Modelling and validation of agricultural and forest biomass potentials for Germany and Austria

Dissertation

Zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

vorgelegt von

Markus Tum

aus Göttingen (Geburtsort)

Göttingen 2012

Gutachter:

1. Prof. Dr. Martin Kappas

2. Prof. Dr. Hans Ruppert

3. Dr. Kurt P. Günther

Tag der mündlichen Prüfung: 23.04.2012

Acknowledgements

Preface

This doctoral thesis comprises the following papers, book entries and manuscripts:

Markus Tum, Kurt P. Günther: Validating modelled NPP using statistical yield data, *Biomass & Bioenergy*, 2011, 35, 4665-4675.

Markus Tum, Marcel Buchhorn, Kurt P. Günther, Bejamin C. Haller: Validation of modelled forest biomass in Germany using BETHY / DLR, *Geoscientific Model Development*, 2011, 4, 1019-1034.

Markus Tum, Kurt P. Günther: Estimating Agricultural Bioenergy Potentials for Germany Using a Process Based Vegetation Model, *Sustainable Bioenergy*, Spinger Book, accepted for publication

Markus Tum, Franziska Strauss, Ian McCallum, Kurt P. Günther, Erwin Schmid: How sensitive are estimates of carbon fixation in agricultural models to input data?, *Carbon Balance and Management*, accepted for publication

Markus Tum, Erik Borg: A Conceptual Remote Sensing based Interception-Infiltration Model for Regional and Global Applications, *Hydrology and Earth System Sciences*, under review

Markus Niklaus, Kurt P. Günther, **Markus Tum**, Michael Bittner: Generation of a global, gap-free SPOT-VGT LAI dataset using spectral analysis techniques, *International Journal of Remote Sensing*, under review

Contents

Chapter 1

Introduction

In the context of the Seventh Framework Programme (EU-FP7) project Ener-GEO (Earth observation for monitoring and assessment of the environmental impact of energy use) strategies shall be developed to:

- assess the impact of increased energy use on the environment and on ecosystems at a global scale, and to

- determine the optimal mix of energy sources to achieve sustainability.

The global observation strategy will assess the impacts of current and future transitions in energy use on the environment by a combination of: 1) models already available for various sources of energy, 2) existing global datasets from which environmental indicators will be derived to quantify changes to earth systems, and 3) existing and currently developed models capable of assessing environmental impacts and costs of energy exploitation. The Biosphere Energy Transfer Hydrology (BETHY/DLR) model, developed at the German Aerospace Center (DLR), will be used to estimate bioenergy potentials for agricultural and forest areas. For pilot studies, Germany and Austria were chosen as areas of investigation and for validation.

This thesis project focuses on three topics:

- The validation of agricultural and forest Net Primary Productivity (NPP) modelled with BETHY/DLR,

- the estimation of sustainable energy potentials, based on modelled NPP, and

- the development of a one-dimensional soil water transport model to improve the soil water balance formulation in vegetation models such as BETHY/DLR.

To provide an overview of the following seven chapters, and to introduce, at a broad scope, the three main topics, a short introduction is given here.

Biomass modeling

Due to ongoing human activity over the past several centuries, many ecosystems have been permanently changed. Current scientific research suggests that these anthropogenic impacts will play an increasingly large role, and challenge to contribute to the acquisition of natural and anthropogenic caused ecosystem changes. Accurate models of the carbon uptake by vegetation (Net Primary Productivity, NPP) are urgently needed to answer questions regarding the carbon exchange between vegetation, the atmosphere, and pedosphere. Since NPP is directly linked to biomass, NPP estimates can be used as proxies for global and regional carbon sinks and sources. For these reasons biomass modelling has gained a prominent position in models of the predicted effects of climate change (IPCC (2007)).

In addition, the estimation of NPP has recently become a fundamental research topic in areas of ecology and environmental science (e.g. Sala and Austin (2000)). The photosynthetic ability of plants is essential to life on Earth; it both builds organic molecules and produces oxygen, given only light and carbon dioxide. Vegetation is a major component of the biosphere, and because it constantly draws carbon dioxide from the surrounding atmosphere, it significantly contributes to the regulation of the global carbon cycle. With

the launch of the first Earth observation satellites in the early 1970s, research studies became possible at a meso-scale level, and the power of these satellites has increased enormously in the decades since. With modern scientific techniques, it is increasingly possible to quantify changes at regional to global scale, and to observe exchanges between subsystems such as the biosphere and atmosphere. Worldwide, due to climate change, the biosphere is in danger of becoming a net carbon source as a result of various positive feedback processes (Treter (2000)). The rapid rise of greenhouse gases is now generally accepted as posing a danger to humanity, and the "climate change discussion" has gained an important role in both research and politics (IPCC (2007)).

The approaches used to assess NPP and biomass range from simple correlations of parameters derived from satellite or airborne sensors to highly complex models which take into account many interactions. In particular, simple approaches have been designed for a broad spectrum of sensors operating at medium to high resolution. The Normalized Differenced Vegetation Index (NDVI), based on optical sensors, is used for such correlation-based approaches (e.g. Tangki and Chapalle (2008), Houghton et al. (2007), Myneni et al. (2001)). NDVI is a parameter that describes vegetative greenness, and ranges from 0 (no vegetation) to 1 (fully covered with healthy vegetation). This parameter can thus be directly linked to the photosynthetic capacity of plant canopies (Myneni et al. (1995), Sellers et al. (1992)).

LIDAR and RADAR sensors can also be used to directly estimate biomass (e.g. Lang et al. (2004), Svoray and Shoshany (2002)); such instruments can be either airborne or spaceborne. Airborne sensors usually have the advantage of relatively high resolution, but because they provide only a small swath width and limited temporal cover, these methods are primarily used for local to regional approaches. In contrast, spaceborne sensors generally have the advantage of providing data on a global scale. The temporal coverage is usually also higher, but this depends on the spatial resolution. Simrad et al. (2011) recently published a new global map of forest canopy height derived using LIDAR data from ICESat (Ice, Cloud, and land Elevation Satellite).

For modelling approaches, information about the development of vegetation and the seasonal greenness of the Earth's surface is essential. This information can be derived both from in situ measurements and from satellite or airborne

remote sensing observations. Sellers (1985) showed that plant-physiological parameters such as the Leaf Area Index (LAI) and the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) can be derived from vegetation indices such as NDVI. NDVI data has itself been derived from sensors such as the AVHRR (Advanced Very High Resolution Radiometer) (Eidenshink (1992)), MODIS (Moderate resolution Imaging Spectroradiometer) (Myneni et al. (2002)), SPOT-VGT (Satellite Pour l'Observation de la Terre - Vegetation) (Baret et al. (2007)), and MERIS (Medium Resolution Imaging Spectrometer) (Günther and Maier (2007), Gobron et al. (2004)). LAI, in particular, plays a major role in most ecosystem and biogeochemical models in which it is used, because of its correlation to Gross Primary Productivity (GPP) (Sellers et al. (1996)).

Many simple models use a description of physical, chemical and plant physiological processes, taking into account interactions with the atmosphere. Photosynthesis is modelled following the Light Use Efficiency (LUE) method of Monsi and Saeki (1953) and Monteith (1965) (e.g. Richters (2005), Williams (1995)). These model approaches are often applied in agriculture: typically, when coupled with the analysis of management practices, they are used to forecast agricultural outputs (e.g.: Li et al. (1992), Jones and Kiniry (1986), Ritchie and Otter (1985)).

In contrast, more sophisticated models exist that take into account the conservation of energy and momentum, based on the approaches of Farquhar et al. (1980) and Collatz et al. (1992) (e.g. Krinner et al. (2005), Wißkirchen (2005), Knorr (1997), Prentice et al. (1992)). Soil-Vegetation-Atmosphere Transfer (SVAT) models such as BETHY/DLR track photosynthesis at the molecular level and take into account environmental conditions that affect its efficiency. Dynamic Global Vegetation Models (DGVM) additionally track competitive vegetation growth and species succession.

For a more detailed overview on modelling approaches and literature see **chapter two (Validating modelled NPP using statistical yield data)** and **chapter three (Validation of modelled forest biomass in Germany using BETHY/DLR)**.

Bioenergy

Due to recent changes in German energy policy, renewable energy sources are receiving greater attention due to their potential to replace nuclear power and fossil fuels. With the latest decision to change the Atomic Energy Law and to finally discontinue nuclear power in 2022, compensatory sources of energy are needed.

The use of biomass as an energy source is highly controversial because of fears that it would compete with food production. However, recent studies have shown that regenerative, non-competitive residual energy potentials are available and could be used Thrän et al. (2010) without affecting important agricultural desiderata such as soil fertility Zeller et al. (2011). Currently 10.9 % of the German energy mix comes from renewable sources, of which the majority, 7.7 %, is from bioenergy (BMU (2011)). These rates compare favourably with many other European countries. However, Germany plans for 18 % of its mix to come from renewables by 2020, and achieving that goal will be challenging. Other countries, such as Sweden and Finland, not only already have a higher share of renewables in their energy mix, but are also closer to their goals for 2020. Globally, in 2008 16.6 % of the energy mix came from renewable sources, dominated by biomass (12.2 %), a trend which has been accelerated by the increased production of liquid energy carriers such as biodiesel (BMU (2011)).

A recent study examined unused bioenergy potentials, concluding that for Germany's forests 52 % of the technical fuel potential, which corresponds to $265\,\mathrm{PJ\,yr^{-1}}$, remains unused (Thrän et al. (2010)). The same study calculated that sustainable use of grain yield residuals could provide another $85\,\mathrm{PJ\,yr^{-1}}$. The technical potential describes the part of the theoretical potential which can be used under given technical limitations and is thus depending on time and location. An overview of definitions of potentials can be found in Kaltschmitt and Hartmann (2001). According to German biomass regulations, several biomass sources may be considered. From the perspective of sustainability, biomass waste sources - mainly yield residuals such as straw and commercially unusable tree parts - are of particular interest (BMU (2012)). However, the assessment of potentials at the country or the continental level is still a challenging task. Such surveys are both costly and time-consuming, and are thus

not usually conducted on an annual basis.

Remote sensing technologies, and especially the combination of these technologies with modelling approaches, have the potential to solve this problem. Straw and forest biomass increases are both directly related to NPP, which can be modelled as described above. In this study it will be shown how bioenergy potentials can be assessed and how they are spatially distributed within the areas studied.

For more detailed information and literature please see **chapter two (Validating modelled NPP using statistical yield data)**, **chapter three (Validation of modelled forest biomass in Germany using BETHY / DLR)** and **chapter four (Estimating agricultural bioenergy potentials for Germany using a process-based vegetation model)**.

Water Transport Models

The realistic prediction of hydrodynamics in vegetation models remains a challenging task, but it is important in order to accurately model physical processes related to the soil water budget. Hydrodynamic processes also govern climate because of its strong connections to the hydrologic cycle (Anderson (1992), Parry (1992)) and the potential of vegetation to cover habitats (Köppen (1936), Guetter and Kutzbach (1990)).

Often simple one-layer soil or "bucket" models are used to predict the soil water balance in SVAT models (e.g. Boulet et al. (2000), Knorr and Heimann (2001), Wißkirchen (2005)) based on the approach taken by Eagleson (1978). The term "bucket" in this context means that soil water characteristics are treated as analogous to the rising and falling level of water in a bucket. These models usually employ a daily timestep and incorporate few climate variables (Evans and Trevisan (1995)). In this approach the water available to plants is equally distributed throughout the bucket; little attention is paid to the effects of surface tension, water adhesion to soil particles, and the natural movement of water through the soil.

To predict the movement of water into and through the unsaturated zone of soils, more detailed approaches are needed. Models based on the Richards equation have proved valuable; however, such models can only be used suc-

cessfully if reliable estimates are available for the soil water retention curve (SWRC) and the unsaturated hydraulic conductivity. The SWRC may be described with different approaches (e.g. van Genuchten (1980), Campbell (1974)), but in general many input parameters are needed. In practice, most SVAT models and DGVMs instead use various multi-layer approaches (Bondeau et al. (2007), Krinner et al. (2005), de Rosnay and Polcher (1998)) to describe soil water characteristics in a more realistic way.

Besides allowing the realistic representation of soil water flows in vegetation modelling, there may be further applications of this approach. More realistic soil water budget models could provide insight into the causes and consequences of the depletion of fossil water reserves and the lowering of ground water table levels, leading to a water-based sustainability perspective on agriculture and forestry. Modelling of other processes affected by soil water movement, such as soil erosion and the soil degradation effects of different land use practices, may also be enabled by this approach.

For more details please see **chapter six (A 1D soil water transport model, driven with van Genuchten parameters and remote sensing data)**.

Introduction to the following chapters

The purpose of this chapter is to give a broad overview of the following chapters, their contents, and their logical connections. To assist in the location of specific main topics, Figure 1.1 illustrates the structure of this thesis.

A detailed introduction to NPP modelling options and their advantages and disadvantages is presented in **chapter two** and **chapter three**. The general model description of BETHY/DLR can also be found here. More details about the geobiochemical modelling approach used in BETHY/DLR are given in Knorr (1997), Knorr and Heimann (2001) and Wißkirchen (2005).

All needed input datasets are described in **chapter two** and **chapter three**. Time series analysis techniques were applied to derive gap-free, outlier-corrected Leaf Area Index (LAI) time series from the CYCLOPES and geoland2 datasets, which are based on SPOT-VEGETATION data and given as 10-day composites. Data gaps may be due to various reasons, mainly cloud obstruction and solar angle. We used Harmonic Analysis (HA) to mathematically fill data gaps
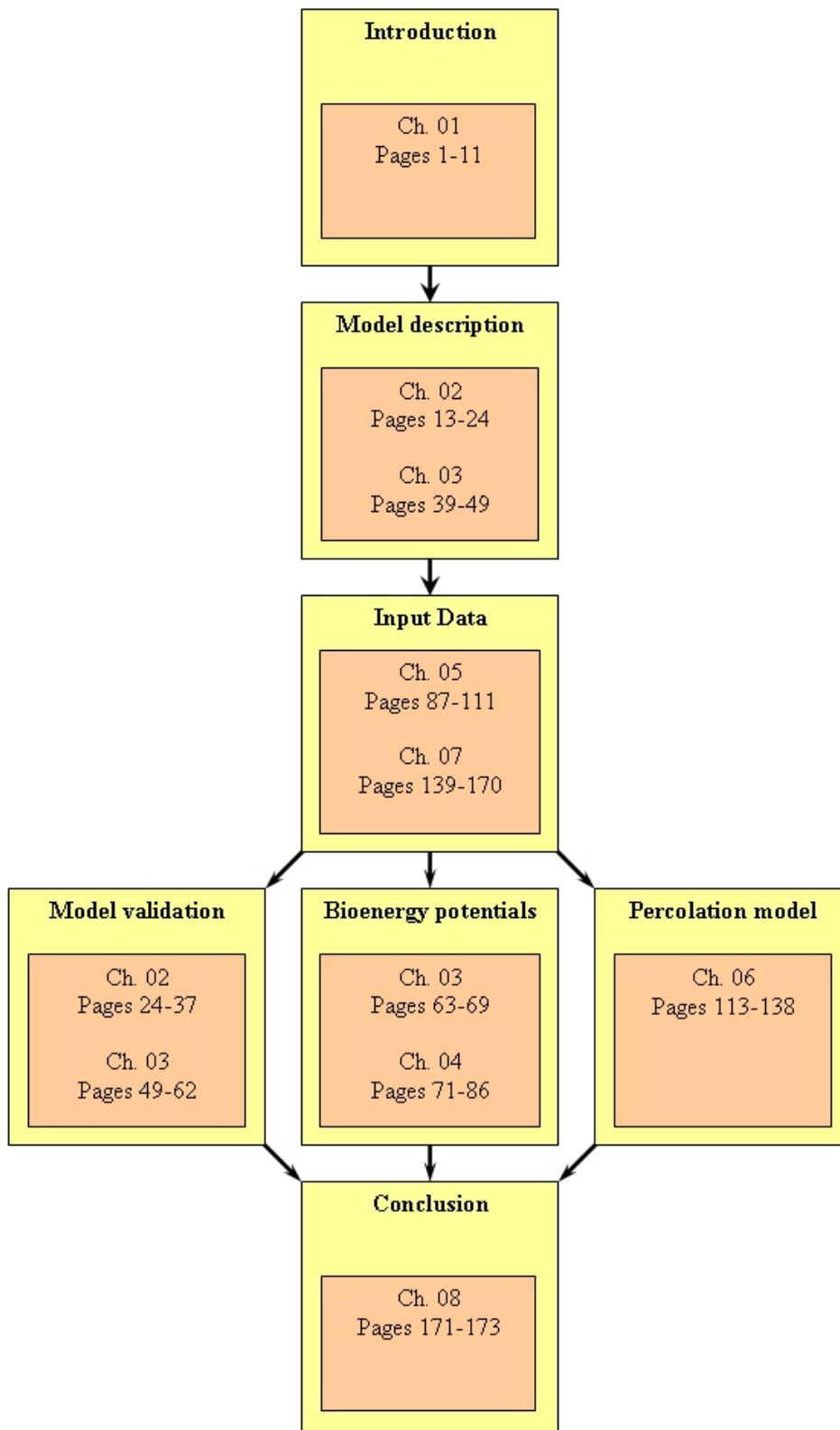
Figure 1.1: *Overview of the dissertation.*

and to detect and eliminate outliers. This was done across the whole time series (1999 to 2010), although a gap remains for all of 2008 and 2009 since no data was available here. A detailed description and validation of this process can be found in **chapter seven**.

Meteorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF) was validated with measured data from 39 stations in Austria. Land cover data (GLC2000) was also compared with another product (CORINE 2000). For this validation, several model runs for a defined area (Marchfeld, Austria) were conducted and crosschecked with the model output of the already-validated EPIC model. More detailed information about this procedure can be found in **chapter five**.

In order to validate the BETHY/DLR model, two comprehensive approaches were developed and applied. For agricultural areas, empirical data on yield and area use, taken from national statistics, were compared with model outputs. The validation was conducted for Germany and Austria over 2000 and 2001. As a first step, the yield data was recalculated to NPP using crop-specific conversion factors such as shoot-to-root and yield-to-straw ratios. This was done at a NUTS-3 resolution. NUTS is an abbreviation for "Nomenclature des Unités Territoriales Statistiques," and is a system of hierarchically organised territorial units used for statistical purposes. Then in a second step the modelled NPP data was aggregated to the same NUTS-3 resolution and directly compared with the empirical NPP estimates. This approach and its results are explained and presented in **chapter two**.

Validation of BETHY/DLR for forests followed a comparable approach, using empirical data on mean annual above-ground biomass increments (MAI). This data was obtained from Germany's national forest inventories, available at NUTS-1 resolution. To allow modelled NPP to be compared to these MAI estimates, the NPP was transformed into current annual above-ground biomass increments (CAI). The CAI data was then aggregated to NUTS-1 resolution for comparison to the MAI data. This validation was conducted for 2000 and 2001, and was done separately for coniferous and deciduous trees by using tree-species-specific conversion factors for above-to-below-ground biomass and carbon content. In addition, a quality check of modelled GPP was carried out using empirical GPP values calculated from measured Net Ecosystem Ex-

change (NEE) time series obtained from two FLUXNET stations. More details on these approaches and results can be found in **chapter three**.

To estimate sustainable bioenergy potentials for both agricultural and forest areas, two approaches were developed. For agricultural areas it was assumed that only grain yield residuals (straw, in this case) would be used for bioenergy production, since the use of harvested grain would compete with the use of grain for food, posing problems for sustainability. Competing uses of straw, such as the humus balance and usage for animal housing, were also taken into account. As a first step, empirical data on land use and mean yields of major crops were used to calculate the potential production of straw, again using conversion factors such as the shoot-to-root and yield-to-straw ratios. The technical energy potential was then estimated using lower heating values. Details on this approach and an example case study for Germany over 2006 and 2007 are presented in **chapter four**.

For forest areas the sustainable theoretical bioenergy potential was also estimated. In this case, competing uses were not taken into account, since the complexity of this problem would have necessitated the development of an entirely new model; the estimated potentials are therefore theoretical maxima, not realizable potentials. To generate these estimates, the available CAI data were transferred to energy potentials using lower heating values. The lower heating values were estimated for each tree class (deciduous and coniferous) and for each NUTS-1 region separately, to respect local variation in tree age and type across Germany. Further details are given in **chapter three**.

During the validation exercises, one model process was identified as having a particularly large potential for improvement. The "bucket model" formulation used in BETHY/DLR for the soil water budget is problematic because of its potential to predict unrealistic water availability, as discussed above. To contribute on the improvement on the prediction of soil water availability, a non-linear percolation model was adapted. It follows the approach of Syring and Kersebaum (1988) who originally developed their model for small scale surveys. It utilizes the Richards equation and the approach of van Genuchten (van Genuchten (1980)). A particular advantage of this new percolation model is that it dynamically treats 128 different FAO soil types. For each soil type, individual depth, layering and texture compositions were calculated, based on

10,000 soil profiles. Van Genuchten parameters were also estimated using the ROSETTA software package (Schaap et al. (2001)). For a quality assessment the percolation model was compared with ECMWF soil water budget data. Further information is given in **chapter six**.

Finally, in **chapter eight** a short conclusion and summary is given reflecting on the findings from this research and proposing future directions for work.

Chapter 2

Validating modelled NPP using statistical yield data

Markus Tum, Kurt P. Günther, *Biomass & Bioenergy*, (2011), 35, 4665-4675

2.1 Abstract

The German Remote Sensing Data Center operates the Biosphere Energy
Transfer Hydrology Model, a process model that estimates the net primary
productivity of agricultural areas. The model is driven by remote sensing data
and meteorological data. Remotely sensed datasets including a time series of
the leaf area index, which describes vegetation condition, and a land cover
classification, which provides information about land use, are needed. Cur-
rently leaf area indices and land cover data derived from the sensor vegetation
are used. Both datasets have spatial resolutions of about $1\,\text{km} \times 1\,\text{km}$ and
are freely available for the area of investigation (Germany and Austria). The
meteorological input parameters are air temperature (at $2\,\text{m}$ height), precip-
itation, cloud cover, wind speed (at $10\,\text{m}$ height) and soil water content (in
the four uppermost soil layers); these are obtained from the European Centre
for Medium-Range Weather Forecasts, with a spatial resolution of about $0.25\,°$
$\times\ 0.25\,°$ and a temporal resolution up to four times daily. The output of the
model, the gross primary productivity, is calculated at daily resolution. By
subtracting the cumulative plant maintenance and growth respiration, the net
primary productivity is then determined. In order to validate the modelled net

primary productivity, crop yield estimates derived from the national statistics of Germany and Austria are used. After estimating above-ground biomass using plant-specific above- to below-ground ratios, conversion factors (corn-to-straw and leaf-to-beet relations) are applied to estimate total biomass. Finally the carbon content of dry matter is estimated. To correlate model results with these statistical data, the modelled data are aggregated to net primary productivity per administrative district. The results show that a process model using remote sensing data as input can deliver reliable estimates of agricultural biomass potential which are highly correlated with statistically derived estimates of actual biomass produced.

2.2 Introduction

In one of the earliest forays into computational prediction of agricultural yield Ritchie and Otter (1985) developed the Crop Environmental Resource Synthesis (CERES) model for simulating the daily growth and development processes of wheat and maize. Later this model was expanded to sorghum, millet, rice and barley. Many factors including environment, nitrogen availability, water stress, pests, genetics and management are considered in CERES to model growth and development. The development processes are differentiated in two stages: the vegetative stage with germination, emergence, end of juvenile and leaf numbers, and the reproductive stage with floral induction, flowering, begin of grain filling and maturity. Stress components, such as water stress, act in different ways depending on the development stage.

The daily growth of plants is modelled in CERES according to the Radiation Use Efficiency (RUE) approach, which is based on the concepts of Monsi and Saeki (1953) and Monteith (1965). In this approach, the potential maximum dry matter production is linearly correlated with the absorbed light. As in most mechanistic models, RUE also varies with temperature, nitrogen and water availability, CO_2 level and fertilization. The allocation of assimilated carbon to particular plant components is modelled, with daily time steps.

Phenology, the timing of biological processes, is driven by temperature, expressed as either thermal temperature or growing-degree-days. In order to calibrate the CERES model, field data are needed, especially the number of

plants planted per unit area and the timing of phenological events such as tilling, stem elongation, and maturation. Grain yield metrics are also mandatory. The CERES model is now integrated in the Crop Simulation Model (CSM) of the Decision Support System for Agrotechnology Transfer (DSSAT) distributed by the International Consortiumfor Agricultural Systems Applications in Honolulu Jones and Kiniry (1986). In its earliest form the DSSAT model was developed to simulate maize growth and development, but in the DSSATCSM, 27 different cropping system models are combined. At a minimum, it needs input data regarding incoming solar radiation, minimum and maximum temperatures, and rainfall. It can additionally utilize several soil-related metrics, such as bulk density, carbon content, and pH, as well as management-related metrics such as planting density, fertilization rates and irrigation data.

Another important crop growth model is the DeNitrification and DeComposition (DNDC) model, originally developed by Li et al. (1992). In DNDC, crop growth is parameterized by generalized crop growth curves together with a crop-specific potential maximum grain yield. The actual grain yield is determined by the availability of nitrogen in the soil. Nitrogen uptake by the plants is controlled by the soil temperature profile and soil moisture. With this approach, the effects of differences in tilling, fertilizer use and irrigation can be taken into account by DNDC, because all of these management practices modify the soil regime and thus affect plant growth. DNDC also integrates crop growth processes with biogeochemical processes by including important nitrogen- and carbon related processes like mineralization, ammonia volatilization, denitrification and nitrification, nitrogen uptake and leaching. The DNDC model, presently implemented with a daily time step, has been validated and used for many subnational and national case studies (e.g.: Stange et al. (2000), Cai et al. (2003), Beheydt et al. (2007)).

The Environment Policy Integrated Climate (EPIC) model is a further Monteith type parametric model which is driven by the International Institute for Applied System Analysis. EPIC was originally designed to quantify the effects of erosion on soil productivity Williams et al. (1984), but has since been expanded into a complex agro-ecosystem model that simulates the growth of crops under complex rotation management operations, such as irrigation, fertilization and tillage Williams (1995). EPIC's main inputs are meteorological

data, provided by the European Center for Medium-Range Weather Forecasts (ECMWF), soil type information from the Food and Agriculture Organisation (FAO) of the United Nations, and field management data.

A further example is the DAYCENT (Daily Century) model, developed by Parton et al. (1994) and used by Del Grosso et al. (2001) and Parton et al. (1998).

These examples show a broad scientific and practical acceptance of the mechanistic modelling approach, particularly when coupled with analysis of management practices, in order to forecast agricultural outputs. However, in contrast to these mechanistic growth and development models, other approaches are typically used to account for the interaction between plants, atmosphere and soil. These so-called dynamic models calculate the uptake of atmospheric CO_2 by plants and the release of CO_2 by plants and soil in a physically consistent way that respects the conservation of energy and momentum. In the literature one can find descriptions of established dynamic vegetation models for use at scales from global to local. Examples are (LPJ), developed by Prentice et al. (1992) and modified by Bondeau et al. (2007), ORCHIDEE, developed by Krinner et al. (2005), and BIOME3, developed by Haxeltine and Prentice (1996). These models are driven by meteorological input data and parameterized for all land cover/landuse classes, such as forest, grassland, shrubland and agricultural areas. The spatial resolution for most dynamic models ranges from a few degrees (global usage, e.g. Bondeau et al. (2007) and Haxeltine and Prentice (1996)), to kilometres (regional usage, e.g. Wißkirchen (2005)). Their main outputs are Gross Primary Productivity (GPP), Net Primary Productivity (NPP), Net Ecosystem Exchange (NEE), Total Ecosystem Respiration (TER), and evapotranspiration. Plant development using plant-specific allocation rules is modelled mainly for global climate change analysis or historic plant development. Yield information of agricultural crops is not an output of these dynamic vegetation models.

We here discuss the Biosphere Energy Transfer Hydrology (BETHY/DLR) model, operated by the German Remote Sensing Data Center (DFD).

BETHY/DLR is based on the formulation of Knorr (1997) and modified by Wißkirchen (2005); a description of the model can also be found in Knorr and Heimann (2001). Besides meteorological input data, BETHY/DLR also

requires land cover/land use maps and Leaf Area Index (LAI) time series as input. These observational data are obtained from satellite images. Thus the LAI time series of a pixel (typically $1\,km^2$) represents the mean phenology of the vegetation of that pixel. It is assumed that management practices as well as plant development are reflected and observed by the LAI time series.

