al.*, 1997; Schulz & Beven, 2003) based on the Penman-Monteith equation is chosen and executed for the 2004 data of MMSF in a slightly modified mode:

$$\lambda E = \frac{\Delta R_N + (\rho_a c_p \delta q)/r_a}{\Delta + \gamma_P \left[1 + (r_s/r_a)\right]} \tag{3.2}$$

with λE [W m^{-2}] being the latent heat flux; Δ [kg kg^{-1} K^{-1}] the rate of change of saturated specific humidity with temperature; R_n [W m^{-2}] the net radiation; ρ_a [kg m^{-3}] the dry air density; c_p [J kg^{-1} K^{-1}] the specific heat of air; δq [kg kg^{-1}] is the specific humidity deficit; γ_P [kg kg^{-1} K^{-1}] the psychrometric constant; and finally the surface and aerodynamic resistances r_s and r_a in [s m^{-1}]. While root zone water storage mainly controls water availability , the canopy and upper soil interception store, the root zone store as well as a variable ground water table act as evapotranspiration sources. Rainfall fills the interception store up to its limit. The interception overflow is assigned to the root zone store whose excess water is transferred to the groundwater table with a soil dependent time delay. The surface resistance r_s is updated every time step according to the actual water availability of the three stores. In this study, BUCKUP is extended by considering LAI seasonality: r_s is assumed to depend inversely proportional to the LAI (Bonan, 2002). Furthermore, the LAI is supposed to alter r_a by means of the displacement height, D, as well as the roughness length, R_L according to following empirical relations (Raupach, 1994):

$$X = 0.2 * LAI \tag{3.3}$$

$$D = H_V * (log(1 + \sqrt[6]{X}) + 0.03 * log(1 + X^6)) \tag{3.4}$$

$$R_L = H_U + 0.3 * H_V * \sqrt{X} \; for X < 0.2) \tag{3.5}$$

$$R_L = 0.3 * H_V * (1 - D/H_V) for X >= 0.2) \tag{3.6}$$

where H_V is the vegetation height and H_U the understorey height. D and R_L are then used to calculate r_a empirically. Details can be found in Schulz *et al.* (1998).

The model is run with hourly time steps using temperature, precipitation and humidity as meteorological boundary conditions. As LAI time series input, the data from Terra, Aqua as well as the combined data set are used. For the combined product, the various spatial aggregations from the mere use of the tower pixel up to the 7x7 pixels are applied.

Since the model needs temporally continuous LAI data, an interpolation method is needed. At the same time, the time series shall be smoothed, since even the time series from taking the mean of all 7x7 pixels appears not to be physiologically sound. Interpolating and smoothing for model purposes has been proposed by other authors, too (Kang *et al.*, 2005; Gu *et al.*, 2006; Fang *et al.*, 2008; Gao *et al.*, 2008; Borak & Jasinski, 2009); they have used spatial averaging, upper LAI envelope smoothing techniques, adaptive Savitzky-Golay filtering, asymmetric and double logistic Gaussian filtering or an ecosystem curve fitting method based on the MODIS vegetation continuous fields product. In this exercise it is intended to give the noisy MODIS LAI data as much weight as possible but at the same time fulfill the contradictory demand of reducing the noise and the influence of outliers. Therefore, we choose a combined temporal-spatial interpolation scheme with a cubic smoothing spline method as implemented in Matlab® ("csaps") is employed. The smoothing spline f minimizes

$$p_S \sum_{\substack{N \\ w(i)}} |y(:,i) - f(x(i))|^2 + (1 - p_S) \int \omega(t) \left| D^2 f(t) \right|^2 dt \tag{3.7}$$

where x in this case represents the time vector and y the LAI data; i runs from 1 to the maximum number of entries of the temporal data vector, N; dt is a time increment; D^2 denotes the second derivate; ω is a weight function and w a weight vector, which is set to 1 in this study. p_S represents a smoothing parameter; it is chosen by the algorithm depending on the LAI values at each time step. Several data points at the same time steps are recognized as the average. csaps is an implementation of the FORTRAN routine "SMOOTH" (Barry, 1973). Figure 3.12a shows the results of this interpolation and smoothing procedure to daily LAI values for the combined Terra and Aqua data and the various quality data sets.
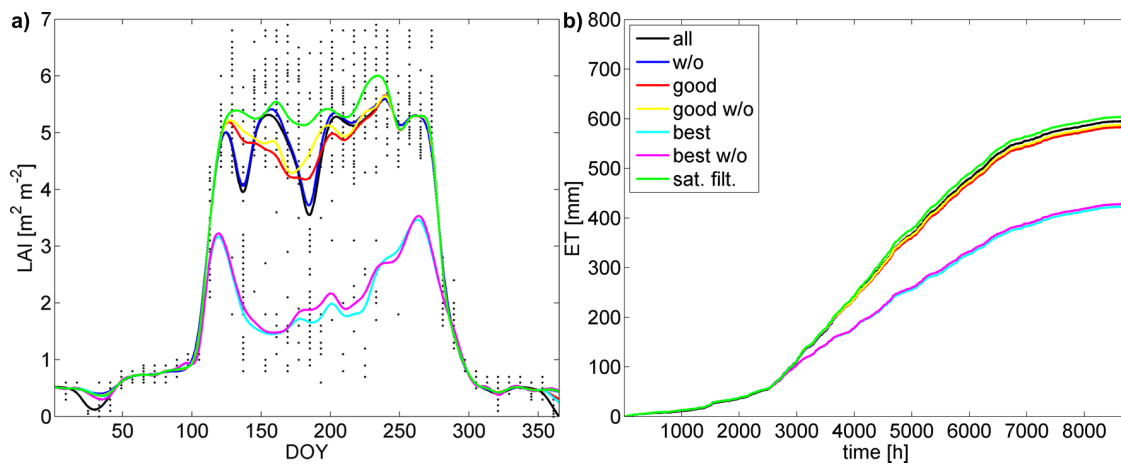
Figure 3.12: (a) MODIS LAI (dots) and interpolated/smoothed LAI time series (lines) of different quality sets at MMSF, 2004. "w/o" means data without clouds, "sat. filt." denotes the data set of "good" values without clouds and without LAI values in summer outside the saturation domain as described above. (b) Cumulative evapotranspiration for the different LAI input data.

The results of the sensitivity analysis illustrated in Figure 3.12b) show that the influence of the sensor choice as well as the quality criteria on the model output is rather small, which is not surprising since the smoothing blurs the differences between the data sets. However there is one drastic exception: The "best" values produce over 30 % lower cumulative amounts of evapotranspiration compared to the measurements and over 25 % compared to the results without using "best" LAI values in summer. This is the case because the model evapotranspiration does not reach the high evapotranspiration values at daytimes in summer due to the much lower LAI values. This behavior will have a large impact on the subsequent prediction of individual components of the water cycle such as groundwater recharge or run off. Also, patterns of surface soil moisture availability will be largely biased and estimates of e.g. net primary production or bio-geochemical processes might be unreliably affected. Fact is, that the LAI time series at temperate forest sites get more unreliable in the sense of physiologically plausible temporal behavior when using the strictest quality filtering "best" and this is true even when exploiting the data sets of both sensors.

3.4   Discussion and Conclusions

MODIS data from Terra and Aqua sensors in combination with ecosystem exchange data collected from FLUXNET have become central to the development, calibration and evaluation of soil-vegetation atmosphere transfer models (SVAT) and their regionalization. MODIS land product subsets provided for long term field measurement sites as from FLUXNET simplify the assimilation of field and remote sensing measurements for practical usage. Despite the frequent application of MODIS LAI subset data in ecosystem models, there is no consensus on their post-processing usage and the employment of the provided QC information. In this study, the consequences of several post-processing user choices for six MODIS LAI subsets at FLUXNET sites (four forests, two grasslands) on the statistical characteristics and temporal behavior of the resulting LAI time series were analyzed. The application of quality assessments as well as differing spatial aggregations from the only use of the tower pixel to the aggregation of all 7x7 pixels have been studied. The resulting various time series have been compared to ground-measured LAI values. The data from both sensors have been compared and their combination has been considered. Finally, the influence of differing post-processing choices on a simple SVAT model has been assessed.

The analysis of the quality criteria showed that most Terra and Aqua LAI values have been retrieved with the main algorithm and without cloud coverage what is an improvement compared to earlier MODIS LAI data collections. This fact has already been shown by Yang *et al.* (2006b) for a provisional collection 5 analysis. The spatial variation within the

7x7 pixels is high and the pixels' time series are often unstable and vary through the whole LAI range from near 0 to over 6 $m^2\ m^{-2}$. Cloud filtered time series and main algorithm time series tend to be smoother than their unfiltered counterparts. Most interesting is however the difference between "best" and "good" values from the main algorithm: "best" values without saturation are significantly lower than main algorithm retrievals in the saturation domain and – by definition – seldom reach for forests realistic summer LAI values greater than 4 $m^2\ m^{-2}$. Cohen *et al.* (2006) mentions this phenomenon for forest sites, too. After the deletion of "best" values in summer, the LAI time series shows – although with a considerable range of dispersion – a physiologically sounder temporal evolution. "best" values in summer compare not as good to field data as the other quality sets. The unfiltered data sets perform well due to compensation effects of too high saturation domain values and too low "best" LAI values. This is especially true for Duke Forest where the MODIS time series exceeds the ground-measured LAI values by over 2 $m^2\ m^{-2}$; one reason for this observation is certainly the land cover classification problem evident at this site (see section 2.3.3) as well as general algorithm problems for coniferous forests (Wang *et al.*, 2005a; Therezien *et al.*, 2007).

Are "best" values in summer at forest sites just an artifact of cloudy conditions, which have not been detected by the algorithm or the consequence of an erratic atmospheric correction? The fact that the most main algorithm retrievals with clouds are also "best" retrievals supports this assumption as well as the fact that cloudy pixels have the tendency to have lower LAI values (Wang *et al.*, 2005a; Yang *et al.*, 2006b). Consequently, the end user has to take care when filtering according to the supposedly most stringent QC, "best"; the semantic meaning of this adjective can be misleading when dealing with pixels covered by forests. Nonetheless, the study showed the benefit of using the provided QC meta data.

The question of spatially aggregating pixel around the tower or not has certainly to be answered from an site specific perspective and depends on the land cover classes involved and their distribution around the tower. If the emphasis is put on reducing the noise and increasing the data basis for interpolation and smoothing algorithms to build a more robust seasonal variation of LAI, as many pixels as the land cover distribution around the tower allows should be selected. The analysis showed that the smoothness, hence the temporal consistency, of the time series is enhanced using more pixels. This is consistent with the findings of Borak & Jasinski (2009): In their data denial exercise to test several interpolation schemes, they used even 9x9 pixels and found it to be as good or even superior compared to 5x5 aggregations with having a higher data coverage at the same time.

The issue of spatial aggregation is tightly connected to three other post-processing options for MODIS Land Product Subsets, namely the selection of pixels according to specific land classes or other criteria as well as the interpolation and smoothing of the time

series. The effects of filtering pixels around the tower on the basis of the MODIS land classification on the resulting LAI time series have not been analyzed in this study; this aspect needs further attention in future work. It will also be worthwhile considering to select pixels with respect to the footprint of the flux tower, which depends mainly on the wind speed and direction. How to interpolate and smooth the LAI time series can certainly be discussed further: as Borak & Jasinski (2009) comments, there is no single interpolation scheme that can serve all purposes and conditions. Further analysis could analyze the benefits of using a weighting factor according to the LAI standard deviation delivered in the MODIS product.

Comparing the Terra and Aqua LAI retrievals at the six test sites, a small bias between the data sets was observed and pixel by pixel comparisons reveal large differences at a small proportion of the data. These observed large discrepancies for a limited number of data values can certainly be attributed to atmospheric conditions which are challenging for the algorithm. The cause for the systematic difference is not as obvious. Yang *et al.* (2006b) also observed a bias in Terra and Aqua reflectances and NDVI values at Harvard Forest. They excluded solar zenith angle and cloud influences and ascribed the discrepancies to erroneous atmospheric corrections, too. It is also known that Aqua suffers of band-to-band misregistrations of its MODIS sensor in the Visible and Infrared (Wang *et al.*, 2007a); it is unclear, however, how this problem affects the LAI retrievals. In spite of the observed bias, Yang *et al.* (2006b) suggested the use of both products. Furthermore, it is argued that the difference between the means of the Terra and Aqua product is small compared to the overall noise of the time series. Consequently, we have taken advantage of both LAI products and combined them as described in section 3.3.4 to further increase the reliability of the LAI time series. The combination of the two data sources increases the percentage of "good" retrievals without cloud coverage by up to 17%. This is consistent with Yang *et al.* (2006b) who found a high quality retrieval increase of 10-20% for woody vegetation. The number of time steps with at least 45 "good" pixels without clouds increases by an amount of up to 40%. However, the temporal smoothness of the LAI time series is not improved (see section 3.3.4).

Also, the sensitivity of a simple SVAT model to the differently post-processed LAI input data has been analyzed. Predicted rates of evapotranspiration confirmed the previous findings: The use of "best" data introduces a large bias of up to 40 % with possibly tremendous consequences on the subsequent prediction of individual components of the water balance. However, these consequences need further investigation and have to be extended to other related research areas.

Merging all analysis results, the conclusion can be drawn that it is advantageous to combine Terra and Aqua to enhance the data coverage and maximize the number of cloud-free main-algorithm retrievals and follow one of two major strategies: to apply no QC filtering and take all pixels from the subset into account to achieve error nullification,

or, alternatively, to make QC adjustments such as deleting back-up retrievals and those from cloudy pixels as well as "best" retrievals belonging to the non-saturation domain when the retrieval can be expected to belong the saturation domain. If the surrounding of the tower allows for taking neighboring pixels into account, a spatial aggregation is recommendable, especially for forests. After these selections and choices an interpolating and smoothing approach such as a spline technique used in this study or a filtering technique as proposed in Gao *et al.* (2008) is recommendable.

A final question with regard to the formulation of a light use efficiency model to be done in this thesis is, if what has been found for the LAI values also holds true for the FPAR values provided in MOD15A2/MYD15A2. A respective analysis showed that this is the fact even if the difference between the "best" retrievals and the values classified as "good" without the latter is somewhat smaller than that of the LAI values. However, the phenomenon is definitively pronounced as Figure 3.13 shows.
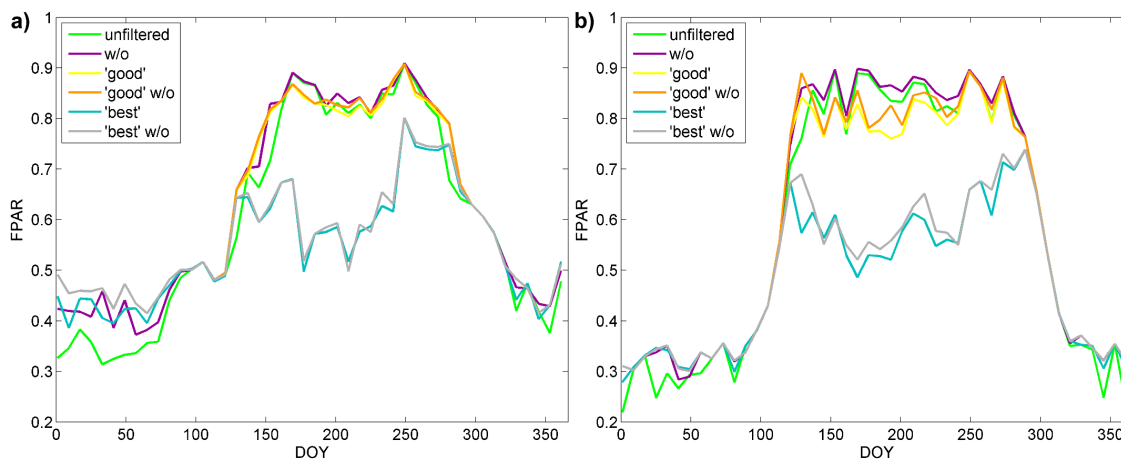
Figure 3.13: Multi-year average MODIS FPAR values for the various quality classes exemplarily shown for the Terra data sets of Harvard (a) and MMSF (b).

In the following two chapters, which concern the formulation of a light use efficiency model and the extrapolation of its parameters, MODIS LAI and FPAR time series are used. The processing of the values as delivered by the product to continuous time series follows the conclusions of this study: MODIS Terra and Aqua data are merged, the "best" values are deleted in summer when larger LAI values can be expected and all subset pixels with a similar land class are taken into account when the final, temporally interpolated time series is derived by the application of smoothing splines though the data points. An exception are sites whose measurement tower is located within a mosaic of different land use classes, what is in particular true for the grassland sites Neustift and Oensingen located in the Alps; here, only the directly surrounding pixels are considered.

4

Formulation of a
Light Use Efficiency Model

4.1   Introduction

The light use efficiency **c**oncept is one of the most widely applied approaches for modeling the uptake of carbon by plants. It has proven to be particularly suitable if the focus is laid on seasonal dynamics, for the use with remote sensing data for applications on regional or larger scales (see Hilker *et al.* (2008) for a review) as well as for data-led modeling approaches utilizing the information content of measurements (Jarvis *et al.*, 2004; Schulz & Jarvis, 2004). The latter studies followed a data-driven modeling approach developed by Young and co-workers (Young, 1998; Young & Pedegral, 1999; Young *et al.*, 2001; Young & Ratto, 2009). This modeling philosophy derives models specifically at the scale of interest beginning with a basic, parsimonious functional relationship. In a second step, this basic equation is iteratively refined by adding additional dynamics on the basis of data analysis results (Figure 4.1b). In this way, observed data is given more weight in the model building process than in purely process-based model building approaches (Figure 4.1a) without disregarding robust mechanistic processes. This leads to hybrid stochastic-mechanistic models with not more complexity than can be supported by the observation data information content. Nonetheless, the model structure and parameters have to fulfill the requirement of being physically or physiologically interpretable.

a) „classic" approach

```
┌─────────────────┐
│ System/Processes │
└─────────────────┘
        │
        ▼
┌─────────────┐        ┌─────────┐
│ Calibration │ ◄────► │  Data   │
└─────────────┘        └─────────┘
        │
        ▼
┌─────────────┐        ┌─────────┐
│ Validation  │ ◄────► │  Data   │
└─────────────┘        └─────────┘
```

b) „data-based" approach

```
┌─────────────────┐        ┌──────────┐
│ System/Processes │ ◄──────┤          │
└─────────────────┘        │          │
        │                   │          │
        ▼                   │          │
┌─────────────┐        ┌─────────┐     │
│ Calibration │ ◄────► │  Data   │─────┘
└─────────────┘        └─────────┘
        │
        ▼
┌─────────────┐        ┌─────────┐
│ Validation  │ ◄────► │  Data   │
└─────────────┘        └─────────┘
```
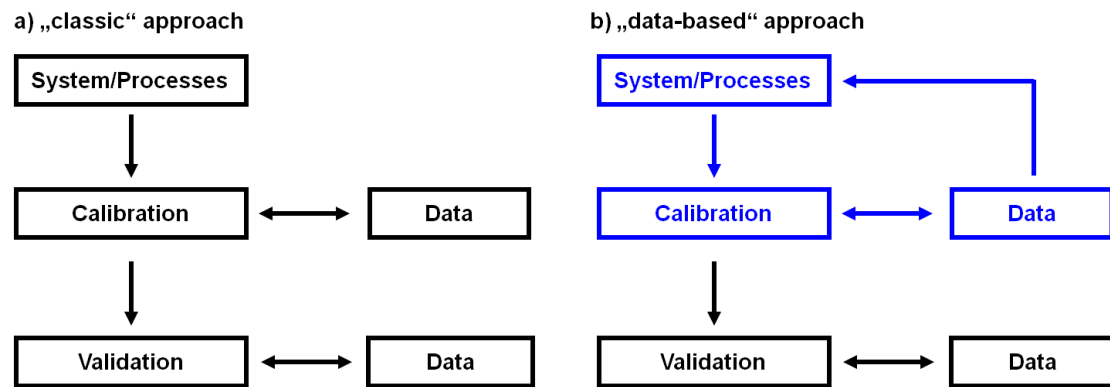
Figure 4.1: a) Scheme of a "classic" model building approach in which data processes are described with a priori knowledge. Data are used to calibrate the model parameters and validate the model. b) The data-based approach in which data are given more weight in the model building process by additionally using them to iteratively refine a basic process description.

It is the intention of this thesis to find such a simple but robust light use efficiency model simulating day-to-day variations of the gross primary production. The model shall allow an application to a broad range of vegetation types and climatic conditions. The approach chosen to reach this aim follows the data-based modeling philosophy discussed above: A model is sought which can be characterized as a "parametrically efficient, low order, dominant mode" model (Young, 1998, p. 1867). Its complexity is intended to be in accordance with the information content of available data. Starting point for this attempt is an oversimplified model formulation, the basic light use efficiency equation (Eq. 1.1).

This study builds upon the work of Jarvis *et al.* (2004). That part of their study dealing with descriptions for the time-varying light use efficiency parameter was a first step towards the final aim of a general gross primary production model and can be considered as a first loop indicated with the color blue in (Figure 4.1b): Jarvis *et al.* (2004) formulated a model – which is called the "Jarvis-model" in the following – on the basis of data available at that time, namely data from two temperate to boreal deciduous forest sites. The Jarvis model was shown to be able to capture and predict the dominant seasonal dynamics of the two sites by solely using the soil temperature as variable influencing the evolution of the light use efficiency. The model is not more complex than necessary and stands out due to low parameter uncertainties; in this sense it fulfills the requirements of the data-based modeling philosophy discussed above.

However, the strength of such a data-driven approach can also be its weakness: It might become necessary to refine the model structure if the model is to be applied to sites with strongly differing boundary conditions or if additional data are available revealing new dynamics. This corresponds to the situation when the Jarvis-model is applied to the 44 study sites now available for this thesis (see section 2.2): As will be seen, the Jarvis-model does not allow to reliably reproduce the gross primary production observed under

warmer and drier climate conditions and for other vegetation types than those analyzed in their study. Consequently, the data analysis and model building loop (colored in blue in Figure 4.1b) has to be repeated. Systematic behavior of the 44 vegetation stands in response to meteorological forcings as expressed in the daily data is extracted. On the basis of these new findings, the Jarvis-model is substantially restructured. A flexible model is developed, which additionally accounts for moisture availability as a major forcing of gross primary production and which is applicable under a broad range of boundary conditions.

4.2   Data Analysis Methods: DLR and SDP

The study of Jarvis *et al.* (2004) proved non-stationary regression frameworks such as the "Captain Toolbox" for Matlab® (Pedregal *et al.*, 2007) to be perfectly suited to derive functional descriptions for the typical seasonal evolution of the respiration and gross carbon flux (F_G) from eddy covariance measurements. These time series do not only depend on environmental conditions in a complex manner, but they are also afflicted with noise. This can hinder a simple signal extraction directly from the time series without filtering and smoothing techniques. Therefore, two of the toolbox's powerful tools, which are based on recursive Kalman filtering and fixed interval smoothing techniques, were employed both in the aforementioned study and are used in this study, too: dynamic linear regression, DLR, and state dependent parameter analysis, SDP. They allow the extraction of systematic trends in the variation of non-constant model parameters directly from measured time series and, hence, enable the objective identification of non-stationarities or state dependencies characterizing these time-varying parameters.

In particular, the underlying regression type model in case of DLR is of the form

$$y(t) = \sum_{i=1}^{N} c_i(t) \cdot x_i(t) + \zeta(t) \tag{4.1}$$

where y is the dependent variable, $x_i(t)$ are the regressors, c_i are time dependent regression parameters and $\zeta(t)$ is the regression model error series assumed to be a serially uncorrelated white noise sequence with a zero mean (Young & Pedegral, 1999). i is the increment running from 1 to the number of regressors, N. DLR extracts the incremental temporal variations in c assuming the parameters to gradually vary with time. The stochastic random walk process follows a white noise sequence ($\eta(t)$) with a zero mean. Each sampling instant depends on the data in its vicinity. A Gaussian weighting function determines the influence of the neighboring data samples on the one currently considered. The "bandwidth" of this Gaussian window function centered at the ith sample instant and declining at either side is determined by the noise-variance ratio (NVR). The NVR is calculated as ratio of the variances of $\eta(t)$ and $\zeta(t)$. A NVR of zero corresponds to

constant parameter values. Large NVR values (e.g. 0.1 and larger) imply a sharp decrease of the weighting function with increasing distance from the considered sample, resulting in rapid changes of the estimated parameter. By using a very large NVR an almost perfect model fit can be achieved; then, however, the estimated SDP model is sensitive to data outliers and anomalies. This is contraproductive if typical, systematic behavior is to be identified to derive model structures for predicting future system behavior (Young, 2000; Young *et al.*, 2001; Young, 2001). The "Captain Toolbox" also provides uncertainty bounds (standard errors) of the fits and of the parameter estimates, which are an important criterion when evaluating the estimated state-dependencies.

As an illustrative example, let the latent heat flux λE be the dependent variable $y(t)$ in Eq. 4.1 and the available energy (the sum of $\lambda E(t)$ [MJ m^{-2} d^{-1}] and the sensible heat flux $H(t)$ [MJ m^{-2} d^{-1}]) the regressor $x(t)$. The regression parameter $c(t)$ is then the well-known evaporative fraction, EF [-]:

$$\lambda E(t) = EF(t) \cdot (\lambda E(t) + H(t)) + \zeta(t). \tag{4.2}$$

Figure 4.2 illustrates the effect of DLR with different NVR values: Panel a) shows the direct quotient of the measured λE and the sum of λE and H time series (black dots). The red line indicates EF obtained by DLR, whereas the NVR value of $4 \cdot 10^{-4}$ was optimized within the DLR algorithm. The grey lines represent the standard error. Note that the plot is a zoom and that there are outliers outside the shown axes range. DLR results in a r^2-value of 0.96 with respect to $\lambda E(t)$. In Panel b) the NVR value is set to a small value of 10^{-5} (green) resulting in EF values changing only slowly with time ($r^2 =0.88$) and to a large NVR value of 0.5 (blue) resulting in rapid changes of EF. The higher r^2 value of 0.99 comes at the cost of larger standard errors (grey), which would render forecasts more uncertain.

The SDP estimation, in contrast, presumes the regression parameters to not only vary with time but also with a state of the considered non-linear system (Young & Pedegral, 1999; Young, 2000; Young *et al.*, 2001; Young, 2001). The state variable, however, also varies with time:

$$y(t) = \sum_{i=1}^{N} c_i(u_i(t)) \cdot x_i(t) + \zeta(t) \tag{4.3}$$

with u_i being variables representing time-varying system states. In the SDP algorithm, c_i is again assumed to evolve in a stochastic random walk process characterized by a white noise sequence with a zero mean. Each sample instant depends on the data in its vicinity in state space, in which all involved variables are sorted with respect to the system states u_i, hence out of temporal order. As with the DLR model, the NVR value determines the weighting of neighboring samples.

This is again illustrated by an example for which the light use efficiency equation is chosen: The basic light use efficiency equation (Eq. 1.1) is transformed to an adequate
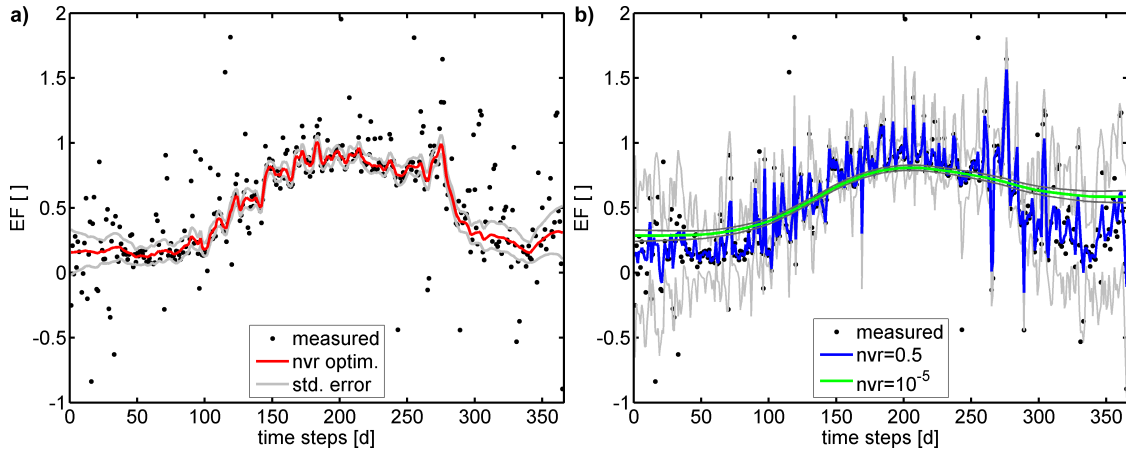
Figure 4.2: Example for EF using DLR. In Panel a the NVR is optimized (NVR= $4 \cdot 10^{-4}$). EF obtained by DLR shows distinct but moderate variations with time, modest standard errors (grey) and a good r^2 of 0.96 with respect to $\lambda E(t)$. In Panel b the NVR value is set to a small (green; $r^2 =0.88$) and to a large NVR value (blue) yielding a higher r^2 of 0.99 but also higher uncertainty bounds (grey). Data from MMSF, 2004.

SDP representation whereas the light use efficiency, ϵ, as a regression parameter is assumed to depend on the soil temperature T_S [°C] as system state (Jarvis *et al.*, 2004):

$$F_G(t) = \epsilon(T_S(t)) \cdot APAR(t) + \zeta(t) \tag{4.4}$$

where APAR [MJ m^{-2} d^{-1}] is the absorbed photosynthetically active radiation as a product of FPAR [-] and PAR [MJ m^{-2} d^{-1}]. In Figure 4.3a, ϵ [gC MJ^{-1}] is plotted as direct ratio of the gross CO_2 flux, F_G [gC m^{-2} d^{-1}]. It is worth noting that the direct quotient of F_G and APAR often results in even noisier time series than in this example. The NVR is optimized ($8 \cdot 10^{-5}$) and the SDP procedure results in a sigmoidal curve (red line) with a quite stable course, acceptable uncertainty bounds (grey) and an r^2 of 0.91. If the NVR value is decreased to 10^{-6} (green line in Figure 4.3b) the curve gets even smoother but some important seasonal variability gets lost and the r^2 drops to a value of 0.88. Enlarging the NVR to a value of 0.01 leads to rapid changes of the regression parameter ϵ (blue line in Figure 4.3b), a r^2 of 0.94 but higher standard errors (grey) of the SDP estimation of ϵ.