The primary objective of this study is to investigate a new approach to the validation of modelled NPP from BETHY/DLR, at $1\,km^2$ spatial resolution, using statistical yield data for major crops. The crops used in validation are the major crops at level 3 of the "Nomenclature des Unités Territoriales Statistiques" (NUTS), a system of hierarchically organised territorial units intended for statistical purposes. For this validation approach, plant-specific yield data and modelled NPP are both downscaled to NPP per NUTS unit, providing a common basis for comparison. The presented validation results were cross-checked with the results of a validated EPIC run for a selected area (Marchfeld) in Austria Schmid et al. (2004). Germany and Austria were selected as test areas due to the availability of detailed statistical data for validation and availability of the EPIC results. Computing time and hard disk storage issues restricted our modelling to the years 2000 and 2001.

2.3   Model and input data

Model

BETHY/DLR models photosynthesis using the combined approach of Farquhar et al. (1980) and Collatz et al. (1992), which parameterizes the enzyme kinetics of photosynthesis at the leaf level. In this context, so called C3 and C4 plants are distinguished because significant differences exist between the carbon fixation strategies of the two classes of plants (C3 and C4). In particular, C4 plants (including corn and sugar cane) can fix more atmospheric carbon dioxide at high temperatures than C3 plants (such as wheat and barley). In either case, in the next step the rate of photosynthesis is extrapolated from leaf to canopy level, taking into account the construction of the canopy as well as interactions between soil, atmosphere and vegetation. Radiation absorption in the canopy is approximated using the two flux scheme of Sellers (1985) with
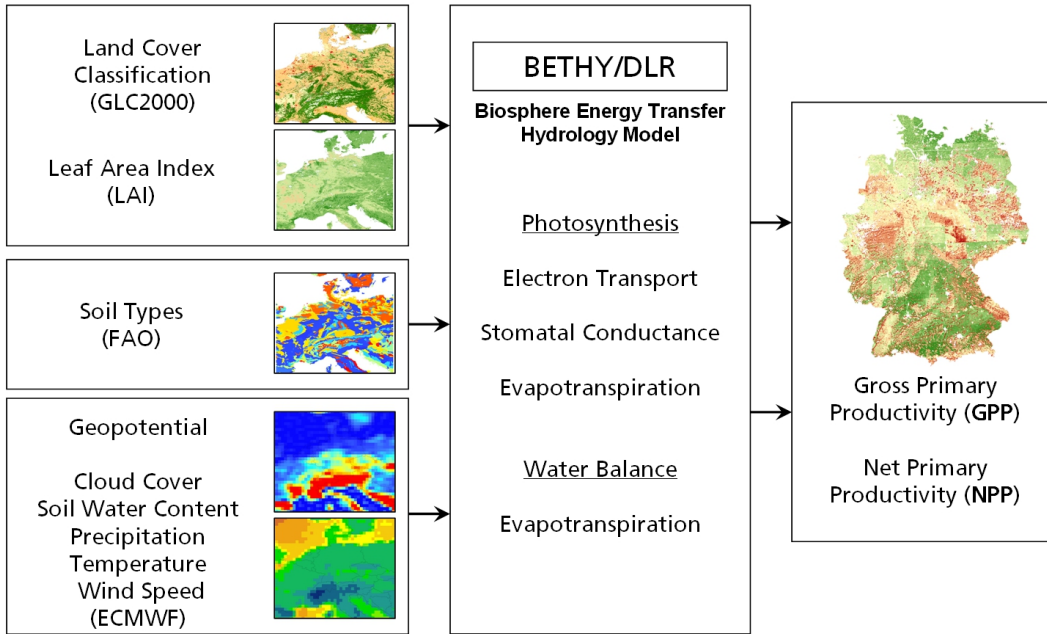
Figure 2.1: *Model schematic for BETHY/DLR, left: input data, middle: internal model processes, right: output data.*

three canopy layers.

Stomatal conductance, evapotranspiration and soil water balance are also included for calculating NPP on an annual basis. The water supply available to plants is considered by calculating the demand for evapotranspiration using the approach of Monteith (1965) against the criteria of Federer (1979), which assumes that evapotranspiration cannot be greater than the possible soil water supply to the roots. Water deficit (or water stress) is thus considered to occur at a soil water content at or below the permanent wilting point (PWP).

Autotrophic respiration is modelled in BETHY/DLR as the sum of the maintenance and growth respiration. Maintenance respiration is mainly determined by the plant-specific dark respiration, while growth respiration is assumed to be proportional to the difference between GPP and maintenance respiration.

The output of BETHY/DLR is a time series, in daily steps, of NPP at the spatial resolution and projection of the land cover classification ($1\,\mathrm{km}^2$, latitude e longitude projection with WGS84 (World Geodetic System 1984) datum). A schematic overview of the currently used input data and the internal model processes is presented in Fig. 2.1.

Input data - meteorology

BETHY/DLR uses remote sensing data and meteorological data to model the photosynthesis of plants, depending on weather and phenological conditions. The meteorological data (see Table 2.1) are derived from the operational processing chain of the ECMWF with temporal resolution up to four times daily and a spatial resolution of $0.25\,°\times 0.25\,°$. The meteorological data used are model analysis of the temperature (at $2\,m$ height), wind speed (at $10\,m$ height), soil water content (in the four uppermost soil layers), and cloud cover. Daily precipitation values are also derived from the ECMWF re-analysis project (ERA-Interim). From this dataset, the daily mean, minimum and maximum temperatures are calculated, as well as the daily mean cloud cover in three strata (high, medium and low) and the water vapour pressure. Daily temperatures are scaled by the difference of ECMWF reference height and global ETOP05 5-minute gridded elevation data and the temperature gradient of the U.S. Standard Atmosphere, which is $-0.65\,K/100\,m$:

$$T' = T_{ECMWF} \times 0.0065\frac{K}{m} \times (h_{ECMWF} - h_{ETOP}) \tag{2.1}$$

T_{ECMWF} represents the reference temperature given by the ECMWF, h_{ECMWF} for the ECMWF reference height (geopotential and h_{ETOP} for the height given by ETOP05 (which has a spatial resolution of about $9\,km^2$).
The daily average photosynthetic active radiation (PAR) is a function of global irradiation, calculated following Burridge and Gadd (1974) from the geographical coordinates, the day and year, and the atmospheric transmission, which depends on the degree of cloudiness. The daily average degree of cloudiness is calculated as a weighted sum of the cloud strata. The advantage of this approach, in contrast to the direct use of ECMWF radiation data, is the use of analysis data of cloud coverage which leads to more exact results than the direct use of radiation forecast data, as shown by Wißkirchen (2005). The global irradiation is calculated for each location for each 1-h time step. The volumetric soil water content was needed to calculate the soil water budged of the model.

Table 2.1: *Summary of meteorological input data (including short names and code numbers) derived from ECMWF.*

Parameter	Short name	Code number
Volumetric soil water layer 1	SWVL1/(SWL1)	039
Volumetric soil water layer 2	SWVL2/(SWL2)	040
Volumetric soil water layer 3	SWVL3/(SWL3)	041
Volumetric soil water layer 4	SWVL4/(SWL4)	042
Geopotential	Z	129
Large scale precipitation	LSP	142
Convective precipitation	CP	143
10 m U-velocity	10U	165
10 m V-velocity	10V	166
2 m temperature	2T	167
Low cloud cover	LCC	186
Medium cloud cover	MCC	187
High cloud cover	HCC	188

Input data - remote sensing

In addition to the meteorological data, the BETHY/DLR model is driven by two satellite remote sensing datasets, time series of the LAI, and detailed and homogeneous land cover / land use information. Phenology of the vegetation is indicated by the LAI time series, which is based on CYCLOPES 10-day compositae datasets downloaded from the POSTEL (Pole d'Observation des Surfaces continentales par Teledetection) databank (www.postel.mediasfrance.org). For each pixel, analysis of the LAI time series is conducted to fill data gaps and eliminate outliers, using harmonic analysis (HA). HA decomposes a time series into a linear combination of suitable trigonometric functions, i.e. sine and cosine oscillations of particular periodicities. The HA technique corresponds to an approximate deconvolution of the power spectrum by iteratively finding and subtracting the highest peak of the time series power spectrum. This method was adapted for the correction of LAI time series data.
CYCLOPES provides land cover and land use information in their GLC2000

dataset. For the derivation of the GLC2000 land cover classes the "Land Cover Classification System (LCCS)" of the FAO was used Bartholome and Belward (2005), DiGregorio and Jansen (2001). In the GLC2000 dataset a classification with 22 different land cover classes is available representative for the year 2000. The global LAI and GLC2000 data are available in tiles of $10°$ by $10°$ as maps in rectangular projection annotated with latitude, longitude, and WGS84 date, with complete coverage of the study area (Germany and Austria). The CYCLOPES dataset was chosen because it is thought to be the most accurate dataset for agricultural areas Garrigues et al. (2008).

In order to use the GLC2000 land use/land cover classification for NPP modelling with BETHY/DLR, the GLC2000 vegetation classes were translated to one of the 33 inherent BETHY/DLR vegetation classes (Table 2.2) which can be regarded as vegetation types. The translation will be discussed in the following section.

Table 2.2: *Vegetation types of the BETHY/DLR model with type number and vegetation parameters after Knorr (1997). V_M: maximum carboxylation rate at 25°C; J_M: maximum electron transport rate at 25°C; height; rooting depth.*

No.	Vegetation types of BETHY/DLR	$V_M[\mu mol\, m^{-2}\, s^{-1}]CO_2$	$J_M[\mu mol\, m^{-2}\, s^{-1}]CO_2$	Height [m]	Rooting depth [m]
1	Trop. BL evergreen trees	62	118	30.0	6.9
2	Trop. BL deciduous trees	90	179	15.0	3.7
3	Temp. BL evergreen trees	41	82	15.0	3.0
4	Temp. BL deciduous trees	35	70	15.0	3.0
5	Evergreen coniferous trees	29	52	15.0	3.9
6	Deciduous coniferous trees	53	95	15.0	1.5
7	Evergreen shrubs	52	102	1.0	3.5
8	Deciduous shrubs	160	266	1.0	3.5
9	C3 short grasses	42	80	0.3	1.8
10	C3 long grasses	42	80	2.0	1.8
11	C4 short grasses	8	140	0.3	1.8
12	C4 long grasses	8	140	2.0	1.8
13	Tundra vegetation	20	37	0.3	0.5
14	Swamp vegetation	20	37	0.6	0.5
15	Arable crops	117	220	0.6	1.8
16	Irrigated crops	123	227	2.0	1.8
17	Trop. tree crops	60	106	2.0	6.9

No.	Vegetation types of BETHY/DLR	$V_M[\mu mol\,m^{-2}\,s^{-1}]CO_2$	$J_M[\mu mol\,m^{-2}\,s^{-1}]CO_2$	Height [m]	Rooting depth [m]
18	Citrus crops	60	106	2.0	3.7
19	Temp. deciduous tree crops	123	227	2.0	3.0
20	Sugar cane	39	700	2.0	1.8
21	Corn	39	700	2.0	1.8
22	Rice	98	190	0.3	0.3
23	Cotton	123	227	2.0	2.1
24	Sugar beet	129	226	0.5	1.8
25	Soy	94	168	0.8	1.8
26	Sunflower	80	213	2.0	2.7
27	Barley	68	169	1.2	1.8
28	Wheat	83	193	1.5	1.8
29	Rapeseed	61	187	1.0	1.8
30	Beech	46	109	15.0	4.0
31	Oak	40	72	15.0	4.0
32	Spruce/Fir	10	24	15.0	2.8
33	Pine	17	30	15.0	4.0

Table 2.3: *Translation of GLC2000 vegetation classes to BETHY/DLR vegetation types with weighting factors.*

GLC2000 class	BETHY/DLR vegetation type	Weighting factor
Cultivated and managed areas (GLC-16)	Arable crops (Type 15)	1.0
Mosaic: cropland/shrub or grass cover (GLC-18)	Arable crops (Type 15)	0.5
	C3 short grasses (Type 9)	0.5

Plant parameters

In BETHY/DLR each vegetation type is linked with biochemical parameters such as the maximum carboxylation rate, the maximum electron transport rate, and other plant-specific photosynthesis related parameters. As can be seen in Table 2.3, it is possible to describe one GLC2000 class using a weighted average of two BETHY/DLR vegetation types. For this study only the two GLC2000 classes which are directly linked with crops are used. The weighting factors are set to 1.0 for the GLC2000 class "Cultivated and managed areas" (GLC-16). This is done under the assumption that this class describes a homogenously crop-covered area. The class "Mosaic: cropland/shrub cover or grass cover" (GLC-18) of GLC2000 was split between arable crops and grass cover using a weighting factor of 0.5 for each. This is done under the assumption that the area is completely vegetated, but only half with crops.

Validation data

To validate the modelled NPP of agricultural crops, empirical estimates of corn yields from the Federal Statistical Office of Germany and from Statistics Austria were used. In both countries farm structure surveys are conducted yearly. The agricultural surveys contain information about arable land, vineyards, horticultural farms, field vegetable farms and commercial fruit plantations, with associated yields. The "NUTS" hierarchical spatial classification starts with the member states of the European Community (EU) (NUTS-0), followed by regions of the EU (NUTS-1), separated to basic administrative

units (NUTS-2) and ends with subdivisions of those basic administrative units
(NUTS-3). As an example, Austria has been divided into three units, East-
ern, Southern and Western Austria (NUTS-1). Each NUTS-1 level comprises
the federal provinces (NUTS-2) such as the "Burgenland". The NUTS-2 level
is split into several NUTS-3 levels; in this case, the "Nordburgenland", "Mit-
telburgenland" and "Südburgenland". Besides these NUTS levels, a further
subdivision is established in the empirical data indicating towns with charters,
political districts and judicial districts.

For Germany and Austria the empirical data are given in NUTS-3 resolution.
For Austria yields for summer rapeseed and grain maize are included, which
are not present in the Germany statistics; otherwise the datasets provide the
same information. As the German dataset contains gaps, necessitating a cri-
terion to fill such gaps; we assumed that gaps for a given crop may be filled
using the mean yield of the given crop from the German NUTS-3 units.

Before validating the modelled results, the $1\,\mathrm{km}^2$ resolution NPP model out-
put must be transferred to a Geographical Information System, taking into ac-
count the equi-rectangular map projection (latitude-longitude projection with
WGS84 datum). Finally, the model results are aggregated to higher NUTS
levels for comparison with the statistical data.

Validation strategy

The yield given by the empirical data does not represent the available biomass
or the biomass potential and thus cannot be compared directly with the mod-
elled yearly NPP sum, which represents the accumulated carbon of pixel over
one year. In order to make a comparison possible, the yield data were used to
estimate the above- and below-ground biomass, using simple growth allocation
schemes. As a first step, it is necessary to calculate the total above-ground
biomass, its dry matter and its carbon content. The literature gives a wide
selection of so-called conversion factors, which give estimated corn-to-straw
or leaf-to-beet ratios (Table 2.4) Köhler and Kolbe (2007), Kaltschmitt and
Hartmann (2001), Jackson et al. (1996), KTBL (2005).

For this study we used values from Köhler and Kolbe (2007), since they rep-
resent the latest available values, and also describe the greatest diversity of
plant species. These relations suggest that, for example, a grain yield of $10\,\mathrm{t}$

of winter wheat will correspond to 11 t of straw. To calculate the dry matter for both the straw and yield fractions, standard estimates of water and carbon content were used, also from Köhler and Kolbe (2007). In this way, the carbon content of the dry matter of straw and yield can be estimated using formulae 2 and 3:

$$NPP_{yi} = yi \times (1 - H_2O_{yi}) \times C_{yi} \qquad (2.2)$$

$$NPP_{stw} = (yi \times \sigma_{stw}) \times (1 - H_2O_{stw}) \times C_{stw} \qquad (2.3)$$

where yi represents the yield of a particular plant species. H_2O_{yi} and H_2O_{stw} represents the water content, and C_{yi} and C_{stw} the carbon content, of the yield and straw fractions, respectively, for that plant species. σ_{stw} represents the plant specific conversion factor of yield-to-straw or leaf-to-beet (Table 2.4). To calculate the total amount of above-ground NPP, NPP_{agb}, one has to aggregate the NPP for the straw and yield fractions for each crop according to formula 4:

$$NPP_{agb} = NPP_{yi} + NPP_{stw} \qquad (2.4)$$

This above-ground NPP is still not comparable with the modelled NPP, however. Because BETHY/DLR gives no information about where the accumulated carbon is stored, it is also necessary to calculate the below-ground NPP, NPP_{bgb}, from the empirical yield data. Simple so-called "shoot-to-root" ratios can be found in, for example, Bolinder et al. (1997) or Jackson et al. (1996). We used those of Bolinder et al. (1997) (Table 2.4), assuming that these estimates, found for crops in Canada, are also representative of Germany and Austria. With these ratios one can estimate NPP_{bgb}:

$$NPP_{bgb} = NPP_{agb} \times \sigma_{agb} \qquad (2.5)$$

σ_{agb} represents the shoot-to-root conversion factor for a specific crop (Table 2.4). The total NPP can now be expressed as:

$$NPP = NPP_{agb} + NPP_{bgb} \qquad (2.6)$$

To calculate the total carbon stored in a NUTS area, this total NPP is integrated over the total cultivation area for each NUTS area and summed over all crops:

$$NPP_{NUTS} = \sum_i NPP_i \times area_i \qquad (2.7)$$

The NPP_{NUTS} per administrative district can now be directly compared with the modelled NPP, also aggregated per NUTS area as previously described.

A comparison with data from eddy covariance towers could not be performed because our area of investigation, Germany and Austria, contains only three agricultural FLUXNET towers. Only one gathered data in 2001, but it had too small a footprint ($<300\,\mathrm{m}$ for wind speeds up to $4\,\mathrm{m\,s^{-1}}$) to quantify CO_2 fluxes on a km^2 scale.

Table 2.4: *Corn-to-straw (leaf-to-beet) ratios and above- to below-ground ratios for selected crops. Carbon and water content for yield and straw fraction.*

Field fruit	Yield/straw (*)	Root/shoot (+)	C_{yi}	C_{stw}	H_2O_{yi} (*)	H_2O_{stw} (*)
Grain	1.13	0.14	0.45 ($)	0.45 ($)	0.14	0.14
Wheat	1.1	0.21	0.455 (#)	0.446 (#)	0.14	0.14
Maize	0.8	0.18	0.456 (#)	0.5 (#)	0.14	0.14
Barley	1.05	0.32	0.45 ($)	0.459 (#)	0.14	0.14
Rye	1.3	0.19	0.475 (#)	0.457 (#)	0.14	0.14
Oat	1.1	0.4	0.45 ($)	0.45 ($)	0.14	0.14
Triticale	1.2	0.19	0.436 (#)	0.436 (#)	0.14	0.14
Beet	0.33	0	0.45 ($)	0.45 ($)	0.88	0.88
Potato	0.2	0	0.45 ($)	0.45 ($)	0.78	0.75
Sugar beet	0.7	0	0.45 ($)	0.45 ($)	0.77	0.84
Oil fruits	1.75	0.14	0.45 ($)	0.605 (#)	0.09	0.14
Rape	2	0.14	0.45 ($)	0.605 (#)	0.09	0.14

+: after Bolinder et al. (1997), *: Köhler and Kolbe (2007), #: Kaltschmitt and Hartmann (2001), $: own estimations.

2.4   Results

Before validating the annual sum of accumulated carbon as modelled with BETHY/DLR for large regions (Germany and Austria), we compared the BETHY/DLR NPP results with the output of the EPIC model over a smaller region.

BETHY/DLR comparison with EPIC

The study site for this comparison was the Marchfeld region of Lower Austria, which is part of the Vienna Basin. With an area of about 100,000 ha, it is one of the largest plains in Austria, and about 75 % of its area is used for agricultural production. The Marchfeld's natural boundaries are the river March to the east (the Austrian border to Slovakia), the hills of Weinviertel to the north, the Bisamberg mountains and the city of Vienna to the west, and the river Danube to the south. The EPIC model has been validated for the Marchfeld Schmid et al. (2004), making it worthwhile to compare BETHY/DLR to EPIC for this region. Since land use practices are not homogenously distributed in this area, five sectors were identified using cluster analysis methods Hofreither et al. (2000). Each sector has an area between $85\,km^2$ and $250\,km^2$. For our analysis one sector was not used, since its land cover is predominantly designated as urban. The NPP of both models for 2000 and 2001 across the four sectors is presented in Fig. 2.2.

Fig. 2.2 shows that BETHY/DLR estimates slightly more NPP (about 15 % higher) than the calibrated EPIC model. Indeed, only in one case, sector 2 in the year 2000, was BETHY/DLR's estimate lower than EPIC's. According to Knorr and Heimann (2001), the calibration of the EPIC model for sugar beets in the Marchfeld has a standard deviation of less than 10 %. Similarly, BETHY/DLR's NPP estimate for 2001 for the NUTS-3 Gänserndorf region (dominated by the Marchfeld) was about 10 % higher than the estimated true NPP (not shown). Looking at all results for the Marchfeld region, it can be concluded that the modelled NPP from BETHY/DLR is in good agreement with both empirical data and the calibrated EPIC results, although with a tendency for minor overestimation.
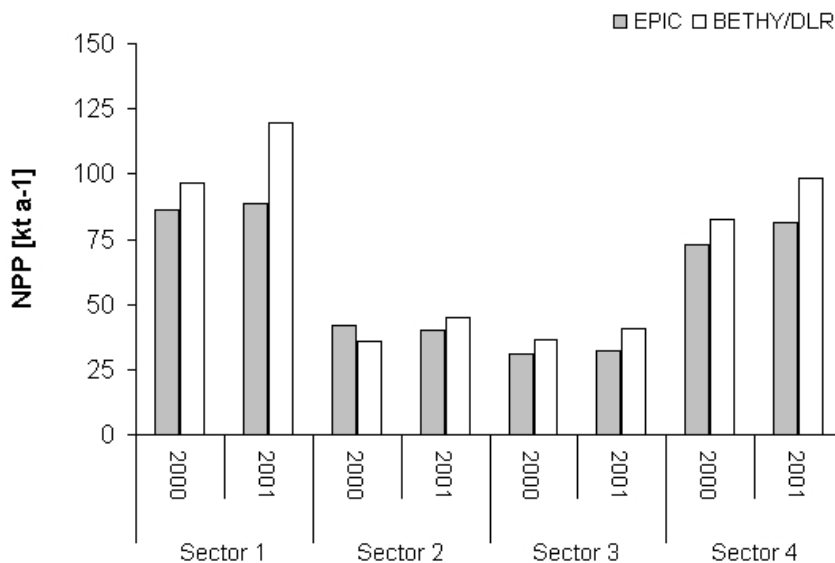
Figure 2.2: *Comparison of modelled NPP derived from BETHY/DLR and EPIC for all four sub-regions of the Marchfeld.*

Validation of BETHY/DLR with statistical data

Modelled NPP, at $1\,\mathrm{km}^2$ resolution, for Austria and Germany was calculated as the annual sum of accumulated carbon for 2000 and 2001 (Fig. 2.3). Yearly NPP is clearly higher in the southern states of Germany than in Germany's northern and eastern regions, in both years. Statistical analysis revealed that the mean annual NPP in carbon (over the whole area of investigation) is 253 $[\mathrm{t\,km^{-2}\,y^{-1}}]$ with a maximum of 662 $[\mathrm{t\,km^{-2}\,y^{-1}}]$ for 2000, and 239 $[\mathrm{t\,km^{-2}\,y^{-1}}]$ with a maximum of 577 $[\mathrm{t\,km^{-2}\,y^{-1}}]$ for 2001. The annual NPP in carbon for Germany is 76.4 Mt for the year 2000 and 73.3 Mt for 2001; for Austria, annual NPP is 7.9 Mt for 2000 and 6.2 Mt for 2001. The conversion of statistical yield data to NPP, as described above, delivers annual sums for Germany of about 67.2 Mt for 2000 and 71.5 Mt for 2001. For Austria these values are about 6.9 Mt for 2000 and 6.3 Mt for 2001. From this it may be seen that the modelled NPP for Germany for both years is overestimated ($\approx$13 % for 2000 and $\approx$2 % for 2001). For Austria the modelled NPP for 2000 is overestimated ($\approx$13 %) for 2000, but underestimated ($\approx$1 %) for 2001. Also notable is the very low annual NPP predicted for parts of eastern Germany (red pixels), particularly for 2001, but also, less strongly, for 2000. On the other hand, in 2001
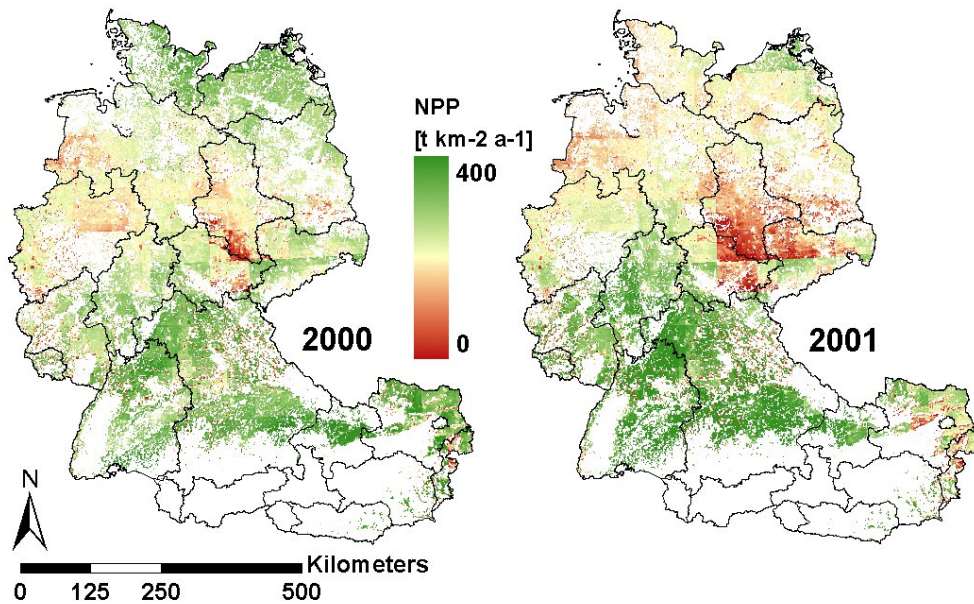
Figure 2.3: *Yearly NPP from BETHY/DLR for agricultural areas in Germany and Austria for 2000 (left) and 2001 (right). High NPP values are green, medium values are beige, and low values are red. White represents areas that do not belong to the GLC2000 classes GLC-16 or GLC-18 (Table 2.3). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)*

modelled NPP in the southern parts of the study area are markedly higher than in 2000.

The sharp boundaries in the NPP maps (clearly visible for the year 2001 in the Saxo-Thuringia region) reflect the coarse pixel size of the ECMWF meteorological input data. This indicates that meteorology has a strong influence on the simulation. Examination of the meteorological input data for both years shows that large differences and leaps are not visible for most parameters; annual precipitation for the Thuringia region, however, is $480 \, \mathrm{mm \, y^{-1}}$ for 2000, and about $760 \, \mathrm{mm \, y^{-1}}$ for 2001.

Fig. 2.3 shows that the alpine regions of Germany and Austria have almost no modelled NPP. In contrast, the statistics of Germany and Austria report yield data for those areas. This is a consequence of the land cover data's spatial

resolution of about $1\,km^2$, which is insufficient to describe the heterogeneous, small-scale structure of mid-European land use practices. Land cover classifications with higher-resolution exist for Europe, such as the CORINE land cover map. But when using high-resolution land cover maps, LAI time series data of the same spatial resolution are mandatory as input for BETHY/DLR. Since no high-resolution LAI time series are available for Austria and Germany, we selected the CYCLOPES LAI product together with the GLC2000, both available at $1\,km^2$, as best practice. Furthermore, the GLC2000 was derived with the same satellite sensor (VEGETATION on SPOT 4) as the LAI time series, providing data homogeneity.

To correlate empirical yields with BETHY/DLR's modelled data, the estimated biomass per pixel was aggregated to biomass per administrative district (NUTS-3 level) as previously described. Linear regression was used to assess the correlation between modelled and empirical yield, separately for 2000 and 2001, and separately for Germany and Austria (Fig. 2.4).
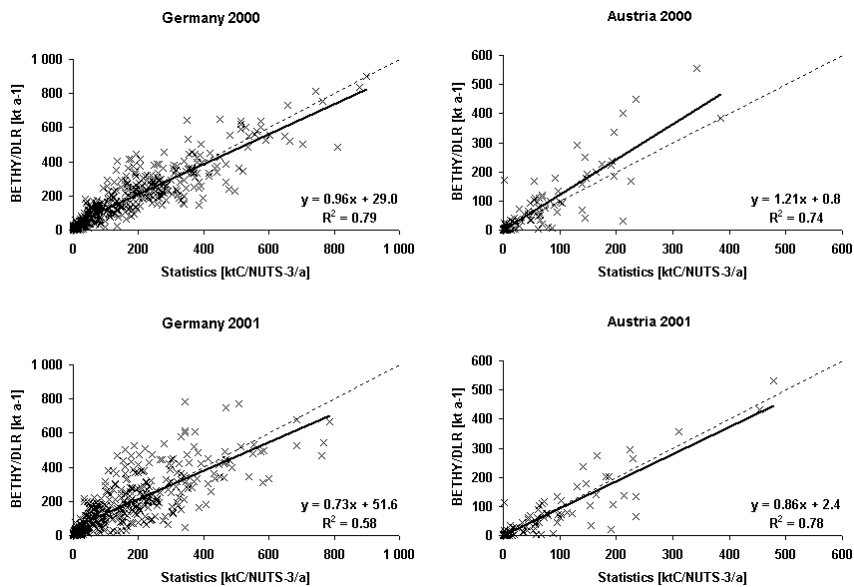

Figure 2.4: *Correlation of modelled NPP with empirical NPP data for Germany (left) and Austria (right) for the years 2000 (top) and 2001 (bottom). Crosses indicate individual NUTS-3 administrative districts. Dotted lines indicate perfect correlation; solid lines indicate the correlation found by linear regression.*

As shown in Fig. 2.4 BETHY/DLR underestimates the NPP for Germany in both years and for Austria in 2001. With a coefficient of determination of about 0.74 for 2000 and 0.78 for 2001, each bound with a slope of 1.21 and 0.86, respectively and an offset of 0.8 and 2.4, respectively, one can speak of a high degree of correlation. For Germany, the coefficient of determination for 2000 is 0.79 and for 2001 0.58. The slopes and offsets are 0.96 and 28.97 for 2000 and 0.73 and 51.56 for 2001.

The different validation results for Germany and Austria might be explained by differences in the distribution of cultivated plants between Germany and Austria, and by the method used to convert yield to NPP. The residual of miscellaneous crops for both statistics are 9 % (Austria) and 13 % (Germany). This is because additional crops (grain maize and summer rapeseed) are reported for Austria. Since the difference in the residuals can be regarded as low, the validation result for 2000 must be argued in a different way. For 2000, NPP in Austria is overestimated by BETHY/DLR for large NUTS-3 units as Hollabrunn, Horn, Mistelbach, and Neusiedel am See by a factor of about 2. A closer look at the empirical data reveals that up to 30 % lower yields are reported for 2000 (in comparison to 2001) for the main crops of these NUTS-3 units. This yield reduction might be explained by a drought starting in April and ending in mid-May when the transition to the reproductive stages begins. Since the input data for BETHY/DLR (meteorological data and LAI time series) do not show large differences between the two years in this region, it is obvious that BETHY/DLR will estimate the NPP for these regions within the same order of magnitude. The ECMWF meteorological data show precipitation of 19.3 mm from April 3rd to May 18th, while the weather station Laa Thaya (Mistelbach, Austria) only reported 4.6 mm precipitation from that period. This water deficit resulted in a reduction in yield which could not be modelled by BETHY/DLR due to the unrealistic precipitation data from ECMWF.

It can be seen in Fig. 2.4 that the scatter for the German data is markedly greater than for the Austrian data, due to the difference in the number of available validation data points (Germany: 412, Austria: 99).

A closer look at the validation results for Germany in 2001 shows two distinct clouds within the scatter plot (Fig. 2.4). Detailed investigation reveals that

most differences between the modelled and empirical biomass data in 2001 occur in regions which experienced low annual precipitation in 2000, especially in the Saxo-Thuringia region. We therefore hypothesize that the parameterization of the water cycle in BETHY/DLR might account for the underestimation of annual NPP.
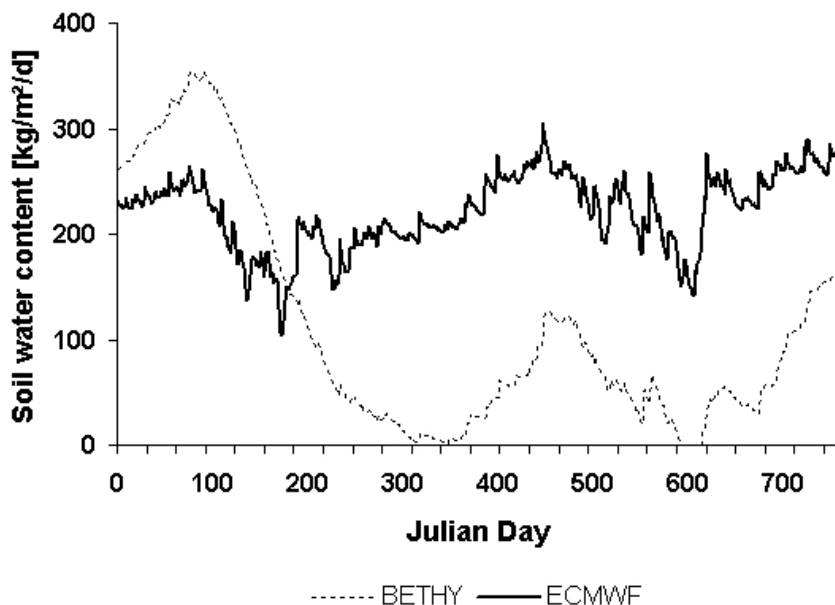

Figure 2.5: *Comparison between the empirical mean plantavailable soil water content derived from ECMWF data (solid line) and computed mean available soil water content of BETHY/DLR (dotted line) aggregated across the NUTS-3 regions Burgenland and Merseburg-Querfurt (Germany, Saxo-Thurinigia region) for the years 2000, 2001 and January 2002. January 1st, 2000, is represented by Julian day 0.*

In BETHY/DLR the soil water budget is tracked using a simple "bucket model" that represents the soil as a single layer. Modelling outputs show that the modelled soil water content diminished to nearly dry conditions (below the PWP) in 2000 in areas with relatively moderate annual precipitation. In contrast, the empirical soil water content data of ECMWF is available as a four-layered unequally spaced soil core. Fig. 2.5 shows a comparison of measured (ECMWF) versus modelled (BETHY/DLR) soil water content for 2000, 2001, and January of 2002. This comparison is aggregated across the NUTS-3 regions "Burgenlandkreis" and "Merseburg-Querfurt", because those

regions are situated in one ECMWF tile. The four-layered ECMWF data was combined and adapted to a single layer with a soil water content having the same soil core depth as used for the BETHY/DLR simulation, to make comparison possible.

The ECMWF soil water content data are clearly different from the BETHY/ DLR results. Differences of up to 100% can be found. The mean difference over the two years is about 51%. With the beginning of the vegetation cycle in 2000 (day 95), BETHY/DLR's soil water content begins a continuous decrease that lasts to the end of the year. As a consequence, the soil water content remains unrealistically low in 2001, and is well below the PWP (zero) at the end of the vegetation cycle (day 600). At the end of 2001 the soil water content starts to recover, due to precipitation and the missing demand from vegetation. This trend is continued in 2002 (not shown in Fig. 2.5). In contrast to this pattern in the modelled soil water content, the ECMWF soil water content shows only a small increase at the beginning of 2000 and a small decrease from Julian Day 95 to Julian Day 180. From Julian Day 400 onward, both datasets show similar patterns.

This indicates that in principle the soil water model of BETHY/DLR exhibits the same patterns in soil water content as the ECMWF, but offset. This offset in soil water content affects the modelled NPP only when the modelled soil water content falls below the PWP, which reduces photosynthesis due to water deficiency. This offset might have been triggered by low annual precipitation rates for the region. As mentioned previously, ECMWF annual precipitation in this region was only $480\,\text{mm}\,\text{m}^2$ in 2000, but jumped to $760\,\text{mm}\,\text{m}^2$ in 2001. We hypothesize that the parameterization of the soil water budget in BETHY/DLR overestimates plant water use or evaporation from the soil, leading to the unrealistically low soil water content at the beginning of 2001 (DOY = 365). Annual precipitation in 2001 in Burgenland and Merseburg-Querfurt was insufficient to support soil water conditions adequate for plant growth. We conclude that the red marked regions in Fig. 2.3 for 2001 may be explained by an excessive loss of modelled soil water content during 2000, as a consequence of low precipitation in 2000. This underestimation of soil water content leads, as described above, to an underestimation of NPP in 2001. The lower correlation cloud found in Fig. 2.4 for Germany in 2001 is strongly linked

with the overestimated decrease of soil water in these regions.

2.5   Conclusions

The Net Primary Productivity for the territories of Germany and Austria for 2000 and 2001 were modelled using the dynamic vegetation model BETHY/ DLR. Inputs for the model were meteorological data from ECMWF, LAI time series from vegetation, and land cover / land use data from GLC2000. We here presented a new approach to validate modelled NPP using empirical data on acreage and averaged grain yields of main crops at the NUTS-3 level. Using conversion factors (corn-tostraw and shoot-to-root ratios), the statistical data were converted to NPP per NUTS unit for comparison. This method yielded high coefficients of determination (R^2 up to 0.74), allowing strong conclusions to be drawn about model validity. For German districts, BETHY/DLR substantially underestimated the NPP (17 %), whereas for Austrian districts a slight overestimation (8 %) was observed.

In areas where the land cover classification (GLC2000) provided insufficient information (particularly in the Alps), modelled NPP was significantly underestimated (even to zero), producing high discrepancies between modelled NPP and empirical data in those regions. This indicates that a spatial resolution of $1\,\mathrm{km}^2$ is insufficient to describe the heterogeneous small-scale structure of mid-European land use practices. To improve modelling results, we recommend the use of a higher-resolution land cover product such as the MERIS GlobCover, with a resolution of $300\,\mathrm{m} \times 300\,\mathrm{m}$ together with LAI time series also derived from MERIS.

In the Saxo-Thuringian Basin very low NPP values were modelled for 2001. Most of the large differences between modelled and observed NPP data for 2001 were found in regions with low annual precipitation in 2000. We demonstrated that the parameterization of the water cycle in BETHY/DLR (in particular, the use of a bucket model for estimating soil water content) was the underlying reason driving low NPP estimates in these regions that had experienced low precipitation the previous year. Finally, we showed that natural drought were not reflected in the model results due to unrealistically precipitation rates reported in the input data.

This study illustrates a novel method of model validation that we believe will be useful in estimating biomass potentials from modelled NPP products on a medium resolution. This method could also be used as a downscaling approach for empirically derived NUTS-level data, since the model results could help to spatially represent the NUTS information.

2.6 Acknowledgements

Chapter 3

Validation of modelled forest biomass in Germany using BETHY / DLR

Markus Tum, Marcel Buchhorn, Kurt P. Günther, Bejamin C. Haller,
Geoscientific Model Development, (2011), 4, 1019-1034

3.1 Abstract

We present a new approach to the validation of modelled forest Net Primary Productivity (NPP), using empirical data on the mean annual increment, or MAI, in above-ground forest stock. The soil-vegetation-atmosphere-transfer model BETHY/DLR is used, with a particular focus on a detailed parameterization of photosynthesis, to estimate the NPP of forest areas in Germany, driven by remote sensing data from VEGETATION, meteorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF), and additional tree coverage information from the MODIS Vegetation Continuous Field (VCF). The output of BETHY/DLR, Gross Primary Productivity (GPP), is converted to NPP by subtracting the cumulative plant maintenance and growth respiration, and then validated against MAI data that was calculated from German forestry inventories. Validation is conducted for 2000 and 2001 by converting modelled NPP to stem volume at a regional level. Our analysis shows that the presented method fills an important gap in methods

for validating modelled NPP against empirically derived data. In addition, we examine theoretical energy potentials calculated from the modelled and validated NPP, assuming sustainable forest management and using species-specific tree heating values. Such estimated forest biomass energy potentials play an important role in the sustainable energy debate.

3.2   Introduction

Models of carbon uptake by plants play an important role in answering questions concerning the mechanisms driving the carbon cycle and the roles of terrestrial carbon sinks and sources (Cox et al. (1999)). Carbon uptake by plants, measured as Gross Primary Productivity (GPP), can be predicted by simple models that describe the physical, chemical, and plant physiological processes of plant development, as well as the interactions between plants and the atmosphere. Such "deterministic" models (sometimes also called "mechanistic" or "Monteith-type" models) calculate photosynthesis following the methods of Monsi and Saeki (1953) and Monteith (1965).

The idea behind these Monteith-type models is that the carbon uptake of sufficiently watered and fertilized plants is linearly correlated with the energy of the incident photosynthetically active radiation (PAR), or more precisely, the fraction of the PAR that is actually absorbed by the plants (fPAR). Following this approach, it is possible to calculate GPP for each vegetation type from the absorbed solar radiation (fPAR) and the light use efficiency (LUE) of the vegetation type. The LUE can be affected by environmental stress factors, particularly temperature, water limitation, and nitrogen availability. Species-specific fPAR values may be estimated by measurement of dry biomass accumulation, or may be derived from satellite data.

GPP, as estimated by such a model, can be converted to NPP by considering temperature-dependent maintenance respiration. Maintenance respiration can be estimated using allometric functions regarding leaf and root distribution following the approach of Ryan et al. (1995), or using the Leaf Area Index (LAI) of the vegetation following Running et al. (2000). In either case, NPP is defined as the remainder after plant maintenance respiration is subtracted from GPP. In a further step, Net Ecosystem Productivity (NEP) can be cal-

culated by subtracting the heterotrophic respiration in an ecosystem from the ecosystem's NPP.

The Monteith-type model architecture has been used many times. For example, the C-Fix model, a Monteith-type parametric model by Veroustraete et al. (1994), was used by Verstraeten et al. (2006) to estimate net ecosystem fluxes for all of Europe. C-Fix is driven by vegetation type data of the Normalized Differenced Vegetation Index (NDVI) and meteorological data (temperature and daily incoming global radiation) obtained from about 800 weather stations administered by the World Meteorological Organization (WMO). Verstraeten et al. (2006) validated their results with eddy covariance flux tower measurements, obtaining an R^2 of 0.84 for pine forests and 0.59 for mixed deciduous forests. The Carnegie-Ames-Stanford Approach (CASA) model introduced by Potter et al. (1993) and validated by Potter et al. (2001) and Potter et al. (2003) is another example of a Monteith-type model. The CASA model is driven by monthly NDVI data from the FASIR database of the Goddard Space Flight Center, monthly temperature and precipitation data from the International Institute for Applied Systems Analysis (IIASA), and monthly PAR data from the Goddard Institute for Space Studies. Validation of CASA was performed against atmospheric CO_2 concentration data from NOAA and the Geophysical Monitoring from Climate Change Flask Sampling Network, and obtained R^2 values between 0.09 and 0.67.

When the LUE approach is integrated into a coupled soil-plant-atmosphere model, such as the Atmosphere Land Exchange (ALEX) model, daily estimates of evapotranspiration and carbon assimilation fluxes can also be obtained (Anderson et al. (1997). Recently, more sophisticated models have begun to be developed that take this integrative approach even further. In computing the uptake of carbon by plants, these so-called "dynamic" models take into account the complex interactions between plants, soil, and the atmosphere, but also account for the carbon released by both plants and soil in a manner that respects the conservation of energy and momentum. At present, only a few dynamic models have been published. Examples are the Lund-Potsdam-Jena (LPJ) model developed by Prentice et al. (1992) and modified by Bondeau et al. (2007), the Equilibrium Terrestrial Biosphere Model (BIOME3) by Haxeltine and Prentice (1996), and the ORganizing Carbon and Hydrology in

Dynamic EcosystEms (ORCHIDEE) model by Krinner et al. (2005). These global models are driven by meteorological input data, and phenology is calculated internally from those inputs using per-species physiological parameters. The spatial resolution for dynamic models can range from degrees, for global models such as Prentice et al. (1992) and Haxeltine and Prentice (1996), to kilometres, for regional models such as Wißkirchen (2005). Model outputs are typically GPP, NPP and NEP, total ecosystem respiration, and evapotranspiration.

This study used the Biosphere Energy Transfer Hydrology (BETHY/DLR) model, a dynamic model based on the Jena Scheme of Atmosphere Biosphere Coupling in Hamburg (JSBACH) by Knorr (1997), which was designed for global applications (see also Knorr and Heimann (2001)). It was modified by Wißkirchen (2005) for application to regional modelling.

Model validation is often conducted using data from devices called eddy covariance flux towers. The relationship between carbon and energy flux has been studied in international networks such as FLUXNET (Baldocchi et al. (2001)) and AmeriFlux, as well as in projects such as EUROFLUX (Valentini (2003)) and CarboEurope. This research has shown that eddy covariance flux tower measurements can be used to quantify NEP at the spatial scale of the footprint of a tower (Baldocchi (1997)). As mentioned above, NEP may also be calculated by subtracting heterotrophic respiration from NPP. Therefore, robust methods have been developed to estimate heterotrophic respiration in order to convert NEP, as measured by eddy towers, into NPP (or, by considering plant maintenance respiration as well, GPP).

For example, the MODIS GPP product (MOD17, C4.5) was validated with eddy tower CO_2 flux estimates across diverse land cover types and climates (Heinsch et al. (2006)). The main test areas were forest ecosystems in North America, but chaparral, oak savannah, northern grassland and Arctic tundra were also included in the investigation. It was found that MODIS overestimated GPP by about 20% to 30%, but this depended strongly on season and ecosystem type. Comparison of annual MODIS GPP (modelled with global meteorological data from NASA's Global Modeling and Assimilation Office) to tower-based GPP measurements yielded an R^2 of 0.72.

The primary objective of this study is to present a new approach to the valida-

tion of modelled NPP. We compare output from BETHY/DLR, run at $1\,km^2$ spatial resolution, with empirical measurements of the mean annual increment (MAI) in above-ground biomass (including bark) observed in forests in Germany. The MAI data are available at a regional scale called the NUTS-1 level; NUTS is an abbreviation for "Nomenclature des Unités Territoriales Statistiques" and is a system of hierarchically organised territorial units used for statistical purposes. The NUTS-1 MAI data were obtained from the National Forest Inventory (NFI) of Germany.

A secondary objective is to use our modelled and validated NPP to estimate theoretical energy potentials, given sustainable forestry practices, for the area of study. Sustainable energy potentials from forests are a key element in planning a sustainable energy economy for Germany (and, of course, the rest of the world), and so developing methods for estimating, and ultimately forecasting, these potentials is of great importance BMVBS (2010).

3.3 Model description

BETHY/DLR is a soil-vegetation-atmosphere-transfer (SVAT) model. SVAT models track the plant-mediated transformation of atmospheric carbon dioxide into energystoring hydrocarbons such as sugars, a process known as carbon fixation. BETHY/DLR models photosynthesis, and takes into account environmental conditions that affect it. Photosynthesis is parameterized following the combined approach of Farquhar et al. (1980) and Collatz et al. (1992), and treats separately the "light" and "dark" reactions of photosynthesis at the leaf level. A benefit of this design is that the photosynthetic rate can be mechanistically limited either by light availability or by the abundance of the carboxylation enzyme Rubisco, the key player in the Calvin cycle that fixes carbon. In addition, so-called C3 and C4 plants are distinguished in BETHY/DLR because significant differences exist between their carbon fixation physiologies; in particular, C4 plants such as sugar beet and corn can fix more atmospheric CO_2 at higher temperatures than can C3 plants such as barley and wheat.

In a second step, the rate of photosynthesis is extrapolated from leaf to canopy level taking into account both canopy structure and the interactions between soil, atmosphere, and vegetation. For closed canopies (trees, shrubs, and

crops) the photosynthetic rate depends on the Leaf Area Index (LAI), a per-species metric of leaf upper surface area per unit ground area. To model self-shading, photosynthetic rate is reduced exponentially from the canopy top to the soil. In this approach, radiation absorption in the canopy is approximated in BETHY/DLR using the so-called "two-flux scheme" of Sellers (1985) with three canopy layers.

Besides photosynthesis, other energy transfers also need to be tracked. The model's energy balance takes into account heat fluxes between the vegetation and the atmosphere above it, as well as the cooling effect of evapotranspiration from the soil and vegetation. Soil heat flux is also estimated, and the storage of heat in the canopy and in the air layer above the canopy is considered.

The coupling of these processes is of great importance, since temperature-dependent photosynthesis transforms light energy into chemical energy, and finally into carbohydrates, using water and CO_2. Water is available for the plants from the soil, while evapotranspiration by plants and soil determine the water vapour deficit in the atmosphere, which is closely linked to the stomatal behaviour of leaves. Thus when considering the dynamic interaction of, for instance, the soil water balance and photosynthesis, the natural behaviour of vegetation can be reflected; this is the motivating idea of the SVAT approach. The water cycle must also be modelled and included in this interaction scheme. In BETHY/DLR three water reservoirs are considered: soil water, snow, and "skin" or "intercepted" water on leaves and other parts of the vegetation. These reservoirs change in time and space depending on precipitation, temperature and evapotranspiration. A "bucket model" is used for calculating soil water dynamics, using the plant rooting depth as the depth of the soil core. In the bucket model, water availability for plants is governed by the soil water content above the permanent wilting point (at which water is so tightly bound by soil particles that it is unavailable to plants) and below the field capacity (at which the soil is full and further water added by precipitation or snowmelt runs off). The distribution of water within the soil is not modelled in BETHY/DLR, although this would be a worthwhile addition. Water limitation is considered by calculating the demand for evapotranspiration using the approach of Monteith (1965) with the criteria of Federer (1979), which state that evapotranspiration cannot be greater than the limit set by the soil water supply and the water

uptake physiology of a plant's roots.

Autotrophic respiration is modelled in BETHY/DLR as the sum of maintenance respiration and growth respiration. Maintenance respiration is determined by plant-specific dark respiration rates, while growth respiration is proportional to the difference between GPP and maintenance respiration. For estimating NEP, heterotrophic soil microbe respiration is calculated as a function of temperature, scaled by the annual NPP and the effect of soil moisture (neglecting, for modelling purposes, other heterotrophic respiration).

The output of BETHY/DLR is a time series of NPP in daily steps, at the resolution and projection of the land cover classification. Here $1\,\mathrm{km}^2$ resolution is used, in a latitude-longitude projection using the WGS84 (World Geodetic System 1984) datum. An overview of the input data used in this study and the internal model processes acting upon them is presented in Fig. 3.1.
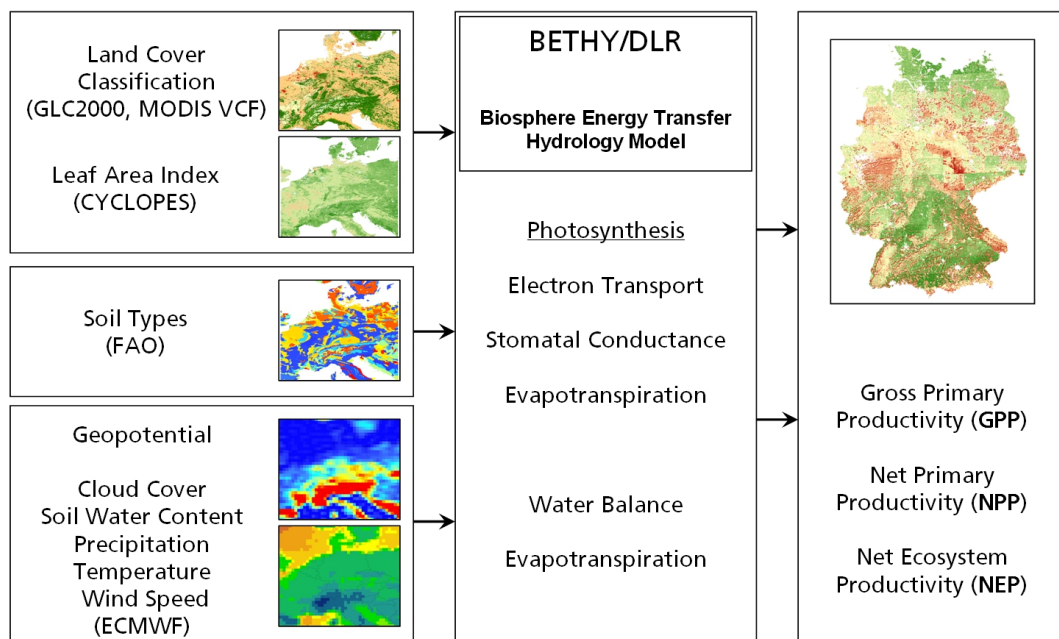

Figure 3.1: *Model overview for BETHY/DLR. Left: input data, middle: internal model processes, right: outputs.*

3.4 Input data

The inputs of the BETHY/DLR model are three remote sensing datasets derived from SPOT-VEGETATION and MODIS, meteorological data provided

by ECMWF, and two further datasets describing soil type and land elevation.

Remote sensing data

Per-pixel Leaf Area Index (LAI) time series were used to determine vegetation phenology. These time series were based on CYCLOPES (Carbon cYcle and Change in Land Observational Products from an Ensemble of Satellites) tenday composite datasets, which can be downloaded from the POSTEL (Pole d'Observation des Surfaces continentales par TELedetection) database. Criteria for the identification of gaps and outliers in the CYCLOPES datasets were defined and crosschecked against a hand-validated dataset. For each identified gap or outlier, harmonic analysis (HA), a type of superpositioning transformation, was applied to estimate a corrected value. Following this procedure, a global meanerror of 9 % was found across a longer time series. Despite these gaps and outliers, the CYCLOPES dataset is the most consistent LAI dataset available (Garrigues et al. (2008)); it was also chosen because it is available in the needed form of ten-day composites.

The CYCLOPES dataset also provides land cover and land use information for the year 2000, available as the product Global Land Cover 2000 (GLC2000). The Land Cover Classification System of the Food and Agriculture Organization of the United Nations (FAO), using 22 different land cover classes, was used to derive GLC2000 (Bartholome et al. (2002); DiGregorio and Jansen (2001)). The CYCLOPES LAI and GLC2000 datasets are both available as $10° \times 10°$ maps in rectangular projection, with latitude and longitude using the WGS84 datum. Complete coverage of the study area (Germany) is available.

In order to make the GLC2000 land use/land cover classification useable for NPP modelling with BETHY/DLR, the GLC2000 vegetation classes were translated to BETHY/DLR's forest-related vegetation types. Each vegetation type in BETHY/DLR is described by biochemical parameters such as the maximum carboxylation rate and the maximum electron transport rate, encapsulating the maximum rates of, respectively, the light and dark reactions of photosynthesis. The internal parameterisation of BETHY/DLR allows a given GLC2000 vegetation class to be represented as a fraction of two BETHY/DLR vegetation types. In the context of this study, only four GLC2000 classes rep-

Table 3.1: *Translation of GLC2000 vegetation classes to BETHY/DLR vegetation types with weighting factors.*

GLC2000 class	BETHY/DLR type	Weighting factor
Tree cover, broadleaved, deciduous closed (glc-2)	Temperate broadleaved deciduous trees	MODIS VCF
Tree cover, broadleaved, deciduous open (glc-3)	Temperate broadleaved deciduous trees	MODIS VCF
Tree cover, needle-leaved, evergreen (glc-4)	Evergreen coniferous trees	MODIS VCF
Tree cover, mixed leaf type (glc-6)	Temperate broadleaved deciduous trees	0.5
	Evergreen coniferous trees	0.5

resented forest types found within Germany; the BETHY/DLR representation of these four classes is shown in Table 3.1.