The examples shown above give an impression of how DLR and SDP can help to identify dominant and systematic behavior of time-varying model parameters with respect to the temporal evolution and other system variables. The DLR as well as the SDP model estimations, however, strongly depend on the applied NVR values (see above), which therefore shall receive further attention: The NVR values in this study are optimized from the data via maximum likelihood prediction error decomposition as proposed by Young & Pedegral (1999) and implemented in the "Captain Toolbox". In all SDP estimations of ϵ described later, the NVR optimization procedure yields typically small values ranging from $1 \cdot 10^{-7}$ to $5 \cdot 10^{-3}$; these values are large enough to confirm that it is justified to treat ϵ as varying parameter, but small enough to prevent an oversensitivity to the existing noise
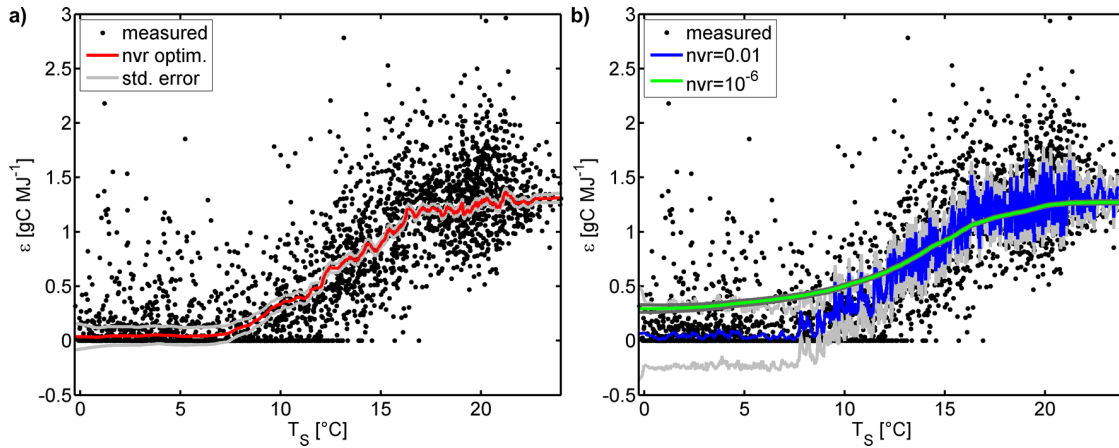
Figure 4.3: Regression with SDP: example for the light use efficiency which is assumed to depend on the soil temperature. In Panel a the NVR is optimized (red), Panel b shows the SDP results if the NVR is set to a smaller value (green) and a larger value (blue). The black dots represent the ϵ as direct quotient of the measured time series F_G and APAR. Grey lines indicate the standard errors. Data from Harvard Forest.

in the measured time series (Richardson *et al.*, 2008). Varying the NVR values manually shows that NVR values larger than $1 \cdot 10^{-3}$ result in estimations too responsive to outliers and anomalies, masking typical seasonal behavior embedded in the measurement data. Therefore, in this study the range of possible NVR values estimated by the toolbox's optimization procedure is constrained by an upper bound of 10^{-3}.

4.3   Model Identification

4.3.1   Evaluation of the Jarvis-model

As mentioned above, Jarvis *et al.* (2004) exploited the information content of data from two temperate forests (Harvard Forest, UMBS) with SDP and, on that basis, derived a functional form for describing ϵ as varying parameter of the following equation, a slight variant of the typical light use efficiency approach:

$$F_{G,J}(t) = \epsilon_J(t) \cdot S_0(t). \tag{4.5}$$

Therein, the above-canopy incident solar radiation (S_0 [MJ m^{-2} d^{-1}]) instead of the absorbed PAR (APAR) was used; therefore, strictly speaking, ϵ_J it is not a light use efficiency and it was therefore termed "radiation capture and utilization coefficient" (Jarvis *et al.*, 2004, p.940). ϵ_J is described by a sigmoidal relationship with T_F, the time-delayed soil temperature (T_S), which is described by following function:

$$\epsilon_J(t) = \frac{\epsilon_{max,J}}{1 + exp(k_{T,J} \cdot (T_F(t) - T_I))} \tag{4.6}$$

where

$$T_F(t) = (1 - \alpha) \cdot T_S(t) + \alpha \cdot T_F(t-1). \tag{4.7}$$

T_I is the inflection point of ϵ between its minimum and maximum level ($\epsilon_{max,J}$) and $k_{T,J}$ the rate of change of this transition. α is the lag parameter for the time-delayed soil temperature, T_F. The mean of T_S-values of the the first 30 days is chosen as starting point for T_F. The soil temperature was filtered because the measured time series showed $\epsilon_{max,J}$ responding with delay to changes of T_S. The four parameters, $\epsilon_{max,J}$, T_I, $k_{T,J}$ and α, were site-specifically calibrated by Jarvis *et al.* (2004) against data of the two study sites. The model was validated at Harvard Forest over a 6-year period. The resulting optimized parameters were shown to be well-defined.

In this thesis, this promising modeling approach serves as starting point for identifying a generalized model scheme applicable to a broader spectrum of vegetation and climate types. To do so, the original Jarvis-model is, as a first step, applied to all study sites regardless of their vegetational and climatological characteristics to test the suitability of the model. The four constant model parameters (ϵ_{max}, T_I, k_T and α) are optimized at each study site individually with the non-linear least squares method. Only one difference is made compared to Jarvis *et al.* (2004): For better comparability with other studies, the global radiation S_0 is changed to PAR, the photosynthetically active part of S_0. This change does not impair the model applicability of the Jarvis-model since PAR is a quite conservative fraction of S_0 (Stigter & Musabilha, 1982).
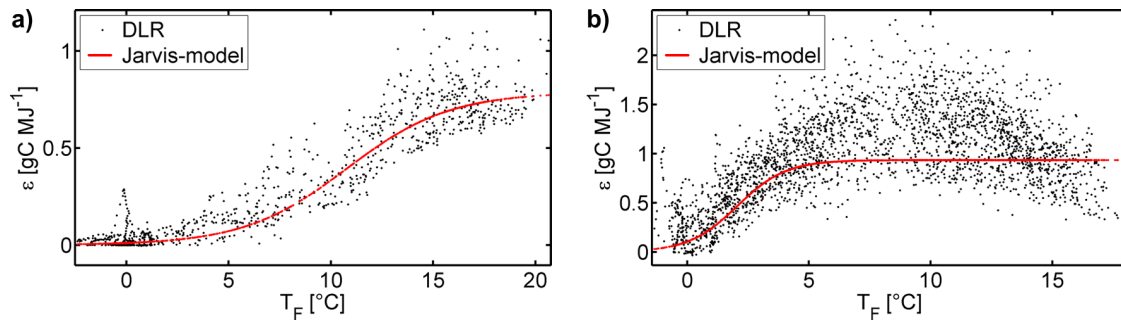
Figure 4.4: The – for better comparability – somewhat noise-reduced ϵ (black) obtained by DLR (Eq. 4.1) compared to ϵ modeled by the Jarvis-model (red) vs. the delayed temperature, T_F. Panel a shows ϵ at the boreal deciduous forest site Sylvania Wilderness where ϵ reaches a plateau at higher temperatures, whereas b) represents a typical example (Tharandt) for coniferous sites or grasslands with a distinct decrease of ϵ at higher temperatures after a peak at mild temperatures.

The Jarvis-model reproduces well the gross CO_2 uptake F_G of boreal and temperate forests in terms of r^2- and NS-values. The model performs particularly well at deciduous forests with strong seasonal dynamics as Figure 4.5 indicates and Figure 4.4a shows. In the latter plot, the sigmoidal function describing ϵ in the Jarvis-model is compared with the

measured ϵ, which is – for better illustration of the dominant behavior of ϵ – somewhat noise-reduced by the application of DLR.

At forests sites in warmer C-climates and at needle-leaf forests, however, the model shows deficiencies in the temperature dependency of ϵ: The model is not able to capture the decrease of ϵ at high temperatures (Figure 4.4b), which is not surprising considering the sigmoidal form of the function. r^2- and NS-values with regard to the measured and modeled F_G time series are quite satisfying for most forests, though (Figure 4.5). However, a comparison with the measured and noise-reduced ϵ such as shown in Figure 4.4b reveals, that these moderate to good model performances are often just a result of the fact that (i) PAR itself explains a large variation of F_G and (ii) a result of the nature of the parameter optimization procedure, which – up to a certain degree – counterbalances shortcomings of model formulations. Specifically, α seems to compensate inappropriate model structures. F_G of forest sites experiencing hot summers as well as most grasslands cannot be simulated by the Jarvis-model at all, because the assumed sigmoidal temperature dependency of ϵ does not exist, letting the conclusion to be drawn that a water availability proxy is lacking. Even at the fully humid study sites analyzed by Jarvis *et al.* (2004), cross-correlations between the model residuals and a water availability measure were found. Consequently, not only the dependency of ϵ to T_S has to be reconsidered, but also appropriate water availability measures have to be identified and their functional relationship to ϵ has to be derived.

4.3.2   Finding New Model Structures

In order to identify the dominant state dependencies of ϵ at all study sites, SDP is applied first with T_S as state variable to systematically examine which pattern ϵ follows with regard to T_S:

$$F_G(t) = \epsilon(T_S(t)) \cdot APAR(t) + \zeta(t) \tag{4.8}$$

confirming that at many sites a distinct decrease of ϵ with increasing T_S occurs (Figure 4.6a). APAR is chosen instead of S_0 or PAR, respectively, for better comparability with other studies and because of the overwhelming evidences for the significance of the LAI or FPAR as scaling-factor for soil-vegetation-atmosphere-transfer processes (Watson, 1958; Monteith, 1977; Tucker & Sellers, 1986; Goetz & Prince, 1999; Gower *et al.*, 1999; Lindroth *et al.*, 2008) which cannot be compensated by other environmental variables used in the light use efficiency modeling approach.

At those sites where the Jarvis-model is not able to properly reproduce the carbon flux dynamics at all, SDP not surprisingly also fails to find a clear temperature dependency indicating some other control on the ϵ-dynamics, presumably the a water availability measure. SDP is therefore used to analyze several water availability measures (W) as

	DBF	MF	ENF	EBF	G
BSh			Yatir		Audubon
BSk					Peck
Csa	Roccarespampani			Castelporziano	Vaira Ranch
Csb			Blodgett / Wind River / Metolius Young / Metolius Interm.	Puechabon	
Cfa	Duke Hardwood / Walker Branch	Duke Loblolly	Donaldson		Goodwin Creek
Cfb	Hesse / Soroe	Brasschaat / Vielsalm	LeBray / Loobos / Griffin		
Dfa	Mssouri Ozark / MMSF		Black Hills		
Dfb	UMBS / Willow Creek / Hainich / Sylvania	Harvard / Howland	Tharandt / Wetzstein / Norunda		Oensingen / Neustift / Lethbridge
Dfc	Bartlett		Hyytiälä / Flakaliden / Niwot Ridge / Boreas / Glees		

Figure 4.5: 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: high (green, $r^2 > 0.8$), moderate (yellow), low (red, $r^2 < 0.5$). Köppen-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).

potential further controls on ϵ:

$$F_G(t) = \epsilon(W(t)) \cdot APAR(t) + \zeta(t) \tag{4.9}$$

including the evaporative fraction (EF) being the fraction of the latent heat and the available energy, the vapor pressure deficit (VPD [kPa]), the soil water content (SWC [%], Figure 4.6b), and the antecedent precipitation index (API [mm], Figure 4.6c). Besides API, these variables have demonstrated before to significantly affect gross primary production and are frequently used in light use efficiency models as moisture availability indicator (Mäkelä *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 (Eq. 4.2). API is calculated by a weighted sum of daily precipitation values in a time window Z before the current time step t (Linsley *et al.*, 1982; Samaniego-Eguiguren, 2003):

$$API = \sum_{d=0}^{Z} \kappa^{-d} \cdot P(t)^{t-d} \tag{4.10}$$

where d denotes the number of time steps before t and κ is a recession constant commonly ranging between 0.85 and 0.98 (Chow *et al.*, 1964). A cosinus function is chosen such, that κ varied between these extremes with the lowest value of κ in summer and the highest in winter to take the higher recession of precipitation events in summer into account. Indeed, in several cases the SDP-analysis shows a distinct dependency on the water availability state variables (Figure 4.6b,c).
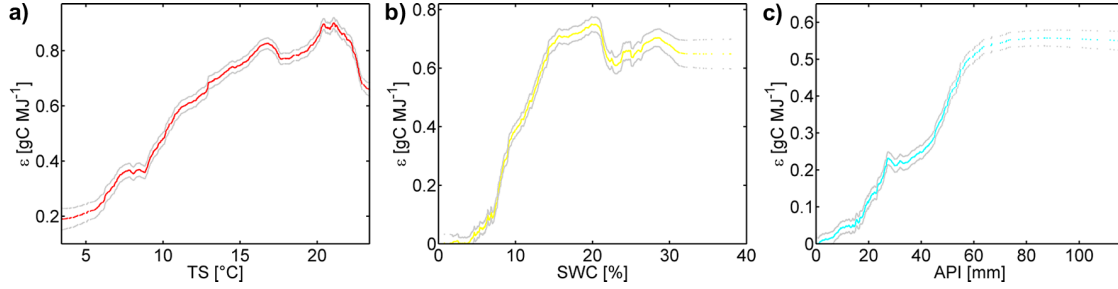
Figure 4.6: ϵ depending on the major driver of F_G as obtained by a SDP-model with one system state (eq. 4.3 with $N = 1$). System states explaining ϵ in these examples were T_S at Duke (a), SWC at Vaira Ranch (b), API and Audubon (c). The grey lines represent the standard errors of the SDP estimation.

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

$$F_G(t) = c_1(T_S(t)) \cdot APAR(t) + c_2(W(t)) \cdot APAR(t) + \zeta(t) \tag{4.11}$$

or respectively

$$F_G(t) = (c_1(T_S(t)) + c_2(W(t))) \cdot APAR(t) + \zeta(t) \tag{4.12}$$

with

$$c_1(T_S(t)) + c_2(W(t)) = \epsilon(t) \tag{4.13}$$

using T_S and one of the mentioned water availability proxies. The model performance varies between the water availability surrogates at the study sites, but none 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 ϵ_{max} and T_S in the additive SDP model either has a sigmoidal form, or it can be described by a (sigmoidal) peak function (Figure 4.7a), or no clear relationship can be identified at all (Figure 4.7c). If a clear relationship between ϵ and the water availability state variable exists – as in the majority of cases – it shows a threshold-like behavior as demonstrated exemplarily in Figure 4.7b and 4.7d.
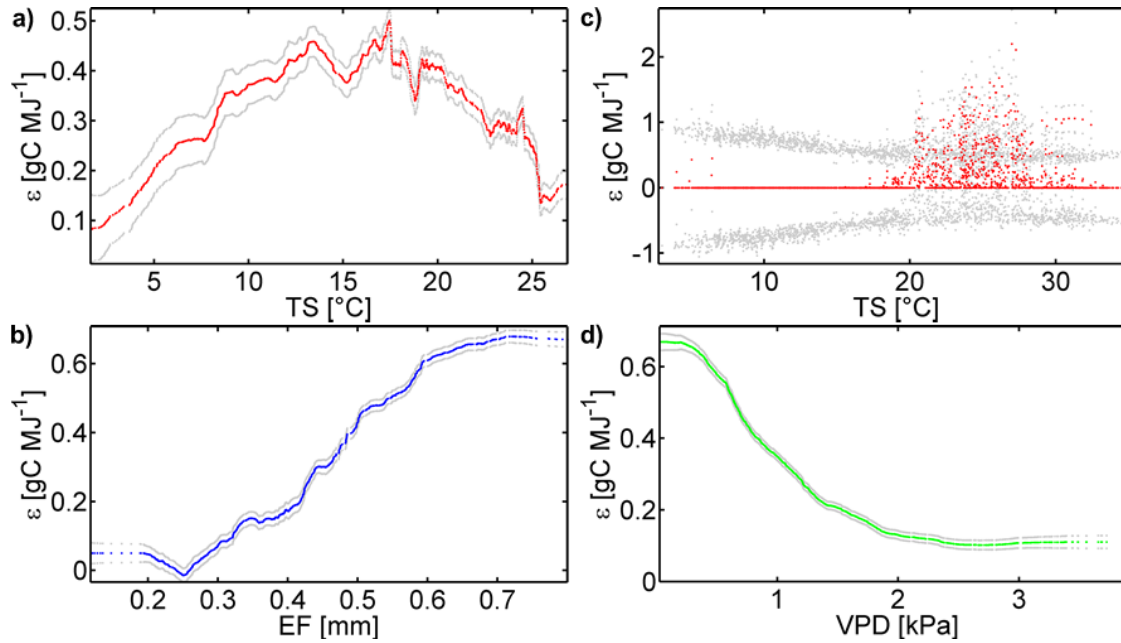
Figure 4.7: ϵ as function of T_S and EF at the Mediterranean site Roccarespampani (a, b) and as function of T_S and VPD at the desert-grassland Audubon (c, d) as determined by the additive SDP-model (Eq. 4.11); at the latter site, no clear T_S-dependency can be determined and the standard error bounds (grey) are consequently large.

4.3.3 Formulating a Generalized Model

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

$$f_T = \frac{4 \cdot exp(-(T_S - T_{opt})/k_T)}{1 + exp(-(T_S - T_{opt})/k_T)^2} \tag{4.14}$$

with k_T [°C $^{-1}$] being the rate of change and T_{opt} [°C] being the temperature at which the function is reaches its maximum.

To allow for the effect of water availability fluctuations, a sigmoidal 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 ϵ; this behavior gets even more obvious when taking additively both temperature and moisture in one SDP-model into account (Figure 4.7). The function allowing for the influence of W on ϵ_{max}, f_W, is therefore chosen to be a sigmoidal function:

$$f_W = \frac{1}{1 + exp(k_W \cdot (W - W_I))} \tag{4.15}$$

with k_W being the rate of change between the minimum and maximum levels of f_W and W_I being the inflection point. Both f_T as well as f_W are scaled to the range between 0 and 1.

To account for lag effects between the response of ϵ to temperature variations the lag-parameter α is again applied to T_S (equation 4.7) as it has proven to be significant in similar light use efficiency model approaches as proposed by Mäkelä *et al.* (2006, 2008) for sites in temperate and boreal climates. However, in cases of W being the main driver of ϵ as it is the case in semi-arid climates, the lag function is applied to W instead of T_S. In applying α only to the main driver, the number of free parameters is minimized, and the lag is only expected to be apparent in a distinctive manner on this 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:

$$F_G = \epsilon_{max} \cdot (p \cdot f_T + (1 - p) \cdot f_W) \cdot APAR \qquad (4.16)$$

where APAR is the absorbed photosynthetically active radiation [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 ϵ_{max} is reached. If no humidity dependency can be detected, because there is always enough water available, and ϵ-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.

4.4   Model Calibration and Evaluation

The final model formulation given by Eq. 4.16 comprises seven constant parameters including k_W, ϵ_{max}, p, T_{opt}, k_T, W_I and α. Before the final calibration, the sensitivity and variability for each of these parameters among different sites is explored. To do so, a set of 750,000 Monte Carlo simulations is executed at each location allowing the seven parameters to vary randomly within predefined (biophysically meaningful) ranges following a uniform distribution. Using the SSE between measured and modeled F_G as performance criteria, the site specific parameters distributions of the respective 1000 best model runs were identified. Distinct minima of the parameter response surface – indicating a unique identifiability – could be observed in most cases for the parameters ϵ_{max} and p as well as for either the parameters of f_T or that of f_W, dependent on the dominant control (temperature or water availability, see below). Analyzing the 1000 very best solutions with regard to all sites reveals that the parameters exploit a wide range

within the assigned upper and lower boundaries. The probability distribution functions (PDF) for the seven parameters are drawn (Figure 4.8). The PDF of the parameter k_W shows the sharpest peak, thus most of the best values at all sites are in a narrow range.
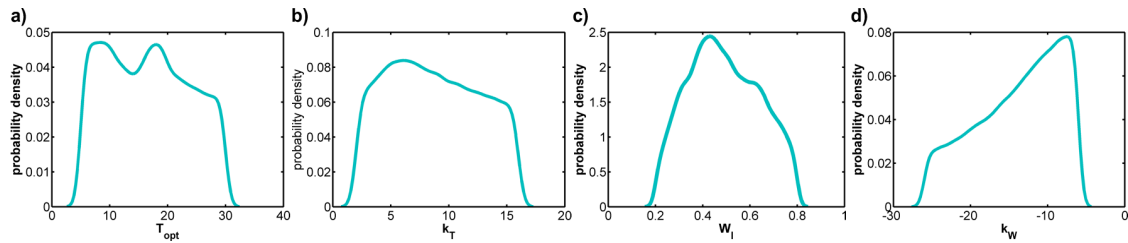
Figure 4.8: Probability density functions of the best parameters from all sites: T_{opt} (a), k_T (b), W_I (c), k_W (d).

Against this background, k_W is set to a constant value at the median of the best k_W-values from the Monte Carlo runs to reduce the parameters to be calibrated. The remaining six parameters, ϵ_{max}, p, T_{opt}, k_T, W_I and α are finally calibrated using a non-linear least-squares optimization routine ("lsqnonlin", MATLAB (Coleman *et al.*, 2010)). The model calibration is performed separately for each candidate for the water availability proxy W (EF, SWC, API and VPD).

The calibration performs well for all model runs: r^2 and NS-values of greater than 0.7 in most cases as well as relatively small biases 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 Figure 4.9 with their 95% uncertainty bounds. These are a consequence of the propagation of error due to parameter calibration uncertainties and were obtained by a Monte Carlo procedure: Instead of a simple drawing of samples from a normal distribution, the samples ($N = 1000$) are generated by a Cholesky decomposition of the covariance matrix of the model residuals. The resulting matrix is multiplied with a matrix of normally distributed random parameter samples. Thereby the interrelations of the parameters are accounted for. While these uncertainty bounds are mostly 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 gC m^{-2} d^{-1}, 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 95% confidence intervals, ii) overall, EF already performed best in the SDP-analysis, iii) it has often been successfully employed as water availability
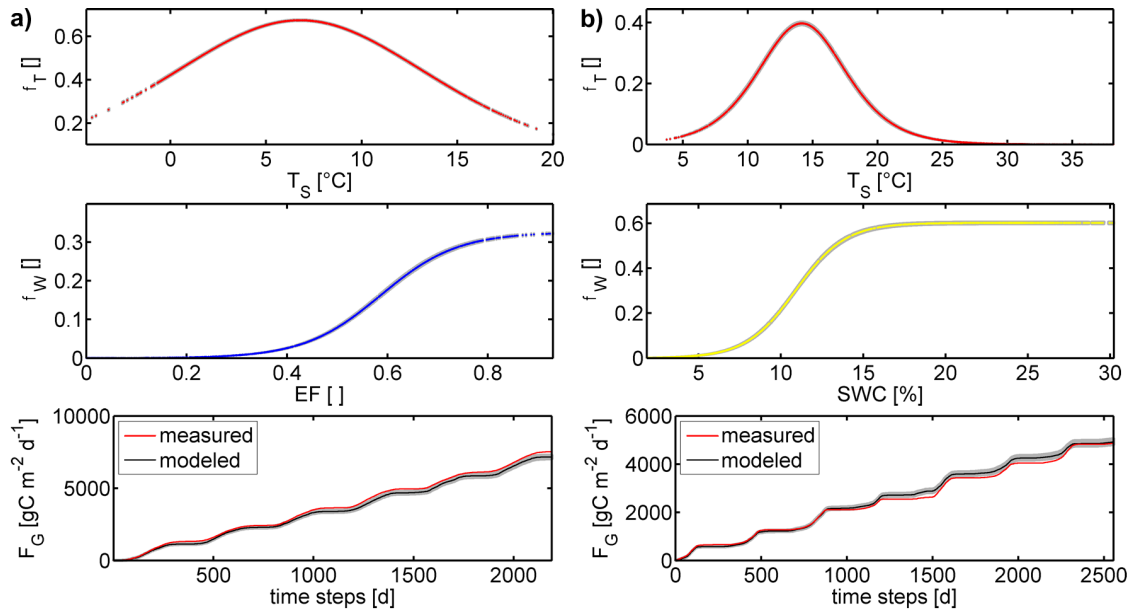
Figure 4.9: f_T, f_W and the cumulative sums of F_G with 95% uncertainty bounds (grey) at Wetzstein (a) and Duke Forest (b), EF was used to model F_G in a) and SWC in b) as water availability measure. The 95% uncertainty bounds are due to the propagation of uncertainties in associated parameter estimates (see text for further details).

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 ϵ in similar analysis (Garbulsky *et al.*, 2010), and iv) it can be retrieved from remote sensing, which is important from a regionalization point of view (Venturini *et al.*, 2004; Wang *et al.*, 2006). Results for the other moisture surrogates are shown exemplarily (Figure 4.9b and 4.13). The resulting parameters of the optimization procedure, their confidence intervals, r^2-values and model efficiencies NS are given in Table A.2 for the EF-model. All parameters are shown in a climate-vegetation-matrix (Figure 4.10).

The parameters of f_T have in general wider confidence intervals than W_I of f_W; one reason for this certainly is that f_T has two free parameters. However, fixation of one parameter of f_T deteriorates the model performance too much. Furthermore, 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 W_I (Figure 4.12a). This fact can also be seen in the SSE-values resulting from the Monte Carlo simulations. These were run again for the final model with the six free model parameters: For high p-values, 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, W_I is better defined (Figure 4.11). This observed characteristic is an advantage
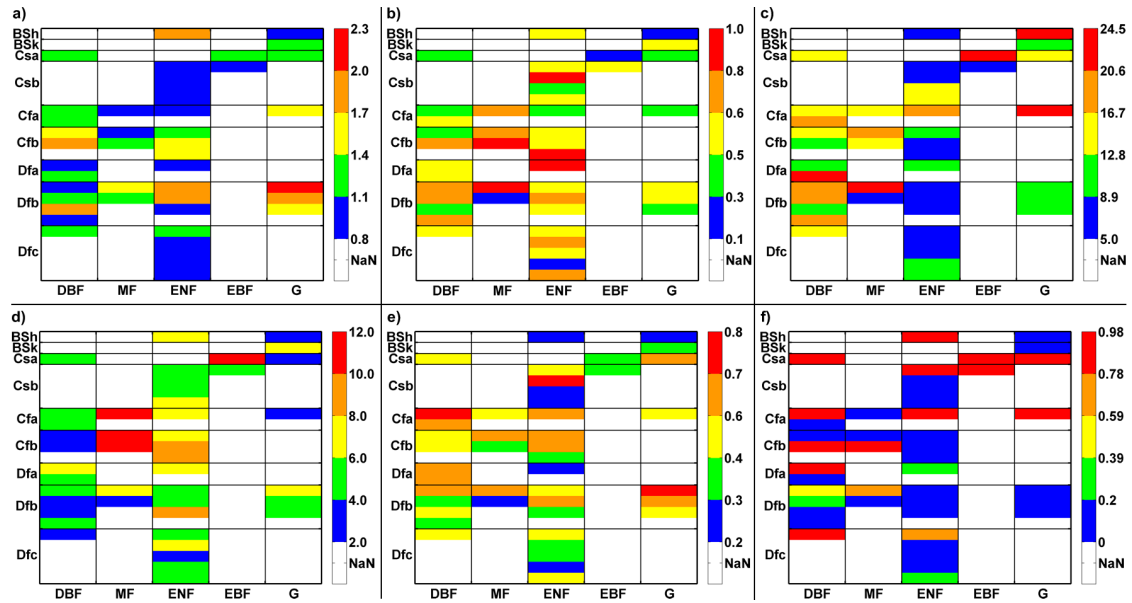
Figure 4.10: The six model parameters ϵ_{max} (a), p (b), T_{opt} (c), k_T (d), W_I (e), α (f) in a vegetation and climate context. See Figure 4.5 for an explanation of the abbreviations.

of the proposed model structure: The model is on the one hand flexible enough to simulate daily fluxes of sites with very different environmental 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 Figure 4.12b 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.
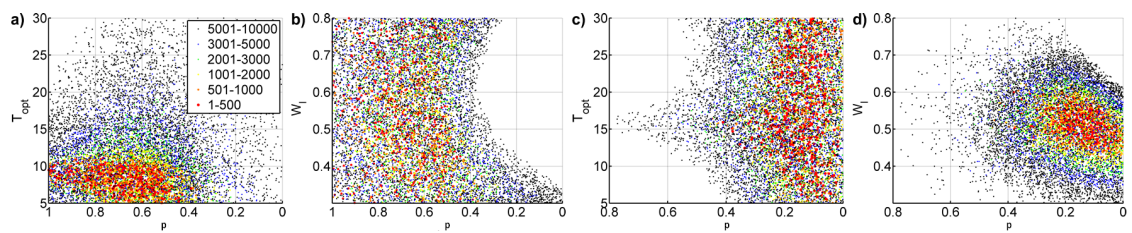
Figure 4.11: The parameter space of p and T_{opt} (a and c) and p and W_I (b and d) resulting from Monte Carlo simulations (1000 best parameter sets according to SSE-values, 750,000 runs) for Wetzstein (a and b) and Lethbridge (c and d). T_{opt} is often better defined than W_I if p is high (greater contribution of f_T) such as at Wetzstein, and Lethbridge is an example for W_I being better defined than T_{opt} if p is low (greater contribution of f_W).

The model parameter ϵ_{max} varied at forest sites between 0.78 gC MJ^{-1} at a needleleaf forest in Canada and 1.93 gC MJ^{-1} at a German needleleaf forest in a temperate climate with mild summers. Deciduous forests form the highest average ϵ_{max} with a mean of 1.25 gC
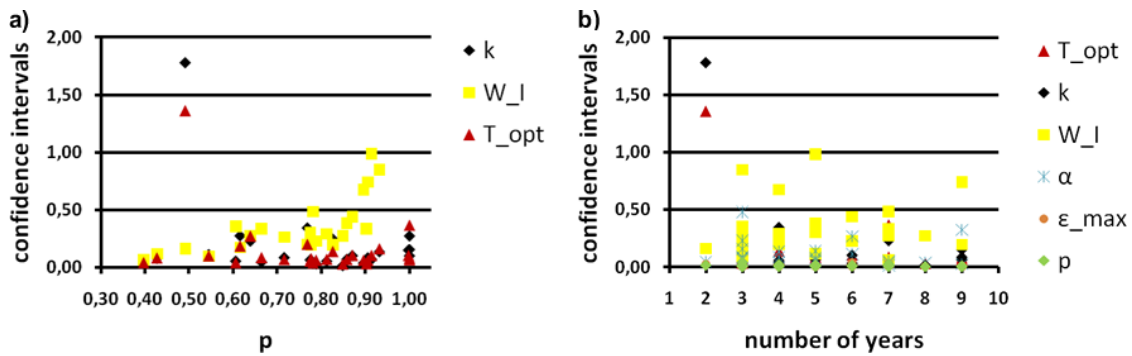
Figure 4.12: Parameter confidence intervals in relation to p (a) and to the number of available measurement years used to optimize the model (b) with *SWC* as water availability measure.