To obtain information about fractional land cover, an additional dataset was used. The MODIS Vegetation Continuous Fields (VCF) dataset (DeFries et al. (2000), Hansen et al. (2002), Hansen et al. (2003)) contains annual global data on percent tree cover at a spatial resolution of 500m square, and is available for 2000 to 2005. For use in BETHY/DLR, the highresolution VCF data were aggregated to match the spatial resolution of the CYCLOPES data using a Geographic Information System (GIS). For GLC2000 class 6 ("Tree cover, mixed leaf type"), a weighting factor of 0.5 between deciduous and coniferous forest types was used (without scaling of those weights by the VCF fractional land cover). This approximation was assumed to be adequate, since only 15 % of German forest areas are described as mixed forest in the GLC2000, and since VCF contains no further information about the proportion of deciduous to coniferous trees. A similar approach was described by Jung (2008) using the AVHRR Continuous Field of Tree Cover dataset.

Meteorological data

In addition to remote sensing data, BETHY/DLR needs meteorological input data (Table 3.2). The ECMWF (Europeanv Centre for Medium-Range Weather Forecasts) provides this data at a spatial resolution of $0.25\,° \times 0.25\,°$ and a temporal resolution of up to four times per day. The data are produced by ECMWF's climate model analysis of meteorological station measurements of air temperature (at $2\,m$ height), wind speed (at $10\,m$ height), soil water content (in the four uppermost layers), and cloud cover. Daily precipitation values were obtained from the ECMWF INTERIM dataset. From these data we calculated daily mean, minimum and maximum temperatures, daily mean cloud cover at three heights, and relative humidity. Daily temperature values were reconstructed at the $1\,km^2$ resolution of the land cover data, adjusting for the elevation difference between the ECMWF data and the elevation of each modelled pixel using the temperature gradient of the international standard atmosphere (-0.65 K per 100 m).

Table 3.2: *Summary of meteorological input data from ECMWF, including short names and code numbers.*

Parameter	Short name	Code number
Volumetric soil water layer 1	SWVL1/(SWL1)	039
Volumetric soil water layer 2	SWVL2/(SWL2)	040
Volumetric soil water layer 3	SWVL3/(SWL3)	041
Volumetric soil water layer 4	SWVL4/(SWL4)	042
Geopotential	Z	129
Large scale precipitation	LSP	142
Convective precipitation	CP	143
10 m U-velocity	10U	165
10 m V-velocity	10V	166
2 m temperature	2T	167
Low cloud cover	LCC	186
Medium cloud cover	MCC	187
High cloud cover	HCC	188

The daily average photosynthetically active radiation (PAR) was calculated
from global radiation following the method of Burridge and Gadd (1974). This
method calculates PAR from the incident sunlight for the given day and year,
modified by atmospheric transmission, which depends on the degree of cloudi-
ness. Daily average cloud cover was calculated using a weighted sum of each
cloud layer. This approach leads to more exact results than the direct use of
radiation forecast data, and is thus preferable to the direct use of ECMWF
radiation data (Wißkirchen (2005)). Global radiation was calculated for each
location at each one-hour time step.

Time series of the volumetric soil water content were needed to calculate the
soil water budget of the model. Information about the soil type was taken
from the International Soil Reference and Information Centre-World Inventory
of Soil Emission Potentials (ISRIC-WISE) dataset, which has a resolution of 5
$\times$ 5 arcminutes. It is a harmonization of the global FAO-UNESCO Soil Map
of the World (FAO (1974)).

3.5   Eddy crosscheck

Before validating BETHY/DLR's modelled NPP for NUTS-1 regions across
all of Germany, we performed a crosscheck of BETHY/DLR GPP results with
eddy covariance flux tower measurements provided by FLUXNET. Two tower
sites were selected, one in the Hainich forest and one in the Tharandt forest.
The Hainich tower is to the west of Jena, Germany, in a mixed deciduous
beech forest, while the Tharandt tower is south of Dresden in a coniferous
forest. For both sites Level 4 data, providing information about GPP, are
available for 2000 and 2001. GPP was calculated by subtracting the estimated
ecosystem respiration from measured NEP) as described in Reichstein et al.
(2005). These data were crosschecked against BETHY/DLR's modelled GPP,
as annual sums at each station (Table 3.3).

Because BETHY/DLR underestimated annual GPP sums by 20 % to 30 % for
both stations over both years, a further analysis of monthly GPP sums was
performed (Fig. 3.2).

BETHY/DLR's GPP estimates qualitatively follow the measured GPP for the
coniferous forest at Tharandt over both years, but with a slight underestima-
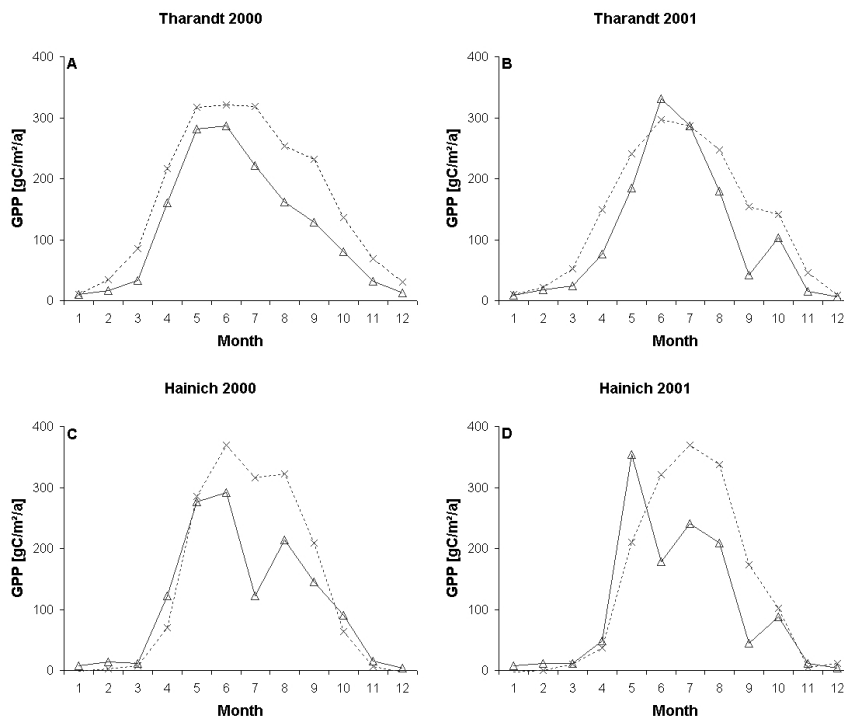
Figure 3.2: *Monthly GPP sums modelled with BETHY/DLR (solid lines) and measured with eddy flux towers (dashed lines) for the Tharandt and Hainich stations for 2000 and 2001.*

tion (Fig. 3.2a and b). This finding accords with the results of Wißkirchen (2005), who also observed good agreement between measured and modelled NEP at Tharandt. For the mixed deciduous forest of the Hainich station in 2000, BETHY/DLR's estimated GPP is good up to May and from September onwards. However, June, July and August exhibit a strong depression of the modelled GPP, responsible for the overall underestimation for this year (Fig. 3.2c). For Hainich in 2001 the modelled GPP is overestimated in May, but underestimated in all other months in the growing season (Fig. 3.2d).

Overall, BETHY/DLR models monthly GPP well for coniferous forest, with an underestimation of yearly GPP of less than 30 %. For deciduous forest the underestimation of yearly GPP is also less than 30 %, but monthly GPP shows an RMSE of up to 25 (Table 3.3). In particular, BETHY/DLR seems to strongly underestimate the peak of productivity during the middle of summer, perhaps due to the more heterogeneous development patterns of the species in this vegetation group.

Table 3.3: *Annual GPP sums in $(gC\,m^{-2}\,a^{-1})$ modelled with BETHY/DLR and measured with eddy flux towers at the Hainich and Tharandt stations for 2000 and 2001. The percent difference and RMSE are provided for each comparison.*

Year		Hainich		
	BETHY/DLR	Tower	Δ	RMSE
2000	1318	1649	20 %	20.9
2001	1210	1576	27 %	25.2

Year		Tharandt		
	BETHY/DLR	Tower	Δ	RMSE
2000	1426	2025	30 %	18.6
2001	1276	1655	23 %	17.9

Comparisons of GPP calculated by other vegetation models, such as BIOME BGC, ORCHIDEE and LPJ, with eddy covariance flux tower measurements revealed an overall RMSE of 30 % for forests in climate zones from boreal to Mediterranean (Jung (2008)). This magnitude of error corresponds well with our findings.

3.6 Validation strategy

To validate BETHY/DLR's modelled NPP, we used empirical data on the mean annual increment (MAI) of timbergrowing stocks. These data describe the change in the aboveground woody parts of trees with a diameter greater than 7 cm, given as solid tree volume (including bark), estimated during the second National Forest Inventory (NFI2) and the first National Forest Inventory (NFI1), divided by the time between the two inventories. We chose to use MAI instead of current annual increments (CAI) because for Germany no empirically-derived data with higher temporal resolution than the NFI is available. For Germany, these timber stock increment data are provided by the second National Forest Inventory, classified by forest, tree species, and age class. The aim of this ongoing large-scale survey is to establish statisti-

cally reliable central monitoring of the development of Germany's forests, to allow assessment of each forest's status and production potential. The NFI survey uses a permanent set of sampling sites, based on a nationwide 4 km × 4 km grid. Samples are taken at randomly chosen sites from this set, using a uniform procedure across all of Germany. This sampling procedure fulfills accuracy requirements at the national level, but more intensive sampling is conducted for greater accuracy in some smaller federal states, such as Bremen and Hamburg. From a statistical point of view the maximum error of the NFI survey regional level data is about 12 % for coniferous forest and below 8 % for deciduous forest, calculated from the data of BMELV (2004).

For each sampling site, the NFI surveys about 150 characteristics, such as tree species composition, timber volume, and growth, using an angle-count sampling method at each corner of the site. Furthermore, sampling circles with defined radii are drawn to survey tree species composition, tree heights, deadwood, ground vegetation and other characteristics. A more detailed description can be found in BMELV (2004).

The first NFI survey was conducted between 1986 and 1988 and was restricted to the ten states of the former West Germany. All data of the NFI1 are referenced to the year 1987. The second NFI survey was carried out in 2001 and 2002 across all federal states of Germany. The reference year for NFI2 is 2002. The NFI data are freely available at the NUTS-1 level (BMELV (2004).

In order to validate the modelled NPP against these NFI surveys, the highly detailed NFI data had to be aggregated. In a first step, the tree species reported in the NFI statistics were grouped into BETHY/DLR's two temperate forest classes, coniferous and deciduous. Coniferous forest in Germany is mainly composed of spruce (≈57 %) and pine (≈33 %), while Germany's deciduous forest is dominated by beech (≈48 %) and oak (≈25 %). It was therefore assumed that all parameters (standing timber stock and mean annual increment of timber stock, in particular) of these two forest classes could be estimated as the sum of the metrics for the two principal species, plus an estimated value for all of the secondary species.

The NFI data were then used to calculate the MAI of the total above-ground biomass (MAI_T), which was then compared to the modelled NPP. To calculate MAI_T for a NUTS region, the increments of the above-ground biomass of

coniferous (MAI_c) and deciduous (MAI_d) trees were summed (Eq. 1).

$$MAI_T = MAI_c + MAI_d \tag{3.1}$$

As described above, we calculated MAI_c and MAI_d as the sum of the increments of the total above-ground biomass of the two dominan species ($\widehat{M}$) plus an estimate for the tertiary species ($\widetilde{M}$). Both MAI_c and MAI_d, represented simply as MAI, were thus calculated following Eq. (2).

$$MAI = \sum_{s=1}^{2} \widehat{M}_s + \widetilde{M} \tag{3.2}$$

The index s represents the two dominant species of the forest class.
Since tree biomass depends upon age, $\widehat{M}$ was calculated using per-species age-dependent biomass increments, β, for the ten age classes in the NFI data, as shown by Eq. (3).

$$\widehat{M}_s = \sum_{a=1}^{10} \beta_{s,a} \times A_{s,a} \tag{3.3}$$

The index a represents the age classes, and A represents the proportional area occupied by each age class (given in the NFI data).
Since values for β are not given by the NFI, the NFI's species- and age-dependent net increments of the outer bark volume V, expressed in $\mathrm{m^3\,a^{-1}\,A^{-1}}$, are used. To convert timber stock biomass into absolute dry biomass, species-dependent conversion factors (ϵ) are needed (Table 3.4). Furthermore, species- and age-dependant conversion factors, P, are needed to estimate the total above-ground biomass. P represents the fraction of total above-ground biomass, including branches of less than 7 cm diameter and needles/leaves, that is outer bark (Table 3.4). Given these values, β was calculated as shown in Eq. (4).

$$\beta_{s,a} = \frac{V_{s,a} \times \epsilon_s}{P_{s,a}} \tag{3.4}$$

Because the first NFI was conducted before 1989, values for V are only available for the ten states of the former West Germany. For the five states of the former East Germany, values for V were taken from yield tables (Erteld

(1963)).

For this study values of ϵ were taken from Dieter and Englert (2001). Since only a single source of values for P was found, the accuracy of these values is unknown.

To calculate $\widetilde{M}$ it was assumed that a weighted average of the biomass increments of the two dominant tree species, $\widehat{M}$, would be representative for the tertiary species (Eq. 5). Statistical analysis indicated that the error of this approach is less than 6 %.

$$\widetilde{M} = \frac{\widetilde{A} \times \widetilde{V} \times \widetilde{\epsilon}}{\widetilde{P}} \tag{3.5}$$

$\widetilde{P}$ and $\widetilde{V}$ represent the arithmetic averages of the weighted average means of P and V, respectively. Since both depend on tree age and species composition, they were thus calculated using Eq. (6) ($\widetilde{P}$ being calculated analogously using P).

$$\widetilde{V} = \frac{\sum_{s=1}^{2} \left(\frac{\sum_{a=1}^{10} V_{s,a} \times A_{s,a}}{\sum_{a=1}^{10} A_{s,a}} \right)}{2} \tag{3.6}$$

$\widetilde{\epsilon}$ was calculated using the arithmetic mean of ϵ, following Eq. (7).

$$\widetilde{\epsilon} = \frac{\sum_{s=1}^{2} \epsilon_s}{2} \tag{3.7}$$

Before the modelled results could be validated, they needed to be aggregated to the NUTS-1 level at which the NFI data are given. To accomplish this, the modelled NPP was first transferred to a GIS, taking into account the equirectangular WGS84 datum map projection. Then, in order to allow comparison of the datasets, the CAI of the above-ground biomass of the modelled NPP was calculated. Because the modelled NPP does not specify which parts of the plant contain the accumulated carbon, the below-ground carbon content had to be estimated and removed. Furthermore, the NPP (in units of carbon) was converted to above-ground biomass (in units of dry weight) by applying a conversion factor (Eq. 8).

Table 3.4: *Per-species percentage of timber stock biomass (without bark) to total above-ground biomass (P) from Kramer and Krüger (1981), max ε from Dieter and Englert (2001), and min ε from Kaltschmitt and Hartmann (2001). /=no value.*

Species	ϵ (t m^{-3})		P per age class									
	min	max	I	II	III	IV	V	VI	VII	VIII	IX	X
Spruce	0.38	0.43	0	0.48	0.25	0.17	0.17	0.13	0.12	0.13	/	/
Pine	0.43	0.49	0	0.33	0.25	0.19	0.19	0.18	0.15	0.16	/	/
Beech	0.56	0.66	0	0.00	0.49	0.32	0.19	0.16	0.14	0.13	0.15	/
Oak	0.58	0.64	0	0.56	0.21	0.19	0.12	0.11	0.12	0.11	0.08	0.08

$$CAI = \frac{NPP \times \lambda}{(1 + R)} \tag{3.8}$$

R represents the ratio of the increment of below- to aboveground biomass, while λ is a conversion factor from NPP to total biomass. Species-specific values for R were taken from Pistorius and Zell (2005). Since the GLC2000 gives no information about tree species distribution, a mean value of R for each of the two temperate forest classes modelled by BETHY/DLR (coniferous and deciduous) was calculated (deciduous: 0.19 ± 0.08, coniferous: 0.23 ± 0.04). To check the numbers upon which these calculations were based, the corresponding allocation factors for above-ground biomass were also calculated using the same dataset (deciduous: 0.81, coniferous: 0.84); these values agree closely with previously reported values (Zhou et al. (2006)), supporting our estimated values for R. The value for λ was set to 2, which is seen as representative for both deciduous and coniferous trees (Houghton et al. (1997)).
After these calculations and conversions, the CAI derived from the model output was compared to the MAI calculated from NFI statistics in order to validate BETHY/DLR's estimates of NPP.

3.7   Results

Figure 3.3 depicts annual modelled NPP for Germany in 2000 and 2001 at this studys spatial resolution of $1\,\mathrm{km}^2$. For both years the major forests of southern Germany (the Spessart, the Palatinate, and the Black Forests) are clearly identifiable from their high NPP values, whereas the northern forested areas, such as the Harz mountains, have substantially lower NPP. The mean annual NPP for 2000 was 139 $(\mathrm{tC\,km^{-2}\,yr^{-1}})$ with a maximum of 547 $(\mathrm{tC\,km^{-2}\,yr^{-1}})$; for 2001 the mean annual NPP was 137 $(\mathrm{tC\,km^{-2}\,yr^{-1}})$ with a maximum of 544 $(\mathrm{tC\,km^{-2}\,yr^{-1}})$. Total annual modelled NPP was thus 21.6×10^6 tC for 2000 and 21.3×10^6 tC for 2001.
Conversion of these NPP values to above-ground biomass as described above gives annual totals of 52.3×10^6 t for 2000 and 51.8×10^6 t for 2001. The value estimated from NFI's data is 82.7×10^6 t (for both of 2000 and 2001). Our modelled NPP thus shows an underestimation of 37 % for both years com-
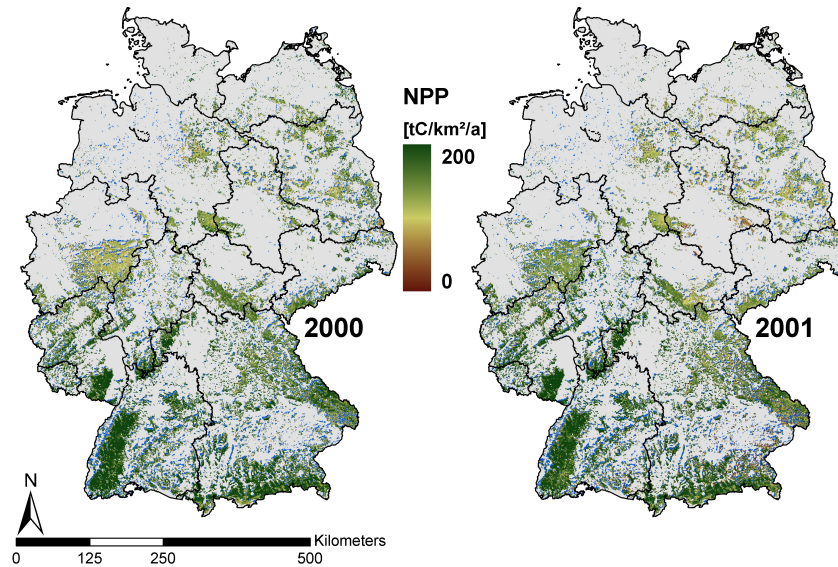
Figure 3.3: *Annual NPP of forest areas in Germany for 2000 and 2001. High NPP is shown in green, moderate NPP in yellow, and low NPP in red. Grey pixels represent areas which do not belong to the GLC2000 classes glc-2, glc-3, glc-4 or glc-6 (see also Table 3.1). Blue pixels represent pixels designated as forest in GLC2000, but that have less than 10 % forest cover according to VCF; their modelled NPP is therefore close to zero despite being considered forest.*

pared to empirical data. Furthermore, large areas with very low NPP can be identified, especially at the borders of larger forests such as the Black Forest of southwestern Germany. This is because these areas are considered to be forest according to GLC-2000, but MODIS VCF indicates very low forest cover fractions (down to one part per thousand). Such areas of conflicting land cover information are shown as blue pixels in Fig. 3.3.

Figure 3.4 shows that BETHY/DLR underestimates the net increment of above-ground biomass for both deciduous and coniferous trees. The R^2 values of 0.74 and 0.76 for deciduous trees indicate a high degree of correlation, however. The correlation for coniferous trees is even stronger, with R^2 values of 0.95 and 0.93, but the underestimation is also higher here. In order to quantify the predictive accuracy of BETHY/DLR's NPP estimates, the root mean square error (RMSE) was calculated for all four panels; for deciduous trees the RMSE is 1.53 (2000) and 1.48 (2001), and for coniferous trees, 1.87 (2000) and 1.93 (2001).
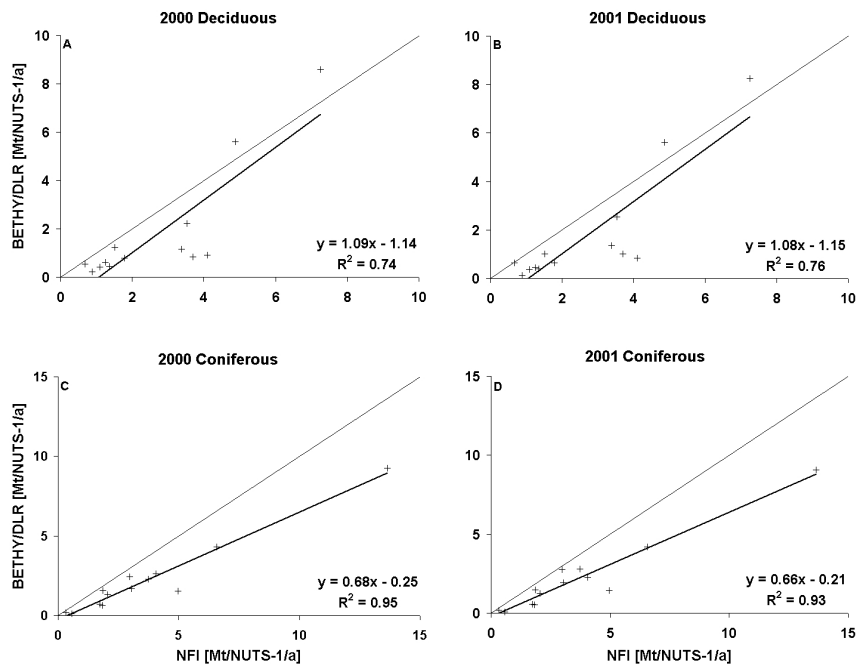
Figure 3.4: *Estimated above-ground biomass increment from modelled NPP (CAI) versus empirical data from the NFI (MAI) for Germany's deciduous and coniferous trees for 2000 and 2001. Each cross represents one NUTS-1 region. Thick lines show linear regressions. Values are given in megatons per NUTS-1 unit per year.*

The MAI of above-ground biomass for deciduous trees in Germany for all NUTS-1 regions is 821.9 tons per km^2 (NFI), but the corresponding value estimated with BETHY/DLR is 530.9 tons per km^2, 35 % less. For coniferous trees these values are 804.7 tons per km^2 (NFI) and 416.0 tons per km^2 (BETHY/DLR), a 48 % underestimate.

A reason for underestimation can be found in the land cover/land use classification used (GLC2000). Figure 3.5 presents a comparison of the forest areas derived from the NFI database, the forest areas of GLC2000, and the forest areas for the intersection of GLC2000 and MODIS VCF.

From Fig. 3.5b it is apparent that NFI and GLC2000 deciduous forest area estimates differ markedly; an underestimation of 66 % for Schleswig-Holstein (SH) and an overestimation of 106 % for Bavaria (BY) is observed, for example. For coniferous forest (Fig. 3.5a) the two area estimates are more comparable, with a mean difference of 20 %. These imbalances are reduced when looking at the total forest areas for deciduous and coniferous trees across all of Germany;
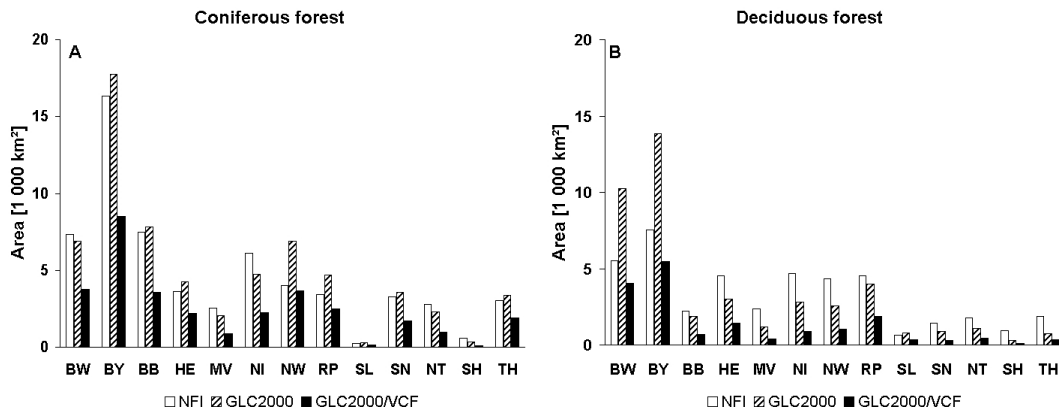
Figure 3.5: *Comparison of coniferous and deciduous forest areas from NFI, from GLC2000, and from the intersection of GLC2000 and MODIS VCF areas. Areas are given in 1 000 km² per NUTS-1 unit. BW: Baden Württemberg, BY: Bavaria, BB: Berlin/Brandenburg, HE: Hesse, MV: Mecklenburg-Western Pomerania, NI: Lower Saxony/Hamburg/Bremen, NW: North Rhine-Westphalia, RP: Rhineland-Palatinate, SL: Saarland, SN: Saxony, NT: Saxony-Anhalt, SH: Schleswig-Holstein, TH: Thuringia.*

total coniferous forest area estimates are 42 400 km² (NFI) and 47 100 km² (GLC2000), and for the deciduous forest, 60 800 km² (NFI) and 61 100 km² (GLC2000). It can also be seen in Fig. 3.5a and b that the GLC2000 underestimates forest areas for the northern states of Germany such as Schleswig-Holstein (SH) and Lower Saxony (NI), whereas it overestimates the forest areas for southern states such as Bavaria (BY) and Baden-Württemberg (BW).

In aggregate, then, GLC2000 represents forest area well, but its spatial distribution is not comparable with the NFI data. We hypothesize that the medium-scale forest structure found in most parts of Germany is not adequately described by the GLC2000, due to the difficulty of accurately classifying a heterogeneous land cover distribution even with a resolution of 1 km² (Mayaux et al. (2006)). According to the Land Cover Classification System (DiGregorio and Jansen (2001)) used in deriving GLC2000, the GLC2000 class "Broadleaved Deciduous Closed to Open (100 to 40) Trees" includes all forest areas with a forest fraction from 40 % to 100 % a very wide range. In order to describe the forest cover fraction more precisely, the MODIS VCF product was combined with the GLC2000 to produce the area estimates used as in-

puts by BETHY/DLR. Figure 3.5a and b show the coniferous and deciduous forest areas that result from this combination of MODIS VCF and GLC2000. Clearly this approach led to underestimations of the forest area in Germany, both for coniferous (47 %) and deciduous (59 %) forest. This underestimation occurs because only areas reported as forested in the GLC2000 were carried forward to be combined with the MODIS VCF coverage data; areas designated as non-forest in GLC2000, but with a non-zero forest cover fraction in VCF, were treated as non-forested. As a result, those forest areas which were underestimated by GLC2000, such as Lower Saxony or Schleswig-Holstein, led to substantial underestimations of the increment of above-ground biomass.

Other land cover datasets with higher resolution, such as CORINE (100 m × 100 m) and MERIS GlobCover (300 m × 300 m), are available for Germany, and their land use structures show a better agreement with the NFI data. Since BETHY/DLR requires land cover and LAI inputs to be at the same spatial resolution, and since no higher-resolution LAI products are available yet for Germany, these finer grained land cover datasets unfortunately could not be used. Exploratory analysis shows, however, that the combination of GlobCover and VCF leads to an underestimation of forest area of 24 %, while the combination of CORINE and VCF yields an underestimation of only 7 %. This agrees with the findings of EEA (2006), which estimated the reliability of the CORINE classes 311 (coniferous forest) and 312 (deciduous forest) at better than 85 %. We observe, therefore, that while area-wide land cover products at high resolution are needed and useful, high-resolution datasets for plant physiology parameters such as LAI must also keep pace if these products are to be of maximal utility.