MJ^{-1} followed by mixed forests (1.18 gC MJ^{-1}) and evergreen needleleaf forests (1.16 gC MJ^{-1}). Evergreen broadleaf forests have with 1.14 gC MJ^{-1} the lowest average ϵ_{max} due to low values in boreal climates and those with dry summers. Regarding the climate classes with more than one forest site, the Cfb-class reveals the highest average ϵ_{max}, closely followed by Dfb; Csb and Dfc have the lowest ϵ_{max}. At grasslands sites, ϵ_{max}-values are surprisingly high: ϵ_{max}-estimations reach 2.25 at Oensingen and 1.80 at Neustift and lead to an average ϵ_{max} of 1.50 gC MJ^{-1}. The highest ϵ_{max}-values at Oensingen are attained in spring and autumn when temperatures are favorable 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 p-value 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-value between 0.4 and 0.8. The low p-values at Hainich and especially at Boreas and Howland can be explained by a an especially high correlation between ϵ and EF in the seasonal course. At Hainich, particularly the distinct summer drought of 2003 (Reichstein *et al.*, 2007) with a strong decrease of ϵ in late summer leads to an higher influence than at the other sites in this climate class. p-values greater than 0.6 are mainly clustered in the forest and fully humid climate classes, whereas p-estimates at forests at summery sites as well as grasslands take values in the medium to lower range.

The temperature T_{opt} at which the sites reach ϵ_{max} is smaller than 14 °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_{opt}-values greater than 14 °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 conditions are dominated by medium

to high T_{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_{opt}-estimates corresponding to the mild average temperatures and highest efficiencies in spring and autumn.

In addition to the information at which temperature ϵ_{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 W_I determining the inflection point of the f_W-function takes in the most cases values between 0.3 and 0.7 with regard to EF as water availability surrogate, but more extreme values are represented, too. The lower W_I-values cluster in the cooler climate classes and those with hot and dry summers, whereas the representatives of the upper third of the W_I-range cluster in the fully humid climate classes with warm and hot summers.

Model parameter α, 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 – strong lag effect and direct immediate reaction of ϵ – are represented. The C-climate grasslands show delay processes whereas the grasslands in semi-arid and hot as well as continental D-climates seem to react rather promptly to the triggering variables.

The parameter p for the further water availability proxies SWC, API and VPD is presented in Fig 4.13. In case of SWC and API it illustrates a more homogeneous pattern of p-values within the vegetation-climate-matrix with lower p-values 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 p-value of 0.15, meaning a contribution of 85 %.

4.5 Discussion

In this study, state dependent parameter estimation was used to determine typical non-parametric relationships between the light use efficiency and relevant state variables. The
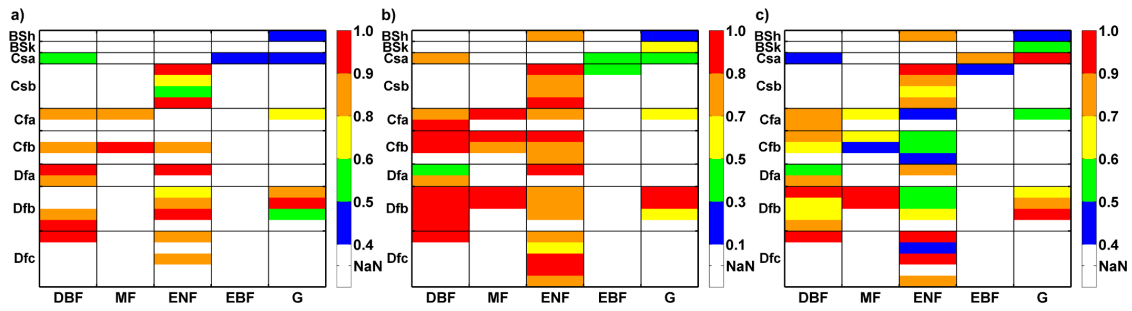
Figure 4.13: Model parameter p with respect to vegetation and climate classes for the model runs with SWC (a), API (b), and VPD (c) as input variables. See Figure 4.5 for an explanation of the abbreviations.

relationship between ϵ and T_S was determined as a sigmoidal peak function. Limiting functions with respect to temperature have been used before: Mäkelä *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 particularly the coniferous forest sites show a relatively small efficiency amplitude during the year with the highest light use efficiencies at lower temperatures and an decrease of ϵ-values at higher temperatures. Potter & Klooster (1999) and Yuan *et al.* (2007) who also modeled F_G across a broad range of conditions applied a peak function, too. They estimated an optimum temperature varying with the geographical latitude and determined the optimum by non-linear optimization merging the data from all study sites, respectively. In this study, however, it is shown that highest actually occurring efficiencies occur at temperatures that vary significantly across study sites. The strongly differing values do not allow the conclusion that a single optimum temperature with an ϵ-maximum for all sites exists. 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 ϵ and EF respectively. Mäkelä *et al.* (2008) represented the relationship between ϵ and VPD by an exponential function and between ϵ and SWC by a Weibull- or sigmoidal function following Landsberg & Waring (1997). In this study, SDP revealed in most cases a threshold-like response of ϵ to all water availability proxies given ϵ was sensitive to them. Overall, SDP-analysis showed similar functional forms of the responses to the driving variables across various boundary conditions and vegetation types. Moreover, these relationships appear to be appropriate means to model F_G if the functions describing the ϵ-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 ϵ-down-regulating scalars determined by predefined functions (e.g. Potter *et al.*, 1993; Running *et al.*, 1999; Mäkelä *et al.*, 2008). This multiplicative approach as

well as a model with a single potential ϵ-maximum (Yuan *et al.*, 2007, 2010), however, can lead to the parameterization of maximum ϵ-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, subfunctions with insensitive variables are prone to high calibration uncertainties of respective parameters; these have less negative impact on the prediction uncertainty if their influence is relativized by a weighting factor as realized in the proposed model. Additionally, maximum or potential ϵ-values even vary if a modifying scalar is added or omitted in a multiplicative approach (Mäkelä *et al.*, 2008). The so-derived potential ϵ is in these cases not more than a calibration artifact and not necessarily physiologically meaningful. Instead, the proposed additive model with a site-specific weighting of the variables' influence on ϵ leads to a maximum ϵ actually realized by the considered vegetation stands. Thus, in contrast to Yuan *et al.* (2007) and Yuan *et al.* (2010) this approach is based on the assumption that maximum ϵ-values and other model parameters on a daily basis vary between forest stands and grasslands (Turner *et al.*, 2003b; Bradford *et al.*, 2005; Schwalm *et al.*, 2006; Kjelgaard *et al.*, 2008; Stoy *et al.*, 2008) and no globally valid maximum ϵ, which is reachable by all vegetation stands under ideal conditions, exists and respective biochemical processes are universal across species. Predictions with the proposed model approach obviously require sufficiently long measurement time series covering optimal periods for vegetation growth to reliably derive a true ϵ_{max}-value for a specific sites. 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 ϵ_{max} varied strongly between sites. Compared to ϵ_{max}-values of several studies presented in the review of Goetz & Prince (1999), the calibrated values appear to be somewhat greater. Compared to the average values per vegetation type obtained by Garbulsky *et al.* (2010), however, ϵ_{max}-values tend to be smaller. However, most vegetation types were not sufficiently well represented in a statistical sense to allow a definite statement. The largest uncertainty factor with respect to the calibrated values, which could possibly lead to errors in ϵ-values, certainly is the input of MODIS LAI/FPAR data. They have often been found to be inaccurate (e.g. Wang *et al.*, 2005b; Pandya *et al.*, 2006; Pisek & Chen, 2007; Horn & Schulz, 2010), especially at needleleaf forests (Wang *et al.*, 2004; Yang *et al.*, 2006a). Additionally , there is a distinct scale mismatch between EC measurements and MODIS data and the linkage between these two data sources is complicated by the variability of the area EC measurements are representative for (Chen *et al.*, 2009, 2010). However, remote sensing data are the only data source for these variables at all study sites, so it is common practice to use this product despite its limitations (e.g. Yuan *et al.*, 2007; Xiao *et al.*, 2008a, 2011) and it is therefore used in this study, too, being aware of its drawbacks.

The highest ϵ_{max}-value reached at Oensingen is equal to the globally optimized potential light use efficiency in Yuan *et al.* (2010). Both relatively high ϵ_{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 ϵ (Ammann, 2010). Schwalm *et al.* (2006) and Garbulsky *et al.* (2010) found a maximum daily ϵ at grasslands, too. Interestingly, in the MODIS database a single value of 0.68 gC 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 gC MJ^{-1} for crop and natural vegetation mosaic. In this study, a relative high mean value of 1.50 gC 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 ϵ-efficiencies on the flux measurement and difficulties calculating FPAR due to mixed MODIS pixels. These problems become evident in a inhomogeneous distribution of MODIS pixels with grass, crop, savanna 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 ϵ_{max}-values among their 35 study sites.

For (northern) forests, Lindroth *et al.* (2008) found maximum daily ϵ-values ranging between 0.7 and 1.4 gC MJ^{-1}, Jung *et al.* (2007) showed maximum ϵ of about 1.5 across European Forests. Compared to these studies the upper boundary of the ϵ_{max}-range of forests in this study was somewhat larger but tend to be lower than in the analysis of Garbulsky *et al.* (2010). This fact presumably reflects the tendency of the model to underestimate F_G-fluxes as apparent by rather positive bias values (see section 4.4). The high ϵ_{max} at Tharandt was also observed in the study of Mäkelä *et al.* (2008) and explained by thinning of the forest stand. In average highest ϵ_{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 NASA's MODIS gross primary production algorithm (Heinsch *et al.*, 2006), a maximum ϵ of 1.01 gC MJ^{-1} is stored in the algorithm's look-up table for evergreen needleleaf forests, compared to a higher mean ϵ_{max} of 1.16 gC MJ^{-1} found in this study; the respective values for deciduous broadleaf forests are 1.16 and 1.25 gC MJ^{-1}, and for mixed forests 1.12 and 1.18 gC MJ^{-1}. Thus the model values of this study are somewhat higher than those applied in the MODIS gross primary production algorithm. That has to be considered against the background of a rather negative bias of the modeled time series (see above) thus probably even higher

ϵ_{max}-values in the measurements – if the error is attributed to the representation of ϵ and not APAR. And it has to be kept in mind, that not in all studies the assigned ϵ_{max}-value is an actually realized value.

The calibration of p shows that temperature has indeed a high influence on ϵ, especially in cooler ecosystems, as shown by numerous studies (Runyon *et al.*, 1994; Chen *et al.*, 1999; Nouvellon *et al.*, 2000; Turner *et al.*, 2003b; Schwalm *et al.*, 2006). SWC as modulating variable had the highest impact at summer-dry sites and grasslands, which reflects the short rooting depth of grass and the low depth at which the *SWC*-measurements were made. VPD appears to influence ϵ not only in dry areas as p values around 0.6 in boreal and temperate forests indicate. Indeed, the interrelation between VPD and ϵ via stomatal conductance has often been shown (Wang & Leuning, 1998; Goetz & Prince, 1999; Lagergren *et al.*, 2005; Katul *et al.*, 2003; McCaughey *et al.*, 2006). This 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 water shortage they can explain ϵ-variations with the lowest p-values, thus the highest contribution of all f_W-functions compared to the other W-variables. API is therefore considered as promising variable. Overall however, the optimization procedure attests EF most often the highest explaining capability on ϵ between the water availability surrogates applied and model runs with EF lead to the best results. This is not surprising considering the correlation of ϵ and EF (Monteith & Greenwood, 1986; Schulz & 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 modeling of managed sites such as Oensingen and Neustift to a certain degree. This behavior 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 λE and F_G. In their correlation analysis Garbulsky *et al.* (2010) determined EF to have the highest explanatory power with respect to ϵ; moreover, EF alone explained ϵ best and their model deteriorated when adding another variable. This contradicts with our calibration, which assigned EF a significant influence but never the only contribution to the variation of ϵ: p-values of nearly 1 occurred but only in one case a p-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 ϵ, but ϵ can not be captured by one variable alone.

The model performance as determined by r^2, NS and the degree of parameter uncertainty militates in favor of the proposed model particularly with regard to the wide variety of ecosystem characteristics represented at the study sites. Throwing a glance at the distribution of the model parameters in the climate-vegetation matrix (Figure 4.10) may lead to the assumption that the optimized parameter values have no bearing on site specific characteristics at all. However, in the majority of cases the parameter 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 ϵ 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 are: Both are located in the same category in the climate-vegetation matrix: mixed forests in a boreal climate with warm summers. Maximum ϵ-values at Howland occur in spring and autumn, whereas the maximum ϵ at Harvard occurs rather in summer. These circumstances explain the higher T_{opt} at Harvard despite lower mean annual temperatures. This fact probably indicates a higher fraction of deciduous trees at Harvard. The higher ϵ at Harvard is also in line with the assumption that the forest at Harvard needs more time to develop green biomass in spring until ϵ_{max} can be reached. This is just one example for existing but sometimes not directly obvious relations between model parameters and site characteristics but illustrates the presumption that the model parameters are applicable to a regionalization strategy by means of characteristic features accounting for meteorological, physiological and phenological attributes.

5

Extrapolation
of Model Parameters

5.1   Introduction

Developing a scheme to spatially extrapolate model parameters is the logical continuation of the previous work and will be the main focus of this chapter. The benefits of relating site-specific calibrated parameters of the derived light use efficiency model (see chapter 4) to site characteristics are explored and these relationships are utilized to spatially extrapolate the model parameter values to sites outside the calibration domain. For this purpose support vector regression (SVR) is used. The application of SVR is beneficial because of its potential to robustly generalize multi-dimensional relationships and represent highly non-linear structures with relatively few training samples (Vapnik *et al.*, 1997; Smola & Schölkopf, 2004). A further advantage of the method is given by the fact that a convex objective function is formulated which guarantees a unique solution (Burges, 1998). In the context of gross primary production modeling, Yang *et al.* (2007) proved SVR to be a powerful technique for spatially upscaling tower measurements. The extrapolation study within this thesis distinguishes itself from their purely empirical approach by not relating the gross primary production directly to remotely sensed land surface characteristics. Instead, an underlying physiological relationship – i.e. the light use efficiency model derived in the last chapter – is used in this study. The calibrated model parameters are extrapolated by means of SVR.

5.2 Methods

5.2.1 Support Vector Regression

SVR is a special application of the support vector machine technique (SVM, Vapnik *et al.* (1997)), which is a supervised statistical learning method (Vapnik, 2000). It is usually used for classification tasks. SVM implements the method of structural risk minimization (Shawe-Taylor *et al.*, 1998) by setting an upper bound on the error rate of a model trained with measurement data rather than solely minimizing the training error itself. This approach leads to a better generalization performance compared to other methods. All SVM methods have in common that they transform the data points into a feature space using a set of nonlinear functions. In this potentially very high dimensional feature space the problem can be solved linearly. SVM methods need only a subset of all training samples: the support vectors.

As with other regression methods, the goal of SVR is to find a function which maps sample data from a space spanned by independent variables into the output domain of a depending variable. This can be a challenging task when being confronted with highly non-linear problems. In these cases SVR is the method of choice since the basic principle behind SVR is that the space of training sample vectors, x, is transformed to a higher dimensional, kernel-induced feature space, in which a linear regression function, y, can be formulated (Gunn, 1998):

$$y = \langle w, x \rangle + b \tag{5.1}$$

with $x \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The function that optimally describes the sample data is the one that minimizes,

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_i (\xi_i^- + \xi_i^+) \tag{5.2}$$

satisfying the conditions

$$y - \langle w, x \rangle - b \leq \epsilon_{SVR} + \xi^- \tag{5.3}$$

$$\langle w, x \rangle + b - y \leq \epsilon_{SVR} + \xi^+$$

with a norm vector w and an offset b. ϵ_{SVR} forms the margins of a band with the width $2 \cdot \epsilon_{SVR}$ wrapping the true output values (Figure 5.1a). Within this band, the optimization is insensitive to deviations of the data points from y and only x-values at the margins or outside this band are recognized by the algorithm and represent the support vectors. Therewith, the magnitude of ϵ_{SVR} has an effect on the complexity of the SVR model and it also affects the number of support vectors and consequently the generalization capability of the SVR model. The training errors as distance to the margins of the ϵ-band are represented by the so-called slack variables $\xi_i^- > 0$ and $\xi_i^+ > 0$ which contribute linearly to the loss function. Hence, this type of SVR combines the structural with the empirical risk, the latter defined by the second term in Eq. 5.2, the so-called "soft-margin" optimization criterion. $C > 0$ is the weighting factor for this training error term and

determines the trade-off between the training error and the dimension of the model; a large value of C leads to a large penalty to training data points outside the ϵ_{SVR}-band and to a more complex model. C and ϵ_{SVR} have to be adjusted for each application.
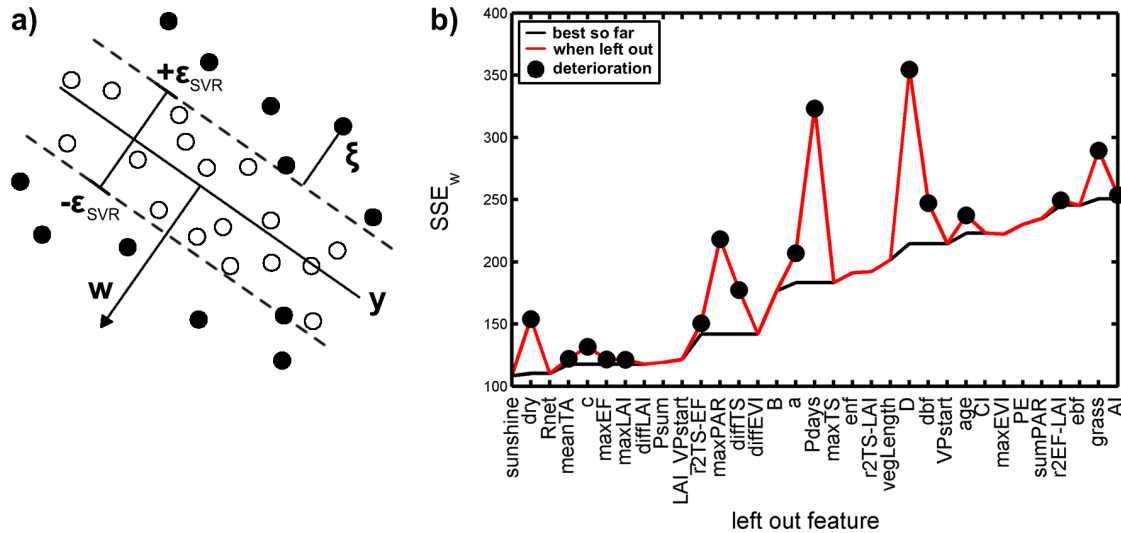
Figure 5.1: a) The SVR model $f(x)$ (black line, with the norm vector w) inside the insensitive band determined by ϵ_{SVR} and the slack variables ξ as distance of the data points (circles) outside the ϵ_{SVR}-band from its margins. b) An example for the evolution of the training quality measure SSE_w during the accomplished attribute selection procedure by leaving – one by one – all attributes successively out. If the SSE_w indicates an equally good or even better result (circles), the left-out attribute is finally removed (without circle).

The mapping into the higher dimensional feature space is done by Kernel functions. The idea behind Kernel functions is to enable operations in the lower dimensional input space rather than the higher dimensional feature space without waiving the advantage of the linear solution in the feature space. The Radial Basis Function (RBF) is such a Kernel-function, and it has shown to be highly flexible (Hsu *et al.*, 2003). As a matter of fact, the best performance was achieved by using the RBF as kernel. It is consequently applied in this study. The RBF uses only one parameter, γ, which has to be identified in addition to the SVR parameters C and ϵ_{SVR}.

To find adequate SVR parameter values, a dynamically dimensioned search (DDS) global optimization algorithm (Tolson & Shoemaker, 2007) instead of a typically applied grid search is used in this study, since it showed to be more efficient in test runs. DDS automatically scales the search within the maximum number of model evaluations and allows the accomplishment of computationally challenging optimization problems. Since it became quickly clear that the applied optimization algorithm (see below) always yields C-values around 1000 and above, and the result is not sensitive to the exact value, C is always set to a value of 2000, and only γ and ϵ_{SVR} are left free in the optimization. SVR is implemented with the LIBSVM-package (Chang & Lin, 2001) for usage with MATLAB®. All data sets are z-transformed to sets with a mean of 0 and a standard deviation of 1.

5.2.2   Finding Explanatory Attributes

Using the described SVR algorithm, the six calibrated light use efficiency model parameters (ϵ_{max}, p, T_I, k_T, W_I and α) shall be spatially extrapolated depending on a bundle of characteristic attributes specific for each parameter. Since only EF as water availability measure is used in this extrapolation exercise W_I is referred to as EF_I in the following. From the data available, attributes in various categories were collected: vegetation (coniferous, deciduous broadleaf, mixed, evergreen broadleaf, grass) and climate classifications (Köppen-Geiger), climate characteristics (temperature, precipitation, radiation, continentality and aridity index), characteristics concerning the physiological status of the vegetation (LAI, EVI, stand age), seasonal characteristics concerning the vegetation period and the seasonal course of climatic and physiological characteristics in relation to each other. Nominal attributes such as the vegetation classes were binarized. Table A.3 lists the variables with a short description.

A SVR model is trained for each parameter separately. Which attributes are relevant for each model parameter is determined by testing the model performance with various attribute combinations. SVR is capable of achieving a very high training accuracy, which often not reflects the model performance for unknown data. Therefore, the cross-validation performance is used to test the several combinations of attributes (Hsu *et al.*, 2003). Considering the small number of study sites with often just very few sites representing a specific climate and vegetation type (see Figure 4.5), a "leave-one-out" cross-validation ($N - 1$ fold, where N is the number of sites) is performed using all data but one as training samples. This is repeated until each site has been used once as validation data set. The performance is measured by the weighted SSE (SSE_W) (see section 2.4) whereas the weighting factor is the light use efficiency model parameter uncertainty resulting from the model calibration. Thus data points having assigned a small confidence interval or a small parameter uncertainty, respectively, receive more consideration in the SVR optimization procedure than those having been retrieved with a higher uncertainty in the light use efficiency model calibration. The model's cross-validation performance is also determined by the usual coefficient of determination, r^2, and the weighted coefficient of determination, r_w^2 (see section 2.4).

Since it is also not known how many attributes are needed to explain the model parameters, many feature combinations are tested for each parameter. However, it is not feasible to test all possible combinations for computational reasons. Furthermore, the explanatory power of a specific attribute can depend on the inclusion of another attribute into the training process, a fact which forbids a cumulative procedure starting with one attribute and adding further attributes step by step. Therefore, the attribute selection and training procedure is done iteratively starting with all attribute candidates (Figure 5.1b): First, a cross-validation is done with all attributes and the resulting SSE_w is stored. In the next step, every attribute is removed and a cross-validation with the remaining attributes is

executed. If the resulting SSE_w is smaller, hence the model performance is better, the left-out attribute is removed definitively, otherwise it is used again. When all attributes have been left out once, the same procedure is done with the remaining set of features. At the latest after five rounds the set of attributes is stable. Since the results of this approach depend on the order in which the attributes are left out, the starting configuration is randomly varied 1000 times. Despite the large number of possible configurations of the attribute matrix, repetitions of the resulting set of attributes occur soon and new results do not appear anymore at the latest after 300 repetitions. The SVR-parameter ϵ_{SVR} is fixed for this procedure to the value 0.06. This is somewhat lower than the default value of 0.1 and was found by test runs; it represents a trade-off between the average number of attributes chosen and the goodness of fit. ϵ_{SVR} is finally optimized when the attributes have been selected. For computing feasibility, the other parameters are optimized during the selection procedure only when the number of features has changed. When all attributes have been selected, each attribute is left out once again and the model trained with the other attributes. The resulting model performance measures show how important the single attributes are within the chosen set of attributes. This selection procedure is executed for each of the six model parameters.

5.3 Results

5.3.1 Selected Attributes

SVR is able to explain the six light use efficiency model parameters in the cross-validation exercise by a combination of seven to twelve attributes (Table 5.1) whereas some of the attributes are binary representatives of one and the same feature; for example, binarization of the vegetation class with five manifestations leads to four binary attributes.

To explain the parameter ϵ_{max} radiation, moisture, temperature and stand characteristics are selected likewise. The radiation influence is represented by the maximum PAR and the average fraction of sunshine, the number of days with precipitation and the aridity index reflect the importance of sufficient water for a high ϵ. The selection of the Köppen-Geiger main type "D" attests a certain importance of a strong seasonality for ϵ_{max}. The distinction between the vegetation classes "deciduous broadleaf forest", "evergreen needleleaf forest" and "grassland" as well as the stand age is also important for SVR to extrapolate ϵ_{max}. Finally, the day of year at which the growing season starts, and the amplitude of EVI and its maximum are exploited by SVR. Leaving out the number of days with precipitation and the binary variables for the Köppen-Geiger climate class "D" and the vegetation type "deciduous broadleaf forest" lead to the greatest decline in the model performance (Figure 5.2).

Table 5.1: Selected attributes to explain the light use efficiency model parameters by means of SVR. "VP" refers to the vegetation period, "CI" and "AI" are the continentality and aridity index. The attributes' names are chosen such that they should be self-explanatory, but an explanation can be found in Table A.3.

ϵ_{max}	p	T_{opt}	k	EF_I	α
maxPAR	meanTA	meanTS	meanTS	meanTA	CI
sunshine	Rnet	meanTA	meanTA	sunshine	sunshine
Pdays	diffLAI	CI	VPstart	diffEF	diffLAI
AI	VPstart	maxPAR	EVI_VPend	maxEF	vegLength
maxEVI	r2EF-LAI	vegLength	r2TS-EF	meanEF	VPstart
diffEVI	r2TS-EF	r2TS-EVI	D	AI	r2TS-EF
age	D	dbf	dbf	VPstart	enf
VPstart	dbf		grass	EVI_VPend	ebf
D				r2EF-LAI	
dbf				c	
enf				ebf	
grass					

Regarding the parameter p, which determines the influence of the temperature and moisture function on ϵ, attributes related to seasonal characteristics come to the fore: In addition to general climate features (mean temperature, average annual net radiation balance and the fraction of sunshine) the differentiation between deciduous and non-deciduous sites, between continental climates with strong seasonal characteristics and those without appears to be essential for the p-differences between sites. Furthermore, the LAI amplitude within the year, the timing of the onset of the growing season and, finally, the correlation between the annual courses of EF and LAI as well as EF and T_S explain p. The omittance of four of the latter variables at a time also deteriorates the extrapolation performance strongly, especially without the LAI amplitude attribute.

The peak of the temperature function T_{opt} can be recaptured by four attributes based on temperature characteristics: the mean soil and air temperature and the continentality index depending on the temperature difference of the coldest and warmest month as well as the relation of T_S to EVI. The maximum PAR is chosen by the selection procedure as a further climate variable. The length of the vegetation period and the variable indicating deciduous broadleaf forests, complement the variables with which SVR can explain the T_{opt}-variations best. Removing the vegetation period attribute has by far the strongest negative effect upon the SVR performance within this combination.

Attributes describing seasonal dynamics account for five of the eight attributes for the parameter k, which indicates the rate of change of the temperature function, f_T (Eq. 4.14): the start of the growing season, the EVI at the end of the vegetation period, the seasonal
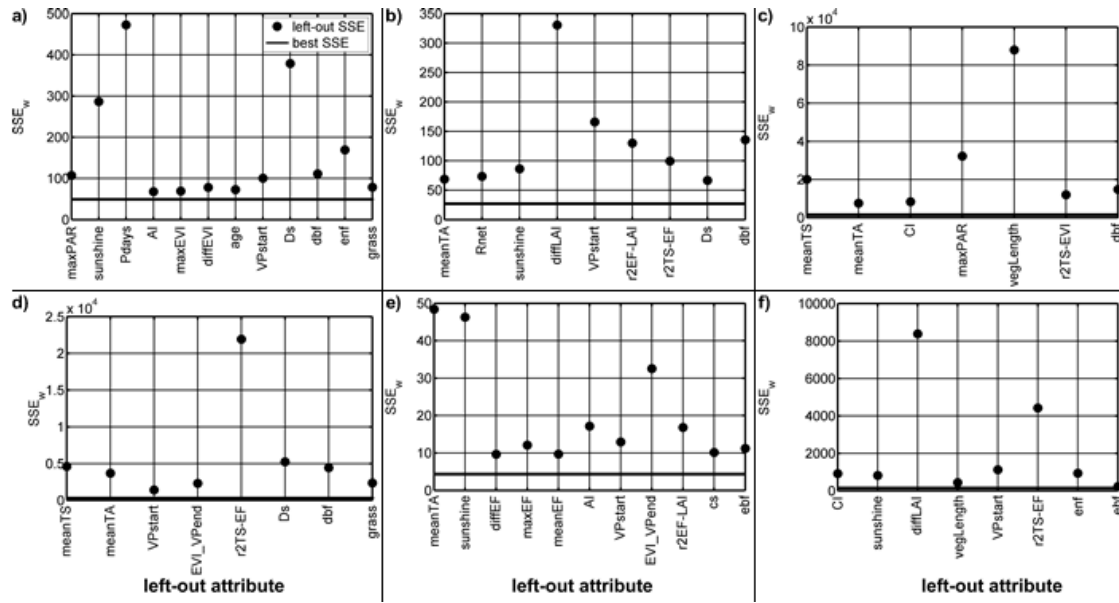
Figure 5.2: The quality criteria SSE_w for each selected attribute when it has been left out for the model parameters a) ϵ_{max}, b) p, c) T_{opt}, d) k, e) EF_I, f) α. The thick black line indicates the SSE_w-value when all attributes have been used.

correlation between T_S and EF as well as the attributes for the Köppen-Geiger main type "D" and deciduous broadleaf vegetation class. The mean air and soil temperatures and the grass vegetation class are additionally selected to extrapolate k. Omitting the temperature-EF-correlation worsens the SVR extrapolation capability most.

Eleven attributes have to be applied to extrapolate the inflection point of the moisture function EF_I whereas four of them are directly related to EF characteristics (amplitude, maximum, mean EF and the correlation between EF and T_S) and one attribute is the aridity index. The two general climate features, mean air temperature and fraction of sunshine, are also among the selected attributes and finally the start of the vegetation period, the EVI at its end, as well as the vegetation class "evergreen broadleaf forest" and the Köppen-Geiger subtype "c" indicating cool and short summers in temperate and continental climates. The mean air temperature and the fraction of sunshine, respectively, have the strongest negative impact on the quality measure SSE_w when applying the SVR extrapolation without them.

The lag parameter α, finally, is extrapolated with five out of eight attributes being related to seasonal dynamics: The continentality index depending on the temperature amplitude, the LAI amplitude, the length and start of the vegetation period, and the correlation between T_S and EF. The average fraction of sunshine and the vegetation class attributes "evergreen needleleaf forest" and "deciduous broadleaf forest" complement the set features explaining α-dynamics between sites. Omitting the LAI amplitude and the T_S-EF-correlation, respectively, deteriorate the extrapolation quality most.

5.3.2   SVR Performance

With the selected features, all six light use efficiency model parameters can be extrapolated with reasonable results in the cross-validation (Figure 5.3). r^2-values range from 0.68 for T_{opt} to 0.90 for EF_I, r_w^2-values from 0.60 for α to 0.94 for ϵ_{max} and EF_I, so an effect of the optimization with SSE_w is evident. The largest deviations occur for T_{opt} at Metolius Intermediate, for k at Oensingen, and for α at various sites, especially at coniferous forests, although their parameter uncertainty with regard to the model calibration is in the medium range. α and T_{opt}, the parameters with the worst SVR cross-validation performance, were also the parameters with the highest uncertainties.
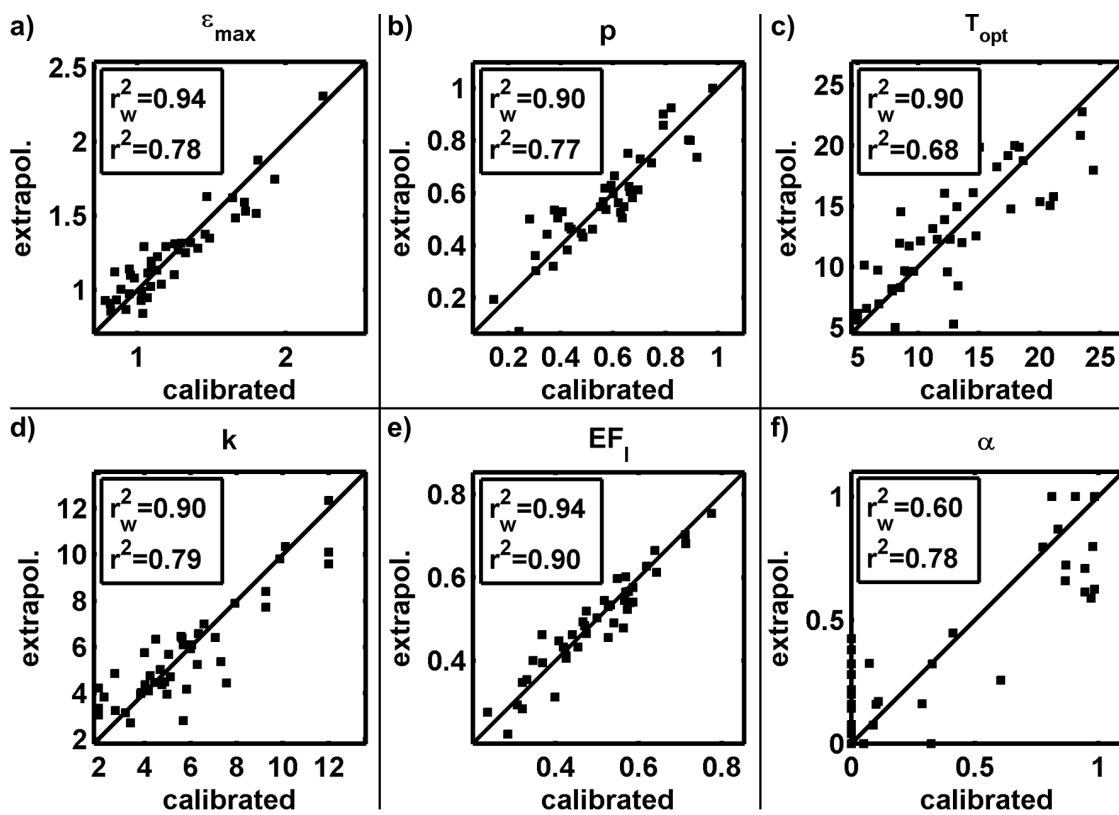
Figure 5.3: The calibrated model parameters compared to the parameters extrapolated with SVR by means of the respective other sites ("leave-one-out" cross-validation).

In addition to the cross-validation, the trained SVR model is run with all sites to determine the number of support vectors since they allow an additional evaluation of the generality capability of the SVR model. Using the SVR-parameter ϵ_{SVR} as determined in the above selection procedure leads to an virtually perfect match of the SVR model applied to all study sites, and the number of support vectors varies between 36 (k) and 41 (T_{opt}, α) support vectors. As a further test, the SVR-parameter ϵ_{SVR} is increased as long as the r^2-value is greater than 0.85, as compromise between precision and generality of the model, and then the number of support vectors is determined. This exercise led to numbers

of support vectors between 17 (EF_I) and 22 (p). Hence, the parameters having yielded better results in the cross-validation also tend to need less support vectors. And data points or sites, respectively, which experienced larger deviations in the cross-validation consequentially, served more often than others as support vectors since they obviously cannot be explained well by the other sites.

5.3.3   Light Use Efficiency Model Performance

More important is the performance of the final light use efficiency model, so it is run with the extrapolated parameters. Comparing the so derived F_G time series with the dynamics of the calibrated model time series reveals a high similarity of the variance indicated by high r^2-values between 0.91 and 1.00 (Table A.4). However, the time series are in some cases biased. The largest bias is found at Sylvania Wilderness with a value of 0.99 gC m^{-2} d^{-1}, accompanied by a RMSE of 1.71 gC m^{-2} d^{-1} with a F_G amplitude of about 13 gC m^{-2} d^{-1}. More positive than negative biases occur indicating an overprediction of fluxes; the median of the bias values is 6.7. RMSE-values normalized by the range of F_G-values are between 0.01 (Willow Creek) and 0.14 (Wind River, Blodgett).

The comparison with the measured F_G time series is certainly more relevant and shows F_G-values derived using the extrapolated parameters explaining large parts of the measured variations with r^2-values between 0.46 at Donaldson and 0.95 at UMBS; these sites had already the lowest and highest calibration performance. Indeed, the r^2-values are – with an average of 0.82 – very similar compared to those of the calibration ($r^2 = 0.84$), only in one case (Wind River), the coefficient of determination differs by more than 0.1 (Figure 5.4a). At this site, the temperature function is given more weight by the extrapolated p-value and its shape leads to higher light use efficiencies at high temperatures. This results in a strong over-prediction of F_G-fluxes in summer and is aggravated by an additional slight over-prediction of ϵ_{max}. Considering all sites, the biases range between 0.01 and 1.12; the latter occurs at Griffin having a F_G-range of about 12 gC m^{-2} d^{-1}. The frequency distribution of all biases yields a mode at 0.00 gC m^{-2} d^{-1} (with a class-width of ± 0.10), and a median of 0.08. RMSE-values range between 0.62 and 2.16 gC m^{-2} d^{-1} with a mean of 1.25 gC m^{-2} d^{-1}. Normalized RMSE-values (nRMSE, see Eq. 2.12) vary between 0.06 and 0.19; the latter is found again at Donaldson. Relating these nRMSE-values to those resulting from the model calibration, however, shows a stronger deterioration of this model performance measure due to the extrapolation process than for the r^2-values (Figure 5.4b). Hence, the correlations of the extrapolated time series are not strongly affected by the extrapolation, but the absolute deviation from the measured time series are, due to observed biases.

Such diagnostic models like the presented light use efficiency model are a suitable tool for carbon budgeting (e.g. Beer *et al.* (2010)). Consequently, not only simulating the temporal evolution of the gross primary production with an adequate accuracy is an important
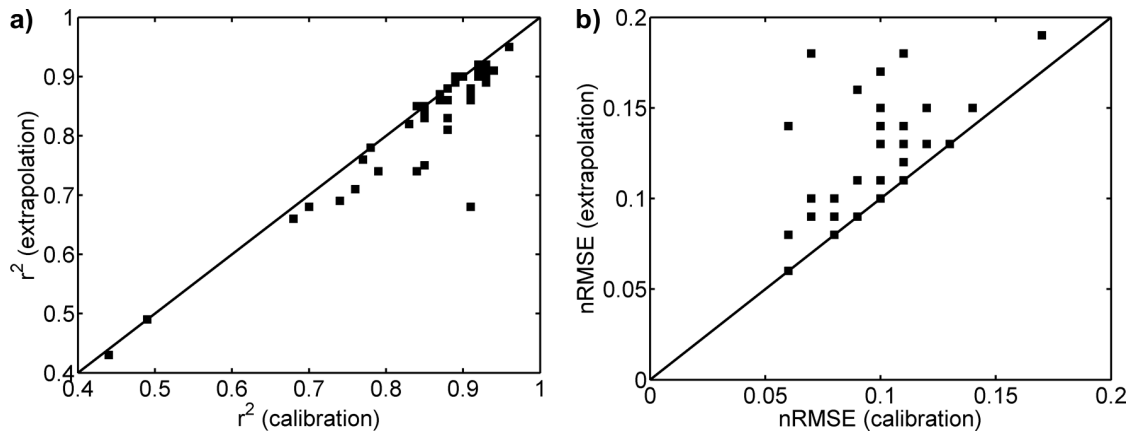
Figure 5.4: Comparison between the model quality measures r^2 and nRMSE of the model runs with the calibrated parameters and with the extrapolated parameters.

criterion for a light use efficiency model, but also its reliability to calculate the cumulative sum of carbon uptake at the considered area in a specific time window. Therefore, the cumulative sums of the modeled and extrapolated F_G time series are compared with the measured ones for the whole time series available; Figure 5.5 shows four examples of this comparison. The relative error of the sum of the carbon uptake per year (RE_M) is calculated: Its values range between -23 % and 37 % difference per year; these values with the lowest performance are found at Sylvania Wilderness and Griffin, both of which show large biases. Griffin already had a relatively low model calibration performance in terms of a quite large bias and wide uncertainty bounds of four parameters including those of the dominating subfunction f_T. At Sylvania Wilderness, in contrast, the model calibration performance was acceptable, only the uncertainty bounds of f_T were somewhat wider than usual; in this case, the extrapolation performance was not satisfying: Especially the parameters of f_T could not be adequately recaptured. Taking all sites into account, however, the mean of the relative differences between the cumulative sums (RE_S) is 1.8 %, so a small negative skewness of the RE_S-distribution is detected. Considering the absolute values of these errors (instead of averaging the positive and negative differences), the relative errors range between 0.4 % and 37 %, the median of these values is 11.8 %.

5.4   Discussion and Conclusions

Using support vector regression, the optimized parameters of a light use efficiency model have been related to climatic and biophysical site characteristics and evaluated with a "leave-one-out" cross-validation. Comparing the extrapolated parameters with the optimized ones at each site shows a good correlation with r_w^2-values between 0.6 and 0.94. The fluxes modeled with these parameters correlated very well with the fluxes
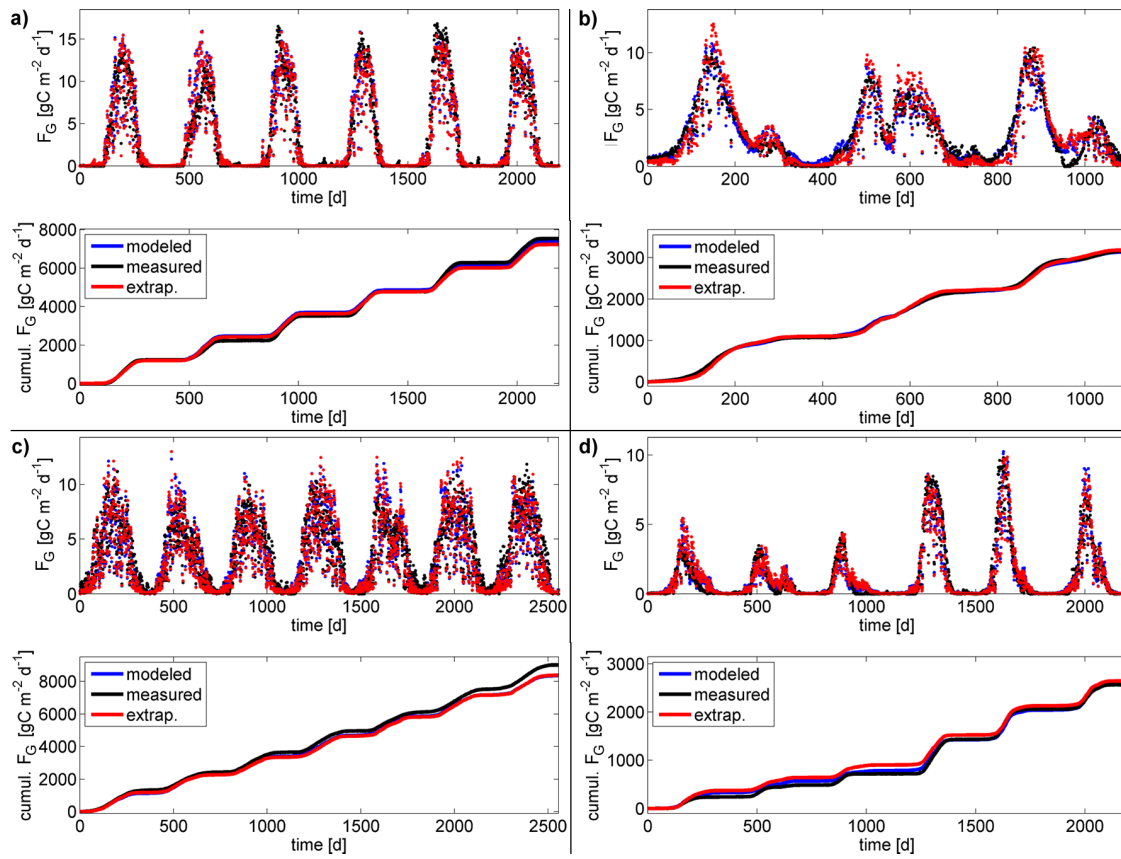
Figure 5.5: The measured (grey) and extrapolated (black) cumulative F_G time series exemplarily shown for four study sites: a) Willow Creek , b) Roccarespampani, c) Wetzstein, d) Lethbridge.

originating from the calibration model results (r^2-values > 0.91) but show biases of 13 % per year in average with respect to the annual cumulative sum. The better correlation of the fluxes compared to the cross-validation performance of the various model parameters themselves is a consequence of the fact that a large part of F_G-dynamics is explained by PAR or APAR itself (Jenkins *et al.*, 2007). The comparison of fluxes between those resulting from model runs with the extrapolated parameters and the measured F_G-values revealed r^2-values in the range of 0.43 and 0.95 what is only marginally lower than the r^2-values of the calibration performance in the most cases. Recapturing the fluxes at Donaldson where the lowest calibration and extrapolation performance occurred revealed also difficulties in the study of Yuan *et al.* (2007). Due to observed biases between the modeled and measured data sets, however, the normalized RMSE-values differ more. Nonetheless, the deterioration of the cumulative F_G-sums per year due to the extrapolation procedure is in average 7 %.

Considering other studies simulating the gross primary production on a daily basis and using similar performance measures, the results of this study compare very well (Table 5.2): Yuan *et al.* (2007) calibrated a light use efficiency model with 12 AmeriFlux sites and validated it with 16 other sites, which yielded in r^2-values of 0.84 and 0.77, respectively;

the relative validation error (RE_M) was about 18%. Yang *et al.* (2007) trained a SVR model with 36 AmeriFlux sites for the years 2000-2003 and predicted the fluxes of 2004 with a r^2-value of 0.71, a RMSE of 1.87 gC m^{-2} d^{-1} and an average error RE_S of 28 %. Both models out-performed NASA's MODIS algorithm for the gross primary production, which is probably the most prominent estimation of the gross primary production offering a global coverage. The calibration of two similar process based photosynthesis models using five consecutive measurement years of a FLUXNET forest site by Verbeeck *et al.* (2008) resulted in r^2-values of 0.72 and 0.73, a RMSE of about 2.3 gC m^{-2} d^{-1} and an error of 16-31 % per year with regard to the cumulative sums of gross primary production. It is clear that the results cannot be compared directly due to differences in the number and type of validation and calibration sites, but they reveal a tendency, though: The presented modeling and extrapolation scheme shows a somewhat better performance than previously proposed models with its average model calibration r^2 of 0.84, a cross-validation r^2 of 0.82, an average RMSE of 1.25 gC m^{-2} d^{-1}, a relative error (RE_S) per year of 13 % with regard to the cumulative sums and of 6 % with regard to the means (RE_M).

Table 5.2: Comparison of the model performance of the proposed model and extrapolation scheme with the calibration and validation performance of other studies using FLUXNET sites. For the various performance measures see section 2.4.

study	r^2		RMSE	RE_S/RE_M		notes
	calibr.	valid.	valid.	calibr.	valid.	
this	0.84	0.82	1.25	5%/6%	13%/5%	42 sites, "leave-one-out" cross-validation
Yuan *et al.* (2007)	0.84	0.77		-/17%	-/18%	12 forest sites for calibration, 16 for validation of a LUE model
Yang *et al.* (2007)		0.71	1.87		28%/-	36 sites, SVR model trained with data from 2000-2003, validated with 2004
Verbeeck *et al.* (2008)	0.78/ 0.80	0.72/ 0.73	2.33/ 2.38	2%/-	16-31% /-	Two leaf scale photosynthesis models applied at one site to 5 years; 1 of these years served for calibration

A particularly interesting part of this study is the automatic attribute selection by the SVR optimization scheme. The attribute sets for the respective parameters show that the selection is definitely not just arbitrary but appears to follow biophysically meaningful patterns. For ϵ_{max}, the maximum PAR and the average annual fraction of sunshine are the most intuitive and directly related explaining attributes; the fraction of sunshine, the

cloud coverage and diffuse PAR, respectively, have been discussed several times in the last years in this context and shown to be an important factor influencing the light use efficiency. The three main vegetation classes were also selected; the vegetation type has often been shown to strongly influence ϵ_{max} (Running *et al.*, 2004). Garbulsky *et al.* (2010) showed in their comprehensive analysis that ϵ_{max} is also determined by the vegetation type, but in the first instance, by precipitation. It is therefore not surprising that two attributes including the number of days with precipitation and the aridity index account for this climatic variable. The inclusion of the two EVI-attributes are supported by studies of Wu *et al.* (2010b) or Sims *et al.* (2008) who found EVI to be capable of capturing light use efficiency variations. The timing of the start of the growing was shown by Falge *et al.* (2002) and Schwalm *et al.* (2006) to influence maximum carbon uptake especially in boreal climates; the respective Köppen-Geiger-class of the latter was also selected for the SVR extrapolation. The effect of stand age, finally, has also often been shown to be important for light use efficiency differences between sites e.g. by Desai *et al.* (2008a).

The outcome of the selection process of the other five parameters cannot be directly compared to other studies since they are model specific. But it can be discussed whether the sets of selected attributes appear to be biophysically meaningful and reveal biophysical characteristics of the respective parameters: Additionally to the three general climate characteristics, mean air temperature, average annual fraction of sunshine and annual radiation net balance, the parameter p balancing temperature and moisture influences is explained by attributes describing the seasonality of the sites as well as the interdependence of temperature, EF and LAI. This selection is consistent with the fact, that EF as moisture surrogate is not a pure moisture indicator but integrates system dynamics: It is not only an index of water deficit and is connected with soil moisture and therewith precipitation, but it is also linked to the temperature gradient between surface and atmosphere and the biophysical process of stomatal carbon exchange (Schwalm *et al.*, 2010). Hence, accounting for the correlations between EF, LAI, and T_S when determining the magnitude of p and therewith the influence of EF and T_S is the logical consequence. Intimately connected with these considerations are attributes indicating strong seasonal dynamics due to dormant periods such as the D-Climate and the vegetation class deciduous forest because in these times the components can be decoupled from each other; i.e. the carbon assimilation of needle leaf forests can be very low due to low temperatures in D-climates despite a LAI greater than zero, and at deciduous forests F_G is not necessarily correlated with EF in the dormant period and there is no photosynthesis when the weather conditions are actually favorable but the trees are still bare-branched.

Four temperature attributes are chosen to explain the variations of the parameter T_{opt}. The attributes indicating the length of the vegetation period and deciduous broadleaf forests can be explained by the pattern observable in the study data: At coniferous forests higher light use efficiencies rather occur in spring or autumn when the temperature conditions are favorable but the solar radiation is not very high: The light use efficiency of deciduous

broadleaf forests, instead, follows more the course of the temperature. Here, the leaf development corresponds to the temperature increase in spring, which leads to higher light use efficiencies occurring at higher temperatures when the leaves have been fully developed. The parameter k as rate of change of the temperature function f_T is closely linked to the parameter T_{opt} and intrinsically connected to seasonality indicators. It is therefore consequential that attributes corresponding to these characteristics dominate this attribute set.

Four attributes predicting site variations of the inflection point EF_I of the moisture function are directly related to EF and aridity. The attribute "evergreen broadleaf forest" is also selected. The appearance of this vegetation class yet in this attribute set can be justified by the adaption of evergreen broadleaf forests to a warmer Mediterranean climate with elevated drought risks by a better water use efficiency (Pereira *et al.*, 2007) and thus a generally lower EF_I than the average.

Finally, all attributes of α but one (fraction of sunshine) are related to seasonal dynamics and indicate large or small seasonal differences of temperature or moisture. This makes sense considered the fact that stronger seasonal differences in environmental conditions tend to lead to lag effects of the reaction of plants to these driving forces. However, the capability of SVR to recapture the parameter α was lowest compared to the other parameters. This is especially true for coniferous forests, even if the parameter calibration uncertainties are considered; these were highest at coniferous and mixed forest sites. Furthermore, a bad reconstruction of α is found at sites having a rather low p-value, thus a higher influence of the EF-function. A reason for this could be that α is not always used by the model parameter optimization process to account for actual lag effects but that α is used to alleviate conceptual model deficiencies. The usage of this parameter has therefore to be reconsidered in further studies.

It has also been analyzed in this study, which attribute has the most negative effect on the performance of the extrapolation scheme when executing the SVR scheme without it. This certainly indicates the importance of the respective parameter but it is clear that this is only true for the specific set of features considered. Another important indicator for the importance of an attribute for a specific parameter is the frequency with which it was selected among all performed selection loops. For example, the mean air temperature and the fraction of sunshine have the largest negative impact on EF_I, but the EF-attributes seem to be at least as important. However, they compensate each other to a certain extent if one of them is left out. In every set of attributes resulting from the selection process there are at least two EF-features, but there are sets without the mean air temperature and even more without the fraction of sunshine.

Overall, as most frequently selected attribute the day of year of the start of the vegetation period appeared. In the only set of attributes it was not a member, the length of the vegetation period was instead. The binary attribute "deciduous broadleaf forest" was

found in four of the six sets. The mean air temperature, the average annual fraction of sunshine, the correlation between the time courses of temperature and EF as well as the continental Köppen-Geiger-climate class "D" explain each three of the six parameters. So amongst the most frequently selected attributes are two related to climate characteristics and four to characteristics related to seasonal dynamics; here, we number "D" and the attribute deciduous forest to this group since they indicate strong seasonality. If we assign all selected attributes to the categories climate, seasonality, physiology and vegetation class, the most fall into the climate category – approximately half of these attributes are related to moisture characteristics – followed by the physiology and seasonality categories with the equal number of matches. If we, however, also classify the attributes "D", the continentality index, the vegetation class deciduous forest and the amplitudes of the variables EF, LAI and EVI as indicators of seasonality, this category clearly dominates the attributes. This outcome is of practical importance for future diagnostic model building exercises.

6

Synthesis and Conclusions

6.1 The Scientific Context

Research efforts to study the global carbon cycle and its interactions with ecosystems and the climate system have not lost their relevance since the beginning of this work. Quite the contrary, describing and quantifying the exchange of carbon with respect to terrestrial ecosystems is more than ever subject to research projects (Jung *et al.*, 2009; Beer *et al.*, 2010; Huemmrich *et al.*, 2010; Mahecha *et al.*, 2010; Rotenberg & Yakir, 2010; Zhao & Running, 2010). The reasons for this tendency cannot only be found in the increasing awareness of the pressing challenges mankind is facing (e.g. Godfray *et al.*, 2010; Immerzeel *et al.*, 2010; Reid *et al.*, 2010; Vörösmarty *et al.*, 2010; Godfray *et al.*, 2011) but also in the increased availability of data provided by international networks such as FLUXNET as well as in advances in the field of remote sensing being made. The growing FLUXNET network makes it possible to address research questions that could not be pursued with data from a single or even a few measurement sites and boosts the exchange of scientific ideas and methodologies. The spectral and spatial resolution of satellite observations with global coverage improves step by step (Schaepman *et al.*, 2009; Ustin & Gamon, 2010). The advantages of these technological advances are obvious: Multiple sensors acquire information on the land surface and atmosphere around the world every day in a relatively economic way and they are furthermore the only source of such comprehensive data collections. Models for large-scale applications in a spatially continuous mode consequently rely on remote sensing data. Ground-based measurement networks, in turn, serve as indispensable data sources for model development, calibration and validation. The integration of remote sensing, ground measurements and modeling enables a powerful interplay of these three cornerstones of ecological research and offers a comprehensive view on ecosystem processes from local to global scales (Running *et al.*, 1999; Knorr & Heimann, 2001b; Turner *et al.*, 2004a; Friend *et al.*, 2007; Chambers *et al.*, 2007).

This thesis set out to exploit this synergetic potential in order to derive a robust, widely applicable gross primary productivity model on basis of FLUXNET data. The model development aimed at offering the possibility for the extrapolation of the calibrated model parameters to larger areas, thus the model was intended to be suitable for the application with remote sensing data.

6.2   Model Development

To achieve this aim, a data-driven model building approach was chosen. It is based on the theses (i) that the temporal and spatial scale of process descriptions used in many ecosystem models often differs significantly from the model application scale, and (ii) that the level of complexity of typical SVAT models can hardly be supported by the information content of available measurement data. The consequences of these mismatches become evident when calibrating the often numerous model parameters against available data: Much of the model functionality is possibly unconstrained by the calibration process rendering any subsequent predictions from these schemes uncertain. This can occur even though the functional form of the process descriptions being implemented are valid at the scale which they were derived at. Equifinality, i.e. the non-uniqueness of identified parameter sets, has proven to be particularly problematic when the aim of model calibration is to derive site-specific parameters that are to be regionalized by means of site characteristics. Against this background, the formulation of a diagnostic light use efficiency model as widely accepted concept for calculating the gross primary production on larger scales was based on a data-driven modeling strategy in this thesis. In this sense, this work is in line with the forefront of the respective research community (Jung *et al.*, 2008; Xiao *et al.*, 2008a; Beer *et al.*, 2009; Houborg *et al.*, 2009; Jung *et al.*, 2009; Turner *et al.*, 2009; Young & Ratto, 2009; Lu & Zhuang, 2010; Rastetter *et al.*, 2010).

Non-linear analysis tools were applied to extract deterministic information as expressed in measured CO_2-fluxes. The tools were explicitly designed to enable the formulation of model structures matching the information content of measurement data afflicted with noise while making minimal prior assumptions. The extraction of deterministic signals of gross primary production in response to environmental forcings as expressed in the available data was supported by splitting the CO_2 flux measurements in a systematic and stochastic component, whereas the first component was used for the analyses. Indeed, a suitable, simple but robust model structure for the gross CO_2-uptake could be deduced from the observations. The derived model is driven by incoming photosynthetically active radiation, its fraction absorbed by vegetation, and two subfunctions depending on the temperature as well as a water availability measure. The latter is represented by either the evaporative fraction, the antecedent precipitation index, vapor pressure

deficit, or the soil moisture. A variable "influence factor" is assigned to each of these two subfunctions for the temperature and water availability and calibrated along with the other model parameters at each study site. In this way, the model is applicable at sites with a strong seasonal temperature variation such as in boreal climates, as well as at arid sites dominated by the absence or availability of water.

The calibration of the model parameters at each study site individually followed the assumptions that (i) the seasonal behavior of canopies varies between vegetation and climate classes, (ii) the influence of explaining variables differs, and that (iii) there is no maximum light use efficiency valid for all ecosystems or vegetation classes. Hence it was assumed that no universal parameter set explaining the variation of CO_2 uptake of all vegetation types in every climate class exists. Overall, the calibration resulted in robust parameter estimates with small uncertainty bounds. However, the parameters of the sub-function describing the non-dominating system state (temperature or water availability, depending on the site characteristics) are more prone to calibration uncertainties than the parameters of the respective other, dominating subfunction. A great advantage of the model is the fact that this secondary subfunction with the rather uncertain parameters has a less influence on the model results due to the smaller contribution factor assigned to it in the additive formulation. Consequently, the results are less corrupted by parameter uncertainties. This characteristic distinguishes the proposed model from other light use efficiency models in which the subfunctions describing limiting environmental conditions are typically multiplied with the maximum or potential light use efficiency nonetheless of their actual explanatory power.

The proposed model demands only variables which can potentially be taken from re-analysis databases like ERA-Interim from the ECMWF, the NASA Data Assimilation Office (DAO) or derived by remote sensing for larger scale applications. Possibilities for the latter are outlined shortly in the following: Photosynthetically active radiation has been estimated with data from the GOES (Geostationary Operational Environmental Satellite) platform (Gu & Smith, 1997) and – for a better spatial resolution – by MODIS data (Liang *et al.*, 2006; Liu *et al.*, 2008). The land surface temperature, often used as a soil temperature surrogate (Sims *et al.*, 2008), can be retrieved from almost all earth observing satellites operating in the near infrared such as MODIS, NOAA-AVHRR, SEVIRI (Spinning Enhanced Visible and Infrared Imager) or GOES (Dash *et al.*, 2002; Sobrino & Romaguera, 2004; Wan, 2008). Explicit soil temperature estimates under vegetation stands are usually retrieved within land surface model and data assimilation frameworks (Anderson *et al.*, 2008); Huang *et al.* (2008) presented an explicit retrieval scheme using MODIS data. A calculation of the evaporative fraction by solely using MODIS atmospheric and land products was proposed by Wang *et al.* (2006); Venturim *et al.* (2008). A proxy for the soil moisture can be estimated with thermal infrared remote sensing data (Crow *et al.*, 2008; Hain *et al.*, 2009); passive and active microwave sensors, finally, serve as data-basis for

direct soil moisture retrievals (Reichle *et al.,* 2007; Wagner *et al.,* 2007; de Jeu *et al.,* 2008; Loew, 2008).