Returning to Fig. 3.4, we note that when the number of observations is small, the slope of a regression line is very sensitive to outliers. In the case of deciduous forests in 2000 and 2001 two outliers can be identified that have a strong influence on the slope of the regression line: Bavaria and Baden Württemberg, the two largest federal states of Germany. In Fig. 3.4 these states have the largest values (on both axes), because of their large areas.

To compensate for the large effect of both: the potential outliers and forest area underestimation, we normalized the CAI and MAI data for each NUTS-1 region by dividing these values by forest area, resulting in units of tons per km^2
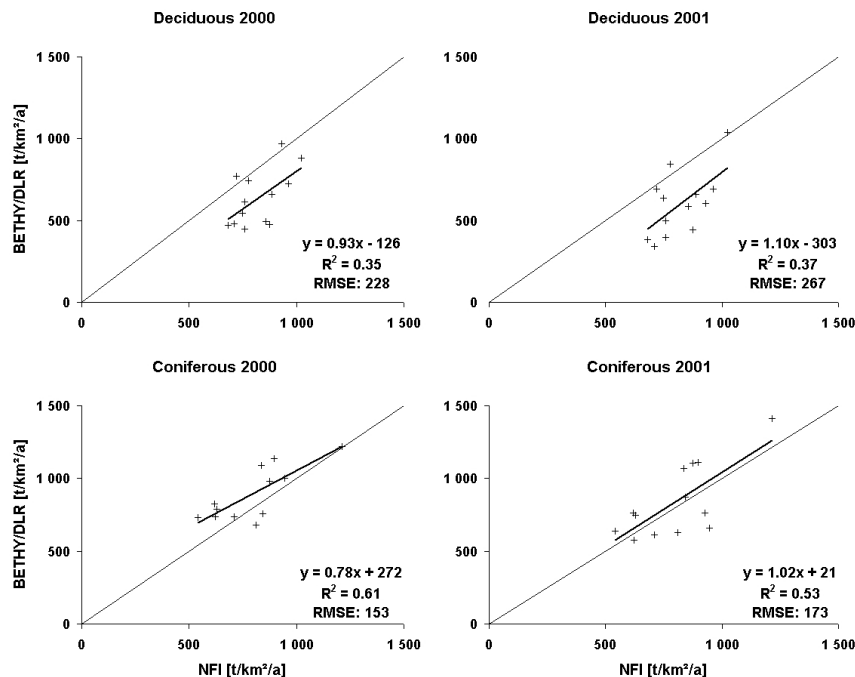
Figure 3.6: *Normalized estimated above-ground biomass increment from modelled NPP versus empirical data from the NFI for Germany's deciduous and coniferous trees for 2000 and 2001. Each cross represents one NUTS-1 region. Thick lines show linear regressions. Values are given in tons km^2*

per NUTS unit. For this the MAI data was devidid by the forest area reported in the NFI and the CAI data by the combined area of GLC2000 and VCF as used for the model run. The results are presented with linear regressions in Fig. 3.6.

Figure 3.6 shows that BETHY/DLR does not exhibit an underestimation for coniferous trees with these area normalized metrics, indicating that the underestimation seen in Fig. 3.4 might indeed be explained by GLC2000's area underestimation, as discussed above. The R^2 values of 0.61 and 0.53 still indicate a reasonable degree of correlation.

The underestimated CAI for deciduous trees, observed in Fig. 3.4 and persisting in Fig. 3.6 after area-normalization, might be explained by the models internal model parameters related to carbon uptake: maximum carboxylation rate and maximum electron transport rate. The values used for these parameters were taken from Knorr (1997) (see also Knorr and Heimann (2001)), where they were used for global carbon assessment; these values thus represent global

mean values. However, forests in Germany are probably more productive than the "global mean" trees simulated by BETHY/DLR using these parameter values, because of their age. The last large reforestation programme in Germany followed World War II, to mitigate the deforestation experienced during the war. According to the NFI, the mean age of Germany's forests is about 67 yr (81 yr for deciduous and 54 yr for coniferous trees), an age class that is expected to exhibit a high rate of increase of timber biomass. Young and old trees differ in their carbon allocation and fixation strategies; in particular, carbon fixation and timber growth decreases with increasing tree age. In old trees, the maintenance respiration rate is nearly as high as the carbon uptake rate, and thus the large majority of GPP in older trees is dedicated to maintenance. The carbon uptake of young trees, on the other hand, is mainly used for growth. Studies show that the transition between these two metabolic regimes occurs at about 60 to 80 yr of age (Zhou et al. (2006)). Therefore it is likely that the values used for the maximum carboxylation rate and the maximum electron transport rate are too low to accurately simulate the tree communities of Germany (see Zähle et al. (2006) for further discussion of this issue).

Underestimations in the modelled NPP could also be the result of the neglect of nitrogen deposition in the model. Luyssaert et al. (2010) showed an increase in modelled NPP of up to 30 % when nitrogen deposition is included in the model formulation.

Uncertainties for deciduous trees are higher (R^2 0.35 and 0.37) than for conifers, which might indicate higher structural variability among deciduous tree species. In particular, deciduous tree species exhibit greater variation in shape than do coniferous tree species in nature.

Finally, it should be taken into account that NFI statistics can only produce MAI values; these values were estimated from the difference between the first NFI survey, in 1987, and the second, in 2002. Until NFI conducts a third survey year, the effects of climatic variability cannot be captured by the NFI statistics.

3.8 Estimation of energy potentials

A further objective of this study is to derive energy potentials both from modelled NPP and from NFI data. The energy potential of forests is of considerable importance to the sustainable energy discussion and the development of sustainable energy policy.

To estimate theoretical energy potentials, species-specific lower heating values (H) can be used to convert from absolute dry above-ground biomass. Heating values represent the maximum energy obtainable from the combustion of biogenic solid fuels, and are given in mega joules per kilogram. Since the GLC2000 gives no information about tree species, but does differentiate between deciduous and coniferous trees, mean heating values representative of each tree class were used. For this study, heating values for deciduous and coniferous trees were calculated for each NUTS-1 unit (Table 3.5), taking into account the relative abundance and age distribution of tree species in each region. Heating values for the main deciduous (oak and beech) and coniferous (pine and spruce) tree species can be found in many sources; see, for example, Kaltschmitt and Hartmann (2001) and Grammel (1989). We assume sustainable management of forest biomass; following this assumption, only the mean annual net increment of forest biomass is used to calculate theoretical energy potential.

Since B (above-ground biomass; see Eq. (8) is determined for absolute dry conditions, we calculated the energy potential (J) as shown in Eq. (9).

$$J = B \times H \tag{3.9}$$

Using this equation, theoretical energy potentials for 2000 and 2001 were estimated from the above-ground biomass previously calculated from modelled NPP (Fig. 3.7). Although the validation previously conducted had demonstrated a systematic underestimation of NPP, no correction was applied here to compensate for this, since such correction would have resulted in an incorrect spatial distribution of estimated energy potentials.

Comparison with Fig. 3.3 shows that those forest areas having the highest NPP values also have the highest theoretical energy potentials. This is also true, mutatis mutandis, for low NPP and energy potential, and is valid for
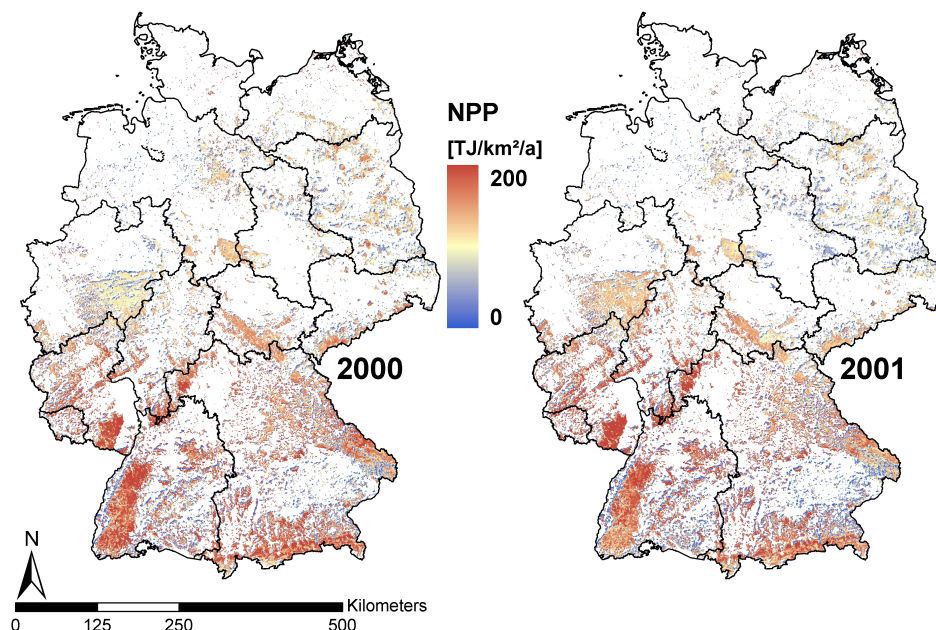
Figure 3.7: *Sustainable theoretically available energy potential, in terajoules per 1 km^2 pixel, of forest areas in Germany for 2000 and 2001. Low energy potentials are shown in blue, intermediate potentials in beige, and high energy potentials in red. White represents, as in Fig. 3.3, areas which are not designated as forested by GLC2000 (see Table 3.1).*

both 2000 and 2001. Analysis revealed that the mean theoretical available energy potential for coniferous forest is 17.5 $(\mathrm{TJ\,km^{-2}\,yr^{-1}})$ for 2000 and 2001, while for deciduous forest these values are 25.0 $(\mathrm{TJ\,km^{-2}\,yr^{-1}})$ (2000) and 24.6 $(\mathrm{TJ\,km^{-2}\,yr^{-1}})$ (2001). Maximum values of 25.7 $(\mathrm{TJ\,km^{-2}\,yr^{-1}})$ (coniferous forest) and 25.4 $(\mathrm{TJ\,km^{-2}\,yr^{-1}})$ (deciduous forest) were found.

The NFI data were also used to estimate empirical energy potentials, using Eq. (10) for all thirteen NUTS-1 units in Germany (Table 3.5). These estimates of energy potential are partitioned into the two main tree classes.

Since it was shown previously that underestimation existed both for the areas of forests (as derived from remote sensing) and for modelled NPP, it is unsurprising that the empirical energy potential estimated from NFI data is 37 % higher than the theoretical estimate from modelled NPP.

Table 3.5: *Total forest areas, estimated increment of above-ground biomass from NFI data, heating values, and empirical energy potentials of woody biomass for Germany's NUTS-1 regions, partitioned into deciduous and coniferous tree classes.*

NUTS-1 region	Overall forest area	Tree Type	AGB increment	Heating Value	Empirical energy potential
	(ha)		(tons yr^{-1})	(MJ kg^{-1})	(PJ yr^{-1})
Hesse	813 092	deciduous	3 384 016	18.4	62.3
		coniferous	3 037 591	19.3	58.7
			6 421 607	18.8	121.0
Schleswig - Holstein	154 602	deciduous	876 626	18.4	16.1
		coniferous	571 957	19.5	11.1
			1 448 583	18.8	27.3
Lower Saxony, Hamburg & Bremen	1 081 248	deciduous	4 091 784	18.4	75.3
		coniferous	4 966 180	19.9	98.9
			9 057 965	19.2	174.2
North Rhine-Westphalia	835 763	deciduous	3 700 153	18.4	68.1
		coniferous	3 733 869	19.1	71.3
			7 434 021	18.7	139.4
Rhineland - Palatinate	794 432	deciduous	3 527 763	18.4	64.9
		coniferous	2 972 838	19.3	57.5
			6 500 601	18.8	122.4

NUTS-1 region	Overall forest area	Tree Type	AGB increment	Heating Value	Empirical energy potential
Baden - Württemberg	1 281 409	deciduous	4 869 976	18.4	89.6
		coniferous	6 564 707	19.2	125.8
			11 434 683	18.8	215.4
Bavaria	2 386 027	deciduous	7 253 796	18.4	133.5
		coniferous	13 642 971	19.3	263.2
			20 896 766	19.0	396.6
Saarland	92 131	deciduous	674 469	18.4	12.4
		coniferous	318 401	19.2	6.1
			992 871	18.7	18.5
Brandenburg & Berlin	973 017	deciduous	1 519 005	18.4	27.9
		coniferous	4 058 284	20.9	84.7
			5 577 289	20.2	112.6
Mecklenburg - Western Pomerania	492 673	deciduous	1 787 216	18.4	32.9
		coniferous	1 817 119	20.3	36.9
			3 604 336	19.4	69.8
Saxony	471 290	deciduous	1 090 393	18.4	20.1
		coniferous	2 057 192	19.7	40.5
			3 147 585	19.2	60.5

NUTS-1 region	Overall forest area	Tree Type	AGB increment	Heating Value	Empirical energy potential
Saxony - Anhalt	454 640	deciduous	1 241 898	18.4	22.9
		coniferous	1 737 999	20.4	35.4
			2 979 897	19.5	58.2
Thuringia	490 276	deciduous	1 356 655	18.4	25.0
		coniferous	1 864 346	19.4	36.1
			3 221 001	19.0	61.1
Germany	10 320 601		82 717 205	19.1	1579.9
Uncertainty				±1.25	±103.5

3.9 Conclusions

For this study we modelled the Net Primary Productivity (NPP) of German forests for 2000 and 2001 using the dynamic biomass model BETHY/DLR. We presented a new method for the validation of modelled NPP using empirical data related to MAI of above ground biomass. Modelled NPP was converted and aggregated to a net increment of above-ground biomass per NUTS-1 unit for comparison to these empirical data. With this method we showed a high degree of correlation between modelled and empirical NPP, although the modelled NPP underestimated the empirical NPP. A comparison with data from two eddy covariance flux towers revealed that BETHY/DLR represents annual productivity patterns well (particularly for coniferous forest) but with substantial underestimation.

In a second step, the sustainable theoretical energy potential of the above-ground biomass was estimated, using heating values to convert estimated above-ground biomass to energy units. For comparison, energy potentials were also calculated from empirical data, which revealed that modelled energy potentials are underestimated by 37%, a consequence of the prior underestimation of modelled NPP.

Reasons for this pattern of underestimation were discussed; in particular, it was shown that GLC2000 does not represent the spatial distribution of forest areas well due to its limited resolution. We thus argue that 1 km^2 resolution is insufficient to describe the heterogeneous small-scale structure of mid-European forests. For future modelling, the use of higher-resolution land cover products might allow more accurate NPP estimation; this should be tested in future research. To facilitate the use of such products, however, there is a need for matching high-resolution datasets for vegetation metrics such as LAI.

Furthermore we hypothesize that the maximum carboxylation rate and maximum electron transport rate is age dependent and thus potentially responsible for the underestimation of modelled NPP, since BETHY/DLR does not take the tree age distribution into account. Further research in this area could thus lead to more exact results.

Modelled NEP, NPP and GPP are typically validated using eddy flux tower measurements, but such measurements are not available in many areas. Our

presented validation method could therefore be helpful in the assessment of model outputs both at a broader spatial scale, and in less developed countries. Our method will, additionally, allow the development of a downscaling procedure for empirically derived NUTS-level data, allowing NUTS data to be partitioned into smaller spatial units. Our MAI-based validation method should, however, be tested in additional countries, and should be comprehensively compared to validation using eddy measurements, so that the benefits and drawbacks of each method are clearly understood.

This new MAI-based validation method will be useful in validating modelled NPP; as we demonstrated here, that also allows the further estimation of other metrics, such as bioenergy potentials. Such estimates of forest energy potentials play an important role in planning for a sustainable economy. More broadly, accurate and precise model results, crosschecked against empirical data, are needed for a better understanding of optimal forest management and the future possibilities of renewable energy.

3.10 Acknowledgements

Chapter 4

Estimating Agricultural Bioenergy Potentials for Germany Using a Process Based Vegetation Model

Markus Tum, Kurt P. Günther, *Sustainable Bioenergy*, Spinger Book, accepted for publication

4.1 Abstract

We present an approach to the estimation of sustainable straw energy potentials using a modelled Net Primary Productivity (NPP) product validated by empirical data on the managed area and mean yields for the main crops in Germany. The Biosphere Energy Transfer Hydrology Model (BETHY/DLR) is the theoretical framework to estimate NPP for agricultural areas in Germany. It is driven by remote sensing data from SPOT-VEGETATION, meteorological data from the European Center for Medium-Range Weather Forecast (ECMWF), and additional static datasets, as a land cover information (GLC2000), a soil map (ISRIC-WISE) and an elevation model (ETOP05). The output of BETHY/DLR, the yearly accumulated NPP, is first converted to straw potentials using simple allocation rules (root-to-shoot and yield-to-straw ratios) and then transferred to energy potentials using species-specific lower heating values. The results for the years 2006 and 2007 are compared with data from literature. With our method to estimating sustainable bio-

energy potentials we found good agreements to established approaches, with only little overestimations (up to 12 %) and high correlations with R^2 of up to 0.78. Our analysis shows that the presented approach fills an important gap in estimating energy potentials from modelled NPP. Such estimated straw biomass energy potentials play an important role in the sustainable energy debate.

4.2   Introduction

During the last sixty years the treatment of straw as a side product of cereal production changed considerably in many developed countries. This is mainly caused by the reduction of on field straw burning, which was done for improving fertilization, pest control and to avoid nitrogen immobilisation (Borresen (1999)). The decreasing demand for straw as bedding litter in feeding lots which is caused by changes in the housing systems (Jordan et al., 2008) also resulted in an increase of available straw on fields in many regions. However along positive effects of letting straw residues on field, like the stabilization of the topsoil, specific crop rotations require the removal of the straw (Zebarth et al., 2009). Cropping systems with a high straw supply rate can thus offer the possibility of straw removal without changing the soil conditions.

The current politically motivated energy discussion has refocused the attention on renewable energy sources and thus on the energetic use of agricultural by-products such as straw. Since no human food competition is related to its use, it has a considerable potential, but is limited by several factors. For Central Europe two major limiting factors can be stressed: animal husbandry and demand for organic material for the humus-balance. During the last ten years several studies have been conducted for Germany, all with the aim to assess total and regional straw potentials (e.g.: Gauder et al. (2011), Zeller et al. (2011), Pacan and Dröge (2010), Thrän et al. (2010), Fritsche et al. (2004)).

The general approach of these studies is to use empirical data on land use and mean yields to estimate theoretical available straw potentials and sustainable energy potentials after respecting use competitions. Thus a spatial limitation to the resolution of the empirical data source is always given.

Besides these empirical approaches, remote sensing driven vegetation models,

established to assess the carbon uptake by plants can also provide information on the straw potential, but with spatial explicit resolution. Vegetation models have become an important tool to answering questions concerning the mechanisms driving the carbon cycle and the roles of terrestrial carbon sinks and sources (Cox et al. (1999)). Models as the Biosphere Energy Transfer Hydrology (BETHY/DLR) model have already been tested to estimate sustainable energy potentials for Germany forests (Tum et al. (2011)) and showed reasonable good results. Reflecting on the current political discussion on renewable energy sources, more detailed information on the local availability of straw potentials are needed, if a sustainable and cost-efficiency use is aimed.

The primary objective of this study is to investigate an approach to the estimation of straw potentials calculated from modelled NPP from BETHY/DLR, at a $1\,km^2$ spatial resolution. Statistical data on the land use and yields of the main crops at level 3 of the "Nomenclature des Unite's Territoriales Statistiques" (NUTS) are used to calibrate the estimated straw potentials. For this plant specific allocation schemes, as the root-to-shoot and yield-to-straw ratios are used. Germany was selected as test areas due to data availability. Computing time and hard disk storage issues restricted our modelling to the years 2006 and 2007.

4.3   Model description

BETHY/DLR is a special soil-vegetation-atmosphere-transfer (SVAT) model that models photosynthesis, and takes into account environmental conditions that affect it. SVAT models track the plant-mediated transformation of atmospheric carbon dioxide into energy-storing hydrocarbons such as sugars, a process known as carbon fixation. BETHY/DLR was originally designed for global applications (Knorr and Heimann (2001)) and was adapted for regional use by Wißkirchen (2005).

The process of photosynthesis is parameterized following the combined approach of Farquhar et al. (1980) and Collatz et al. (1992). Dark and light reactions of photosynthesis are calculated on leaf level and treated separately. With this approach the photosynthesis rate can be limited either by light availability or the carboxylation enzyme Rubisco, the key player in the Calvin cycle that

fixes carbon. Because of their significant differences in their carbon fixation physiologies so-called C3 and C4 plants are distinguished in BETHY/DLR. C4 plants such as sugar beet and corn can fix more atmospheric carbon dioxide at higher temperatures than can C3 plants such as barley and wheat.

To extrapolate photosynthesis from leaf to canopy level, the canopy structure, and the soil-atmosphere-vegetation interaction is taken into account. For closed and open canopies (forests, shrubs, grassland and crops) the photosynthetic rate depends on the Leaf Area Index (LAI). Self shading is considered by reducing the photosynthetic rate from canopy top to soil using the "two-flux scheme" of Sellers (1985) with three canopy layers.

Besides photosynthesis, other energy transfers, such as heat fluxes between vegetation and the atmosphere and the cooling effect of evapotranspiration, are also considered. Furthermore the soil heat flux and the storage of heat in the canopy is taken into account. The coupling of these processes is of great importance, since temperature-dependent photosynthesis transforms light energy into chemical energy, and finally into carbohydrates, using water and CO_2.

The water cycle is also modelled and included in the interaction scheme. Three reservoirs are considered: soil water, snow, and "skin" or "intercepted" water on leaves and other parts of the vegetation, which change in time and space. Soil water is available for vegetation, while evapotranspiration from vegetation and evaporation from soil determine the water loss to the atmosphere. Water limitation is modelled by calculating the demand for evapotranspiration using the approach of Monteith (1965) with the criteria of Federer (1979), assuming that evapotranspiration cannot be greater than the limit set by the soil water supply and the water uptake of a plant's roots. Thus when considering the dynamic interaction of, for instance, the soil water balance and photosynthesis, the natural behaviour of vegetation can be reflected, which is the motivating idea of the SVAT approach.

Autotrophic respiration is modelled in BETHY/DLR as the sum of maintenance respiration and growth respiration. Maintenance respiration is limited by vegetation-specific dark respiration rates. Growth respiration is assumed to be a constant fraction of NPP. The model output of BETHY/DLR is given as a time series of NPP in daily steps, at the resolution and projection of the

land cover classification. For this study the Global Landcover Classification 2000 (GLC2000) with a $1\,\mathrm{km}^2$ resolution is used.

4.4 Input data

The inputs of the BETHY/DLR model are two remote sensing datasets derived from SPOT VEGETATION, meteorological time series data provided by ECMWF, and two static datasets describing soil type and land elevation.

Meteorological data

BETHY/DLR needs meteorological time series with a temporal resolution of at least once per day. The European Centre for Medium-Range Weather forecast (ECMWF) provides this data with a spatial resolution of $0.25\,^\circ \times 0.25\,^\circ$ and a temporal resolution of up to four times a day. The ECMWF INTERIM dataset contains a broad variety of parameters from which air temperature (at $2\,\mathrm{m}$ height), wind speed (at $10\,\mathrm{m}$ height), soil water content (in the four uppermost layers), cloud cover and precipitation are used. The INTERIM reanalysis combines meteorological station, satellite and airborne based measurements. From these data we calculated daily mean, minimum and maximum temperatures, daily mean cloud cover at three heights. Daily temperatures were reprocessed at the $1\,\mathrm{km}^2$ resolution of the model output, adjusting for the elevation difference between the ECMWF data and the elevation of each model pixel, using a $1\,\mathrm{km}^2$ elevation map and the temperature gradient of the international standard atmosphere (-0.65 K per 100 m).

We calculated the daily average photosynthetically active radiation (PAR) from global radiation following the method of Burridge and Gadd (1974). Using this method PAR was calculated using the incident sunlight for the given day and year, limited by atmospheric transmission, which depends on the degree of cloudiness. Daily average cloud cover was calculated using a weighted sum of each cloud layer. The advantage of this approach is to be seen in the more exact results than the direct use of radiation forecast data (Wißkirchen (2005)).

Daily volumetric soil water content data was needed to calculate the soil water budged of the model. Soil type information was taken from the International

Soil Reference and Information Centre-World Inventory of Soil Emission Potentials (ISRIC-WISE) dataset, which is a harmonization of the global FAO-UNESCO Soil Map of the World (FAO (1974)) and is available with a 5×5 arcminutes resolution.

Remote Sensing Data

In addition to meteorological data BETHY/DLR is driven by two remote sensing based datasets. These are time series of the LAI and a detailed and homogenous land cover/land use product. LAI time series are used to indicate the phenology of vegetation and are based on the CYCLOPES 10-day composites dataset, provided by the POSTEL (Pole d'Observation des Surfaces continentales par Teledetection) database.

For each $1\,km^2$ pixel time series analysis, namely the harmonic analysis, was applied to fill data gaps and eliminate outliers. The harmonic analysis decomposes a time series into a linear combination of suitable trigonometric functions, (sine and cosine oscillations) of particular periodicities. In principal the power spectrum is deconvolved by iteratively finding and subtracting the highest peak of the time series power spectrum.

The CYCLOPES database also provides a land cover and land use information, given as the GLC2000 (Fritz et al. (2003), Bartholome and Belward (2005)). For the derivation of the GLC2000 land cover classes the "Land Cover Classification System (LCCS)" of the FAO was used (DiGregorio and Jansen (2001)). The GLC2000 dataset is representative for the year 2000 and includes 22 different land cover classes. The CYCLOPES / VITO dataset was chosen because it is thought to be the most accurate dataset for agricultural areas (Garrigues et al. (2008)).

4.5 Energy Potentials

The main objective of this study is to derive sustainable straw energy potentials from modelled and validated NPP (Tum and Günther (2011)) for agricultural areas in Germany, and to compare these with recently published estimates. Straw energy potentials are of considerable importance to the sustainable energy discussion and the development of sustainable energy policy.

Before estimating the energy content of straw the modelled NPP needs to be transferred to dry above ground biomass. This can be done by using simple crop specific allocation schemes. Because the GLC2000 only gives information about the general land use an additional dataset, describing the area use and yields for the main crops was needed, to differentiate straw crops, such as wheat, barley and non-straw crops such as sugar beet and potatoes. We used empirical data from the Federal Statistical Office of Germany. This farm structure survey is conducted yearly. It contains yield and area use information for the main crops on NUTS-3 resolution. The "NUTS" hierarchical spatial classification starts with the member states of the European Community (EU) (NUTS-0), followed by regions of the EU (NUTS-1), separated to basic administrative units (NUTS-2) and ends with subdivisions of those basic administrative units (NUTS-3). As the dataset contains gaps, a criterion was needed to fill these. We assumed that gaps for a given crop may be filled using the mean yield of the given crop from the German NUTS-3 units.

In a first step the modelled NPP of BETHY/DLR was aggregated to NUTS-3 units and put in relation to NPP values per NUTS-3 unites calculated from empirical data, as described in Tum and Günther (2011). To calculate the NPP of straw-providing crops (NPP_s) the NPP of non-straw-providing crops (NPP_{ns}) was subtracted from the aggregated modelled NPP per NUTS area (NPP_N) considering the percentage land use as described in the empirical dataset.

$$NPP_s = NPP_N - NPP_{ns} \qquad (4.1)$$

The remaining NPP_s was then transferred to above ground NPP (NPP_a) by subtracting the below ground NPP part (NPP_b) using crop specific root to shoot ratios:

$$NPP_a = NPP_s - NPP_b \qquad (4.2)$$

In a next step the straw content (NPP_{st}) of NPP_a was calculated by subtracting the yield content (NPP_{yi}) using crop specific yield to straw ratios:

$$NPP_{st} = NPP_a - NPP_{yi} \tag{4.3}$$

The final sustainable straw potential (S_{pot}) was then calculated by adding non carbon ($_{non}C$) and water (H_2O) contents to NPP_{st}. The empirical factor of 0.29 was taken from Gauder et al. (2011), to respect use competitions of the harvested straw.

$$S_{pot} = (NPP_{st} + H_2O +_{non} C) \times 0.29 \tag{4.4}$$

Crop specific root to shoot and yield to straw ratios, water and non carbon contents were taken from Tum and Günther (2011).