Overall, the proposed light use efficiency model proved to be a simple scheme which is attractive in its generality and simplicity. With its help, the gross flux of carbon uptake by desert grasslands as well as by boreal forests can be assessed likewise. Despite its simplicity, the model showed to capture a major proportion of the day-to-day variations in the gross CO_2 uptake at 44 forest and grassland sites with largely well defined parameters. In this sense, the derived model fulfills the primary requirements placed on the sought model. Amongst the water availability surrogates tested, the evaporative fraction calculated using dynamic linear regression turned out to be the most successful variable in explaining the light use efficiency parameter. It appears to incorporate more information on ecosystem processes than just information on the current water availability.

Due to its empirical nature, the parameter sets will get even more robust the longer available time series are. However, if longer time series are analyzed, new dynamics can become apparent. These have to be captured by the flexible modeling methodology, which enables the incorporation of new information as an inherent characteristic. In both cases, yet a further performance increase with regard to predictions can be expected with more data available. Generally, the benefit of a refinement of the model structure and the integration of additional internal and external drivers could be tested in a future study to achieve an even higher explanatory power for all study sites and a significantly improved performance for those sites at which currently the model achieves only a moderate performance. Possible model variables include the often discussed influence of leaf nitrogen concentrations (Sinclair & Horie, 1989; Dewar, 1996; Kergoat *et al.,* 2008; Mäkelä *et al.,* 2008; Ollinger *et al.,* 2008) as intrinsic variable, the saturating behavior of ϵ for high PAR values (Ruimy *et al.,* 1995; Turner *et al.,* 2003b; 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). Using the evaporative fraction as system state variable already takes the tight coupling of photosynthesis and transpiration into account. Further exploiting this relationship could form the basis for a strategy to further improve the model structure for example via the water use efficiency (Schulz & Jarvis, 2004; Stoy *et al.,* 2008; Jassal *et al.,* 2009; Beer *et al.,* 2009; Lu & Zhuang, 2010).

6.3   Extrapolation

The calibrated parameters of the derived light use efficiency model showed physiologically meaningful patterns, a fact which led to the conclusion that the model is suitable to serve as basis for regionalization strategies to perform the step from the point to the area. So, as a last working step this thesis attempted to relate the model parameters to site

characteristics to extrapolate them to unobserved sites. This serves the final aim of regionalizing the model parameters in a spatially continuous way in order to avoid a coarse grouping of the parameters to a few vegetation and climate classes. To do so, support vector regression, a supervised machine learning technique exploiting structural risk minimization, was applied. The model parameters were related to site characteristics obtained by remote sensing or extracted from the FLUXNET data base; the latter features were selected such that they can potentially also be retrieved by remote sensing data.

A cross-validation carried out showed the applied extrapolation scheme to be suitable to extrapolate the optimized parameters of the proposed light use efficiency model. The variations of the model parameters between the study sites could be recaptured in a cross-validation with reasonable precision in most cases. The resulting time series of carbon uptake modeled with the extrapolated parameters yielded in good correlations to the measured values. The model performance deteriorated slightly compared to the original model with the optimized parameters. A bias was in some cases introduced leading to deviations of the annual sum of assimilated carbon of 12 % in average. However, this bias is smaller than that observed for comparable models.

The composition of the attribute sets retrieved by an automatic selection algorithm appear to be reasonable. The sets are not of a mere empirical nature, but also seem to carry a particular biophysical meaning. Those attributes related to seasonality characteristics dominate. This fact supports the extrapolation scheme applied and, in turn, attests a certain biophysical meaning to the parameters of the light use efficiency model and shows them not purely being of empirical nature. The selection procedure showed that the extrapolation scheme could also have performed well with some other sets of attributes for each parameter, but it became also obvious from the outcome that SVR cannot extrapolate the parameters with arbitrarily chosen sets of attributes. This fact undermines the argument SVR is not suitable for such an extrapolation task, because it can perform with any variable attribute combination via the high dimensional feature space.

The exercise which challenged the SVR model with an increase of the SVR parameter responsible for the generalization capability shows a promising potential of the proposed extrapolation methodology in this regard. But it became evident that the number of sites is at its minimum for the applied extrapolation scheme in order to adequately capture the dynamics in response to the manifold boundary conditions across the different sites. Certainly, an evaluation with an x-fold cross-validation rather than a mere "leave-one-out" cross-validation is desired; the obtained results indicate that this is possible with an increase of the number of study sites. So this thesis arouses curiosity about performing this exercise with a larger and more comprehensive data set.

Every year, this task is more feasible: The FLUXNET data base comprises meanwhile over 500 towers worldwide (Figure 6.1). Most of the towers are still located in North

America and Europe but more and more measurement sites are established on the other continents, too. Thus FLUXNET successively comprises combinations of vegetation types and environmental conditions that have not been covered so far. These circumstances facilitate an up-scaling of the measured fluxes to larger scales with data-driven methods in the future. What is equally important, the measurement time series become longer. Since the beginning of this thesis, the number of sites with at least three measurement years increased by at least 60 %. These developments make data-driven approaches more and more practical, since these methodologies suffer from the limitation that derived models and parameters are prone to uncertainty when applied outside the range of environmental conditions the observations were made. Since data-led modeling approaches not only depend on a data basis covering sufficient environmental conditions but also on the quality and comparability of the data from different sources, another great potential to be explored in this context is given by the harmonized and standardized FLUXNET "La Thuile" data set. The provision of ancillary site information such as biological, disturbance and soil data by principle investigators also continues to evolve within FLUXNET. Such data are of particular importance for extrapolation purposes. As a matter of principle, not only the establishment of more towers and the generation of prolonged measurement time series but also the willingness to share the data with the scientific community will be the key to success of FLUXNET and comparable synthesis projects.
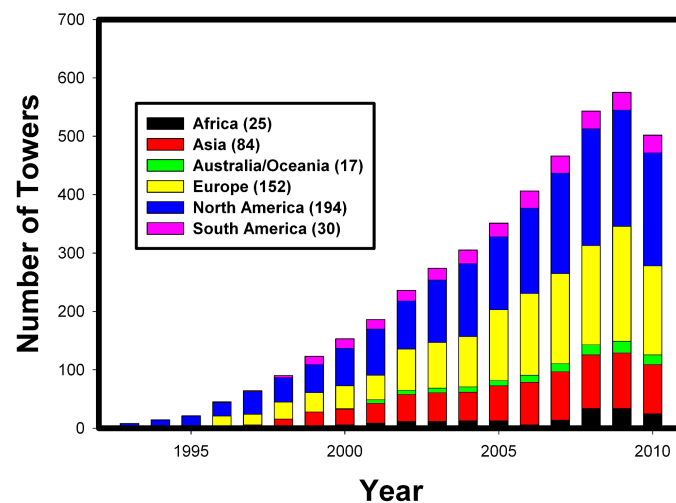
Figure 6.1: The growth of FLUXNET illustrated by the numbers of active towers per year and continent as of March 2010.

6.4 Remote Sensing

For the proposed model as well as for the subsequent extrapolation scheme, the inclusion of the fraction of photosynthetically active radiation, FPAR, and the leaf area index, LAI, which were obtained from the MODIS sensor, was essential. Therefore, this thesis also dealt with post-processing issues of the respective MODIS product containing subsets regridded specifically for the usage with FLUXNET data (Horn & Schulz, 2010). The product utilization was discussed from an end-user point of view. Usage options with regard to quality filtering, spatial aggregation and sensor choice were analyzed and their effects on the resulting LAI time series were examined. It turned out that spatial aggregations within the subset around the measurement tower and consideration of both MODIS sensors are helpful in terms of a more reliable and somewhat less noisy LAI time series. It is furthermore recommended to handle the quality criteria that come along with the product with care. This is particularly true for the values classified as "best" in case of forests. But nevertheless, even if product problems and possible pitfalls for the end user have been revealed by this and other studies, this extensive MODIS database belongs without doubt to the most valuable information sources on vegetation dynamics the model community has to date.

As elaborated above, it is important to use the MODIS LAI/FPAR product with caution. As a matter of fact, this is true for all remote sensing products and therefore, the same study should also be carried out for the MODIS vegetation index products used in this study. Furthermore, beside the analyzed effects from spatial aggregation, sensor choice and quality criteria application on the LAI/FPAR time series, the selection of the land classes considered for the aggregation needs further attention: In this thesis, various land classes were considered as similar against the background of an often inexact classification by the applied MODIS product. The benefits of applying a stricter selection could be examined in a further study. In this context it shall be noted that hyperspectral remote sensing (Figure 6.2b) – as will be provided by the German satellite mission "EnMAP" (Stuffler *et al.*, 2007) – makes an enhanced approach possible: Instead of the categorization of a few broad vegetation classes as retrieved by MODIS data, hyperspectral data offer novel ways of identifying optically detectable plant functional types (Ustin & Gamon, 2010) particularly for the discrimination of processes and patterns regulating the CO_2 uptake (Shaver *et al.* (2007), Figure 6.2b).

The potential offered by remote sensing has not been fully tapped in this thesis. Firstly, this would simply have gone beyond the scope of it, and secondly, data with a higher resolution were not available. Remotely sensed parameters and indices offer relations to the functionally and status of ecosystems and thus photosynthesis and should be tested in a further study within the proposed extrapolation framework. Among them are estimates of the content of leaf pigments such as carotenoide, anthocyanide and chlorophyll (Feret *et al.*, 2008; Rascher & Pieruschka, 2008; Ustin & Gamon, 2010). The
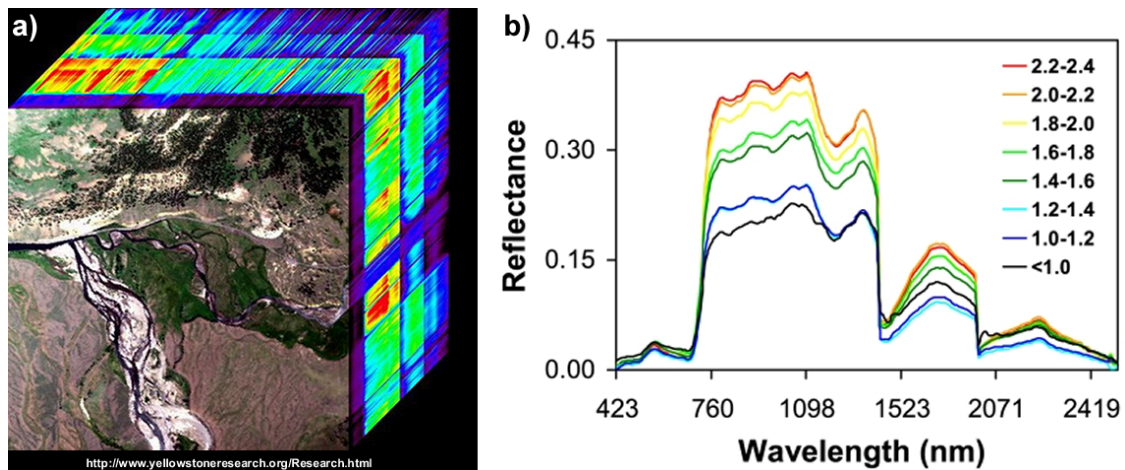
Figure 6.2: a) A "data cube" of the airborne hyperspectral sensor HyMap. The sensor has 126 bands between 0.45 and 2.5 μm and is often considered as best hyperspectral airborne sensor currently available. The EnMap sensor is expected to provide such high quality data from a satellite platform and therewith opening up entirely new possibilities for ecosystem observation and modeling. b) Exemplary reflectance spectra of forest canopies with different nitrogen contents (%) obtained from the airborne AVIRIS (Airborne Visible/Infrared Imaging Spectrometer; adapted from Ollinger *et al.* (2008)). New satellite imaging spectrometers could offer the possibility of exploiting information on leaf nitrogen and pigments over spatial and temporal continuous scales. This information can help to distinguish between specific plant functional types as compromise between broad vegetation classes and species types.

photochemical reflectance index (PRI) is of increasing interest to ecologists: It is sensitive to activities of the xanthophyll cycle responsible for the dissipation of excess energy and is therefore a good indicator for stress; it has been directly related to the light use efficiency (Barton & North, 2001; Garbulsky *et al.*, 2008; Hilker *et al.*, 2009, 2010). The advent of hyperspectral satellite remote sensing and therewith more precise information promises to improve these relationships (Rahman *et al.*, 2001, 2003; Rahman & Gamon, 2004; Fuentes *et al.*, 2006; Schaepman *et al.*, 2009) and thus also the presented modeling and extrapolation scheme. Combining the mentioned remotely sensed parameters and indices with information on water availability and other ecosystem information holds an enormous potential to find ways to derive detailed maps of carbon uptake. Producing those maps for spatially continuous areas defines the next research step beyond this thesis; advances in remote sensing offer the best conditions for this task. However, to fully exploit this potential, a more thorough integration of remote sensing techniques and ecological theory than it generally has been the case is required in future (Ustin & Gamon, 2010; Ollinger, 2010).

6.5   Further Research Needs and Outlook

A regionalization strategy to spatially continuous scales also implies the consideration of vegetation classes other than forests and grasslands. In further research activities, modeling the carbon uptake of these vegetation classes has to be addressed; these include shrublands, peatlands, wetlands, tundra, steppe dominated by C4-plants, croplands, and tropical forests holding giant reservoirs of carbon and being crucial to the world's climate (Zhao & Running, 2010).

The light use efficiency concept was first developed in the tropics (Monteith, 1972) but research within the FLUXNET community has primarily focused on temperate, boreal and Mediterranean forests, where the majority of the measurement towers are located and which cover large areas of the Earth. However, rainforests are a key element of the global carbon balance (Zhao & Running, 2010). Only in the last ten years longterm measurements above tropical vegetation have been started within the FLUXNET community; longterm data records for model building purposes and validation therefore have just not been available so far (Fisher *et al.*, 2009). With measurements being especially challenging, these regions might particularly benefit from hyperspectral satellite remote sensing campaigns (Papeş *et al.*, 2010; Asner & Alencar, 2010).

In the first years after its development, the light use efficiency concept was mostly adopted in the agricultural context for estimations of the efficiency of crop production (Monteith, 1977). The interest in the light use efficiency of crops is unbroken (Rascher & Pieruschka, 2008; Amthor, 2010; Rascher *et al.*, 2010). On the contrary, as formulated in the U.N.'s Millennium Development Goals, increasing the agricultural productivity – and thus the light and water use efficiency – is an imperative for science in the face of climate change and increasing water scarcity expected in many regions (Ziska & Bunce, 2007; Murchie *et al.*, 2009; Godfray *et al.*, 2010). The expanding role of biomass as an energy source and the increased market demand for energy crops competing with food production intensifies the need of understanding the conversion of light to biomass (Karp & Shield, 2008; Edgerton, 2009; Hill, 2009) and requirements for enhancing yields by developing more efficient ways to harness plants for meeting the growing demands (McCormick & Tjian, 2010). Agricultural areas furthermore contribute significantly to the release of greenhouse gases (Searchinger *et al.*, 2008). Monitoring the carbon and water exchange of agricultural lands and predicting crop carbon uptake and yields should therefore receive more attention in the FLUXNET community (Wang *et al.*, 2005c; Moureaux *et al.*, 2008; Aubinet *et al.*, 2009; Morgan *et al.*, 2010) despite the fact that more measurement towers are located in forests.

As stated above (see chapter 1), light and water use efficiency approaches are based on resource optimization theories. Recent studies have shown that plants and canopies can also be viewed from a thermodynamic point of view: Plants can be considered as members of a wider class of systems far from a thermodynamic equilibrium. The

principle of maximum entropy production (MEP) based on statistical mechanics and information theory states that non-equilibrium processes dissipate energy and produce entropy at the maximum possible rate (Kleidon, 2010). Following this MEP-principle, recent studies were able to unify different optimization theories (Dewar, 2010). MEP is therefore expected to contribute to an improved understanding of the biosphere and to lead to novel modeling approaches (Kleidon *et al.*, 2010; Schymanski *et al.*, 2010).

When spending efforts to quantify the global carbon cycle and make predictions for the future, the respiratory release of CO_2 from ecosystems as opposed flux to the CO_2 uptake, with which it sums up to the net carbon flux, has of course to be addressed, too. Moreover, better model descriptions for the respiration dynamics are still to be found, as e.g. Portner *et al.* (2009) made clear recently: They stated that there is still no consensus on the choice of the form of the response function that is used to describe the sensitivity of respiratory processes of terrestrial ecosystems to temperature. Furthermore, they noted a "lack of confidence in the parameter estimates of the temperature response" (Portner *et al.*, 2009, p. 3669), particularly for higher temperatures. Despite strong evidence for comparable short-term temperature sensitivities between ecosystems, their respiratory behavior was additionally found to differ significantly from each other on longer time scales. Furthermore, a complex low-frequency influence of photosynthetic carbon uptake on carbon release dynamics, which is not accounted for in typical models, was detected (Mahecha *et al.*, 2010). Data-led model building approaches resulting in a controlled model complexity matching the information content exploitable from the rich FLUXNET data base could contribute to these raised research needs, too.

By showing the potential of data-driven model development and a powerful extrapolation scheme, this thesis also paved the way for modeling evapotranspiration using these principles. With plants trading water for CO_2, evapotranspiration is closely connected to the carbon uptake addressed in this thesis and a basic component of the hydrological cycle, which is likewise a crucial global cycle of fundamental importance to mankind.

Bibliography

Agarwal D, Baldocchi D, Goode M, *et al.* (2008) An evolving La Thuile Fluxnet dataset and support infrastructure. In: *Geophysical Research Abstracts*, vol. 10. URL `http://bwc.berkeley.edu/Presentations/EGU2008-A-04835.pdf`.

Al-Saadi J, Soja AJ, Pierce RB, *et al.* (2008) Intercomparison of near-real-time biomass burning emissions estimates constrained by satellite fire data. *Journal Of Applied Remote Sensing*, **2**, 021504. doi:10.1117/1.2948785. URL `http://link.aip.org/link/?JRS/2/021504/1`.

Allard V, Ourcival JM, Rambal S, Joffre R, Rocheteau A (2008) Seasonal and annual variation of carbon exchange in an evergreen Mediterranean forest in southern France. *Global Change Biology*, **14**, 714–725. doi:10.1111/j.1365-2486.2008.01539.x.

AmeriFlux Network (Accessed 04/2009) American regional carbon flux network. URL `http://public.ornl.gov/ameriflux/`.

Ammann C (2010) personal comment.

Ammann C, Spirig C, Leifeld J, Neftel A (2009) Assessment of the nitrogen and carbon budget of two managed temperate grassland fields. *Agriculture, Ecosystems & Environment*, **133**, 150–162. doi:10.1016/j.agee.2009.05.006.

Amthor J (2010) From sunlight to phytomass: on the potential efficiency of converting solar radiation to phyto-energy. *New Phytologist*, **188**, 939–959. doi:10.1111/j.1469-8137.2010.03505.x.

Andersen J, Dybkjaer G, Jensen K, Refsgaard J, Rasmussen K (2002) Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. *Journal of Hydrology*, **264**, 34–50. doi:10.1016/S0022-1694(02)00046-X.

Anderson MC, Norman JM, Kustas WP, Houborg R, Starks PJ, Agam N (2008) A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. *Remote Sensing of Environment*, **112**, 4227–4241. doi:10.1016/j.rse.2008.07.009.

Anthoni P, Unsworth M, Law B, Irvine J, Baldocchi D, Tuyl S, Moore D (2002) Seasonal differences in carbon and water vapor exchange in young and old-growth ponderosa pine ecosystems. *Agricultural and Forest Meteorology*, **111**, 203–222. doi:10.1016/S0168-1923(02)00021-7.

Arneth A, Harrison S, Zaehle S, *et al.* (2010) Terrestrial biogeochemical feedbacks in the climate system. *Nature Geoscience*, **3**, 525–532. doi:10.1038/ngeo905.

Arya S (2001) *Introduction to micrometeorology*. Academic Press, San Diego, CA, 2nd edn.

Asner G, Alencar A (2010) Drought impacts on the Amazon forest: the remote sensing perspective. *New Phytologist*, **187**, 569–578. doi:10.1111/j.1469-8137.2010.03310.x.

Aubinet M, Chermanne B, Vandenhaute M, Longdoz B, Yernaux M, Laitat E (2001) Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes. *Agricultural and Forest Meteorology*, **108**, 293–315. doi:10.1016/S0168-1923(01)00244-1.

Aubinet M, Moureaux C, Bodson B, *et al.* (2009) Carbon sequestration by a crop over a 4-year sugar beet/winter wheat/seed potato/winter wheat rotation cycle. *Agricultural and Forest Meteorology*, **149**, 407–418. doi:10.1016/j.agrformet.2008.09.003.

Baldocchi D (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, **9**, 479–492. doi:10.1046/j.1365-2486.2003.00629.x.

Baldocchi D (2008) Turner review No. 15. Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*, **56**, 1–26. doi:10.1071/BT07151.

Baldocchi D, Falge E, Gu LH, *et al.* (2001) FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin Of The American Meteorological Society*, **82**, 2415–2434. doi: 10.1175/1520-0477.

Baldocchi D, Valentini R, Running S, Oechel W, Dahlman R (1996) Strategies for measuring and modelling carbon dioxide and water vapour fluxes over terrestrial ecosystems. *Global Change Biology*, **2**, 159–168. doi:10.1111/j.1365-2486.1996.tb00069.x.

Baldocchi DD, Hicks BB, Meyers TP (1988) Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology*, **69**, 1331–1340. doi:10.2307/1941631.

Baret F, Morissette JT, Fernandes RA, *et al.* (2006) Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: Proposition of the CEOS-BELMANIP. *IEEE Transactions on Geoscience and Remote Sensing.*, **44**, 1794–1803. doi:10.1.1.70.7416.

Barr AG, Black TA, Hogg EH, *et al.* (2007) Climatic controls on the carbon and water balances of a boreal aspen forest, 1994-2003. *Global Change Biology*, **13**, 561–576. doi: 10.1111/j.1365-2486.2006.01220.x.

Barry JM (1973) SPLINS and SMOOTH: Two Fortran routines smoothing routines. Tech. rep., Australian Atomic Energy Comission (AAEC), Research Establisment Lucas Heights.

Barton CVM, North PRJ (2001) Remote sensing of canopy light use efficiency using the photochemical reflectance index - model and sensitivity analysis. *Remote Sensing of Environment*, **78**, 264–273. doi:10.1016/S0034-4257(01)00224-3.

Beer C, Ciais P, Reichstein M, *et al.* (2009) Temporal and among-site variability of inherent water use efficiency at the ecosystem level. *Global Biogeochemical Cycles*, **23**, GB2018. doi:10.1029/2008GB003233.

Beer C, Reichstein M, Tomelleri E, *et al.* (2010) Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science*, **329**, 834–838. doi:10.1126/science.1184984.

Berbigier P, Bonnefond J, Mellmann P (2001) CO2 and water vapour fluxes for 2 years above Euroflux forest site. *Agricultural and Forest Meteorology*, **108**, 183–197. doi:10.1016/S0168-1923(01)00240-4.

Betts RA, Collins M, Hemming DL, Jones CD, Lowe JA, Sanderson MG (2011) When could global warming reach 4°c? *Philosophical Transactions of the Royal Society A: Mathematical,Physical and Engineering Sciences*, **369**, 67–84. doi:10.1098/rsta.2010.0292.

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

Boer E, de Beurs K, Hartkamp A (2001) Kriging and thin plate splines for mapping climate variables. *International Journal of Applied Earth Observation and Geoinformation*, **3**, 146–154. doi:10.1016/S0303-2434(01)85006-6.

Bonan G (2002) *Ecological climatology: concepts and applications*. Cambridge University Press, Cambridge.

Bonan GB (1993) Importance of leaf-area index and forest type when estimating photosynthesis in boreal forests. *Remote Sensing of Environment*, **43**, 303–314. doi:10.1016/0034-4257(93)90072-6.

Borak JS, Jasinski MF (2009) Effective interpolation of incomplete satellite-derived leaf-area index time series for the continental United States. *Agricultural and Forest Meteorology*, **149**, 320–332. doi:10.1016/j.agrformet.2008.08.017.

Bosveld F, Beljaars A (2001) The impact of sampling rate on eddy-covariance flux estimates. *Agricultural and Forest Meteorology*, **109**, 39–45. doi:10.1016/S0168-1923(01)00257-X.

Bradford J, Hicke J, Lauenroth W (2005) The relative importance of light-use efficiency modifications from environmental conditions and cultivation for estimation of large-scale net primary productivity. *Remote Sensing of Environment*, **96**, 246–255. doi:0.1016/j.rse.2005.02.013.

Breda NJJ (2003) Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany*, **54**, 2403–2417. doi:10.1093/jxb/erg263.

Budyko M (1958) *The heat balance of the earth's surface*. US Dept. of Commerce, Washington.

Burges C (1998) A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, **2**, 121–167.

CarboEuropeIP (Accessed 04/2009) Ecosystem component database. URL `http://gaia.agraria.unitus.it/database/carboeuropeip/`.

Carrara A, Janssens I, Curiel Yuste J, Ceulemans R (2004) Seasonal changes in photosynthesis, respiration and NEE of a mixed temperate forest. *Agricultural and Forest Meteorology*, **126**, 15–31. doi:10.1016/j.agrformet.2004.05.002.

Carrara A, Kowalski A, Neirynck J, Janssens I, Yuste J, Ceulemans R (2003) Net ecosystem CO2 exchange of mixed forest in Belgium over 5 years. *Agricultural and Forest Meteorology*, **119**, 209–227. doi:10.1016/S0168-1923(03)00120-5.

Chambers J, Asner G, Morton D, *et al.* (2007) Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests. *Trends in Ecology & Evolution*, **22**, 414–423. doi:10.1016/j.tree.2007.05.001.

Chang C, Lin C (2001) LIBSVM: a library for support vector machines. Tech. rep., Department of Computer Science, National Taiwan University, Taipei 106, Taiwan. URL `http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html`.

Chase T, Pielke R, Kittel T, Nemani R, Running S (1996) Sensitivity of a general circulation model to global changes in leaf area index. *Journal of Geophysical Research – Atmospheres.*, **101**, 7393–7408. doi:10.1029/95JD02417.

Chen B, Black T, Coops N, Hilker T, Trofymow J, Morgenstern K (2009) Assessing tower flux footprint climatology and scaling between remotely sensed and eddy covariance measurements. *Boundary-Layer Meteorology*, **130**, 137–167. doi:10.1016/j.agrformet.2010.09.005.

Chen B, Ge Q, Fu D, Yu G, Sun X, Wang S, Wang H (2010) A data-model fusion approach for upscaling gross ecosystem productivity to the landscape scale based on remote sensing and flux footprint modelling. *Biogeosciences*, **7**, 2943–2958. doi:10.5194/bg-7-2943-2010.

Chen JM, Black TA (1992) Foliage area and architecture of plant canopies from sun-fleck size distributions. *Agricultural and Forest Meteorology*, **60**, 249–266. doi:10.1016/0168-1923(92)90040-B.

Chen W, Black T, Yang P, *et al.* (1999) Effects of climatic variability on the annual carbon sequestration by a boreal aspen forest. *Global Change Biology*, **5**, 41–53. doi:10.1046/j.1365-2486.1998.00201.x.

Chen X, Vierling L, Deering D, Conley A (2005) Monitoring boreal forest leaf area index across a Siberian burn chronosequence: a MODIS validation study. *International Journal of Remote Sensing*, **26**, 5433–5451. doi:10.1109/TGRS.2005.853936.

Cheng Y, Zarco-Tejada P, Riaño D, Rueda C, Ustin S (2006) Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between AVIRIS and MODIS indexes. *Remote Sensing of Environment*, **105**, 354–366. doi:10.1016/j.rse.2006.07.005.

Chow V, *et al.* (1964) *Handbook of applied hydrology*. McGraw-Hill, New York.

Chuvieco E, Giglio L, Justice C (2008) Global characterization of fire activity: toward defining fire regimes from Earth observation data. *Global Change Biology*, **14**, 1488–1502. doi:10.1111/j.1365-2486.2008.01585.x.

Clement R, Moncrieff J, Jarvis P (2003) Net carbon productivity of Sitka spruce forest in Scotland. *Scottish Forestry*, **57**, 5–10.

Cleugh HA, Leuning R, Mu Q, Running SW (2007) Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sensing of Environment*, **106**, 285. doi:10.1016/j.rse.2006.07.007.

Cohen W, Maiersperger T, Turner D, *et al.* (2006) MODIS land cover and LAI collection 4 product quality across nine sites in the western hemisphere. *IEEE Transactions on Geoscience and Remote Sensing.*, **44**, 1843–1857. doi:10.1109/TGRS.2006.876026.

Cohen WB, Maiersperger TK, Yang ZQ, *et al.* (2003) Comparisons of land cover and LAI estimates derived from ETM+ and MODIS for four sites in North America: a quality assessment of 2000/2001 provisional MODIS products. *Remote Sensing of Environment*, **88**, 233–255. doi:10.1016/j.rse.2003.06.006.

Coleman T, Branch MA, Grace A (2010) Optimization toolbox™ 5 – user's guide.

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

Conrad V (1950) *Methods in climatology*. Harvard University Press, Cambridge, Mass., USA, 2nd edn.

Cook B, Davis K, Wang W, *et al.* (2004) Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA. *Agricultural and Forest Meteorology*, **126**, 271–295. doi:10.1016/j.agrformet.2004.06.008.

Coops NC, Black TA, Jassal RPS, Trofymow JAT, Morgenstern K (2007) Comparison of MODIS, eddy covariance determined and physiologically modelled gross primary production (GPP) in a Douglas-fir forest stand. *Remote Sensing of Environment*, **107**, 385–401. doi:10.1016/j.rse.2006.09.010.

Coops NC, Waring RH, Law BE (2005) Assessing the past and future distribution and productivity of ponderosa pine in the Pacific Northwest using a process model, 3-PG. *Ecological Modelling*, **183**, 107–124. doi:10.1016/j.ecolmodel.2004.08.002.

Cowling S, Jones C, Cox P (2009) Greening the terrestrial biosphere: simulated feedbacks on atmospheric heat and energy circulation. *Climate Dynamics*, **32**, 287–299. doi:10.1007/s00382-008-0481-8.

Cox P, Betts R, Jones C, Spall S, Totterdell I (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, **408**, 184–187. doi:10.1038/35041539.

Cox PM, Huntingford C, Harding RJ (1998) A canopy conductance and photosynthesis model for use in a GCM land surface scheme. *Journal Of Hydrology*, **212-213**, 79. doi:10.1016/S0022-1694(98)00203-0.

Crabtree R, Potter C, Mullen R, *et al.* (2009) A modeling and spatio-temporal analysis framework for monitoring environmental change using NPP as an ecosystem indicator. *Remote Sensing of Environment*, **113**, 1486–1496. doi:10.1016/j.rse.2008.12.014.

Cramer W, Bondeau A, Woodward F, *et al.* (2001) Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**, 357–373. doi:10.1046/j.1365-2486.2001.00383.x.

Crow W, Kustas W, Prueger J (2008) Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model. *Remote Sensing of Environment*, **112**, 1268–1281. doi:10.1016/j.rse.2006.11.033.

Dash P, Göttsche F, Olesen F, Fischer H (2002) Land surface temperature and emissivity estimation from passive sensor data: theory and practice-current trends. *International Journal of Remote Sensing*, **23**, 2563–2594. doi:10.1080/01431160110115041.

Dawson TP, North PRJ, Plummer S, Curran PJ (2003) Forest ecosystem chlorophyll content: implications for remotely sensed estimates of net primary productivity. *International Journal of Remote Sensing*, **24**, 611–617. doi:10.1080/01431160304984.

de Jeu R, Wagner W, Holmes T, Dolman A, van de Giesen N, Friesen J (2008) Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surveys in Geophysics*, **29**, 399–420. doi:10.1007/s10712-008-9044-0.

Demarty J, Chevallier F, Friend A, Viovy N, Piao S, Ciais P (2007) Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model. *Geophys. Res. Lett.*, **34**, L15402. doi:10.1029/2007GL030014.

Desai A, Bolstad P, Cook B, Davis K, Carey E (2005) Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA. *Agricultural and Forest Meteorology*, **128**, 33–55. doi:10.1016/j.agrformet.2004.09.005.

Desai A, Noormets A, Bolstad P, *et al.* (2008a) Influence of vegetation type, stand age and climate on carbon dioxide fluxes across the Upper Midwest, USA: implications for regional scaling of carbon flux. *Agricultural and Forest Meteorology*, **148**, 288–308. doi:10.1016/j.agrformet.2007.08.001.

Desai A, Richardson A, Moffat A, *et al.* (2008b) Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. *Agricultural and Forest Meteorology*, **148**, 821–838. doi:10.1016/j.agrformet.2007.11.012.

Dewar R (2010) Maximum entropy production and plant optimization theories. *Philosophical Transactions B*, **365**, 1429. doi:10.1098/rstb.2009.0293.

Dewar RC (1996) The correlation between plant growth and intercepted radiation: An interpretation in terms of optimal plant nitrogen content. *Annals of Botany*, **78**, 125–136. doi:10.1006/anbo.1996.0104.

Dolman A, Moors E, Elbers J (2002) The carbon uptake of a mid latitude pine forest growing on sandy soil. *Agricultural and Forest Meteorology*, **111**, 157–170. doi:doi:10.1016/S0168-1923(02)00024-2.

Doraiswamy PC, Sinclair TR, Hollinger S, Akhmedov B, Stern A, Prueger J (2005) Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment*, **97**, 192–202. doi:10.1016/j.rse.2005.03.015.

Duro D, Coops NC, Wulder MA, Han T (2007) Development of a large area biodiversity monitoring system driven by remote sensing. *Progress in Physical Geography*, **31**, 235–260. doi:10.1177/0309133307079054.

Edgerton M (2009) Increasing crop productivity to meet global needs for feed, food, and fuel. *Plant Physiology*, **149**, 7. doi:10.1104/pp.108.130195.

Falge E, Baldocchi D, Olson R, *et al.* (2001a) Gap filling strategies for long term energy flux data sets. *Agricultural and Forest Meteorology*, **107**, 71–77. doi:10.1016/S0168-1923(00)00235-5.

Falge E, Baldocchi D, Olson et al R (2001b) Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology*, **107**, 43–69. doi:10.1016/S0168-1923(00)00225-2.

Falge E, Baldocchi D, Tenhunen J, *et al.* (2002) Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. *Agricultural and Forest Meteorology*, **113**, 53–74. doi:10.1016/S0168-1923(02)00102-8.

Fan S, Gloor M, Mahlman J, Pacala S, Sarmiento J, Takahashi T, Tans P (1998) A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models. *Science*, **282**, 442. doi:10.1126/science.282.5388.442.

Fang HL, Liang SL, Townshend JR, Dickinson RE (2008) Spatially and temporally continuous LAI data sets based on an integrated filtering method: Examples from North America. *Remote Sensing of Environment*, **112**, 75–93. doi:10.1016/j.rse.2006.07.026.

Farquhar G, Caemmerer S, Berry J (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta*, **149**, 78–90. doi:10.1007/BF00386231.

Feret J, François C, Asner G, *et al.* (2008) PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, **112**, 3030–3043. doi:10.1016/j.rse.2008.02.012.

Field CB (1991) *Ecological scaling of carbon gain to stress and ressource gain*, pp. 35–65. Academic Press, San Diego, CA.

Fisher J, Malhi Y, Bonal D, *et al.* (2009) The land-atmosphere water flux in the tropics. *Global Change Biology*, **15**, 2694–2714. doi:10.1111/j.1365-2486.2008.01813.x.

Flanagan L (2009) Phenology of Plant Production in the Northwestern Great Plains: Relationships with Carbon Isotope Discrimination, Net Ecosystem Productivity and Ecosystem Respiration. *Phenology of Ecosystem Processes*, pp. 169–185. doi:10.1007/978-1-4419-0026-5_7.

Flanagan LB, Wever LA, Carlson PJ (2002) Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland. *Global Change Biology*, **8**, 599–615. doi:10.1046/j.1365-2486.2002.00491.x.

FLUXNET (Accessed 10/2010) A global flux measurement network. URL `http://www.fluxnet.ornl.gov/fluxnet/graphics.cfm`.

Frank D, Esper J, Raible C, Büntgen U, Trouet V, Stocker B, Joos F (2010) Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. *Nature*, **463**, 527–530. doi:10.1038/nature08769.

Franks S, Beven K, Quinn P, Wright I (1997) On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. *Agricultural and Forest Meteorology*, **86**, 63–75. doi:10.1016/S0168-1923(96)02421-5.

Friedl M, McIver D, Hodges J, *et al.* (2002) Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, **83**, 287–302. doi:10.1016/S0034-4257(02)00078-0.

Friedlingstein P, Prentice IC (2010) Carbon-climate feedbacks: a review of model and observation based estimates. *Current opinion in environmental sustainability*, **2**, 251–257. doi:{10.1016/j.cosust.2010.06.002}.

Friend AD, Arneth A, Kiang NY, *et al.* (2007) FLUXNET and modelling the global carbon cycle. *Global Change Biology*, **13**, 610–633. doi:10.1111/j.1365-2486.2006.01223.x.

Fuentes DA, Gamon JA, Cheng Y, *et al.* (2006) Mapping carbon and water vapor fluxes in a chaparral ecosystem using vegetation indices derived from AVIRIS. *Remote Sensing of Environment*, **103**, 312. doi:10.1016/j.rse.2005.10.028.

Funk J, Vitousek P (2007) Resource-use efficiency and plant invasion in low-resource systems. *Nature*, **446**, 1079–1081. doi:10.1038/nature05719.

Gamier P (2006) Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling & Software*, **21**. doi:10.1016/j.envsoft.2005.05.007.

Gamon JA, Rahman AF, Dungan JL, Schildhauer M, Huemmrich KF (2006) Spectral Network (SpecNet) - What is it and why do we need it? *Remote Sensing of Environment*, **103**, 227. doi:10.1016/j.rse.2006.04.003.

Gao F, Morisette J, Wolfe R, *et al.* (2008) An algorithm to produce temporally and spatially continuous MODIS-LAI time series. *IEEE Geosciences and Remote Sensing Letters.*, **5**, 60. doi:10.1109/LGRS.2007.907971.

Garbulsky M, Peñuelas J, Papale D, *et al.* (2010) Patterns and controls of the variability of radiation use efficiency and primary productivity across terrestrial ecosystems. *Global Ecology and Biogeography*, **19**, 253–267. doi:10.1111/j.1466-8238.2009.00504.x.

Garbulsky MF, Penuelas J, Papale D, Filella I (2008) Remote estimation of carbon dioxide uptake by a Mediterranean forest. *Global Change Biology*, **14**, 2860–2867. doi:10.1111/j.1365-2486.2008.01684.x.

Gholz H, Clark K (2002) Energy exchange across a chronosequence of slash pine forests in Florida. *Agricultural and Forest Meteorology*, **112**, 87–102. doi:10.1016/S0168-1923(02)00059-X.

Giglio L, Van Der Werf G, Randerson J, Collatz G, Kasibhatla P (2006) Global estimation of burned area using MODIS active fire observations. *Atmospheric Chemistry and Physics*, **6**, 957–974. doi:10.5194/acp-6-957-2006.

Godfray H, Beddington J, Crute I, *et al.* (2010) Food security: the challenge of feeding 9 billion people. *Science*, **327**, 812. doi:10.1126/science.1185383.

Godfray HCJ, Pretty J, Thomas SM, Warham EJ, Beddington JR (2011) Linking Policy on Climate and Food. *Science*. doi:10.1126/science.1202899.

Goetz SJ, Prince SD (1999) Modelling terrestrial carbon exchange and storage: Evidence and implications of functional convergence in light-use efficiency. In: *Advances in Ecological Research*, vol. 28 of *Advances In Ecological Research*, pp. 57–92. Academic Press.

Goldstein A, Hultman N, Fracheboud J, *et al.* (2000) Effects of climate variability on the carbon dioxide, water, and sensible heat fluxes above a ponderosa pine plantation in the Sierra Nevada (CA). *Agricultural and Forest Meteorology*, **101**, 113–129. doi:10.1016/S0168-1923(99)00168-9.

Gough C, Vogel C, Schmid H, Curtis P (2008) Controls on annual forest carbon storage: lessons from the past and predictions for the future. *Bioscience*, **58**, 609–622. doi:10.1641/B580708.

Goulden ML, Winston GC, McMillan AMS, Litvak ME, Read EL, Rocha AV, Rob Elliot J (2006) An eddy covariance mesonet to measure the effect of forest age on land-atmosphere exchange. *Global Change Biology*, **12**, 2146–2162. doi:10.1111/j.1365-2486.2006.01251.x.

Gower S, Kucharik C, Norman J (1999) Direct and indirect estimation of leaf area index, f (APAR), and net primary production of terrestrial ecosystems. *Remote Sensing of Environment*, **70**, 29–51. doi:10.1016/S0034-4257(99)00056-5.

Granier A, Bréda N, Longdoz B, Gross P, Ngao J (2008) Ten years of fluxes and stand growth in a young beech forest at Hesse, North-eastern France. *Annals of Forest Science*, **65**, 704–704. doi:10.1051/forest:2008052.

Grieser J, Gommes R, Bernardi M (2006) New LocClim-the Local Climate Estimator of FAO. In: *Geophysical Research Abstracts*, vol. 8, p. 08305. doi:10.1111/j.1365-2486.2007.01455.x.

Groenendijk M, Dolman A, van der Molen M, *et al.* (2011) Assessing parameter variability in a photosynthesis model within and between plant functional types using global

Fluxnet eddy covariance data. *Agricultural and Forest Meteorology*, **151**, 22–38. doi: 10.1016/j.agrformet.2010.08.013.

Grünwald T, Bernhofer C (2007) A decade of carbon, water and energy flux measurements of an old spruce forest at the Anchor Station Tharandt. *Tellus B*, **59**, 387–396.

Gu J, Smith E (1997) High-resolution estimates of total solar and PAR surface fluxes over large-scale BOREAS study area from GOES measurements. *Journal of Geophysical Research*, **102**, 29685. doi:10.1029/96JD03706.

Gu L, Meyers T, Pallardy S, *et al.* (2006) Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site. *Journal of Geophysical Research – Atmospheres*, **16102**. doi: 10.1029/2006JD007161.

Gu LH, Meyers T, Pallardy SG, *et al.* (2007) Influences of biomass heat and biochemical energy storages on the land surface fluxes and radiative temperature. *Journal Of Geophysical Research – Atmospheres*, **112**. doi:10.1029/2006JD007425.

Gunn S (1998) Support vector machines for classification and regression. Tech. rep., University of Southampton, UK.

Hain C, Mecikalski J, Anderson M (2009) Retrieval of an available water-based soil moisture proxy from thermal infrared remote sensing. Part I: methodology and validation. *Journal of Hydrometeorology*, **10**, 665–683. doi:10.1175/2008JHM1024.1.

Hanan NP, Begue A (1995) A method to estimate instantaneous and daily intercepted photosynthetically active radiation using a hemispherical sensor. *Agricultural and Forest Meteorology*, **74**, 155–168. doi:10.1016/0168-1923(94)02196-Q.

Hao X, Qu JJ (2007) Saharan dust storm detection using moderate resolution imaging spectroradiometer thermal infrared bands. *Journal of Applied Remote Sensing*, **1**, 013510. doi:10.1117/1.2740039. URL `http://link.aip.org/link/?JRS/1/013510/1`.

He K, Zhang J, Zhang Q (2009) Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *Acta Oecologica*, **35**, 14–21. doi:10.1016/j.actao.2008.07.006.

Heimann M, Reichstein M (2008) Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*, **451**, 289–292. doi:10.1038/nature06591.

Heinsch FA, Zhao MS, Running et al SW (2006) Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. *IEEE Transactions on Geoscience and Remote Sensing.*, **44**, 1908–1925. doi: 10.1016/S0034-4257(02)00078-0.

Hess L, Ratana P, Huete A, Potter C, Melack J (2010) Use of MODIS enhanced vegetation index to detect seasonal patterns of leaf phenology in central Amazon várzea forest. In: *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, vol. 4. doi:10.1109/IGARSS.2009.5417550.

Hilker T, Coops NC, Wulder MA, Black TA, Guy RD (2008) The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements. *Science of the Total Environment*, **404**, 411–423. doi:10.1016/j.scitotenv.2007.11.007.

Hilker T, Hall F, Coops N, *et al.* (2010) Remote sensing of photosynthetic light-use efficiency across two forested biomes: Spatial scaling. *Remote Sensing of Environment*. doi:10.1016/j.rse.2010.07.004.

Hilker T, Lyapustin A, Hall FG, Wang Y, Coops NC, Drolet G, Black TA (2009) An assessment of photosynthetic light use efficiency from space: Modeling the atmospheric and directional impacts on pri reflectance. *Remote Sensing of Environment*, **113**, 2463–2475. doi:10.1016/j.rse.2009.07.012.

Hill J (2009) Environmental costs and benefits of transportation biofuel production from food-and lignocellulose-based energy crops: a review. *Sustainable Agriculture*, pp. 125–139. doi:10.1051/agro:2007006.

Hollinger D, Aber J, Dail B, *et al.* (2004) Spatial and temporal variability in forest–atmosphere CO_2 exchange. *Global Change Biology*, **10**, 1689–1706. doi:10.1111/j.1365-2486.2004.00847.x.

Hollinger DY, Goltz SM, Davidson EA, Lee JT, Tu K, Valentine HT (1999) Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. *Global Change Biology*, **5**, 891–902. doi:10.1046/j.1365-2486.1999.00281.x.

Horn JE, Schulz K (2010) Post-processing analysis of MODIS leaf area index subsets. *Journal of Applied Remote Sensing*, **4**, 043557. doi:10.1117/1.3524265.

Houborg R, Anderson M, Norman J, Wilson T, Meyers T (2009) Intercomparison of a 'bottom-up' and 'top-down' modeling paradigm for estimating carbon and energy fluxes over a variety of vegetative regimes across the US. *Agricultural and forest meteorology*, **149**, 1875–1895. doi:10.1016/j.agrformet.2009.10.002.

Houborg R, Soegaard H, Boegh E (2007) Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data. *Remote Sensing of Environment*, **106**, 39–58. doi:10.1016/j.rse.2006.07.016.

Hsu C, Chang C, Lin C, *et al.* (2003) A practical guide to support vector classification. URL http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Huang C, Geiger EL, Van Leeuwen WJD, Marsh SE (2009) Discrimination of invaded and native species sites in a semi-desert grassland using MODIS multi-temporal data. *International Journal Of Remote Sensing*, **30**, 897–917. doi:10.1080/01431160802395243.

Huang C, Li X, Lu L (2008) Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sensing of Environment*, **112**, 1320–1336. doi:10.1016/j.rse.2007.03.028.

Huemmrich K, Gamon J, Tweedie C, *et al.* (2010) Remote sensing of tundra gross ecosystem productivity and light use efficiency under varying temperature and moisture conditions. *Remote Sensing of Environment*, **114**, 481–489. doi:10.1016/j.rse.2009.10.003.

Huete A, Didan K, Miura T, Rodriguez E, Gao X, Ferreira L, *et al.* (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, **83**, 195–213. doi:10.1016/S0034-4257(02)00096-2.

Hyman A (1996) Information presentation for new sensors: a focus on selected sensors of the Earth Observing System (EOS). *Prog. Phys. Geog.*, **20**, 146. doi:10.1177/030913339602000202.

Immerzeel WW, van Beek LPH, Bierkens MFP (2010) Climate Change Will Affect the Asian Water Towers. *Science*, **328**, 1382–1385. doi:10.1126/science.1183188.

Janzen HH (2004) Carbon cycling in earth systems–a soil science perspective. *Agriculture, Ecosystems & Environment*, **104**, 399–417. doi:10.1016/j.agee.2004.01.040.

Jarvis AJ, Stauch VJ, Schulz K, Young PC (2004) The seasonal temperature dependency of photosynthesis and respiration in two deciduous forests. *Global Change Biology*, **10**, 939–950. doi:10.1111/j.1365-2486.2004.00743.x.

Jassal R, Black T, Spittlehouse D, Brümmer C, Nesic Z (2009) Evapotranspiration and water use efficiency in different-aged Pacific Northwest Douglas-fir stands. *Agricultural and Forest Meteorology*, **149**, 1168–1178. doi:10.1016/j.agrformet.2009.02.004.

Jenkins JP, Richardson AD, Braswell BH, Ollinger SV, Hollinger DY, Smith ML (2007) Refining light-use efficiency calculations for a deciduous forest canopy using simultaneous tower-based carbon flux and radiometric measurements. *Agricultural and Forest Meteorology*, **143**, 64–79. doi:10.1016/j.agrformet.2006.11.008.

Juarez RIN, da Rocha HR, Figueira A, Goulden ML, Miller SD (2009) An improved estimate of leaf area index based on the histogram analysis of hemispherical photographs. *Agricultural and Forest Meteorology*, **149**, 920–928. doi:10.1016/j.agrformet.2008.11.012.

Julien Y, Sobrino JA (2009) Global land surface phenology trends from GIMMS database. *International Journal Of Remote Sensing*, **30**, 3495–3513. doi:10.1080/01431160802562255.

Jung M, Reichstein M, Bondeau A (2009) Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, **6**, 2001–2013. doi:10.5194/bg-7-2351-2010.

Jung M, Verstraete M, Gobron N, *et al.* (2008) Diagnostic assessment of European gross primary production. *Global Change Biology*, **14**, 2349–2364. doi:10.1111/j.1365-2486.2008.01647.x.

Jung M, Vetter M, Herold M, *et al.* (2007) Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models. *Global Biogeochemical Cycles*, **21**, 1–B4021. doi:10.1029/2006GB002915.

Justice CO, Vermote E, Townshend JRG, *et al.* (1998) The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing.*, **36**, 1228–1249. doi:10.1109/36.701075.

Kaminski T, Knorr W, Rayner P, Heimann M (2002) Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Global Biogeochemical Cycles*, **16**, 1066. doi:10.1029/2001GB001463.

Kang S, Running SW, Zhao M, Kimball JS, Glassy J (2005) Improving continuity of MODIS terrestrial photosynthesis products using an interpolation scheme for cloudy pixels. *International Journal of Remote Sensing*, **26**, 1659–1676. doi:10.1080/01431160512331326693.

Kanniah KD, Beringer J, Hutley LB, Tapper NJ, Zhu X (2009) Evaluation of Collections 4 and 5 of the MODIS Gross Primary Productivity product and algorithm improvement at a tropical savanna site in northern Australia. *Remote Sensing of Environment*, **113**, 1808–1822. doi:doi:10.1016/j.rse.2009.04.013.

Karp A, Shield I (2008) Bioenergy from plants and the sustainable yield challenge. *New Phytologist*, **179**, 15–32. doi:10.1111/j.1469-8137.2008.02432.x.

Katul G, Leuning R, Oren R (2003) Relationship between plant hydraulic and biochemical properties derived from a steady-state coupled water and carbon transport model. *Plant Cell and Environment*, **26**, 339–350. doi:10.1046/j.1365-3040.2003.00965.x.

Keenan T, García R, Friend A, Zaehle S, Gracia C, Sabate S (2009) Improved understanding of drought controls on seasonal variation in Mediterranean forest canopy CO2 and water fluxes through combined in situ measurements and ecosystem modelling. *Biogeosciences*, **6**, 1423–1444. doi:10.5194/bg-6-1423-2009.

Kergoat L, Lafont S, Arneth A, Le Dantec V, Saugier B (2008) Nitrogen controls plant canopy light-use efficiency in temperate and boreal ecosystems. *Journal Of Geophysical Research – Biogeosciences*, **113**. doi:10.1029/2007JG000676.

King MD, Kaufman YJ, Menzel WP, Tanre D (1992) Remote sensing of cloud, aerosol, and water-vapor properties from the moderate resolution imaging spectrometer (MODIS). *IEEE Transactions on Geoscience and Remote Sensing.*, **30**, 2–27. doi:10.1109/36.124212.

Kjelgaard JF, Heilman JL, McInnes KJ, Owens MK, Kamps RH (2008) Carbon dioxide exchange in a subtropical, mixed C-3/C-4 grassland on the Edwards Plateau, Texas. *Agricultural and Forest Meteorology*, **148**, 953–963. doi:10.1016/j.agrformet.2008.01.006.

Kleidon A (2010) A basic introduction to the thermodynamics of the Earth system far from equilibrium and maximum entropy production. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1303. doi:10.1098/rstb.2009.0310.

Kleidon A, Malhi Y, Cox P (2010) Maximum entropy production in environmental and ecological systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1297. doi:10.1098/rstb.2010.0018.

Klein R, Schipper E, Dessai S (2005) Integrating mitigation and adaptation into climate and development policy: three research questions. *Environmental Science & Policy*, **8**, 579–588. doi:10.1016/j.envsci.2005.06.010.

Knorr W (2000) Annual and interannual CO_2 exchanges of the terrestrial biosphere: process-based simulations and uncertainties. *Global Ecology and Biogeography*, **9**, 225–252. doi:10.1046/j.1365-2699.2000.00159.x.

Knorr W, Heimann M (2001a) Uncertainties in global terrestrial biosphere modeling, Part I: a comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme. *Global Biogeochemical Cycles*, **15**, 207–225. doi:10.1029/1998GB001059.

Knorr W, Heimann M (2001b) Uncertainties in global terrestrial biosphere modeling, Part II: global constraints for a process-based vegetation model. *Global Biogeochemical Cycles*, **15**, 227–246. doi:10.1029/1998GB001060.

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

Knyazikhin Y, Glassy J, Privette JL, *et al.* (1999) MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document. URL `http://modis.gsfc.nasa.gov/data/atbd/atbd_mod15.pdf`.

Krause P, Boyle DP, Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, **5**, 89–97. doi:10.5194/adgeo-5-89-2005.

Kulinskaya E, Staudte R (2006) Interval estimates of weighted effect sizes in the one-way heteroscedastic ANOVA. *British Journal of Mathematical and Statistical Psychology*, **59**, 97–111. doi:10.1348/000711005X68174.

Kustas WP, Perry EM, Doraiswamy PC, Moran MS (1994) Using satellite remote sensing to extrapolate evapotranspiration estimates in time and space over a semiarid rangeland basin. *Remote Sensing of Environment*, **49**, 275. doi:10.1016/0034-4257(94)90022-1.

Lagergren F, Eklundh L, Grelle A, Lundblad M, Mölder M, Lankreijer H, Lindroth A (2005) Net primary production and light use efficiency in a mixed coniferous forest in Sweden. *Plant, Cell & Environment*, **28**, 412–423. doi:10.1111/j.1365-3040.2004.01280.x.

Lahsen M (2009) A science–policy interface in the global south: the politics of carbon sinks and science in Brazil. *Climatic change*, **97**, 339–372. doi:10.1007/s10584-009-9610-6.

Landsberg JJ, Waring RH (1997) A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management*, **95**, 209–228. doi:10.1016/S0378-1127(97)00026-1.

Law BE, Waring RH (1994) Combining remote-sensing and climatic data to estimate net primary production across oregon. *Ecological Applications*, **4**, 717–728. doi:10.2307/1942002.

Lefsky M, Cohen W, Parker G, Harding D (2002) Lidar remote sensing for ecosystem studies. *Bioscience*, **52**, 19–30. doi:10.1016/S0169-5347(03)00070-3.

Legates D, McCabe Jr G (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, **35**, 233–241. doi:10.1029/1998WR900018.

Leuning R, Cleugh HA, Zegelin SJ, Hughes D (2005) Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates. *Agricultural and Forest Meteorology*, **129**, 151–173. doi:10.1016/j.agrformet.2004.12.004.

Leuning R, Kelliher FM, Depury DGG, Schulze ED (1995) Leaf nitrogen, photosynthesis, conductance and transpiration - scaling from leaves to caopies. *Plant Cell and Environment*, **18**, 1183–1200. doi:10.1111/j.1365-3040.1995.tb00628.x.

Leuning R, Zhang Y, Rajaud A, Cleugh H, Tu K (2008) A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. *Water Resources Research*, **44**, W10419. doi:10.1029/2007WR006562.

Li Z, Yu G, Xiao X, *et al.* (2007) Modeling gross primary production of alpine ecosystems in the Tibetan Plateau using MODIS images and climate data. *Remote Sensing of Environment*, **107**, 510–519. doi::10.1016/j.rse.2006.10.003.

Liang S, Zheng T, Liu R, Fang H, Tsay S, Running S (2006) Estimation of incident photosynthetically active radiation from MODIS data. *Journal of Geophysical Research – Atmospheres*, **111**. doi:10.1029/2005JD006730.

Lindroth A (1993) Aerodynamic and canopy resistance of short-rotation forest in relation to leaf area index and climate. *Boundary Layer Meteorology*, **66**, 265–279. doi:10.1007/BF00705478.

Lindroth A, Lagergren F, Aurela M, *et al.* (2008) Leaf area index is the principal scaling parameter for both gross photosynthesis and ecosystem respiration of Northern deciduous and coniferous forests. *Tellus Ser. B*, **60**, 129–142. doi:10.1111/j.1600-0889.2007.00330.x.

Linsley K, Kohler M, Joseph L (1982) *Hydrology for Engineers*. McGraw Hill, New York.

Liu R, Liang S, He H, Liu J, Zheng T (2008) Mapping incident photosynthetically active radiation from MODIS data over China. *Remote Sensing of Environment*, **112**, 998–1009. doi:10.1016/j.rse.2007.07.021.

Loew A (2008) Impact of surface heterogeneity on surface soil moisture retrievals from passive microwave data at the regional scale: The Upper Danube case. *Remote Sensing of Environment*, **112**, 231–248. doi:10.1016/j.rse.2007.04.009.

Lu X, Liu R, Liu J, Liang S (2007) Removal of noise by wavelet method to generate high quality temporal data of terrestrial MODIS products. *Photogrammetric Engineering and Remote Sensing*, **73**, 1129. doi:0099-1112/07/7310Ű1129.

Lu X, Zhuang Q (2010) Evaluating evapotranspiration and water-use efficiency of terrestrial ecosystems in the conterminous United States using MODIS and AmeriFlux data. *Remote Sensing of Environment*, **114**, 1924Ű1939. doi::10.1016/j.rse.2010.04.001.

Luyssaert S, Inglima I, Jung M, *et al.* (2007) CO2 balance of boreal, temperate, and tropical forests derived from a global database. *Global Change Biology*, **13**, 2509–2537. doi:10.1111/j.1365-2486.2007.01439.x.

Ma S, Baldocchi D, Xu L, Hehn T (2007) Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California. *Agricultural and Forest Meteorology*, **147**, 157–171. doi:10.1016/j.agrformet.2007.07.008.

Mahecha M, Reichstein M, Carvalhais N, *et al.* (2010) Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science*, **329**, 838. doi:10.1126/science.1189587.

Mäkelä A, Kolari P, Karimaki J, Nikinmaa E, Peramaki M, Hari P (2006) Modelling five years of weather-driven variation of GPP in a boreal forest. *Agricultural and Forest Meteorology*, **139**, 382–398. doi:10.1016/j.agrformet.2006.08.017.

Mäkelä A, Pulkkinen M, Kolari P, *et al.* (2008) Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. *Global Change Biology*, **14**, 92–108. doi:10.1111/j.1365-2486.2007.01463.x.

Maseyk K, Grünzweig J, Rotenberg E, Yakir D (2008) Respiration acclimation contributes to high carbon-use efficiency in a seasonally dry pine forest. *Global Change Biology*, **14**, 1553–1567. doi:10.1111/j.1365-2486.2008.01604.x.

Massman W, Clement R (2005) *Uncertainty in eddy covariance flux estimates resulting from spectral attenuation*. Springer, New York, NY, USA, 67–99 pp.

Massman W, Lee X (2002) Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges. *Agricultural and Forest Meteorology*, **113**, 121–144. doi:10.1016/S0168-1923(02)00105-3.

McCarthy H, Oren R, Finzi A, Ellsworth D, Kim H, Johnsen K, Millar B (2007) Temporal dynamics and spatial variability in the enhancement of canopy leaf area under elevated atmospheric CO_2. *Global Change Biology*, **13**, 2479–2497. doi:10.1111/j.1365-2486.2007.01455.x.

McCarthy HR, Oren R, Kim HS, Johnsen KH, Maier C, Pritchard SG, Davis MA (2006) Interaction of ice storms and management practices on current carbon sequestration in forests with potential mitigation under future CO_2 atmosphere. *Journal of Geophysical Research – Atmospheres*, **111**, 10. doi:10.1029/2005JD006428.

McCaughey JH, Pejam MR, Arain MA, Cameron DA (2006) Carbon dioxide and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada. *Agricultural and Forest Meteorology*, **140**, 79. doi:10.1016/j.agrformet.2006.08.010.

McCormick S, Tjian R (2010) A New Focus on Plant Sciences. *Science*, **330**, 1021. doi:10.1126/science.1198153.

McMurtrie R, Gholz H, Linder S, Gower S (1994) Climatic factors controlling the productivity of pine stands: a model-based analysis. *Ecological Bulletins*, **43**, 173–188.

Moffat AM, Papale D, Reichstein M, *et al.* (2007) Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology*, **147**, 209–232. doi:10.1016/j.agrformet.2007.08.011.

Monteith J (1981) Evaporation and surface temperature. *Quarterly Journal of the Royal Meteorological Society*, **107**, 1–27. doi:10.1256/smsqj.45101.

Monteith J, Greenwood D (1986) How do crops manipulate water supply and demand? *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **316**, 245–259.

Monteith JL (1972) Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, **9**, 747–766.

Monteith JL (1977) Climate and efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **281**, 277–294.