The now available S_{pot} values per NUTS-3 region can directly be used to estimate sustainable straw energy potentials. For this species-specific lower heating values (H) are needed to convert dry above ground biomass to energy. Heating values represent the maximum energy output from burning biogenic solid fuels and are given in mega joules per kilogram. Since the GLC2000 gives no information about crop species a mean heating value per NUTS-3 unit was calculated ($\langle H \rangle$). For this study, heating values for rye-, wheat-, barley- and rapeseed straw from Kaltschmitt and Hartmann (2001)) were used.

For each NUTS-3 unit we calculated the energy potential (J_n) as shown in Eq. 5.

$$J_n = S_{pot} \times \langle H \rangle \tag{4.5}$$

In a last step the energy potentials per NUTS-3 units were spatially reallocated using the modelled NPP values. For this we assumed high NPP values of the model output to be representative for high energy potentials and vice versa. For each pixel (i) we calculated the energy content (J_i) as presented in Eq. 6.

$$J_i = \frac{NPP_i}{NPP_N} \times J_n \tag{4.6}$$

With this approach we have assumed that the percentage of straw-providing

crops for each pixel is the same as for the full NUTS-3 region.

4.6 Results and Discussion

Figure 4.1 depicts estimated annual straw energy potentials for Germany in 2006 and 2007 at the study's spatial resolution of $1\,\mathrm{km}^2$. For both years areas with high energy potential values can be identified central Germany. This area is located in the Magdeburger Börde which is well known for its extensive agricultural use. Other areas as the Münsterland in Northwest Germany and parts of Southeast Germany show a high variability over the two years of observation. Overall the calculated energy potentials for 2006 are lower than for 2007, which we assume to be caused by climate conditions. The mean annual energy potential for 2006 was $0.52\,[\mathrm{TJ\,km^{-2}\,y^{-1}}]$ with a maximum of $2.85\,[\mathrm{TJ\,km^{-2}\,y^{-1}}]$; for 2007 the mean annual energy potential was $0.70\,[\mathrm{TJ\,km^{-2}\,y^{-1}}]$ with a maximum of $2.75\,[\mathrm{TJ\,km^{-2}\,y^{-1}}]$. Total annual estimated energy potential was thus $156\,\mathrm{PJ}$ for 2006 and $217\,\mathrm{PJ}$ for 2007.

Our estimates are in good agreement to values calculated from mean straw potentials reported in literature (Zeller et al. (2011)). They used three different methods to estimate annual straw potentials for Germany and the 16 federal states taking into account the humus balance which is required for sustainable crop and soil management and which is the basis for the direct payment obligation - accounting regulation. Depending on the method of estimation, a mean annual energy potentials of 112 to $186\,[\mathrm{PJ\,y^{-1}}]$ is calculated for Germany applying a mean heating value H of $14.05\,\mathrm{MJ\,kg^{-1}}$. The heating value is representative for dry matter with $14\,\%$ moisture.

In addition to the annual sum we analyzed the correlation of the modelled sustainable energy potential of both years with the mean sustainable energy straw potential for each Federal State of Germany as given in Zeller et al. (2011). Therefore the biomass potential was transferred into energy potentials using a mean H value of $14.05\,\mathrm{MJ\,kg^{-1}}$ as before. The results are presented as linear regressions in Figure 4.2.

Figure 4.2 shows for both years that the calculated sustainable energy potential from BETHY/DLR model tends to slightly overestimate on the Federal State level. The R^2 values of 0.78 and 0.70 indicate a high degree of correlation
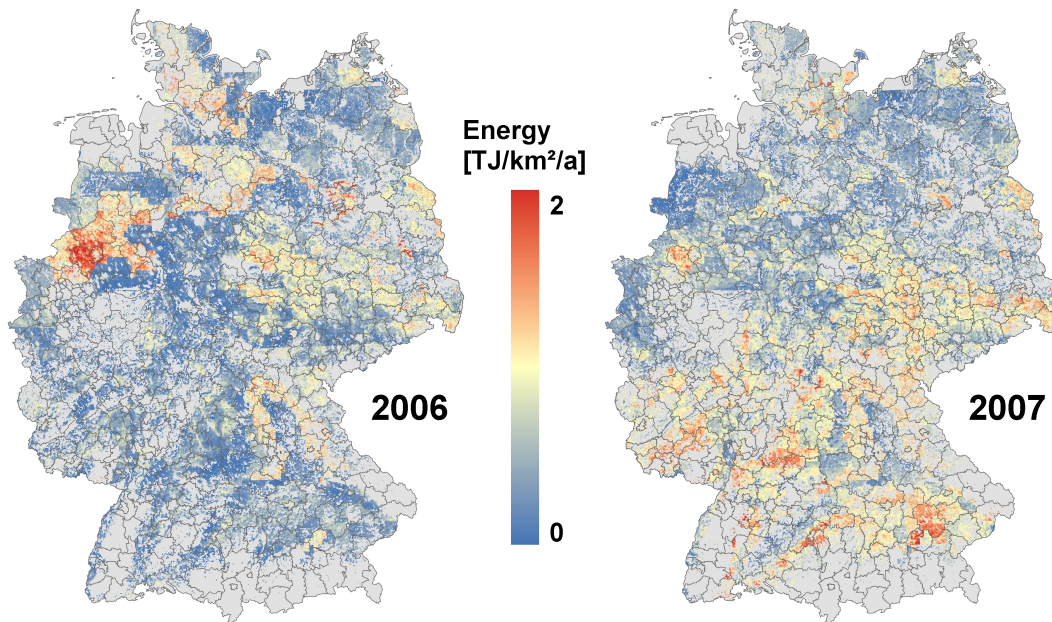
Figure 4.1: *Sustainable energy potential, in terajoules per $1\,km^2$ pixels of agricultural areas in Germany for 2006 and 2007, modelled by BETHY/DLR. Low energy potentials are shown in blue, intermediate in beige, and high energy potentials in red. Grey represents areas which are not designated as managed areas by GLC2000.*
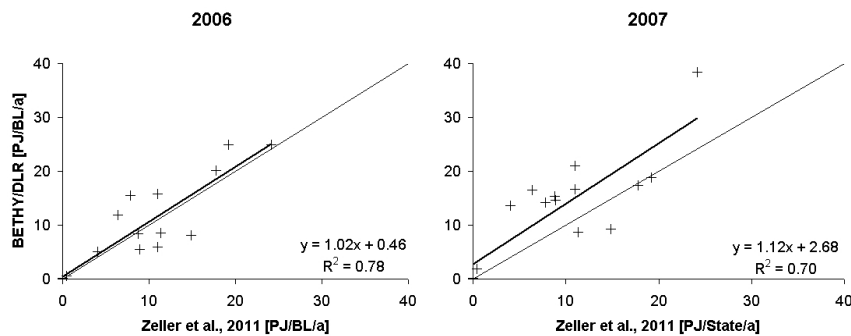
Figure 4.2: *Correlation of sustainable energy potentials for the 16 Federal States of Germany derived from modelled NPP with data from Zeller et al. (2011). Energy potentials are modelled for 2006 and 2007. Data points indicate energy potential of the Federal States. Doted lines indicate perfect correlation; solid lines indicate the correlation found by linear regression. Energy potentials are given in PJ per Federal State (FS) and year.*

for both years. In order to quantify the agreement of our estimations to the literature data, the root mean square error (RMSE) was calculated for both years; for 2006 the RMSE is $3.9\,\mathrm{PJ\,FS^{-1}\,y^{-1}}$ and for 2007 $6.5\,\mathrm{PJ\,FS^{-1}\,y^{-1}}$. It is obvious from Figure 4.1 that the sustainable energy potentials for most regions of Germany are reduced compared to 2007. Our assumption that this finding is related to meteorological conditions is supported by a note in the agro meteorological bulletin of the MARS (Monitoring Agriculture with Remote Sensing) project (JRC (2006)). There the year 2006 is characterized as "a below-average cereal season explained by hot and dry summer followed by over-wet conditions at harvest". For Germany a mean yield for wheat (including soft and durum wheat) of 6.6 tons per hectare is reported while the 5 years moving average is estimated to 7.4 tons per hectare. This corresponds to a reduction of about 11 %. On the other hand barely and grain maize show the same yield as for the previous years. Thus a slight reduction of the total cereal yield for Germany is observed for 2006 compared to the 5 years average. In the year 2007 the cereal production in "Germany was again limited by wet conditions at harvest (winter cereals) but not on the same amplitude as in 2006" (JRC (2007)). The yield forecast for wheat in 2007 is in the range of the 5 years average. For barely the forecast is about 3 % lower while for grain maize the yield forecast is about 8 % higher than the average yield. In total, 2007 is a more productive year than 2006 according to the MARS bulletins. This finding is supported by our results.

Looking again to figure 4.1, it is obvious an interesting feature with high energy potentials can be identified in central Germany, namely the Magdeburger Börde. This region is also an area with extensive agricultural use. For this area the ISRIC-WISE dataset reports cambisols and chernozems. Especially chernozems are highly fertile soils and thus favourite agricultural areas. Thus constantly high straw potential are expected as seen for both years.

In Northwestern Germany, namely the Münsterland, high energy potentials are observed for 2006 and also to minor degree in 2007. In the Münsterland the precipitation sum for the full year 2006 (732mm) is considerably lower than for 2007 (831 mm) and even lower than the 10 years average (781 mm). A similar pattern for the precipitation is seen in the area around Landshut in southeast Germany for 2006 and 2007. For the Landshut area the precipitation for the

Table 4.1: *Precipitation rates given in milimeters and mean temperature given in*
°C for the Münsterland and Landshut area for 2006 and 2007.

	Münsterland		Landshut area	
10 year mean precipitation sum [mm]	781		779	
10 year mean precipitation sum [mm] for the growing season (15. Mar. - 30. Sept.)	432		422	
10 year mean temperature (15. Mar-30. Sept.) [°C]	15.2		14.7	
	2006	**2007**	**2006**	**2007**
Precipitation sum (1. Jan.-31. Dec.) [mm]	732	831	736	971
Mean temperature (1. Jan.-31. Dec.) [°C]	11.4	11.4	9.4	10.1
Precipitation (15. Mar-30. Sept.) [mm]	475	496	427	558
Mean temperature (15. Mar-30. Sept.) [°C]	14.9	15.4	13.9	15.0

full year 2006 is about 736 mm while for 2007 971 mm of precipitation was
measured as presented in Table 4.1. When discussing the energy potential of
straw and in turn the biomass development the most important meteorological
parameters are the precipitation and the mean temperature in the growing sea-
son. For our analysis we defined the growing season from 15. April till end of
September. In the growing season the precipitation of both years is higher than
the 10 years average, for both regions. An analysis of monthly mean temper-
atures, which we calculated from daily values taken from ECMWF data, was
performed for both regions to investigate potential warming or cooling effects
on the plant growth and finally on the straw energy potential. The time series
of the mean monthly temperature for the Münsterland and Landshut area are
presented in Figure 4.3 for both years. The monthly mean temperatures for
2006 and 2007 are significantly different in the non productive time period
(January to mid March and October to December) for both regions. However
for the growing season (mid March to September) small changes can be found.
The mean temperature during the growing period from mid March to Septem-
ber is lower in 2006 than in 2007 for the Landshut region and equal for the
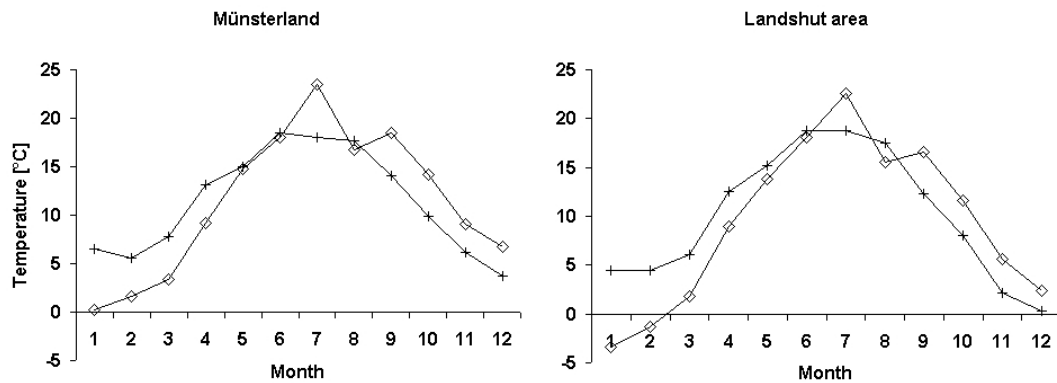Münsterland region. Compared to the 10 years average, the mean temperature

Figure 4.3: *Monthly mean temperatures in °C for the Münsterland (left) and Landshut area (right) for 2006 (triangles) and 2007 (diamonds).*

for 2006 is lower for both regions. An explanation for the high energy potentials in the Münsterland in 2006 and in the Landshut region in 2007 can be found when looking at the scatterplot of mean temperature and precipitation in the growing season as shown in Figure 4.4 for the years 1999 to 2010. All mean values of meteorological parameters are based on daily ECMWF data. It becomes evident that for the Landshut region the growing season of 2006 is relative cold and wet while the growing season of 2007 is relative warm and wet (compared to the 10 years average, shown in Figure 4.4 as an open diamond). For the Münsterland, the growing season of 2006 is only a little bit colder and wetter than the 10 years average but the growing season of 2007 is a little bit warmer and significantly wetter than the 10 years average. These findings can be summarized that the meteorological conditions in the Münsterland in 2006 are more favourable than in 2007. For the Landshut region, the cold conditions in 2006 reduce biomass growth and thus energy potential of straw.

When looking at the mean yields for two NUTS-3 units which are representative for the two described regions it becomes evident that our modelled NPP data reproduce lower yields and thus lower straw potentials (Table 4.2) for the Landshut region for 2006 and for the Münsterland for 2007. The mean yield for the two NUTS-3 units comes from the agricultural statistical survey.
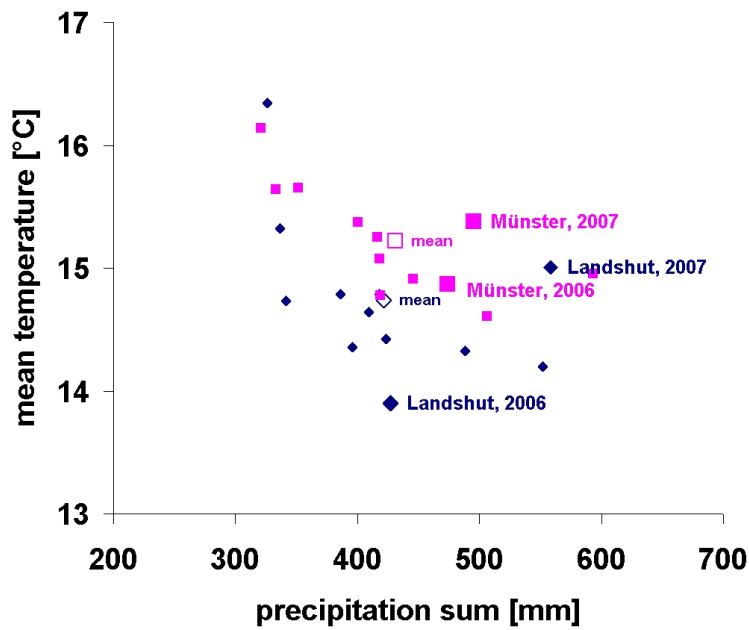
Figure 4.4: *Scatterplot of the precipitation sum [mm] and the mean temperature [°C] for the growing season (15. March - 30. September) for the 11 years (1999 - 2010). Data for the Münsterland are presented as squares (magenta) while data for the Landshut region are given as diamonds (blue). The average values for both regions are presented as open symbols.*

Table 4.2: *Mean yields for prominent straw-providing crops for two NUTS-3 units in Germany for 2006 and 2007. Steinfurt is representative for the Münsterland area while Landshut represents the Landshut area. Values are given in $dt\,ha^{-1}$.*

	Winter wheat	Rye	Winter barley	Summer barley	Oats	Triticale	Other Cereals
Landshut							
2006	70.4	54.3	56.8	46.7	49.6	68.8	57.8
2007	78.6	62.1	65.3	51.0	50.0	76.2	63.9
Steinfurt							
2006	68.4	60.4	59.2	44.5	39.7	54.7	54.5
2007	60.7	40.7	48.5	34.3	38.3	47.3	45.5

4.7 Conclusion

Sustainable straw energy potentials for Germany for 2006 and 2007 were estimated using modelled NPP from the vegetation model BETHY/DLR. Inputs for the model were LAI time series from the VEGETATION satellite, meteorological data from ECMWF and land cover/land use data from the GLC2000. We here present an approach to estimate sustainable energy potentials using empirical data on average grain yields and acreage of main crops on NUTS-3 level. Using conversion factors (root-to-shoot and corn-to-straw ratios), the modelled NPP data was converted to harvested straw per NUTS unit, taking into account NUTS specific land use practices. Compared to recently published straw potential values this method yielded reasonable high coefficients of determination (R^2 up to 0.78) combined with a slight overestimation (up to 12 %), allowing strong conclusions to be drawn about the usability of the presented method.

For individual areas changes in the rate of precipitation and mean annual temperature could be shown. We furthermore proofed lower mean temperatures and wet conditions, especially for the growing season to be correspondent with lower mean grain yields. We hypothesize significant cooler mean temperatures for the growing seasons combined with high precipitation rates to be the cause of yield losses. In the corresponding signal of our calculated sustainable energy potential this phenomena can also be found. In this we see another good indicator on the usability of our method.

This study illustrates an approach of calculating sustainable straw energy potentials that we believe will be useful in estimating energy potentials from modelled NPP products on a medium resolution. This method could also be used as a downscaling approach for empirically derived straw potential data on NUTS level, since the model results could help to spatially represent the NUTS information.

4.8 Acknowledgements

Chapter 5

How sensitive are estimates of carbon fixation in agricultural models to input data?

Markus Tum, Franziska Strauss, Ian McCallum, Kurt P. Günther, Erwin Schmid, *Carbon Balance and Management*, in revision

5.1 Abstract

Background

Process based vegetation models are central to understand the hydrological and carbon cycle. To achieve useful results at regional to global scales, such models require various input data from a wide range of earth observations. Since the geographical extent of these datasets varies from local to global scale, data quality and validity is of major interest when they are chosen for use. It is important to assess the effect of different input datasets in terms of quality to model outputs. In this article, we reflect on both: the uncertainty in input data and the reliability of model results. For our case study analysis we selected the Marchfeld region in Austria. We used independent meteorological datasets from the Central Institute for Meteorology and Geodynamics and the European Centre for Medium-Range Weather Forecasts (ECMWF). Land cover / land use information was taken from the GLC2000 and the CORINE

2000 products.

Results

For our case study analysis we selected two different process based models: the Environmental Policy Integrated Climate (EPIC) and the Biosphere Energy Transfer Hydrology (BETHY/DLR) model. Both process models show a congruent pattern to changes in input data. The annual variability of NPP reaches 36 % for BETHY/DLR and 39 % for EPIC when changing major input datasets. However, EPIC is less sensitive to meteorological input data than BETHY/DLR. The ECMWF maximum temperatures show a systematic pattern. Temperatures above 20 °C are overestimated, whereas temperatures below 20 °C are underestimated, resulting in an overall underestimation of NPP in both models. Besides, BETHY/DLR is sensitive to the choice and accuracy of the land cover product.

Discussion

This study shows that the impact of input data uncertainty on modelling results need to be assessed: whenever the models are applied under new conditions, local data should be used for both input and result comparison.

5.2 Background

Modelling the net carbon uptake by vegetation (Net Primary Productivity, NPP) and estimating the yields of agricultural plants have become important tools to study the mechanisms of carbon exchange between the atmosphere and vegetation, as well as issues of food security. Different approaches are currently tracked which can be grouped to their approaches how photosynthesis is modelled.

Models describing the chemical, physical and plant physiological processes of plant development and the interaction of plants with the atmosphere can be applied to simulate the rate of carbon dioxide uptake of the plant through photosynthesis (called Gross Primary Productivity, GPP). These models follow the concept of Monsi and Saeki (1953) and Monteith (1965) to simulate

the process of photosynthesis. Moreover, carbon uptake of well-watered and fertilized annual plants is linearly related to the amount of absorbed Photosynthetically Active Radiation (PAR), which can be derived from satellite data (i.e. the fraction of PAR which is absorbed by the canopy; cp. McCallum et al. (2010) or calculated by the accumulation of dry matter.

NPP is defined as the difference between GPP and autotrophic respiration. Therefore, it is important to estimate the autotrophic respiration of plants following the determination of GPP. Autotrophic respiration is defined as the oxidation of organic compounds found in roots, stems and leaves, to CO_2 or water. In the literature, different approaches to estimate autotrophic respiration are discussed, taking into account the actual biomass or GPP (e.g. Knorr and Heimann (2001), Running et al. (2000), Goetz et al. (1999)). When the Light Use Efficiency (LUE) approach is integrated in a coupled soil - plant - atmosphere model as in the EPIC (Environment Policy Integrated Climate) model, daily estimates of evapotranspiration and carbon assimilation fluxes can be obtained Williams (1995).

In contrast to these models, more sophisticated approaches are in use and under development. These models track photosynthesis on the molecule level. They take into account the interaction between plants, atmosphere and soil by simulating the uptake and release of carbon by plants and soil in a physically consistent way including conservation of energy and momentum.

In the literature one can find descriptions of established vegetation models for use on different scales Bondeau et al. (2007), Krinner et al. (2005), Haxeltine and Prentice (1996), Prentice et al. (1992). Each of these models is driven by meteorological input data and parameterized for global use with special focus on the long-term competition between the plant functional types when natural disturbance and succession driven by light competition occur. Models with a spatial resolution of kilometres and a time horizon of some years as e.g. the Examples are soil-vegetation-atmosphere-transfer (SVAT) model BETHY/DLR (Biosphere Energy Transfer Hydrology Model) Wißkirchen (2005) can be used for regional assessments of NPP or biomass development.

During the last decades, the use of both modelling approaches was often met with resistance, mainly because of the need of calibration, validation and determination of the level of uncertainty (e.g.: Battaglia et al. (2004), Sands

et al. (2000), Vancley and Skovsgaard (1997)). Furthermore for many users, i.e. policy makers, it is difficult to judge whether the model outputs are within acceptable levels of uncertainty or not, mainly due to their limited background in model development Jakeman et al. (2006). However, in this context it is of importance to the policy maker to understand the validity of the model results and their associated uncertainties.

Since empirical research traditionally advances in its data accuracy and validity - in contrast - process-based models do not always provide comparable outputs, it is difficult to judge on the quality of modelled data, especially with the traditional criteria for assessing scientific outcomes van Oijen (2002). However, regardless of the data's source, there will always be some uncertainty associated with it.

To address these issues, we have assessed the variability of the soil-vegetation-atmosphere-transfer model BETHY/DLR (Wißkirchen (2005)) and the biophysical process model EPIC (Williams (1995)) on three different meteorological input datasets and two land cover maps. Since the two models were designed for different specific purposes, we do not intend to discuss advantages or disadvantages but place special attention on the investigation of model sensitivity to the spatial resolution of the input datasets. The Austrian Marchfeld region has been chosen as case study analysis because many datasets (table 5.2) are readily available. The period of investigation is 2000 to 2003. It is important to note that this study is not a classical sensitivity analysis for assessing systematically the responses of models to changes in input data and model parameters (e.g. Lamboni et al. (2009), Larocque et al. (2008), White et al. (2000), Recknagel (1984)), but a model variability analysis.

5.3   Methods

Biophysical process models

EPIC is a comprehensive model under continuous development since 1981, capable of simulating many agricultural processes that occur as a result of climate forcing, landscape characteristics, soil conditions and crop management schemes (Williams (1995), Izaurralde et al. (2006), Williams et al. (1984)).

Biophysical processes simulated with EPIC include among others plant and crop growth, hydrology, wind and water erosion, and nutrient cycling. These processes are simulated with daily time steps or smaller. EPIC contains algorithms that allow for a complete description of the hydrological balance at the small watershed scale (up to 100 ha) including snowmelt, surface runoff, infiltration, soil water content, percolation, lateral flow, water table dynamics, and evapotranspiration. Daily weather can be endogenously generated for precipitation, temperature, solar radiation, wind, and relative humidity or it can be input exogenously.

EPIC uses the concept of radiation-use efficiency by which a fraction of daily photosynthetically active radiation is intercepted by the plant canopy and converted into plant biomass. The leaf area index is simulated as a function of heat units, crop stress and development stages. Daily gains in plant biomass are affected by vapor pressure deficits and atmospheric CO_2 concentration (Stockle et al. (1992)). Crop yield is simulated using the harvest index which is affected by the heat unit factor and includes the amount of the crop removed from the field as well as the above-ground biomass. Stress indices for water, temperature, nitrogen, phosphorus and aeration are calculated daily using the value of the most severe of these stresses to reduce potential plant growth and crop yield. Similarly, stress factors for soil strength, temperature, and aluminum toxicity are used to adjust potential root growth (Jones et al. (1991)).

The soil water balance depending on the potential water use, the root zone depth and the water use distribution parameter is applied in a general water use function where any water deficit can be overcome if a layer that is encountered has adequate water storage. The potential water use is reduced when the soil water storage is less than 25 % of plant-available soil water by using dependencies on the soil water contents at field capacity and wilting point (Williams (1995)).

BETHY/DLR belongs to the family of SVAT models, which track the transformation of atmospheric carbon dioxide into energy storing sugars, a process known as photosynthesis. BETHY/DLR is based on the Jena Scheme of Atmosphere Biosphere Coupling in Hamburg (JSBACH) by Knorr and Heimann (2001) and was modified by Wißkirchen (2005). The JSBACH model was originally considered for global usage and computes the biosphere-atmosphere

exchange within the Global Circulation Model ECHAM5 (European Centre Hamburg). BETHY/DLR as well as JSBACH use the combined approach to integrate photosynthesis (Farquhar et al. (1980), Collatz et al. (1992)), which means that the enzyme kinetics are parameterized on the leaf level. In this context, C3 and C4 plants are distinguished because of significant differences in the way of their carbon-fixation: C4 plants (e.g. corn and sugar cane) are able to fix more atmospheric carbon dioxide at high temperatures than C3 plants (e.g. wheat and barley). Thus, the photosynthesis of C3 plants is saturated at higher temperatures. In a second step, the rate of photosynthesis is extrapolated from leaf to canopy level by taking into account both, the canopy structure as well as the interaction of the plant between soil, atmosphere and vegetation. The two-flux scheme of Sellers (1985) which includes three canopy layers, is used to approximate the radiation absorption in the canopy. Evapotranspiration, stomatal conductance and the soil water balance is included in the model formulation. To compute NPP on an annual basis snow is included in the water budget. Water stress is considered by calculating the demand for evapotranspiration using the approach of Monteith (1965) limited by the criteria of Federer (1979). Here it is assumed, that evapotranspiration can not be higher than a certain soil water supply via roots. Autotrophic respiration is evaluated as the sum of maintenance and growth respiration. The plant specific dark respiration determines the maintenance respiration, while growth respiration is assumed to be proportional to the difference between GPP and maintenance respiration. The main outputs of BETHY/DLR are given by time series of GPP, NPP, evapotranspiration, and of soil water content in daily steps with the spatial resolution of the respective land cover classification. A more detailed model description can be found in Wißkirchen (2005).

The general characteristics as e.g. main outputs and the general formulation to compute NPP of the two models BETHY/DLR and EPIC are presented in table 5.1.

Table 5.1: *General characteristics of the biophysical process models EPIC and BETHY/DLR.*

	BETHY/DLR	EPIC
Abbreviation	Biosphere Energy Transfer Hydrology Model	Environmental Policy Integrated Climate
References	Knorr and Heimann (2001); Wißkirchen (2005)	Williams et al. (1984); Williams (1995); Izaurralde et al. (2006)
Model type	Dynamic vegetation model	Deterministic crop model
Time step	Daily	Up to < 1 day
Main simulation processes	GPP, NPP, NEP, evapotranspiration, soil water content	plant and crop growth, heat and water balance, wind and water erosion, nutrient cycling
General formulation to compute NPP	NPP = GPP - autotrophic respiration	NPP = (yield + straw + roots) - (water content + non carbon fraction)

5.4   Framework of Case Study Analysis

The Austrian Marchfeld region serves as case study area to assess the variability of the two biophysical process models on alternative input datasets. The EPIC model has already been applied and validated here Schmid et al. (2004), and the data necessary for our study is readily available (see table 5.2). The Marchfeld region is located in Lower Austria, part of the Vienna Basin, and forms with around 100,000 ha one of the largest plains in Austria. Around 75 % of the area is used for agricultural production. The natural boundaries are to the East the river March (the Austrian border to Slovakia), to the North the hills of the Weinviertel, to the West the mountain range of Bisamberg and the city of Vienna, and to the South the river Danube. For locating the region a map is presented in figure 5.1.