Monteith JL, Unsworth MH (2008) *Principles of environmental physics*. Chapman and Hall, New York, 3rd edn., XII+291P pp.

Morgan J, Follett R, Allen L, *et al.* (2010) Carbon sequestration in agricultural lands of the United States. *Journal of Soil and Water Conservation*, **65**, 6A. doi:10.2489/jswc.65.1.6A.

Morisette J, Privette J, Justice C (2002) A framework for the validation of MODIS land products. *Remote Sensing of Environment*, **83**, 77–96. doi:10.1016/S0034-4257(02)00088-3.

Moureaux C, Debacq A, Hoyaux J, *et al.* (2008) Carbon balance assessment of a Belgian winter wheat crop (Triticum aestivum L.). *Global Change Biology*, **14**, 1353–1366. doi: 10.1111/j.1365-2486.2008.01560.x.

Mund M, Kutsch W, Wirth C, Kahl T, Knohl A, Skomarkova M, Schulze E (2010) The influence of climate and fructification on the inter-annual variability of stem growth and net primary productivity in an old-growth, mixed beech forest. *Tree Physiology*, **30**, 689–704. doi:10.1093/treephys/tpq027.

Murchie E, Pinto M, Horton P (2009) Agriculture and the new challenges for photosynthesis research. *New Phytologist*, **181**, 532–552. doi:10.1111/j.1469-8137.2008.02705.x.

Myneni R (2009) MODIS land team validation status for: LAI/FPAR (MOD15). URL `http://landval.gsfc.nasa.gov/ProductStatus.php?ProductID=MOD15`.

Myneni RB, Hoffman S, Knyazikhin Y, *et al.* (2002) Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, **83**, 214–231. doi:10.1016/S0034-4257(02)00074-3.

Nash J, Sutcliffe J (1970) River flow forecasting through conceptual models part I–A discussion of principles. *Journal of Hydrology*, **10**, 282–290. doi:10.1016/0022-1694(70)90255-6.

Nemani R, Hashimoto H, Votava P, *et al.* (2009) Monitoring and forecasting ecosystem dynamics using the Terrestrial Observation and Prediction System (TOPS). *Remote Sensing of Environment*, **113**, 1497–1509. doi::10.1016/j.rse.2008.06.017.

Nouvellon Y, Begue A, Moran MS, *et al.* (2000) PAR extinction in shortgrass ecosystems: effects of clumping, sky conditions and soil albedo. *Agricultural and Forest Meteorology*, **105**, 21–41. doi:10.1016/S0168-1923(00)00194-5.

Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) (Accessed 08/24/2009) Modis subsetted land products, collection 5. URL `http://daac.ornl.gov/MODIS/modis.html`.

Ollinger S (2010) Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, **188**. doi:10.1111/j.1469-8137.2010.03536.x.

Ollinger S, Richardson A, Martin M, *et al.* (2008) Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. *Proceedings of the National Academy of Sciences*, **105**, 19336. doi:10.1073/pnas.0810021105.

Pandya M, Singh R, Chaudhari K, Bairagi G, Sharma R, Dadhwal V, Parihar J (2006) Leaf area index retrieval using IRS LISS-III sensor data and validation of the MODIS LAI product over central India. *Geoscience and Remote Sensing, IEEE Transactions on*, **44**, 1858–1865. doi:10.1109/TGRS.2006.876028.

Papale D, Reichstein M, Aubinet M, *et al.* (2006) Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, **3**, 571–583. doi:10.5194/bg-3-571-2006.

Papale D, Valentini R (2003) A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology*, **9**, 525–535. doi:10.1046/j.1365-2486.2003.00609.x.

Papeş M, Tupayachi R, Martínez P, Peterson A, Powell G (2010) Using hyperspectral satellite imagery for regional inventories: a test with tropical emergent trees in the Amazon Basin. *Journal of Vegetation Science*, **21**, 342–354. doi:10.1111/j.1654-1103.2009.01147.x.

Parry M (2007) *Climate Change 2007: impacts, adaptation and vulnerability: contribution of Working Group II to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Pedregal D, Taylor C, Young P (2007) System identification, time series analysis and forecasting: the Captain Toolbox handbook. Tech. rep., Centre for Research on Environmental Systems and Statistics (CRES), Lancaster University.

Pereira J, Mateus J, Aires L, *et al.* (2007) Net ecosystem carbon exchange in three contrasting Mediterranean ecosystems? the effect of drought. *Biogeosciences*, **4**, 791–802. doi:10.5194/bg-4-791-2007.

Piao S, Fang J, Ciais P, Peylin P, Huang Y, Sitch S, Wang T (2009) The carbon balance of terrestrial ecosystems in China. *Nature*, **458**, 1009–1013. doi:10.1038/nature07944.

Pilegaard K, Mikkelsen T, Beier C, Jensen N, Ambus P, Ro-Poulsen H (2003) Field measurements of atmosphere-biosphere interactions in a Danish beech forest. *Boreal Environment Research*, **8**, 315–333.

Pisek J, Chen JM (2007) Comparison and validation of MODIS and VEGETATION global LAI products over four BigFoot sites in North America. *Remote Sensing of Environment*, **109**, 81–94. doi:10.1016/j.rse.2006.12.004.

Portner H, Bugmann H, Wolf A (2009) Temperature response functions introduce high uncertainty in modelled carbon stocks in cold temperature regimes. *Biogeosciences Discussions*, **6**, 8129–8165. doi:10.5194/bg-7-3669-2010.

Potter C, Boriah S, Steinbach M, Kumar V, Klooster S (2008) Terrestrial vegetation dynamics and global climate controls. *Climate Dynamics*, **31**, 67–78. doi:10.1007/s00382-007-0339-5.

Potter CS, Klooster SA (1999) Dynamic global vegetation modelling for prediction of plant functional types and biogenic trace gas fluxes. *Global Ecology and Biogeography*, **8**, 473–488. doi:10.1046/j.1365-2699.1999.00152.x.

Potter CS, Randerson JT, Field CB, Matson PA, Vitousek PM, Mooney HA, Klooster SA (1993) Terrestrial ecosystem production - a process model-based on global satellite and surface data. *Global Biogeochemical Cycles*, **7**, 811–841. doi:10.1029/93GB02725.

Prince SD, Goetz SJ, Goward SN (1995) Monitoring primary production from earth observing satellites. *Water Air and Soil Pollution*, **82**, 509–522. doi:10.1007/BF01182860.

Rahman A, Gamon J (2004) Detecting biophysical properties of a semi-arid grassland and distinguishing burned from unburned areas with hyperspectral reflectance. *Journal of Arid Environments*, **58**, 597–610. doi:10.1016/j.jaridenv.2003.12.005.

Rahman A, Sims D, Cordova V, El-Masri B (2005) Potential of MODIS EVI and surface temperature for directly estimating per-pixel ecosystem C fluxes. *Geophysical Research Letters*, **32**, L19404. doi:10.1029/2005GL024127.

Rahman AF, Gamon JA, Fuentes DA, Roberts DA, Prentiss D (2001) Modeling spatially distributed ecosystem flux of boreal forest using hyperspectral indices from AVIRIS imagery. *Journal of Geophysical Research – Atmospheres*, **106**, 33579–33591. doi:10.1029/2001JD900157.

Rahman AF, Gamon JA, Sims DA, Schmidts M (2003) Optimum pixel size for hyperspectral studies of ecosystem function in southern California chaparral and grassland. *Remote Sensing of Environment*, **84**, 192–207. doi:10.1016/S0034-4257(02)00107.

Rascher U, Biskup B, Leakey A, McGrath J, Ainsworth E (2010) Altered physiological function, not structure, drives increased radiation-use efficiency of soybean grown at elevated co$_2$. *Photosynthesis Research*, **105**, 15–25. doi:10.1007/s11120-010-9548-6.

Rascher U, Pieruschka R (2008) Spatio-temporal variations of photosynthesis: the potential of optical remote sensing to better understand and scale light use efficiency and stresses of plant ecosystems. *Precision Agriculture*, **9**, 355–366. doi:10.1007/s11119-008-9074-0.

Rastetter E, Williams M, Griffin K, *et al.* (2010) Processing arctic eddy-flux data using a simple carbon-exchange model embedded in the ensemble Kalman filter. *Ecological Applications*, **20**, 1285Ű–1301. doi:10.1890/09-0876.1.

Raupach M (1994) Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index. *Boundary Layer Meteorology*, **71**, 211–216. doi:10.1007/BF00709229.

Rayner P, Scholze M, Knorr W, Kaminski T, Giering R, Widmann H (2005) Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochemical Cycles*, **19**, GB2026. doi:10.1029/2004GB002254.

Rebmann C, Zeri M, Lasslop G, Mund M, Kolle O, Schulze E, Feigenwinter C (2010) Treatment and assessment of the co$_2$-exchange at a complex forest site in thuringia, germany. *Agricultural and Forest Meteorology*, **150**, 684–691. doi:10.1016/j.agrformet.2009.11.001.

Reichle R, Koster R, Liu P, Mahanama S, Njoku E, Owe M (2007) Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research*, **112**, D09108. doi:10.1029/2006JD008033.

Reichstein M, Ciais P, Papale D, *et al.* (2007) Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Global Change Biology*, **13**, 634–651. doi:10.1111/j.1365-2486.2006.01224.x.

Reid W, Chen D, Goldfarb L, *et al.* (2010) Earth System Science for Global Sustainability: Grand Challenges. *Science*, **330**, 916. doi:10.1126/science.1196263.

Richardson AD, Mahecha MD, Falge E, *et al.* (2008) Statistical properties of random CO_2 flux measurement uncertainty inferred from model residuals. *Agricultural and Forest Meteorology*, **148**, 38–50. doi:10.1016/j.agrformet.2007.09.001.

Roehrig J, Laudien R (2009) Evaluation of agricultural land resources by implementing a computer-based spatial decision support system for national deciders in Benin, West

Africa. *Journal of Applied Remote Sensing*, **3**, 033502. doi:10.1117/1.3079033. URL `http://link.aip.org/link/?JRS/3/033502/1`.

Rosati A, Dejong TM (2003) Estimating photosynthetic radiation use efficiency using incident light and photosynthesis of individual leaves. *Annals of Botany*, **91**, 869–877. doi:10.1093/aob/mcg094.

Rotenberg E, Yakir D (2010) Contribution of Semi-Arid Forests to the Climate System. *Science*, **327**, 451. doi:10.1126/science.1179998.

Roughgarden J, Running S, Matson P (1991) What does remote sensing do for ecology? *Ecology*, **72**, 1918–1922. doi:10.2307/1941546.

Ruimy A, Jarvis PG, Baldocchi DD, Saugier B, Begon M, Fitter AH (1995) CO_2 Fluxes over Plant Canopies and Solar Radiation: A Review. In: *Advances in Ecological Research*, vol. 26, pp. 1–68. Academic Press.

Running S (2008) Ecosystem disturbance, carbon, and climate. *Science*, **321**, 652–653. doi:10.1126/science.1159607.

Running S, Hunt E (1993) *Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models*, pp. 141–158. Academic Press, San Diego, CA.

Running SW, Baldocchi DD, Turner DP, Gower ST, Bakwin PS, Hibbard KA (1999) A Global Terrestrial Monitoring Network Integrating Tower Fluxes, Flask Sampling, Ecosystem Modeling and EOS Satellite Data. *Remote Sensing of Environment*, **70**, 108–127. doi:10.1016/S0034-4257(99)00061-9.

Running SW, Justice CO, Salomonson V, *et al.* (1994) Terrestrial remote-sensing science and algorithm planned for EOS MODIS. *International Journal of Remote Sensing*, **15**, 3587–3620. doi:10.1080/01431169408954346.

Running SW, Nemani RR, Heinsch FA, Zhao MS, Reeves M, Hashimoto H (2004) A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, **54**, 547–560.

Runyon J, Waring RH, Goward SN, Welles JM (1994) Environmental limits of net primary production and light-use efficiency across the Oregon transect. *Ecological Applications*, **4**, 226–237. doi:10.2307/1941929.

Sacks W, Schimel D, Monson R, Braswell B (2006) Model-data synthesis of diurnal and seasonal CO2 fluxes at Niwot Ridge, Colorado. *Global Change Biology*, **12**, 240–259. doi:10.1111/j.1365-2486.2005.01059.x.

Samaniego-Eguiguren L (2003) *Hydrological Consequences of Land Use/Land Cover and Climatic Changes in Mesoscale Catchments*. Ph.D. thesis, Universitätsbibliothek Stuttgart.

Schaepman M, Ustin S, Plaza A, Painter T, Verrelst J, Liang S (2009) Earth system science related imaging spectroscopy–An assessment. *Remote Sensing of Environment*, **113**, S123–S137. doi:10.1016/j.rse.2009.03.001.

Schimel D, House J, Hibbard K, *et al.* (2001) Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature*, **414**, 169–172. doi:10.1038/35102500.

Schmid HP, Grimmond CSB, Cropley F, Offerle B, Su HB (2000) Measurements of CO_2 and energy fluxes over a mixed hardwood forest in the mid-western United States. *Agricultural and Forest Meteorology*, **103**, 357–374. doi:10.1016/S0168-1923(00)00140-4.

Schmitt M, Bahn M, Wohlfahrt G, Tappeiner U, Cernusca A (2009) Land use affects the net ecosystem CO_2 exchange and its components in mountain grasslands. *Biogeosciences Discussions*, **6**, 11435–11462. doi::10.5194/bg-7-2297-2010.

Schnur M, Xie H, Wang X (2010) Estimating Root Zone Soil Moisture at Distant Sites Using MODIS NDVI and EVI in a Semi-Arid Region of Southwestern USA. *Ecological Informatics*, **5**, 400–409. doi:10.1016/j.ecoinf.2010.05.001.

Schulz K, Beven K (2003) Data-supported robust parameterisations in land surface-atmosphere flux predictions: towards a top-down approach. *Hydrological Processes*, **17**, 2259–2277. doi:10.1002/hyp.1331.

Schulz K, Franks S, Beven K (1998) TOPUP - A TOPMODEL based SVAT model to calculate evaporative fluxes between the landsurface and the atmosphere. Program documentation, version 1.1, IENS, Department of Environmental Sciences, Lancaster University, Lancaster,UK.

Schulz K, Jarvis A, Beven K (2001) The predictive uncertainty of land surface fluxes in response to increasing ambient carbon dioxide. *Journal of Climate*, **14**, 2551–2562. doi:10.1175/1520-0442(2001)014.

Schulz K, Jarvis AJ (2004) Environmental and biological controls on the seasonal variations in latent heat fluxes derived from flux data for three forest sites. *Water Resources Research*, **40**. doi:10.1029/2004WR003155.

Schwalm C, Williams C, Schaefer K, *et al.* (2010) Assimilation exceeds respiration sensitivity to drought: A FLUXNET synthesis. *Global Change Biology*, **16**, 657–670. doi: 10.1111/j.1365-2486.2009.01991.x.

Schwalm CR, Black TA, Arniro BD, *et al.* (2006) Photosynthetic light use efficiency of three biomes across an east-west continental-scale transect in canada. *Agricultural and Forest Meteorology*, **140**, 269–286. doi:10.1016/j.agrformet.2006.06.010.

Schymanski S, Kleidon A, Stieglitz M, Narula J (2010) Maximum entropy production allows a simple representation of heterogeneity in semiarid ecosystems. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, **365**, 1449. doi:10.1098/rstb.2009.0309.

Searchinger T, Heimlich R, Houghton R, *et al.* (2008) Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science*, **319**, 1238. doi:10.1126/science.1151861.

Sellers P, Dickinson R, Randall D, *et al.* (1997) Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, **275**, 502–509. doi:10.1126/science.275.5299.502.

Sellers PJ (1985) Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, **6**, 1335–1372. doi:10.1016/0034-4257(87)90051-4.

Seufert G, Bartzis J, Bomboi T, *et al.* (1997) An overview of the Castelporziano experiments. *Atmospheric Environment*, **31**, 5–17. doi:10.1016/S1352-2310(97)00334-8.

Shaver G, Street L, Rastetter E, VAN W, *et al.* (2007) Functional convergence in regulation of net CO2 flux in heterogeneous tundra landscapes in Alaska and Sweden. *Journal of Ecology*, **95**, 802–817. doi:10.1111/j.1365-2745.2007.01259.x.

Shaw D, Franklin J, Bible K, Klopatek J, Freeman E, Greene S, Parker G (2004) Ecological setting of the Wind River old-growth forest. *Ecosystems*, **7**, 427–439. doi:10.1007/s10021-004-0135-6.

Shawe-Taylor J, Bartlett P, Williamson R, Anthony M (1998) Structural risk minimization over data-dependent hierarchies. *IEEE Transactions on Information Theory*, **44**, 1926–1940. doi:10.1109/18.705570.

Sims DA, Rahman AF, Cordova VD, *et al.* (2005) Midday values of gross CO_2 flux and light use efficiency during satellite overpasses can be used to directly estimate eight-day mean flux. *Agricultural and Forest Meteorology*, **131**, 1–12. doi:10.1016/j.agrformet.2005.04.006.

Sims DA, Rahman AF, Cordova VD, *et al.* (2008) A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *Remote Sensing of Environment*, **112**, 1633–1646. doi:10.1016/j.rse.2007.08.004.

Sinclair TR, Horie T (1989) Leaf nitrogen, photosynthesis, crop radiation use efficiency - a review. *Crop Science*, **29**, 90–98. doi:10.3146/i0095-3679-20-1-11.

Siqueira M, Katul G, DA SAMPSON P, JUANG J, MCCARTHY H, OREN R (2006) Multi-scale model intercomparisons of CO2 and H2O exchange rates in a maturing southeastern US pine forest. *Global Change Biology*, **12**, 1189–1207. doi:10.1111/j.1365-2486.2006.01158.x.

Smola A, Schölkopf B (2004) A tutorial on support vector regression. *Statistics and Computing*, **14**, 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

Sobrino J, Romaguera M (2004) Land surface temperature retrieval from MSG1-SEVIRI data. *Remote Sensing of Environment*, **92**, 247–254. doi:10.1016/j.rse.2004.06.009.

Soja AJ, Al-Saadi J, Giglio L, *et al.* (2009) Assessing satellite-based fire data for use in the National Emissions Inventory. *Journal of Applied Remote Sensing*, **3**, 031504. doi: 10.1117/1.3148859. URL http://link.aip.org/link/?JRS/3/031504/1.

Soudani K, le Maire G, Dufrene E, Francois C, Delpierre N, Ulrich E, Cecchini S (2008) Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Remote Sensing of Environment*, **112**, 2643–2655. doi:10.1016/j.rse.2007.12.004.

Stauch V, Jarvis A (2006) A semi-parametric gap-filling model for eddy covariance CO2 flux time series data. *Global Change Biology*, **12**, 1707–1716. doi:10.1111/j.1365-2486.2006.01227.x.

Stauch V, Jarvis A, Schulz K (2008) Estimation of net carbon exchange using eddy covariance CO2 flux observations and a stochastic model. *Journal of Geophysical Research*, **113**, D03101. doi:10.1029/2007JD008603.

Stigter C, Musabilha V (1982) The conservative ratio of photosynthetically active to total radiation in the tropics. *Journal of Applied Ecology*, **19**, 853–858.

Still C, Randerson J, Fung I (2004) Large-scale plant light-use efficiency inferred from the seasonal cycle of atmospheric co_2. *Global Change Biology*, **10**, 1240–1252. doi: 10.1111/j.1365-2486.2004.00802.x.

Stitt M (2006) Rising co_2 levels and their potential significance for carbon flow in photosynthetic cells. *Plant, Cell & Environment*, **14**, 741–762. doi:10.1111/j.1365-3040.1991.tb01440.x.

Stoeckli R, Lawrence DM, Niu GY, *et al.* (2008) Use of FLUXNET in the community land model development. *Journal of Geophysical Research – Biogeosciences*, **113**. doi: 10.1029/2007JG000562.

Stoy P, Katul G, Siqueira M, *et al.* (2005) Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis. *Tree Physiology*, **25**, 887. doi:10.1093/treephys/25.7.887.

Stoy P, Palmroth S, Oishi A, *et al.* (2007) Are ecosystem carbon inputs and outputs coupled at short time scales? A case study from adjacent pine and hardwood forests using impulse–response analysis. *Plant, Cell & Environment*, **30**, 700–710. doi:10.1111/j.1365-3040.2007.01655.x.

Stoy P, Richardson A, Baldocchi D, *et al.* (2009) Biosphere-atmosphere exchange of CO_2 in relation to climate: a cross-biome analysis across multiple time scales. *Biogeosciences*, **6**, 2297–2312. doi:10.5194/bg-6-2297-2009.

Stoy PC, Katul GG, Siqueira MBS, *et al.* (2008) Role of vegetation in determining carbon sequestration along ecological succession in the southeastern United States. *Global Change Biology*, **14**, 1409–1427. doi:10.1111/j.1365-2486.2008.01587.x.

Stuffler T, Kaufmann C, Hofer S, *et al.* (2007) The EnMAP hyperspectral imager–An advanced optical payload for future applications in Earth observation programmes. *Acta Astronautica*, **61**, 115–120. doi:10.1016/j.actaastro.2007.01.033.

Suni T, Rinne J, Reissell A, *et al.* (2003) Long-term measurements of surface fluxes above a Scots pine forest in Hyytiala, southern Finland, 1996-2001. *Boreal Environment Research*, **8**, 287–302.

Tait A, Henderson R, Turner R, Zheng X (2006) Thin plate smoothing spline interpolation of daily rainfall for new zealand using a climatological rainfall surface. *International Journal of Climatology*, **26**, 2097–2115. doi:10.1002/joc.1350.

Tan B, Hu J, Huang D, *et al.* (2005) Assessment of the broadleaf crops leaf area index product from the Terra MODIS instrument. *Agricultural and Forest Meteorology*, **135**, 124–134. doi:10.1016/j.agrformet.2005.10.008.

Taylor C, Pedregal D, Young P, Tych W (2007) Environmental time series analysis and forecasting with the Captain toolbox. *Environmental Modelling & Software*, **22**, 797–814. doi:10.1016/j.envsoft.2006.03.002.

Therezien M, Palmroth S, Brady R, Oren R (2007) Estimation of light interception properties of conifer shoots by an improved photographic method and a 3D model of shoot structure. *Tree physiology*, **27**, 1375. doi:10.1093/treephys/27.10.1375.

Thornton P, Doney S, Lindsay K, *et al.* (2009) Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model. *Biogeosciences*, **6**, 2099–2120. doi:10.5194/bg-6-2099-2009.

Tolson B, Shoemaker C (2007) Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, **43**, W01413. doi:10.1029/2005WR004723.

Tucker C, Sellers P (1986) Satellite remote sensing of primary production. *International Journal of Remote Sensing*, **7**, 1395–1416. doi:10.1016/S0304-3800(99)00140-4.

Turner D, Ritts W, Cohen W, *et al.* (2003a) Scaling gross primary production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation. *Remote Sensing of Environment*, **88**, 256–270. doi:10.IO16,~j.rse.2003.06.005.

Turner D, Ritts W, Wharton S, Thomas C, Monson R, Black T, Falk M (2009) Assessing FPAR source and parameter optimization scheme in application of a diagnostic carbon flux model. *Remote Sensing of Environment*, **113**, 1529–1539. doi:10.1016/j.rse.2009.03.003.

Turner DP, Guzy M, Lefsky MA, Ritts WD, Van Tuyl S, Law BE (2004a) Monitoring forest carbon sequestration with remote sensing and carbon cycle modeling. *Environmental Management*, **33**, 457–466. doi:10.1016/j.rse.2006.02.017.

Turner DP, Ollinger SV, Kimball JS (2004b) Integrating remote sensing and ecosystem process models for landscape- to regional-scale analysis of the carbon cycle. *Bioscience*, **54**, 573–584.

Turner DP, Ritts WD, Cohen WB, *et al.* (2006) Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment*, **102**, 282–292. doi:10.1016/j.rse. 2006.02.017.

Turner DP, Urbanski S, Bremer D, Wofsy SC, Meyers T, Gower ST, Gregory M (2003b) A cross-biome comparison of daily light use efficiency for gross primary production. *Global Change Biology*, **9**, 383–395. doi:10.1046/j.1365-2486.2003.00573.x.

Urbanski S, Barford C, Wofsy S, *et al.* (2007) Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest. *Journal of Geophysical Research*, **112**, G02020. doi:10.1029/2006JG000293,2007.

Ustin S, Gamon J (2010) Remote sensing of plant functional types. *New Phytologist*, **186**, 795–816. doi:10.1111/j.1469-8137.2010.03284.x.

Vapnik V (2000) *The nature of statistical learning theory*. Springer Verlag, New York, NY, USA, 2nd edn.

Vapnik V, Golowich S, Smola A (1997) Support vector method for function approximation, regression estimation, and signal processing. In: *Advances in Neural Information Processing Systems*, pp. 281–287. NIPS.

Venturim V, Islam S, Rodriguez L (2008) Estimation of evaporative fraction and evapotranspiration from MODIS products using a complementary based model. *Remote Sensing of Environment*, **112**, 132–141. doi:10.1016/j.rse.2007.04.014|.

Venturini V, Bisht G, Islam S, Jiang L (2004) Comparison of evaporative fractions estimated from AVHRR and MODIS sensors over South Florida. *Remote Sensing of Environment*, **93**, 77. doi:10.1016/j.rse.2004.06.020.

Verbeeck H, Samson R, Granier A, Montpied P, Lemeur R (2008) Multi-year model analysis of GPP in a temperate beech forest in France. *Ecological Modelling*, **210**, 85–103. doi: 10.1016/j.ecolmodel.2007.07.010.

Verger A, Baret F, Weiss M (2008) Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products. *Remote Sensing of Environment*, **112**, 2789–2803. doi:10.1016/j.rse.2008.01.006.

Veroustraete F, Sabbe H, Eerens H (2002) Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. *Remote Sensing of Environment*, **83**, 376. doi:10.1016/S0034-4257(02)00043-3.

Vörösmarty CJ, McIntyre PB, Gessner MO, *et al.* (2010) Global threats to human water security and river biodiversity. *Nature*, **467**, 555–561. doi:10.1038/nature09440.

Wagner W, Naeimi V, Scipal K, De Jeu R, Martínez-Fernández J (2007) Soil moisture from operational meteorological satellites. *Hydrogeology Journal*, **15**, 121–131. doi: Soilmoisturefromoperationalmeteorologicalsatellites.

Wahba G (1979) Convergence rates of" thin plate" smoothing splines wihen the data are noisy. . In: *Lecture Notes in Mathematics*, vol. 757, pp. 233–245. Springer. doi:10.1007/BFb0098499.

Wallin G, Linder S, Lindroth A, Rantfors M, Flemberg S, Grelle A (2001) Carbon dioxide exchange in Norway spruce at the shoot, tree and ecosystem scale. *Tree Physiology*, **21**, 969–976.

Wan Z (2008) New refinements and validation of the MODIS land-surface temperature/emissivity products. *Remote Sensing of Environment*, **112**, 59–74. doi:10.1016/j.rse.2006.06.026.

Wang K, Li Z, Cribb M (2006) Estimation of evaporative fraction from a combination of day and night land surface temperatures and ndvi: A new method to determine the priestley-taylor parameter. *Remote Sensing of Environment*, **102**, 293. doi:10.1016/j.rse.2006.02.007.

Wang L, Qu JJ, Xiong X, Hao X (2009) Analysis of seven-year moderate resolution imaging spectroradiometer vegetation water indices for drought and fire activity assessment over Georgia of the United States. *Journal of Applied Remote Sensing*, **3**, 033555. doi:10.1117/1.3256138. URL `http://link.aip.org/link/?JRS/3/033555/1`.

Wang L, Xiong X, Qu J, Xie Y, Hao X, Che N (2007a) Impact assessment of Aqua MODIS band-to-band misregistration on snow index. *Journal of Applied Remote Sensing*, **1**. doi:10.1117/1.2786934. URL `http://spie.org/x648.html?product_id=776985`.

Wang Q, Ni J, Tenhunen J (2005a) Application of a geographically-weighted regression analysis to estimate net primary production of chinese forest ecosystems. *Global Ecology and Biogeography*, **14**, 379–393. doi:10.1111/j.1466-822X.2005.00153.x.

Wang Q, Tenhunen J, Dinh NQ, Reichstein M, Otieno D, Granier A, Pilegarrd K (2005b) Evaluation of seasonal variation of MODIS derived leaf area index at two European deciduous broadleaf forest sites. *Remote Sensing of Environment*, **96**, 475–484. doi: 10.1016/j.rse.2005.04.003.

Wang Q, Watanabe M, Ouyang Z (2005c) Simulation of water and carbon fluxes using BIOME-BGC model over crops in China. *Agricultural and Forest Meteorology*, **131**, 209–224. doi:10.1016/j.agrformet.2005.06.002.

Wang Y, Leuning R (1998) A two-leaf model for canopy conductance, photosynthesis and partitioning of available energy I: Model description and comparison with a multi-layered model. *Agricultural and Forest Meteorology*, **91**, 89–111. doi:10.1016/S0168-1923(98)00061-6.

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

Wang Y, Woodcock CE, Buermann W, *et al.* (2004) Evaluation of the MODIS LAI algorithm at a coniferous forest site in Finland. *Remote Sensing of Environment*, **91**, 114 – 127. doi:DOI:10.1016/j.rse.2004.02.007.

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

Watson D (1958) The dependence of net assimilation rate on leaf-area index. *Annals of Botany*, **22**, 37.

Watson DJ (1947) 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 years. *Annals of Botany*, **11**, 41–76.

Weiss M, Baret F, Smith GJ, Jonckheere I, Coppin P (2004) Review of methods for in situ leaf area index (LAI) determination Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology*, **121**, 37–53. doi:10.1016/j.rse.2007.03.001.

White M, Thornton P, Running S, Nemani R (2000) Parameterization and sensitivity analysis of the BIOME–BGC terrestrial ecosystem model: net primary production controls. *Earth Interactions*, **4**, 1–85.

Williams M, Richardson AD, Reichstein M, *et al.* (2009) Improving land surface models with Fluxnet data. *Biogeosciences*, **6**, 1341–1359. doi:10.5194/bg-6-1341-2009.

Williams M, Schwarz P, Law B, Irvine J, Kurpius M (2004) An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, **11**, 89–105. doi:10.1111/j.1365-2486.2004.00891.x.

Wilson TB, Meyers TP (2007) Determining vegetation indices from solar and photosynthetically active radiation fluxes. *Agricultural and Forest Meteorology*, **144**, 160–179. doi:10.1016/j.agrformet.2007.04.001.

Wohlfahrt G, Hammerle A, Haslwanter A, Bahn M, Tappeiner U, Cernusca A (2008) Seasonal and inter-annual variability of the net ecosystem CO_2 exchange of a temperate mountain grassland: Effects of weather and management. *Journal of Geophysical Research – Atmospheres*, **113**. doi:10.1029/2007JD009286.

Wu C, Munger J, Niu Z, Kuang D (2010a) Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest. *Remote Sensing of Environment*. doi:10.1016/j.rse.2010.07.012.

Wu C, Niu Z, Gao S (2010b) Gross primary production estimation from MODIS data with vegetation index and photosynthetically active radiation in maize. *Journal of Geophysical Research*, **115**, D12127. doi:10.1029/2009JD013023.

Xiao J, Zhuang Q, Law B, *et al.* (2011) Assessing net ecosystem carbon exchange of US terrestrial ecosystems by integrating eddy covariance flux measurements and satellite observations. *Agricultural and Forest Meteorology*, **151**, 60–69. doi:10.1016/j.agrformet.2010.09.002.

Xiao JF, Zhuang QL, Baldocchi DD, *et al.* (2008a) Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology*, **148**, 1827–1847. doi:10.1016/j.agrformet.2008.06.015.

Xiao JF, Zhuang QL, Baldocchi et al D (2008b) Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology*, **148**, 1827–1847. doi:10.1016/j.agrformet.2008.06.015.

Xiao X, Hollinger D, Aber J, Goltz M, Davidson EA, Zhang Q, Moore IIIB (2004a) Satellite-based modeling of gross primary production in an evergreen needleleaf forest. *Remote Sensing of Environment*, **89**, 519. doi:10.1016/j.rse.2004.08.015.

Xiao X, Zhang Q, Braswell B, *et al.* (2004b) Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, **91**, 256. doi:10.1016/j.rse.2004.03.010.

Xiao X, Zhang Q, Saleska S, *et al.* (2005) Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest. *Remote Sensing of Environment*, **94**, 105. doi:10.1016/j.rse.2004.08.015.

Xiao ZQ, Liang SL, Wang JD, Song JL, Wu XY (2009) A Temporally Integrated Inversion Method for Estimating Leaf Area Index From MODIS Data. *IEEE Transactions on Geoscience and Remote Sensing.*, **47**, 2536–2545. doi:10.1109/TGRS.2009.2015656.

Xiong X, Wenny BN, Barnes WL (2009) Overview of NASA Earth Observing Systems Terra and Aqua moderate resolution imaging spectroradiometer instrument calibration algorithms and on-orbit performance. *Journal of Applied Remote Sensing*, **3**, 032501. doi:10.1117/1.3180864. URL `http://link.aip.org/link/?JRS/3/032501/1`.

Xu LK, Baldocchi DD (2004) Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California. *Agricultural and Forest Meteorology*, **123**, 79–96. doi:10.1016/j.agrformet.2003.10.004.

Yan H, Fu Y, Xiao X, Huang HQ, He H, Ediger L (2009) Modeling gross primary productivity for winter wheat-maize double cropping System using MODIS time series and CO_2 eddy flux tower data. *Agriculture Ecosystems & Environment*, **129**, 391–400. doi:10.1016/j.agee.2008.10.017.

Yang F, Ichii K, White MA, *et al.* (2007) Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach. *Remote Sensing of Environment*, **110**, 109. doi:10.1016/j.rse.2007.02.016.

Yang W, Huang D, Tan B, *et al.* (2006a) Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the terra MODIS sensor: 2000-2005. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1829–1842. doi:10.1109/TGRS.2006.871214.

Yang W, Shabanov NV, Huang D, *et al.* (2006b) Analysis of leaf area index products from combination of MODIS Terra and Aqua data. *Remote Sensing of Environment*, **104**, 297–312. doi:10.1016/j.rse.2006.04.016.

Yebra M, Chuvieco E, Riaño D (2008) Estimation of live fuel moisture content from MODIS images for fire risk assessment. *Agricultural and forest meteorology*, **148**, 523–536. doi:10.1016/j.agrformet.2007.12.005.

Yi C, Ricciuto D, Li R, *et al.* (2010) Climate control of terrestrial carbon exchange across biomes and continents. *Environmental Research Letters*, **5**, 034007. doi:10.1088/1748-9326/5/3/034007.

Young P (1998) Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software*, **13**, 105–122. doi:10.1016/S1364-8152(98)00011-5.

Young P (2000) *Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. IN: Nonlinear and nonstationary signal processing.* Cambridge University Press, Cambridge, UK, 1st edn., 74–114 pp.

Young P (2001) *The Identification and Estimation of Nonlinear Stochastic Systems. IN: Nonlinear Dynamics and Statistics.* Birkhauser, Boston.

Young P, McKenna P, Bruun J (2001) Identification of non-linear stochastic systems by state dependent parameter estimation. *International Journal of Control*, **74**, 1837–1857. doi:10.1080/00207170110089824.

Young P, Pedegral D (1999) Recursive and en-bloc approaches to signal extraction. *Journal of Applied Statistics*, **26**, 103–128. doi:10.1080/02664769922692.

Young P, Ratto M (2009) A unified approach to environmental systems modeling. *Stochastic Environmental Research and Risk Assessment*, **23**, 1037–1057. doi:10.1007/s00477-008-0271-1.

Yuan W, Liu S, Yu G, *et al.* (2010) Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. *Remote Sensing of Environment*. doi:10.1016/j.rse.2010.01.022.

Yuan WP, Liu S, Zhou GS, *et al.* (2007) Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agricultural and Forest Meteorology*, **143**, 189–207. doi:10.1016/j.agrformet.2006.12.001.

Zaehle S, Friedlingstein P, Friend AD (2010) Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophysical Research Letters*, **37**. doi:10.1029/2009GL041345.

Zaehle S, Sitch S, Smith B, Hatterman F (2005) Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. *Global Biogeochemical Cycles*, **19**, GB3020. doi:10.1029/2004GB002395.

Zhang LM, Yu GR, Sun XM, *et al.* (2006) Seasonal variations of ecosystem apparent quantum yield (alpha) and maximum photosynthesis rate (P-max) of different forest ecosystems in China. *Agricultural and Forest Meteorology*, **137**, 176–187. doi:10.1016/j.agrformet.2006.02.006.

Zhang XY, Friedl MA, Schaaf CB, *et al.* (2003) Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, **84**, 471–475. doi:10.1016/S0034-4257(02)00135-9.

Zhang YQ, Yu Q, Jiang J, Tang YH (2008) Calibration of Terra/MODIS gross primary production over an irrigated cropland on the North China Plain and an alpine meadow on the Tibetan Plateau. *Global Change Biology*, **14**, 757–767. doi:10.1111/j.1365-2486.2008.01538.x.

Zhao B, Yan Y, Guo H, He M, Gu Y, Li B (2009) Monitoring rapid vegetation succession in estuarine wetland using time series MODIS-based indicators: An application in the Yangtze River Delta area. *Ecological Indicators*, **9**, 346–356. doi:10.1016/j.ecolind.2008.05.009.

Zhao M, Heinsch F, Nemani R, Running S (2005) Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, **95**, 164–176. doi::10.1016/j.rse.2004.12.011.

Zhao M, Running S (2008) *Remote sensing of terrestrial primary production and carbon cycle*, pp. 423–444. Springer, New York, NY, USA.

Zhao M, Running S (2010) Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science*, **329**, 940. doi:10.1126/science.1192666.

Zhao WG, Qualls RJ (2006) Modeling of long-wave and net radiation energy distribution within a homogeneous plant canopy via multiple scattering processes. *Water Resources Research*, **42**. doi:10.1029/2005WR004581.

Zheng G, Moskal LM (2009) Retrieving leaf area index (LAI) using remote sensing: Theories, methods and sensors. *Sensors*, **9**, 2719–2745. doi::10.3390/s90402719. URL `http://www.mdpi.com/1424-8220/9/4/2719/pdf`.

Ziska L, Bunce J (2007) Predicting the impact of changing CO2 on crop yields: some thoughts on food. *New Phytologist*, **175**, 607–618. doi:10.1111/j.1469-8137.2007.02180.x.

A

Appendix

A.1 FLUXNET study sites

Table A.1: Name, vegetation (veg.) and climate class, used measurement years, as well as references for each study site. Vegetation classes: deciduous broadleaf forest (DBF), mixed (MF), evergreen needleleaf (ENF), evergreen broadleaf (EBF), grass (G). Köppen-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).

Name	veg.	climate	years	reference
Black Hills (US-Blk)	ENF	Dfa	2004–2006	Wilson & Meyers (2007)
Blodgett (US-Blo)	ENF	Csb	2002–2006	Goldstein *et al.* (2000)
Boreas (CA-Man)	ENF	Dfc	1995–2005	Goulden *et al.* (2006)
Donaldson (US-SP3)	ENF	Cfa	2001–2004	Gholz & Clark (2002)
Flakaliden (SE-Fla)	ENF	Dfc	2000-2002	Wallin *et al.* (2001)
GLEES (US-GLE)	ENF	Dfc	2006–2008	Massman & Clement (2005)
Griffin (UK-Gri)	ENF	Cfb	1998,2000–2001	Clement *et al.* (2003)
Hyytiälä (Fl-Hyy)	ENF	Dfc	1997–2006	Suni *et al.* (2003)
Le Bray (FR-LBr)	ENF	Cfb	2001–2003	Berbigier *et al.* (2001)
Loobos (NL-Loo)	ENF	Cfb	1997–2006	Dolman *et al.* (2002)
Metolius Interm. (US-Me2)	ENF	Csb	2002–2005,2007	Anthoni *et al.* (2002)
Metolius Young (US-Me5)	ENF	Csb	2002–2002	Anthoni *et al.* (2002)
Niwot Ridge (US-NR1)	ENF	Dfc	1999–2006	Sacks *et al.* (2006)
Norunda (SE-Nor)	ENF	Dfb	1996–2005	Lagergren *et al.* (2005)
Tharandt (DE-Tha)	ENF	Dfb	1997–2003	Grünwald & Bernhofer (2007)
Wetzstein (DE-Wet)	ENF	Dfb	2002–2008	Rebmann *et al.* (2010)

Table A.1: ...continued

Name	veg.	climate	year	reference
Wind River (US-Wrc)	ENF	Csb	1999–2004,2006	Shaw *et al.* (2004)
Yatir (IL-Yat)	ENF	BSh	2001-2002,2005	Maseyk *et al.* (2008)
Bartlett (US-Bar)	DBF	Dfc	2004–2007	Jenkins *et al.* (2007)
Duke Hardwood (US-Dk2)	DBF	Cfa	2001–2005	Stoy *et al.* (2005, 2007)
Hainich (DE-Hai)	DBF	Dfb	2000–2007	Mund *et al.* (2010)
Hesse (FR-Hes)	DBF	Cfb	1997–2007	Granier *et al.* (2008)
MMSF (US-MMS)	DBF	Dfa	1999-2006	Schmid *et al.* (2000)
Missouri Ozark (US-MOz)	DBF	Dfa	2005–2008	Gu *et al.* (2006, 2007)
Roccarespampani (IT-Ro1)	DBF	Csa	2001–2003	Keenan *et al.* (2009)
Soroe (DK-Sor)	DBF	Cfb	1997–2005	Pilegaard *et al.* (2003)
Sylvania Wilderness (US-Syv)	DBF	Dfb	2002–2004	Desai *et al.* (2005)
UMBS (US-UMB)	DBF	Dfb	1999–2003	Gough *et al.* (2008)
WalkerBranch (US-WBW)	DBF	Cfa	1995-1999	Wilson & Meyers (2007)
Willow Creek (US-WCr)	DBF	Dfb	2000–2006	Cook *et al.* (2004)
Castelporziano (IT-Cpz)	EBF	Csa	2002–2003	Seufert *et al.* (1997)
Puechabon (FR-Pue)	EBF	Csb	2001–2008	Allard *et al.* (2008)
Audubon (US-Aud)	G	BSh	2004–2008	Wilson & Meyers (2007)
Goodwin Creek (US-Goo)	G	Cfa	2004–2006	Wilson & Meyers (2007)
Lethbridge (CA-Let)	G	Dfb	1999–2004	Flanagan (2009)
Neustift (AT-Neu)	G	Dfb	2002,2005-2007	Wohlfahrt *et al.* (2008)
Oensingen (CH-Oe1)	G	Dfb	2002–2007	Ammann *et al.* (2009)
Peck (US-FPe)	G	BSk	2000–2006	Wilson & Meyers (2007)
Vaira Ranch (US-Var)	G	Csa	2001–2007	Ma *et al.* (2007)
Brasshaat (BE-Bra)	MF	Cfb	1997–2008	Carrara *et al.* (2003, 2004)
Duke (US-Dk3)	MF	Cfa	1999–2002	Siqueira *et al.* (2006)
Harvard (US-Ha1)	MF	Dfb	1992–2007	Urbanski *et al.* (2007)
Howland (US-Ho3)	MF	Dfb	1996–2004	Hollinger *et al.* (2004)
Vielsalm (BE-Vie)	MF	Cfb	2000–2008	Aubinet *et al.* (2001)

A.2 Light use efficiency model

Table A.2: Optimized light use efficiency 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 A.1.

Name	ϵ_{max}	p	T_{opt}	k_T	W_I	α	r^2	EC
US-Blk	0,95 (0,01)	0,89 (0,04)	11,60 (0,08)	6,58 (0,23)	0,24 (0,05)	0,29 (0,04)	0,78	0,78
US-Blo	0,82 (0,00)	0,52 (0,01)	5,59 (0,09)	5,59 (0,10)	0,45 (0,00)	0,91 (0,00)	0,74	0,73
CA-Man	0,78 (0,00)	0,30 (0,01)	8,91 (0,09)	4,01 (0,11)	0,29 (0,00)	0,05 (0,05)	0,85	0,84
US-SP3	1,07 (0,01)	0,44 (0,02)	17,99 (0,13)	6,01 (0,20)	0,58 (0,01)	0,97 (0,00)	0,44	0,27
SE-Fla	0,89 (0,01)	0,66 (0,01)	5,00 (0,14)	6,05 (0,17)	0,41 (0,01)	0,00 (0,24)	0,85	0,85
US-GLE	0,86 (0,01)	0,79 (0,01)	10,20 (0,08)	5,12 (0,08)	0,47 (0,01)	0,32 (0,02)	0,88	0,88
UK-Gri	1,64 (0,01)	0,98 (0,01)	8,59 (0,44)	9,87 (0,77)	0,42 (0,16)	0,00 (0,42)	0,89	0,85
Fl-Hyy	1,16 (0,01)	0,55 (0,00)	7,90 (0,05)	4,87 (0,04)	0,54 (0,00)	0,61 (0,02)	0,92	0,92
FR-LBr	1,13 (0,01)	0,62 (0,01)	12,41 (0,20)	7,07 (0,21)	0,64 (0,01)	0,00 (0,10)	0,76	0,75
NL-Loo	1,66 (0,01)	0,63 (0,00)	6,70 (0,27)	9,27 (0,25)	0,62 (0,00)	0,00 (0,13)	0,87	0,83
US-Me2	0,92 (0,00)	0,57 (0,01)	12,94 (0,08)	7,94 (0,12)	0,32 (0,01)	0,07 (0,05)	0,85	0,85
US-Me5	0,82 (0,01)	0,41 (0,01)	13,30 (0,08)	4,49 (0,13)	0,32 (0,01)	0,00 (0,04)	0,84	0,84
US-NR1	0,82 (0,00)	0,59 (0,01)	8,52 (0,04)	3,86 (0,05)	0,44 (0,00)	0,00 (0,04)	0,84	0,84
SE-Nor	0,95 (0,01)	0,60 (0,01)	7,84 (0,20)	9,27 (0,23)	0,42 (0,00)	0,00 (0,28)	0,85	0,85
DE-Tha	1,93 (0,01)	0,57 (0,00)	8,11 (0,05)	4,25 (0,04)	0,53 (0,00)	0,00 (0,04)	0,88	0,87
DE-Wet	1,82 (0,01)	0,67 (0,00)	6,78 (0,05)	4,76 (0,03)	0,59 (0,00)	0,00 (0,02)	0,89	0,88
US-Wrc	0,98 (0,02)	0,82 (0,02)	5,77 (0,09)	5,64 (0,05)	0,71 (0,02)	0,00 (0,03)	0,70	0,65
IL-Yat	1,72 (0,01)	0,50 (0,01)	7,19 (0,42)	7,39 (0,20)	0,22 (0,00)	0,98 (0,00)	0,91	0,91
US-Bar	1,14 (0,01)	0,59 (0,01)	14,52 (0,08)	3,84 (0,05)	0,47 (0,01)	0,87 (0,01)	0,93	0,93
US-Dk2	1,25 (0,00)	0,43 (0,01)	14,55 (0,11)	5,04 (0,10)	0,71 (0,00)	0,98 (0,00)	0,93	0,93
DE-Hai	1,72 (0,01)	0,39 (0,00)	11,23 (0,04)	2,24 (0,03)	0,53 (0,00)	0,00 (0,03)	0,93	0,93
FR-Hes	1,46 (0,00)	0,43 (0,01)	13,63 (0,04)	2,75 (0,04)	0,52 (0,00)	0,00 (0,07)	0,85	0,85
US-MMS	1,29 (0,00)	0,49 (0,01)	21,19 (0,19)	5,67 (0,15)	0,57 (0,00)	0,00 (0,07)	0,91	0,91
US-MOz	0,95 (0,00)	0,56 (0,00)	9,63 (0,26)	7,33 (0,16)	0,59 (0,00)	0,98 (0,00)	0,92	0,92
IT-Ro1	1,19 (0,01)	0,37 (0,00)	15,05 (0,11)	4,99 (0,08)	0,47 (0,00)	0,81 (0,00)	0,93	0,93
DK-Sor	1,73 (0,01)	0,75 (0,01)	12,64 (0,05)	3,17 (0,03)	0,50 (0,01)	0,87 (0,01)	0,87	0,87
US-Syv	0,85 (0,01)	0,79 (0,01)	20,06 (0,25)	5,83 (0,16)	0,35 (0,02)	0,10 (0,04)	0,94	0,93
US-UMB	1,07 (0,00)	0,71 (0,01)	17,42 (0,06)	4,17 (0,05)	0,57 (0,01)	0,41 (0,02)	0,96	0,96
US-WBW	1,10 (0,01)	0,48 (0,01)	18,35 (0,10)	4,43 (0,08)	0,58 (0,00)	0,09 (0,08)	0,88	0,88
US-Wcr	1,28 (0,00)	0,67 (0,02)	18,66 (0,06)	3,39 (0,09)	0,40 (0,01)	0,33 (0,03)	0,90	0,90
IT-Cpz	1,25 (0,01)	0,14 (0,01)	23,40 (3,39)	12,00 (2,78)	0,37 (0,00)	0,95 (0,00)	0,77	0,61
FR-Pue	1,03 (0,00)	0,60 (0,00)	8,51 (0,10)	5,69 (0,05)	0,43 (0,00)	0,84 (0,01)	0,79	0,75
US-Aud	1,04 (0,01)	0,24 (0,01)	24,45 (0,13)	2,00 (0,12)	0,33 (0,00))	0,00 (0,05)	0,68	0,68
US-Goo	1,47 (0,01)	0,31 (0,01)	23,56 (0,08)	2,00 (0,08)	0,53 (0,00)	0,95 (0,00)	0,88	0,87
CA-Let	1,49 (0,01)	0,35 (0,00)	12,19 (0,12)	4,68 (0,08)	0,47 (0,00)	0,00 (0,04)	0,92	0,92
AT-Neu	1,80 (0,01)	0,64 (0,02)	12,21 (0,08)	4,04 (0,12)	0,64 (0,01)	0,00 (0,08)	0,83	0,82
CH-Oe1	2,25 (0,01)	0,54 (0,01)	9,27 (0,17)	7,56 (0,14)	0,78 (0,00)	0,00 (0,07)	0,88	0,86
US-FPe	1,09 (0,02)	0,63 (0,01)	9,13 (0,31)	6,30 (0,19)	0,43 (0,01)	0,00 (0,10)	0,49	0,48
US-Var	1,36 (0,00)	0,38 (0,00)	13,24 (0,05)	2,71 (0,04)	0,57 (0,00)	0,91 (0,00)	0,91	0,91
BE-Bra	1,05 (0,01)	0,66 (0,00)	17,66 (0,42)	10,13 (0,41)	0,57 (0,01)	0,00 (0,16)	0,87	0,87
US-Dk3	1,03 (0,01)	0,69 (0,01)	16,52 (0,28)	12,00 (0,68)	0,55 (0,01)	0,11 (0,51)	0,84	0,82
US-Ha1	1,41 (0,02)	0,89 (0,01)	20,88 (0,36)	6,34 (0,17)	0,56 (0,02)	0,78 (0,01)	0,91	0,91
US-Ho3	1,32 (0,01)	0,28 (0,00)	5,00 (0,06)	2,00 (0,05)	0,31 (0,00)	0,00 (0,09)	0,88	0,88
BE-Vie	1,09 (0,01)	0,92 (0,01)	14,81 (1,61)	12,00 (1,60)	0,37 (0,03)	0,98 (0,00)	0,87	0,85

A.3　Parameter extrapolation

Table A.3: Site characteristics used to regionalize the model parameters. See Table A.1 for explanations of the Köppen-Geiger-climate and vegetation classes.

attribute	description
vegetation and climate classes:	
B,C,D	Main type of the Köppen-Geiger classification
f,s	Subtype 1 of the Köppen-Geiger classification
a,b,c	Subtype 2 of the Köppen-Geiger classification
ENF, DBF, MF, EBF, G	Vegetation class
climate characteristics:	
meanTS	mean annual soil temperature (T_S) [°C]
diffTS	amplitude of T_S [°C]
maxTS	maximum T_S [°C]
meanTA	mean annual air temperature (T) [°C]
sumP	average sum of precipitation per year [mm]
Pdays	number of days with precipitation [d]
maxPAR	maximum photosynthetically acitve radiation (PAR) [MJ m^{-2} d^{-1}]
sumPAR	average annual cumulative sum of PAR [MJ m^{-2} d^{-1}]
sunshine	average fraction of sunshine per year (according to LocClim) [%]
Rnet	annual net radiation balance [MJ m^{-2} d^{-1}]
maxEF	average maximum evaporative fraction (EF)
minEF	average minimum evaporative fraction (EF)
diffEF	amplitude of the EF
meanEF	average EF
PE	Potential monthly evaporation [mm] (according to LocClim)
AI	aridity index (Budyko, 1958): $AI = 100 \cdot Psum / (Rnet \cdot \lambda)$
CI	continentality index (Conrad, 1950): $CI = 1.7 \cdot (T_{max} - T_{min}) / sin(\phi + 10) - 14$
physiological characteristics:	
maxLAI	maximum leaf area index (LAI) [m^2 m^{-2}]
diffLAI	amplitude of the LAI [m^2 m^{-2}]
meanLAI	average LAI [m^2 m^{-2}]
maxEVI	maximum enhanced vegetation index (EVI)
minEVI	minimum EVI
diffEVI	amplitude of the EVI
meanEVI	average EVI
age	age of the vegetation stand [a]
seasonal characteristics:	
vegLength	average length of the vegetation period [d]
VPstart	average day of year of the start of the vegetation period
VPend	average day of year of the end of the vegetation period
VPstartEVI	average EVI at the start of the vegetation period
VPendEVI	average EVI at the end of the vegetation period
VPstartLAI	average LAI at the start of the vegetation period [m^2 m^{-2}]
VPendLAI	average LAI at the end of the vegetation period [m^2 m^{-2}]
r2TS-LAI	correlation between the TS and LAI time series
r2EF-LAI	correlation between the EF and LAI time series
r2TS-EF	correlation between the TS and EF time series
r2EF-EVI	correlation between the EF and EVI time series
r2TS-EVI	correlation between the TS and EVI time series

Table A.4: SVR extrapolation performance related to the calibrated model (columns 2 to 5) and related to the measured time series of carbon uptake (remaining columns). For the various performance measures see section 2.4. RMSE and bias are given in units of gC m^{-2} d^{-1}, RE$_S$ in %.

site	r^2	$RMSE$	$nRMSE$	bias	r^2	$RMSE$	$nRMSE$	bias	RE$_S$
US-Blk	0,99	0,52	0,07	-0,41	0,78	1,02	0,15	-0,36	-13,76
US-Blo	0,92	1,06	0,14	-0,67	0,69	1,33	0,17	-0,61	-20,42
CA-Man	0,99	0,77	0,11	-0,54	0,83	1,02	0,15	-0,44	-31,11
US-SP3	0,97	0,75	0,09	-0,59	0,43	1,63	0,19	-0,35	-6,83
SE-Fla	1,00	0,26	0,03	-0,13	0,85	0,79	0,09	-0,01	-0,38
US-GLE	0,98	0,43	0,05	0,06	0,86	0,88	0,11	0,06	2,69
UK-Gri	0,99	0,77	0,06	0,57	0,90	1,54	0,12	1,12	23,37
FI-Hyy	1,00	0,21	0,02	0,13	0,92	0,71	0,08	0,23	11,25
FR-LBr	0,96	0,82	0,09	0,56	0,71	1,34	0,14	0,71	21,27
US-Me2	0,99	0,64	0,06	0,50	0,87	1,21	0,11	0,85	22,97
US-Me5	0,91	0,86	0,09	0,27	0,75	1,28	0,14	0,36	10,28
NL-Loo	0,92	0,43	0,08	-0,11	0,74	0,75	0,14	-0,06	-2,83
US-NR1	1,00	0,25	0,04	-0,20	0,85	0,79	0,11	-0,25	-13,89
SE-Nor	1,00	0,67	0,07	-0,46	0,85	1,14	0,11	-0,29	-12,29
DE-Tha	0,97	0,80	0,06	0,29	0,83	1,49	0,11	0,59	14,31
DE-Wet	0,99	0,31	0,03	-0,01	0,89	1,07	0,09	0,25	7,04
US-Wrc	0,97	1,19	0,14	-0,91	0,68	1,55	0,18	-0,65	-23,07
US-Bar	0,96	0,77	0,06	0,11	0,90	1,27	0,10	0,22	7,17
US-Dk2	1,00	0,41	0,03	0,24	0,92	1,06	0,08	0,37	8,68
DE-Hai	0,98	0,66	0,05	-0,27	0,91	1,31	0,09	-0,10	-2,89
FR-Hes	0,99	0,63	0,04	0,30	0,84	1,77	0,10	0,57	15,67
US-MMS	0,97	0,93	0,06	0,37	0,88	1,62	0,11	0,45	11,78
US-MOz	0,99	0,36	0,03	0,19	0,91	0,88	0,08	0,33	10,96
IT-Ro1	0,96	0,68	0,07	-0,04	0,89	0,98	0,09	-0,02	-0,74
DK-Sor	1,00	0,63	0,03	0,39	0,87	1,58	0,09	0,69	19,79
US-Syv	0,97	1,71	0,13	-0,99	0,91	1,87	0,14	-0,90	-37,46
US-UMB	0,99	0,74	0,06	0,47	0,95	1,09	0,08	0,48	16,95
US-WBW	1,00	0,46	0,03	-0,18	0,88	1,37	0,09	0,03	0,68
US-WCr	1,00	0,19	0,01	0,06	0,90	1,52	0,09	0,14	4,09
IT-Cpz	1,00	0,12	0,02	0,04	0,76	0,88	0,13	0,28	6,74
FR-Pue	0,94	0,66	0,09	-0,35	0,74	1,10	0,15	-0,11	-3,66
US-Aud	0,98	0,33	0,04	0,18	0,66	0,89	0,10	0,08	12,42
US-Goo	0,99	0,73	0,06	-0,56	0,86	1,28	0,11	-0,32	-9,26
CA-Let	0,98	0,28	0,03	-0,03	0,90	0,62	0,06	-0,04	-3,05
AT-Neu	0,98	0,88	0,05	0,55	0,82	2,16	0,13	0,98	18,58
CH-Oe1	0,94	1,08	0,07	0,40	0,83	1,83	0,11	0,80	15,16
US-FPe	0,99	0,19	0,03	-0,06	0,49	0,82	0,15	0,07	7,58
US-Var	0,95	0,76	0,07	-0,51	0,86	1,12	0,10	-0,50	-26,12
BE-Bra	0,99	1,26	0,13	-0,96	0,86	1,49	0,16	-0,78	-31,56
US-Dk3	0,99	0,81	0,07	0,68	0,85	1,48	0,13	0,99	21,44
US-Ha1	0,97	0,87	0,05	-0,36	0,87	1,61	0,09	-0,28	-7,34
US-Ho3	0,94	0,85	0,08	0,08	0,81	1,43	0,13	0,30	8,97
VE-Vie	0,99	0,66	0,06	0,50	0,87	1,28	0,11	0,81	22,52

Curriculum vitae

Personal and contact information

Name: Judith Elisabeth Horn

Date of birth: 30.04.1978

Place of birth: Offenburg

Nationality: German

Postal address: Egerlandstr. 5a, 76228 Karlsruhe

email address: judith.horn@email.de

Education

1984–1988:   Elementary School in Ohlsbach

1988–1997:   Secondary School "Klostergymnasium" in Offenburg

1998–1999:   Studies of meteorology, Technical University of Karlsruhe

1999–2006:   Studies of geoecology, Technical University of Karlsruhe

                 Main subjects: Landscape ecology, soil science, remote sensing, hydrology, meteorology, geoinformatics

2004–2005:   Research studies at the Department of Land, Air and Water Resources, University of California, Davis; DAAD visiting-scholarship

2005:   Diploma thesis: "Numerical simulation of septic tank effluents and their impact on domestic drinking water wells in the Modesto area, California"

Academic career

2006-2007:   PhD studies at the Center for Environmental Research – UFZ, Leipzig

2007-2009:   Continuation of the PhD studies at the Department of Geography at the Ludwig-Maximilians-Universität (LMU), München (LMU)

2010-2011:   Scientific coordination of the proposed Cluster of Excellence "Hydrosphere", Institute for Hydromechanics, Karlsruhe Institute of Technology (KIT)

2011:   Scientific assistant, Institute of Photogrammetry and Remote Sensing, KIT

Scientific publications

Horn, J.E., Harter, Th., 2009: Domestic well capture zone and influence of the gravel pack length. Ground Water 47(2), 277-286.

Horn, J.E., Schulz, K., 2010: Post-processing analysis of MODIS leaf area index subsets. Journal of Applied Remote Sensing 4, 043557.

Horn, J.E., Schulz, K., 2011: Identification of a general light use efficiency model for gross primary production. Biogeosciences 8(4), 999-1021.

Horn, J.E., Harter, Th., 2011: Domestic wells have high probability of pumping septic tank leachate. Hydrology and Earth System Sciences Discussions 8, 5701-5732.

Horn, J.E., Schulz, K., 2011: Spatial extrapolation of light use efficiency model parameters to predict gross primary production. Journal of Advances in Modeling Earth Systems (submited, in revision).