Since land use practices are not homogenously distributed in this area, five
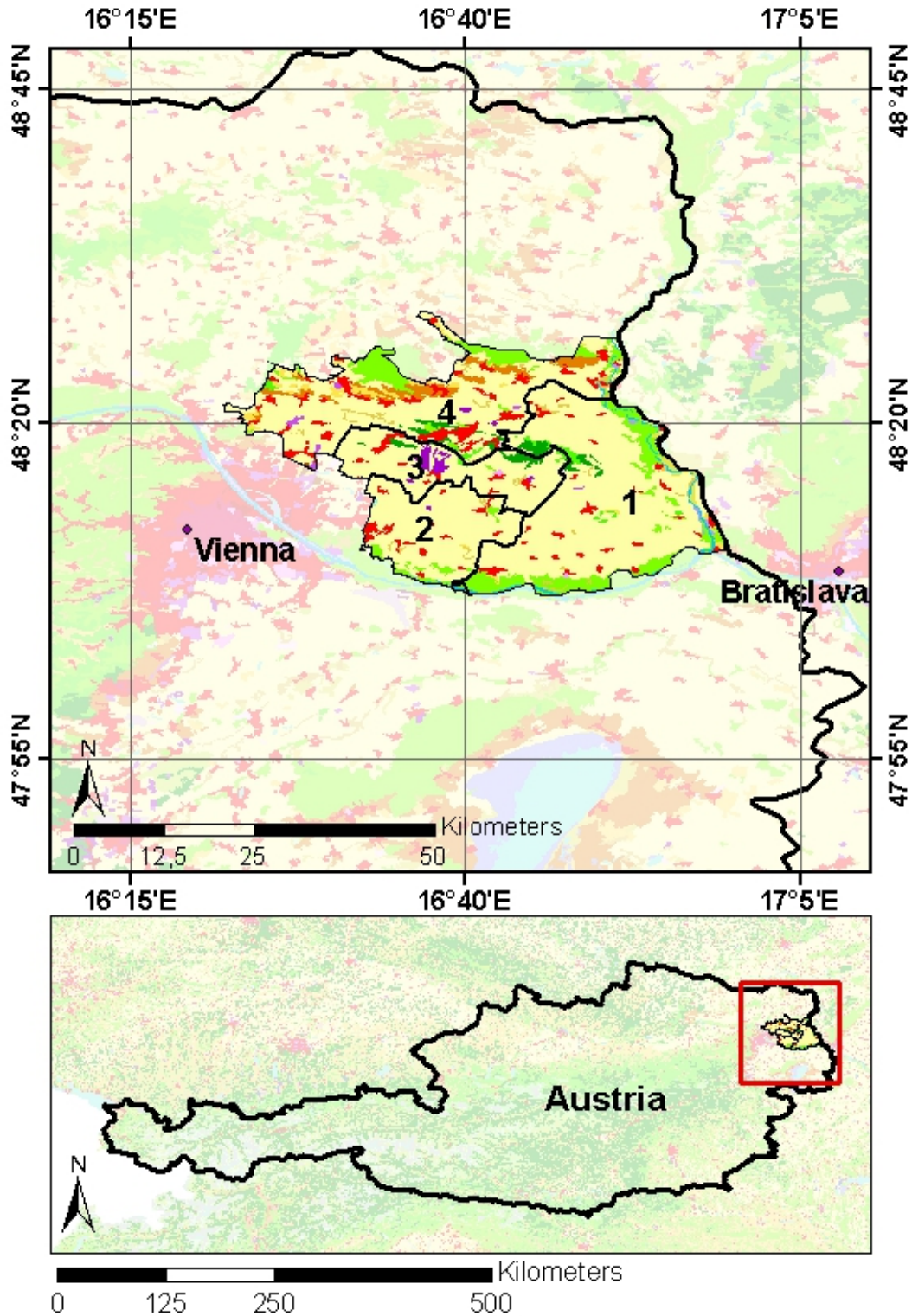
Figure 5.1: *The case study area Marchfeld with the four sub-regions (upper figure), with underlying CORINE land cover dataset 2000. Green pixels represent forest, red and violet pixels urban areas, brown pixels shrub land, and yellow pixels agricultural areas. The lower figure highlights the location of the Marchfeld region. The red square represents the map extract of the upper figure.*

sub-regions have been identified using the cluster analysis methods Hofreither et al. (2000). Each sub-region has an area of in between $85\,\mathrm{km}^2$ and $250\,\mathrm{km}^2$. The urban land cover as well as forest and shrub lands have not been taken into account in the variability analysis. Five typical soils have been selected with respect to majority criteria for the agricultural land cover (four different Chernozems and one black earth; Schmid et al. (2004)).

The biophysical process models have been applied with different meteorological inputs (table 5.2) from the period 2000 to 2003. We have used meteorological observations from weather stations of the Central Institute for Meteorology and Geodynamics (ZAMG) in the Marchfeld region, reallocated meteorological data from weather stations across Austria of ZAMG Strauss et al. (2012), and meteorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF).

The meteorological observations (ZAMG) are from the weather station in Gross Enzersdorf, and provide daily values of six weather parameters including minimum and maximum temperatures, relative humidity, wind speed precipitation and solar radiation.

Strauss et al. (2012) developed a reallocated meteorological dataset comprising climate data for Austria and the period from 1975 to 2007 with temporal and spatial resolution of one day and $1\,\mathrm{km}^2$. In addition climate change scenarios have been developed for the period 2008 to 2040. They processed daily data from 34 weather stations of ZAMG to 60 spatial climate clusters with homogeneous climates relating to mean annual precipitation sums and mean annual temperatures from the period 1961-1990. Based on these precipitation and temperature classes four climate clusters describe the climate in the Marchfeld region (cluster 1: mean annual precipitation sums smaller than 500 mm and mean annual temperatures between $8.5\,°\mathrm{C}$ and $9.5\,°\mathrm{C}$; cluster 2: mean annual precipitation sums smaller than 500 mm and mean annual temperatures between $9.5\,°\mathrm{C}$ and $10.5\,°\mathrm{C}$; cluster 3: mean annual precipitation sums between 500 mm and 600 mm and mean annual temperatures between $8.5°\mathrm{C}$ and $9.5\,°\mathrm{C}$; cluster 4: mean annual precipitation sums between $500\,\mathrm{mm}$ and $600\,\mathrm{mm}$ and mean annual temperatures between $9.5°\mathrm{C}$ and $10.5°\mathrm{C}$). For each homogenous climate cluster, Strauss et al. (2012) performed regression model

Table 5.2: *Meteorological, land cover, and other data.*

Datatype	Period used	Resolution of space and time	Parameters used	Characteristics	References
Meteorology	2000-2003	weather stations; daily	Precipitation; Minimum temperature; Maximum temperature; Wind Speed; Radiation; Relative Humidity	Measured time series	Central Institute for Meteorolo6gy and Geodynamics (ZAMG)
Meteorology	2000-2003	$1\,km^2$ grid; daily	Precipitation; Minimum temperature; Maximum temperature; Wind Speed; Radiation; Relative Humidity	Reallocated time series (from now on 'ZAMG reallocated')	Strauss et al. (2012)
Meteorology	2000-2003	$0.25\,°$; up to 4 times a day	Precipitation; Minimum temperature; Maximum temperature; Wind Speed; Cloud cover; Soil Water Content	Time series of model re-analysis (ERA-40)	European Centre for Medium-Range Weather Forecasts (ECMWF)

Datatype	Period used	Resolution of space and time	Parameters used	Characteristics	References
Vegetation Indices	2000-2003	1 km^2 grid; 36 time steps per year	Leaf Area Index (LAI)	Global coverage	Pôle d'Observation des Surfaces continentals par TELédétection (POSTEL)
Landcover	2000	1 km^2 grid, year 2000	Land cover / land use information; 22 classes	Global coverage (GLC2000)	Bartholome et al. (2002); DiGregorio and Jansen (2001)
Landcover	2000	1 ha grid, year 2000	Land cover / land use information; 44 classes	European coverage (CORINE 2000)	Bossard et al. (2000)
Census	1999 'Agrarstrukturerhebung'	Marchfeld sectors, year 1999	Agricultural land use information; Main soil type distribution Land use data of farms aggregated to municipalities	Schmid et al. (2004); Schmid et al. (2007)	

analyses primarily to compute a set of daily climate data for the time period 2008 to 2040. This method has also been applied for the time period 1975 to 2007 to provide a consistent dataset. The integral parts of the regression model are i) the consideration (extrapolation in the period 2008 to 2040, respectively) of the observed linear temperature trend from 1975 to 2007 derived from a homogenized dataset, and ii) the repeated bootstrapping of temperature residuals and of observations for solar radiation, precipitation, relative humidity, and wind speed to ensure consistent spatial and temporal correlations. We have also used these reallocated data for the period 2000 to 2003 in our variability analysis.

The third dataset is derived from ECMWF data and has a temporal resolution of up to four times a day and a spatial resolution of $0.25\,° \times 0.25\,°$. It includes model analysis data of $2\,m$ air temperature, cloud cover, soil water content of the four upper layers and wind speed at $10\,m$ above ground. From this dataset the daily mean, as well as minimum and maximum temperatures and the daily mean of cloud cover in all three strata (high, medium, low) are used. The daily temperature values are scaled with the difference between ECMWF reference height and the global ETOP05 (Earth Topography and Ocean Bathymetry Database) 5-minute gridded elevation data by using the temperature gradient of the U.S. Standard Atmosphere (-0.65 K per $100\,m$) in order to downscale the ECMWF temperature data to km^2 resolution. Precipitation values are derived twice a day from the ECMWF re-analysis project (ERA-40). PAR is not used directly from the corresponding ECMWF product data as it is only available as forecast data and therefore rather uncertain. Thus, daily PAR is determined from global radiation which is computed following the approach of Burridge and Gadd (1974) taking into account the geographical coordinates of the day, and using a transmission, which depends on the degree of cloudiness. The degree of cloudiness is calculated as a weighted sum of each cloud strata for each day, and the global radiation is calculated for each location in the time step of one hour. The advantage of this approach is the use of analysis data of cloud coverage to compute PAR data which leads to more exact results than directly using the PAR forecast data Wißkirchen (2005).

Hence the BETHY/DLR model needs an initial soil water content, the ECMWF soil water dataset is used only for the transient phase of the model. Afterwards

the model simulates the soil water content independently, according to the hydrological boundary conditions. Investigations of Wißkirchen (2005) have shown that in most cases sufficient hydrological boundary conditions are available after a transient phase of about one year.

In addition to the meteorological data, the BETHY/DLR model is driven by two sets of remote sensing data. Detailed and homogenous land cover /land use information are used to get information about the vegetation types the model is run for. Vegetation is represented by time series of the Leaf Area Index (LAI). Time series of LAI were used from the "Carbon cycle and Change in Land Observational Products from an Ensemble of Satellites" (CYCLOPES) 10 day composite datasets of POSTEL (Pole d'Observation des Surfaces continentales par TELedetection), which have a spatial resolution of $1\,km \times 1\,km$. For each of the grid cells, time series analysis has been applied in order to eliminate data gaps and outliers. In the framework of this study the harmonic analysis has been used. The method of the harmonic analysis is based on the method of superposition such as the Fourier transformation. This method (Bittner (1993)) is used to process LAI time series at the German Remote Sensing Data Center.

The CYCLOPES dataset additionally contains information of land cover and land use and is available as GLC2000 (valid for the year 2000). The Land Cover Classification System of the Food and Agriculture Organization of the United Nations has been used to derive land cover classes of GLC2000 resulting in 22 different land cover classes (Bartholome et al. (2002), DiGregorio and Jansen (2001)).

A translation of the GLC2000 vegetation classes had to be performed in order to use the GLC2000 land use / land cover classification to model NPP with BETHY/DLR. The actual model setup of BETHY/DLR includes 33 inherent vegetation classes which can be regarded as vegetation types. Each vegetation type is linked with biochemical parameters as i.e. the maximum electron transport rate and the maximum carboxylation rate, and other vegetation specific parameters as maximum height and rooting depth. These parameters describe the mechanism of photosynthesis of vegetation. In this study only the GLC2000 class 16 "Cultivated and managed areas" has been used and translated to the BETHY/DLR vegetation type "arable land" as no further

detailed information about the land use (e.g. crop rotation) is available from the GLC2000.

In addition to the GLC2000 dataset the Coordinated Information on the European Environment (CORINE) 2000 land cover / land use classification has been used, to validate the GLC2000 dataset. The CORINE 2000 data was derived from LANDSAT satellite images and is also available for the year 2000 (Bossard et al. (2000)). The CORINE 2000 is available as raster datasets in spatial resolutions of $100\,\text{m} \times 100\,\text{m}$, $250\,\text{m} \times 250\,\text{m}$ and $1\,\text{km} \times 1\,\text{km}$ for 32 European countries, including Austria. For this study the dataset with resolution $100\,\text{m} \times 100\,\text{m}$ has been used. The CORINE 2000 provides information about 44 vegetation classes which had also to be translated to BETHY/DLR vegetation types. We assumed that only the CORINE 2000 class "Non-irrigated arable land" contains the needed information about agricultural land, since all other classes which are available for the Marchfeld region report different land use (e.g. forests and urban areas). The CORINE 2000 class "Non-irrigated arable land" is then translated to the BETHY/DLR class "arable land".

5.5 From Yield to NPP

The crop yields of EPIC for the thirteen crops in the Marchfeld region have been converted to NPP values (table 5.1) for comparison with the BETHY/DLR outputs, which are given as time series of NPP. For this purpose, conversion factors of the relation between yield and straw as well as the above- and below-ground biomass are used. Empirical conversion factors about the relations between crop yield and straw yield can be found in e.g. Köhler and Kolbe (2007), Kaltschmitt and Hartmann (2001). In a first step, the above-ground biomass is computed for each crop using these empirical conversion factors. In a second step the below-ground biomass is computed with the use of conversion factors about the ratio of above- to below- ground biomass which are described in Bolinder et al. (1997). These conversion factors which originally have been derived for crops in Canada are assumed to be valid for the area of interest as well, as it already was proposed by Tum and Günther (2011). After calculating the biomass of the whole plant, the remaining water content and the non carbon content have to be subtracted, following crop specific values, which are

also reported in e.g. Köhler and Kolbe (2007). A detailed description of the approach and the used factors can be found in Tum and Günther (2011).

In order to compare the now available NPP per crop and sub-cluster of EPIC with the BETHY/DLR results, statistical data about the land use of each of the four sub-regions is used to aggregate the NPP of EPIC. These statistical data provided by Schmid et al. (2004) and Schmid et al. (2007) give detailed information about the distribution of agricultural area over the thirteen main crops as well as the distribution of the five main soils being representative for the Marchfeld region. The results of BETHY/DLR have been aggregated to annual sums per sub-region with a Geographic Information System (GIS) tool, taking into account the equi-rectangular projection (latitude - longitude, WGS84 (World Geodetic System 1984) of the data.

5.6   Results and Discussion

The variability analysis consists of seven model setups to compare model response to different input datasets. Three model simulations with the EPIC model have been performed and four with the BETHY/DLR model. The model setups are presented in table 5.3.

Table 5.3: *Model setups for the variability analysis.*

Model	Meteorological input	Land cover classification	Short Name
BETHY/DLR	ZAMG	CORINE 2000	BETHY(1)
	ZAMG reallocated	CORINE 2000	BETHY(2)
	ECMWF	CORINE 2000	BETHY(3)
	ECMWF	GLC2000	BETHY(4)
EPIC	ZAMG	-	EPIC(1)
	ZAMG reallocated	-	EPIC(2)
	ECMWF	-	EPIC(3)

The EPIC model requires homogeneity with respect to data input (i.e. soil, topography, weather, crop management) such that the model has been applied for all combinations of climate, soil, and crop management, separately.

Thus, the variability analysis has been conducted mainly for the meteorological datasets. In total 60 different model runs have been performed with EPIC for each crop. In contrast, the BETHY/DLR model is driven with the two different land cover classifications as well as the three different meteorological input data sets. For the Marchfeld region the FAO soil map of the world, which is used as input data for BETHY/DLR, reports one major soil type (Haplic Chermozem) which occupies 89 % of the area and four additional soil types for the rest of the area. The EPIC model setup EPIC(1) is interpreted as reference, as it represents the already validated model setup (Schmid et al. (2004)). In figure 5.2, all model results fare compared to the EPIC(1) results (table 5.3). The values of NPP are given in kilotonnes carbon per sub-region and year.

Depending on the model setup, the NPP results of BETHY/DLR show a variability of overestimations of up to 32 % and underestimations of up to 12 %, linked with coefficients of determination between 0.94 and 0.63, respectively. The highest overestimation of NPP (32 %) is modelled when using the GLC2000 and meteorological input data from ECMWF (figure 5.2D). Figure 5.2D represents the results of both models with the typical setup which was used in previous investigations (default setup). This overestimation is combined with a high coefficient of determination of about 0.94. When changing the land cover classification from GLC2000 to CORINE 2000 (while the meteorological input remains unchanged) an underestimation of about 12 % has been found (figure 5.2C). From figure 5.2C it is evident that only 4 BETHY/DLR results determine the underestimation and thus the coefficient of determination of about 0.77. These four data points are all representative for sub-region 4, whereas the rest is close to the 45° line. Using measured meteorological data from ZAMG results in an overestimation of NPP of about 11 % (figure 5.2A), which is combined with the highest variability within the sub-regions and years for all four model setups of BETHY/DLR. Nevertheless a high coefficient of determination of about 0.68 is achieved. When using the reallocated ZAMG data of Strauss et al. (2012) for BETHY/DLR combined with CORINE as land cover an overestimation of the modelled NPP of about 15 % (figure 5.2B) has been found. A strong correlation of the simulation years is observed, which indicates homogeneity in the meteorological data.
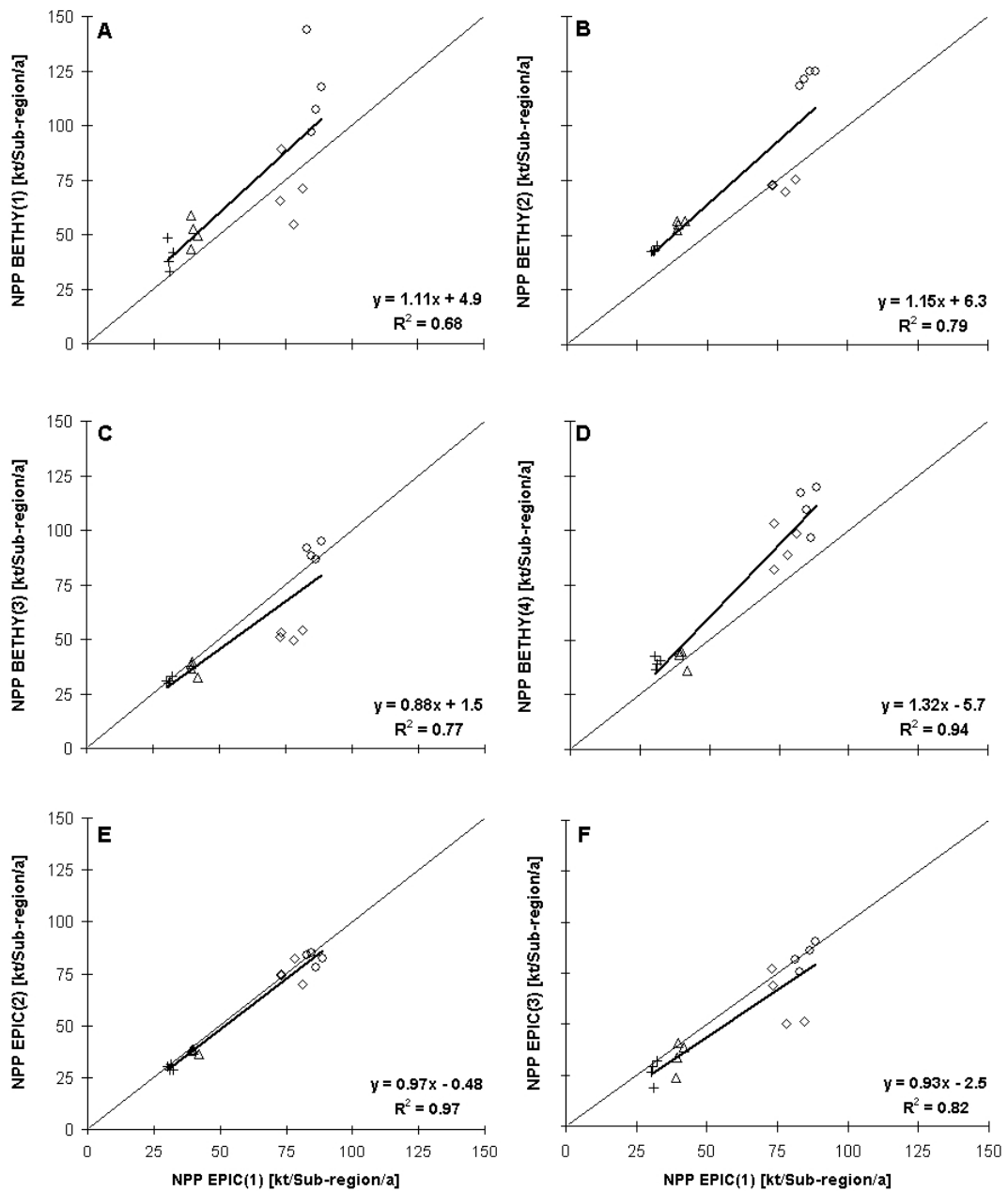
Figure 5.2: *Comparison of the model results (NPP) of BETHY/DLR and EPIC for the four Marchfeld sub-regions and the period 2000 to 2003. The nomenclature follows the scheme of table 5.3. Circles represent sub-region 1, triangles sub-region 2, crosses sub-region 3 and diamonds sub-region 4.*

The comparison between EPIC results with different weather input reveals that the ECMWF data affects the EPIC model to underestimate NPP by 8 % (figure 5.2F). The use of the reallocated meteorological dataset (figure 5.2E) results in a little underestimation, linked with the highest coefficient of determination (0.97). Figure 5.2E demonstrates that EPIC is not very sensitive to measured or homogenized meteorological input data just in contrast to BETHY/DLR which can be seen in figure 5.2A and 5.2B. Measured meteorological data during the four years result in a high variability of the annual NPP of sub-region 1 and 4 while the homogenized meteorological data cluster the annual NPP of all sub-regions resulting in low variability for all sub-regions.

Figure 5.2D and figure 5.2F show that the EPIC model as well as the BETHY/ DLR model react in a similar way when alternating between ECMWF and ZAMG data. The BETHY/DLR model simulates 23 % more NPP when using the ZAMG data, and the EPIC model simulates around 7 % more NPP when using the ZAMG data.

A reason for investigating the influence of different land cover classifications (GLC2000 versus CORINE 2000) is the higher spatial resolution of CORINE 2000. It is expected that CORINE 2000 represents the small scale land use structure of the Marchfeld region better than the GLC2000 classification. In figure 5.3 the agricultural areas reported in the statistical source (Schmid et al. (2004), Schmid et al. (2007)), the GLC2000 and CORINE 2000 are presented for all four Marchfeld sub-regions.
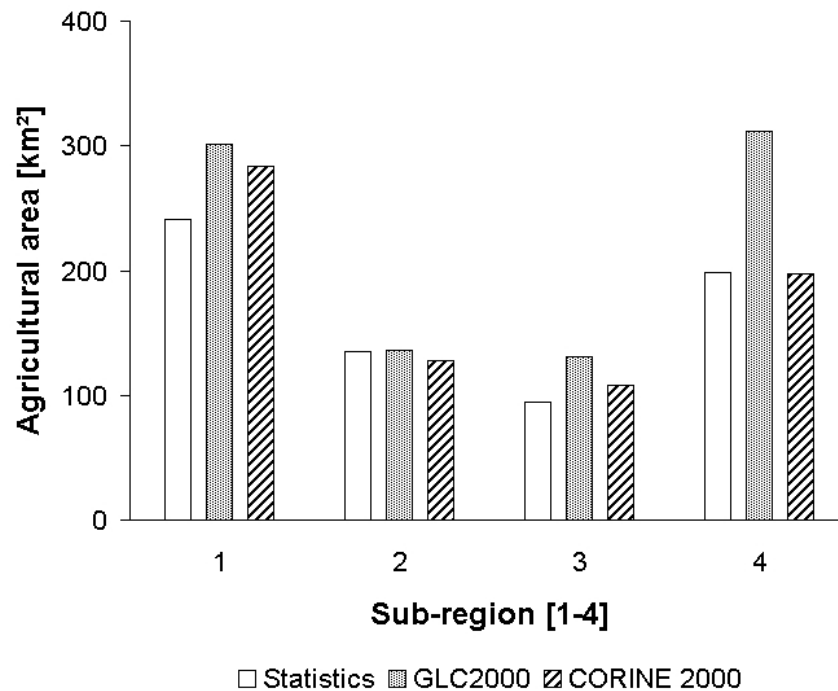
Figure 5.3: *Comparison of agricultural areas described by statistical sources (Schmid et al. (2004), Schmid et al. (2007)), GLC2000 and CORINE 2000 in km^2 for the four sub-regions of the Marchfeld region.*

The agricultural areas presented in GLC2000 and CORINE 2000 have been computed using GIS tools. As shown in figure 5.3, the GLC2000 considerably overestimates the agricultural areas (sub-regions one, three and four) by 25 % to 57 % compared to the statistical information. On the other hand, CORINE 2000 slightly over- (17 %) or underestimates (6 %) the agricultural areas compared to the statistical sources. However, approximately the same agricultural area is found for sub-region two for each land cover classification. For all sub-regions of the Marchfeld region the statistical data report an agricultural area of around $670\,km^2$, GLC2000 of $881\,km^2$, and CORINE 2000 of $718\,km^2$. As the difference in agricultural area between CORINE 2000 and the statistical data is smaller than the difference between GLC2000 and the statistical data, we conclude that the CORINE 2000 land cover represents the real situation more precisely than GLC2000. The differences of the results described in figure 5.2D and figure 5.2C showing an NPP decrease when changing from GLC2000 to CORINE 2000 can thus be explained by the fact that the BETHY/DLR
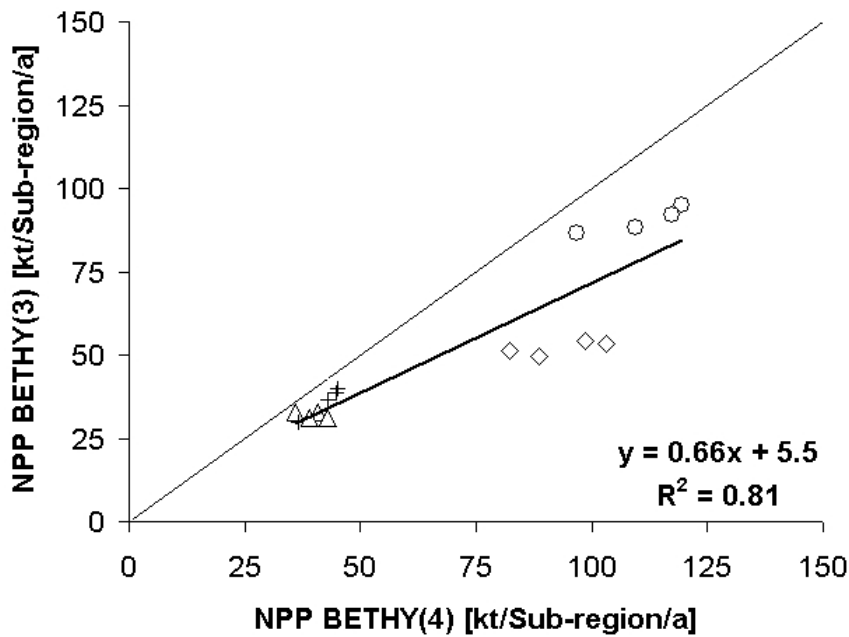
Figure 5.4: *Comparison of the model results (NPP) of the BETHY/DLR runs BETHY(3) and BETHY(4) for the four Marchfeld sub-regions and the period 2000 to 2003. The nomenclature follows the scheme of table 5.3. Circles represent sub-region 1, triangles sub-region 2, crosses sub-region 3 and diamonds sub-region 4.*

model was driven for a smaller agricultural area.

To proof this, the results for BETHY(3) and BETHY(4) are presented in figure 5.4 as a linear correlation. For both model setups meteorology was fix (ECMWF), but the land cover classification was changed. With this direct comparison it becomes clear that the reason for the highly different model results presented in figure 5.2C and figure 5.2D lays in the uncertainty in the two land covers.

When comparing the ECMWF data with the measured ZAMG data it is obvious that the ECMWF data underestimates the maximum and minimum temperature (see figure 5.5). The comparison of daily weather measurements is conducted for two of the 34 ZAMG weather stations which are situated closest to the Marchfeld (Schwechat and Gross Enzersdorf) and for the time period 2000 to 2003.

For both stations the maximum temperature of the ECMWF data is underestimated by about 21 % expressed by a high coefficient of correlation of up to
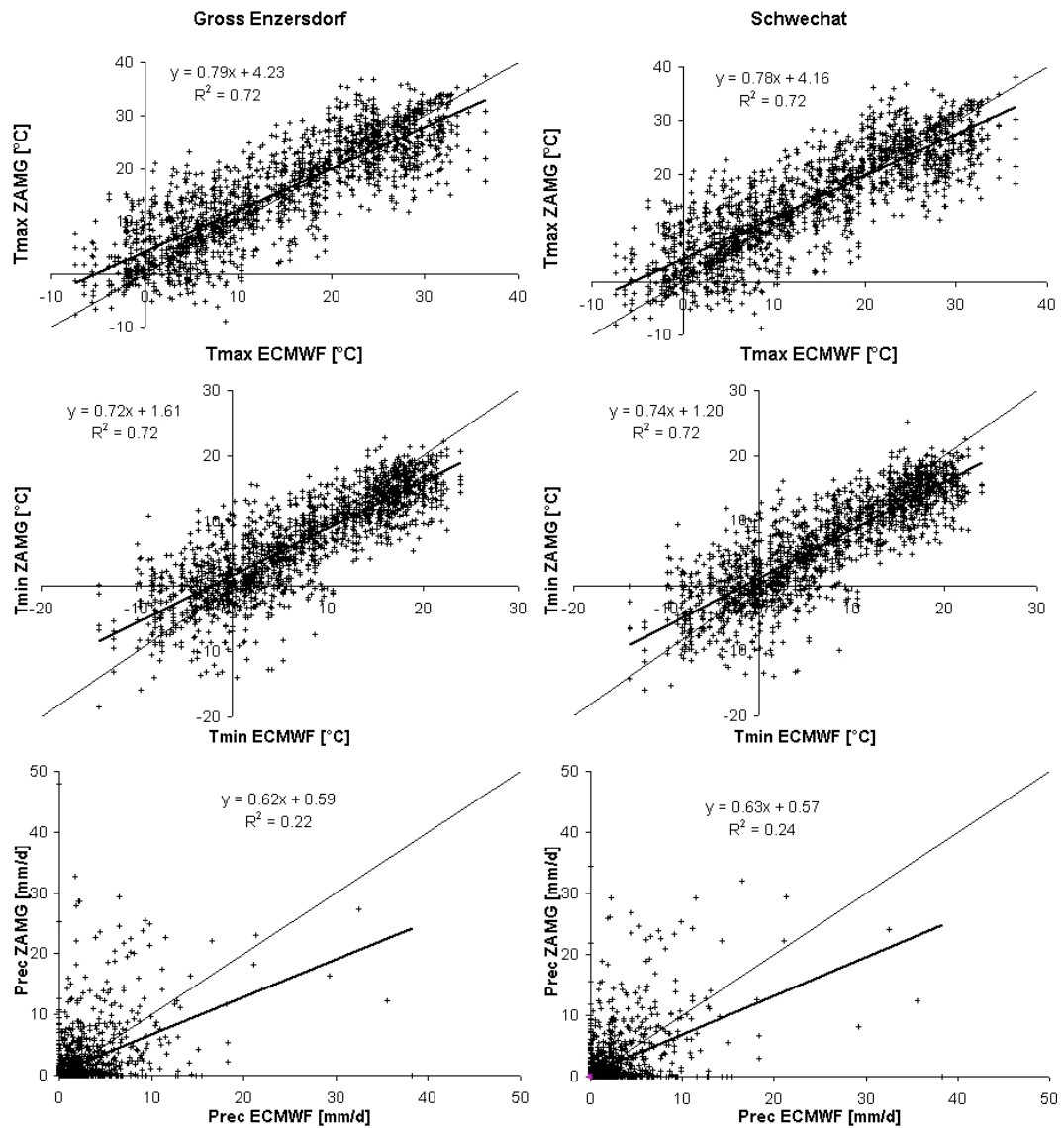
Figure 5.5: *Comparison of the ECMWF time series of minimum and maximum temperature as well as precipitation with the corresponding daily measured data of the ZAMG stations Gross Enzersdorf and Schwechat in the period 2000 to 2003.*

0.72. The minimum temperature is underestimated even slightly higher (up to 28 %) but again combined with a high coefficient of correlation (up to 0.74). In contrast, precipitation is not represented very well by ECMWF data as the correlation for the precipitation reveals high uncertainties. Hence a comparison of the ECMWF data for only two measurement stations is not very meaningful. Therefore the analysis has been expanded to all of the 34 available ZAMG weather stations. The analysis shows that the mean maximum and minimum temperatures of ECMWF data averaged over daily values in the period 2000 to 2003 are about 24 % and 29 % lower, respectively, than the temperatures recorded by the 34 ZAMG weather stations. However, minimum and maximum temperatures are both linked with a coefficient of determination of about 0.65, which is in good correspondence with the two presented observation stations in figure 5.5. The comparison between sums of annual precipitation between the ECMWF and the ZAMG data reveals over- and underestimations of up to 90 % for single stations. The daily precipitation rates averaged over all ZAMG observation stations show a coefficient of determination of about 0.27. This very low coefficient corresponds with the presented stations in figure 5.5 and indicates poor agreement of measured and simulated precipitation.

As ECMWF data significantly underestimate temperature, the increase of NPP when using ZAMG data could be explained by longer vegetation periods in the ZAMG data. We investigated the vegetation period by computing the growing-degree-days (GDD). The basic equation is:

$GDD = [(T_{MAX} + T_{MIN})/2] - T_{Base}$, where TMAX and TMIN are daily maximum and minimum temperatures, respectively and TBASE is the base temperature which can be fixed with at $10\,°C$ McMaster and Wilhelm (1997). Furthermore, the growing period in Austria is assumed to be from mid March to mid October. The mean GDD averaged over all 34 ZAMG stations in Austria and the years 2000 to 2003 is about 1186.2, which is about 136.1 ($\approx$11.5 %) more than the corresponding ECMWF GDD value (1050.1). In a third model setup both models are driven with the reallocated ZAMG data to test the model response to homogenized trend data. Figure 5.2B and figure 5.2E show that both, the EPIC and the BETHY/DLR models respond in a consistent way, concerning their annual variability, to the reallocated ZAMG data. The variability in the NPP over the four years within one sub-region is about 4 %

(EPIC) and 3 % (BETHY/DLR), respectively.

To give information about the annual variability of NPP within the model results, annual sums of NPP over the whole area of investigation are presented in figure 5.6. The values are given in kilotonnes carbon per year.
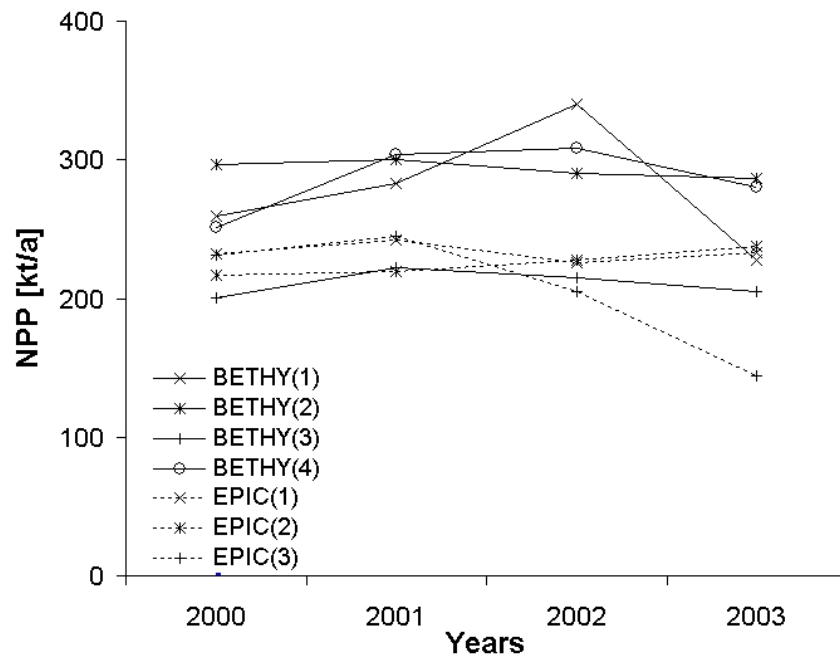

Figure 5.6: *Annual sums of NPP in kilotonnes carbon for the Marchfeld region in the period 2000 to 2003 simulated with the models BETHY/DLR and EPIC. The nomenclature follows the scheme of table 5.3.*

The nomenclature in figure 5.6 follows the scheme of table 5.1. When using the reallocated weather data, the annual variability of NPP is very low for both models BETHY(2) and EPIC(2), which can also be seen from figure 5.2B and figure 5.2E. This is not surprising since they represent trend data with lower inter-annual variability. When looking at the model setup for BETHY(1) with measured ZAMG data, BETHY/DLR strongly responds to the climate data. This is very prominent for the year 2003, for which a water stress situation for the Marchfeld region is reported StartClim (2004). In comparison to the NPP sum calculated for 2002, the annual NPP in 2003 is lower by about 23 %. However, this model response cannot be seen in the EPIC output, which might be due to the reason that for one of the four climate clusters, which is representative for most of the area of the Marchfeld region, higher crop yields have

been simulated especially for winter crops in 2003. With the use of ECMWF data in model setup EPIC(3), the EPIC output shows a massive NPP decrease in 2003 compared to 2002. This could again be explained with the lower GDD of the ECMWF. In addition, the ECMWF data represent around 8.5 % less precipitation over the days which have been counted as GDD. The reason for the non equidistant annual differences between the BETHY/DLR model runs BETHY(3) and BETHY(4) might be that the misclassified pixels of GLC2000 represent non agricultural areas which react in different ways to climate conditions than agricultural areas.

It is notable that the variability of the model outputs can be as large as 36 % for BETHY/DLR and 39 % for EPIC when changing major input datasets. Furthermore, it is remarkable that both models response similarly when using the same datasets. For instance, all three model setups with the ECMWF data show for all four sub-regions a relative increase of NPP from 2000 to 2001 followed by a decrease in 2002 and again in 2003.

5.7 Conclusion

Net-Primary-Productivity (NPP) has been modelled using the SVAT model BETHY/DLR and the biophysical process model EPIC for the Austrian March-feld region and the period 2000 to 2003. Both models seem to be robust but respond differently on alternative input datasets (i.e. meteorological and land cover /use data). We have used meteorological data from the ECMWF and the ZAMG as well as a reallocated dataset based on ZAMG weather observations. Land cover / land use information have been taken from the GLC2000 and the CORINE 2000 products. With these datasets, we have performed a variability analysis with the two models BETHY/DLR and EPIC with respect to their output responses. We show that lower NPP values were modelled when using ECMWF data as an input compared to ZAMG data. This is confirmed by both models. The reason is traced to the underestimation of the GDD of about 11.5 % in the ECMWF data. We observe that both models response similarly to changes in input data, albeit with a different magnitude. For single years, variabilities in the NPP of up to 36 % for BETHY/DLR and of up to 39 % for EPIC can occure with alternative input data.

Besides the variability analysis of alternative model input data sources, we have also analysed the accuracy of the input data. We have found that the GLC2000 land cover classification overestimates the agricultural area of the Marchfeld region by 24 %, whereas the CORINE 2000 dataset overestimates land cover classification by only 7 %. With this finding preference for land cover datasets with higher resolution is recommended. The ECMWF data has been compared with measured data from ZAMG. We have found high uncertainties in the daily precipitation and small ones in daily maximum and minimum temperatures, which is confirmed by other studies.

For further investigations in other regions, the finding of the bias in the ECMWF data should be taken into account and crosschecked with local weather station data. In addition, more detailed land cover products should be considered with respect to spatial resolution and reported land use practices. Thus whenever the models (or any model) are applied under new conditions, local data (if applicable) should be used for both input and result comparison.

This study shows that especially for process-based modelling approaches, not only comprehensive validation and calibration approaches need to be applied, but also knowledge of input data uncertainty and variability of the modelling results need to be assessed. Process-based models have a potentially valuable role for various applications. However their validity must be determined where possible, especially when used for decision making processes.

5.8 Acknowledgements

112

this study. Special thanks are given to Dr Georgii Alexandrov for inspiring suggestions during the review process.

Chapter 6

A Conceptual Remote Sensing based Interception-Infiltration Model for Regional and Global Applications

Markus Tum, Erik Borg, *Hydrology and Earth System Sciences*, under review

6.1 Abstract

We present a remote sensing driven modelling approach to simulate the one dimensional water transport in the vadose zone of unsaturated soils on a daily basis, which can be used for regional to global applications. Our model needs van Genuchten parameters to calculate the hydraulic conductivity, which we estimated using the ISRIC-WISE Harmonized Global Soil Profile Dataset Ver. 3.1 and the Rosetta programme. We calculated all needed parameters for 26 global main soil types and 102 soils of second order, which are based on the original, global FAO 1974 soil classification. Soil depth and the layering of one to six layers were defined for each soil. The parameters for the main soils are presented in this paper. Interception by vegetation is also considered using remote sensing calculated Leaf Area Index (LAI) time series from SPOT-VEGETATION. Precipitation is based on daily time series from the European Centre for Medium-Range Weather Forecasts (ECMWF). For Germany we

compared our model output with soil moisture data from the ECMWF, which is based on the same precipitation dataset. We found a good agreement for the general characteristics of our modelled plant available soil water with this dataset, especially for soils which are close to the standard characteristics of the ECMWF. Disagreements were found for soils under stagnant moisture and for shallow soils, which are not considered in the ECMWF model scheme, but can be distinguished with our approach. The proposed approach for combining established model formulations for interception and one-dimensional vertical water transport with time-series of remote sensing data intends to contribute to the realistic parameterization of the soil water budged. This is especially needed for the global and regional assessment of e.g. net primary productivity which can be calculated with vegetation models.

6.2   Introduction

The prediction of hydrodynamics in unsaturated soils remains a challenging task in the topic of soil physics and is important for modelling physical processes which are related to the soil water content. During the last years the development of models capable to simulate the water flow in soils has gained an important role. In this context, computer models based on the numerical solution of Richards' equation has proved as being valuable. Their application is often restricted by a lack of hydraulic property information involving the soil water retention curve (SWRC) and the unsaturated hydraulic conductivity. For modelling the SWRC many diverse empirical approaches can be found (e.g.: Gardner (1958), Brooks and Corey (1964), Campbell (1974), van Genuchten (1980), Hutson and Cass (1987), Russo (1988)). Usually many input parameters are required to describe the soil processes. Due to inherent temporal and spatial variability of hydraulic properties in nature, large numbers of samples are generally required to properly characterize the spatial distribution of these hydraulic properties. Therefore, direct measurements are time-consuming and expensive. In contrast, indirect methods are increasingly used to provide estimates. Presuming Richards' equation can be applied, the most crucial point is the exact measurement and description of hydraulic properties, or to be more precise: the soil water retention curve $\Theta(h)$ and the

hydraulic conductivity function $k(\Theta)$, where Θ is the volumetric water content, h the pressure head and k the hydraulic conductivity. To solve this problem diverse pedotransfer functions (PTFs) have been developed (e.g. Vereecken et al. (1989), Wosten (1997), Mayer and Jarvis (1999), Minasny and McBratney (2000), Schaap et al. (2001), Jarvis et al. (2000), Tomasella et al. (2003), Weynants et al. (2009), Wessolek et al. (2011)). A comprehensive overview of developed pedotranfer functions in soil hydrology can be found in Pachepsky and Rawls (2005),allowing a good understanding of state-of-the-art modelling approaches with respect to advantages and restrictions of pedotransfer functions to predict hydrological soil properties.

To couple these complex physics of water transport in soil with the atmosphere and vegetation an interception model has to be considered. The capacity of vegetation to intercept water is of great importance, since the rate of evaporation from wet canopy is higher than from dry canopy conditions (Stewart (1977)).Thus rainfall interception and its following rainfall evaporation may result in a net loss to the system, but depending on the surrounding conditions (e.g.: coastal or mountain fog belts), could also lead to a net gain (Bruijnzeel (1990)).As a consequence the presence or absence of vegetation strongly affects the amount of rainfall reaching the soil surface. Model formulations to describe interception of vegetation have been developed (e.g.: Rutter et al. (1971), Gash (1979), Massman (1983), Xiao et al. (2000)) which use parameters to describe the threshold amount of rain that can be stores in the canopy and a descriptive parameter for the canopy structure. The Braden (1985) and Calder (1986) models are further examples which use information of Leaf Area Index (LAI) to describe the canopy instead of a descriptive parameter. A review about approaches to model interception was recently published by Muzylo et al. (2009).

The second interaction of vegetation with soil water, namely water suction via roots and its following evapotranspiration by vegetation is a further influencing factor to the available soil water content. However in this study we will not focus on the description of this process.

The primary objective of this study is to introduce our soil water transport model, which is suitable to calculate the soil water balance on a regional to global scale. It is driven by van Genuchten parameters and remote sensing

data and distinguishes 128 soils, which follow the system of the original FAO-UNESCO legend (FAO (1974)). We compared our model results with modelled time series taken from the European Center for Medium Range Weather Forecast (ECMWF), which are based on the same precipitation dataset we used. With our approach to combine established model formulations with the use of remote sensing data, we see the potential in our model to be applied in remote sensing based vegetation models. In vegetation models the realistic parameterisation of the soil water budged is a challenging task, but of major interest, since it is usually very simplified.

6.3   Theoretical background

In our model the water balance is considered regarding the two reservoirs which influence the water availability and is affected by vegetation: soil water and intercepted water on leaves and other parts of vegetation. These reservoirs change in time and space depending on precipitation, temperature and evapotranspiration. Evaporation from soil is calculated daily following the approach of Ritchie (1972). Transpiration is not considered in our model formulation. The processes of interception and percolation, on which we focused on in this study will be discussed in more detail in the next sections.

Interception

Interception (P_i) is considered following the concept of Braden (1995). The general assumption of this approach is that P_i is empirically related to the Leaf Area Index (LAI, Λ), the fraction of canopy closure (f_c) and the precipitation sum (P):

$$P_i = f_c[1 - (1 + b_0 \left(P - P_{sn}\right)/f_c)^{-1}] \tag{6.1}$$

where b_0 is the fraction of soil covered by plants and P_{sn} the share of precipitation which reaches the ground as snow, which is not considered as available for soil infiltration. Snowfall is calculated as linear function according to Wigmosta et al. (1994), falling linearly from P to 0 between -1.1 C and 3.3 C. Since

b_0 and f_c are both related to Λ one might express b_0 and f_c following Eq.: 6.2 and Eq.: 6.3:

$$b_0 = 1 - \exp(-0.5\Lambda) \tag{6.2}$$

$$f_c = \frac{\Lambda}{\Lambda_{lim} f_{cmax}} \tag{6.3}$$

where Λ_{lim} is the limiting LAI, which is set to 3 assuming LAIs greater than 3 do not influence the canopy fraction by raising it. f_{cmax} represents the maximum share of canopy fraction and is set to 0.9, assuming the maximum fraction of cover per grid cannot be higher than 90 %.

Evaporation from canopy is calculated taking into account both: the dropping loss of water from leafs (P_d) the disposition of water in the skin reservoir (W_s), but is limited by the evaporation from canopy (E_c):

$$P_d = W_s + P_i - 0.1\Lambda \; with \; P_d \geq 0 \tag{6.4}$$

E_c is calculated following the scheme of Philip (1957), assuming limitation at the maximum potential evaporation rate (E_{cpot}):

$$E_c = \begin{cases} P_i - P_d + W_s & with \; E_c < E_{cpot} \\ E_{cpot} & with \; E_c \geq E_{cpot} \end{cases} \tag{6.5}$$

The skin reservoir is considered as a cumulative reservoir which is filled and drained over time by precipitation, throughfall and evaporation from the canopy. We assume the skin reservoir to be empty at time step $j = 0$:

$$W_s^j = \begin{cases} W_s^{j-1} + P_i - P_d - E_c & with \; W_s^j < f_c, \; j \geq 1 \\ f_c & with \; W_s^j \geq f_c, \; j \geq 1 \end{cases} \tag{6.6}$$

Finally the ground reaching precipitation without being interfered by vegetation i.e. throughfall (P_t) can be expressed as:

$$P_t = \begin{cases} P_d + P - P_i & with W_s < f_c \\ P_d + P - P_i + (W_s^{j-1} + P_i - P_d - E_c - f_c) & with W_s = f_c \end{cases} \tag{6.7}$$

Infiltration

The process of water, penetrating the soil surface is defined as infiltration. It has a dominant role among the components of hydrological processes of catchment areas (Dyck et al. (1980)). We base our approach on an adaption of the one-dimensional vertical transport algorithm described by Syring and Kersebaum (1988), which is based on the theoretical approach of Dyck et al. (1980) and Anlauf et al. (1987). The vertical transport within the soil is calculated using a combination of the Darcy-equation (Eq. 6.8) with the local balance (or continuity) equation (Eq. 6.9).

$$q_w = -k \left(\frac{d\psi}{dz} - 1 \right) \tag{6.8}$$

$$\frac{\partial \Theta}{\partial t} = -\frac{\partial q}{\partial z} + A\left(t, z\right) \tag{6.9}$$

Here q_w represents the water flux from a layer to its subjacent layer. k is the hydraulic conductivity of the soil water flux, which is dependant on the matric potential ψ and the depth of layer z. The local balance equation describes the relation of time (t) dependant volumetric soil water content (Θ) and soil layer depth depending water flux. Parameter A represents the source or sink term as function of depth and time. When combining Eq. 6.8 and Eq. 6.9 the problem can be described by Richard's-equation:

$$\frac{\partial \Theta}{\partial t} = -\frac{\partial[k\left(\frac{\partial \psi}{\partial z} - 1\right)]}{\partial z} + A\left(t, z\right) \tag{6.10}$$

The matric potential depending on volumetric soil water content Θ_ψ and hydraulic conductivity k_ψ can be calculated following the approach proposed by van Genuchten (1980):

$$\Theta_\psi = \Theta_r + \left(\Theta_s - \Theta_r\right)\left[1 + \left(\alpha|\psi|\right)^n\right]^{\left(1 - \frac{1}{n}\right)} \tag{6.11}$$

$$k_\psi = k_s \frac{\left(1 - \left(\alpha|\psi|\right)^{n-1} \left(1 + \left(\alpha|\psi|\right)^n\right)^{\left(1 - \frac{1}{n}\right)}\right)^2}{\left(1 + \left(\left(\alpha|\psi|\right)^n\right)^{\frac{\left(1 - \frac{1}{n}\right)}{2}}\right)} \tag{6.12}$$

where Θ_r is the volumetric soil water content at the permanent wilting point (PWP), Θ_s, the volumetric soil water content and k_s the hydraulic conductivity at saturation. The parameters α and n represent two form parameters which are needed for the van Genuchten parameters. Since α, n, k_2, Θ_r and Θ_s are highly empirically derived parameters it is discussed as challenging to derive van Genuchten parameters on a broader scale Schaap et al. (2001) We used soil core measurements and the ROSETTA program to define these parameters, which will be discussed in more detail later. Because Θ_ψ and k_ψ are functions of the matrix potential ψ Eq. 10 can be used to determine ψ. Equation 10 itself is a differential equation of second order.

In a first step the starting conditions of the upper and lower boundary have to be defined. At its upper boundary the whole amount of throughfall can infiltrate until the soil dependant saturated condition is reached. The top layer of each soil type is set to $3\,\mathrm{cm}$, since it is assumed that only the upper $3\,\mathrm{cm}$ can directly react to precipitation and thus evaporate water. At saturation ψ is set to 0:

$$z = 0, t > 0 : \; P_t = k \left(\frac{\partial \psi}{\partial z} - 1 \right) \; and \; \psi \leq 0 \tag{6.13}$$

At the lower boundary ψ_L is set to zero in the case that a soil under stagnant moisture condition as for instance gleyic soils is modelled. Otherwise ψ_L is set to -15,000 to simulate a soil layer with dry conditions and to respect water run off processes, or to be more precise: to allow water to leave the system.

In a second step for the layers (i) between the upper and lower boundary, with the specific thickness Δz, the water flux q_i is calculated:

$$q_i = k_{i+0.5} \left(\frac{(\psi_{i+1} - \psi_i)}{\Delta z} - 1 \right) \tag{6.14}$$

To calculate the hydraulic conductivity of the inner compartments we followed the approach of Syring and Kersebaum (1988), who used the arithmetic mean of the two surrounded compartments are considered:

$$k_{i+0.5} = 0.5[k\left(\psi_i\right) + k\left(\psi_{i+1}\right)] \tag{6.15}$$

Thus for each inner layer n at time step t, ψ can be expressed as:

$$\psi_t = \psi_{t-1} - 0.5\Delta_z \left(1 - \frac{q_n}{k_{0.5}}\right) \; with \; \psi_t \leq 0 \qquad (6.16)$$

To solve this equation a variation of the Newton-algorithm (Remson et al. (1971)) was considered to calculate the function f of ψ:

$$f\left(\psi_i\right) = \frac{\left(\Theta_i^j - \Theta_i^{j-1}\right)}{\Delta t} + \frac{\left(\Theta_i^j - \Theta_{i-1}^j\right)}{\Delta z} - A_t, z \qquad (6.17)$$

Following this one may approximate ψ as:

$$\psi_i^* \approx \psi_i - \frac{f\left(\psi_i\right)}{\left(\frac{\partial f(\psi_i)}{f\psi_i}\right)} \qquad (6.18)$$

6.4   Input data

van Genuchten Parameter

For our modelling approach we used, as described above, van Genuchten parameters, which were estimated by using the Rosetta program (Schaap et al. (2001)). Rosetta contains a neuronal network to predict van Genuchten parameters which are based on estimates on grain size distribution (sand, silt and clay content) of a soil. The original FAO legend distinguishes 26 main soil types and 102 soils of second order. In order to estimate the mean grain size distributions for the 128 FAO soils we used the ISRIC-WISE Harmonized Global Soil Profile Dataset Ver. 3.1 Batjes (2009)), which contains data of 10,253 soil profiles and is classified following the system of original and revised FAO-UNESCO legends (FAO (1974); FAO (1988)).

Since for the reported soils individual measurements had wide ranges of total layers (2-12 layers) we decided to calculate - in a first step - the median of reported soil layers for each soil type. This was done to minimize the complexity of soils and to minimize the computational effort. In a second step we calculated the average grain size distribution and layer depth for the soil profiles which were selected in the first step. Other reported soil profiles with

more or less layers than the median were not used for the calculation. For the main soils these values and the correspondent van Genuchten parameters are presented in 6.1.

To obtain spatial information for the global soil type distribution we used the Harmonized World Soil Database (HWSD) provided by the International Institute for Applied Systems Analysis (IIASA). The HWSD is freely available as grid with 30 arc seconds resolution in a latitude-longitude projection using the WGS84 (World Geodetic System 1984) datum (FAO/IIASA (2009)). It contains information about the dominant soil type, following the systems of the FAO from 1974, 1985 and 1990, depending on the location on earth. Since we estimated our van Genuchten parameters for the FAO '74 soil classification we harmonized the HWSD dataset to this classification scheme, by transforming newer classifications to the 74' standard.

Table 6.1: *Soil properties for 26 FAO '74 main soils, including total soil depth and layering (d), sand (sa), silt (si) and clay (cl) content permanent wilting point (Θ_r), field capcity (Θ_s), van Genuchten parameters (α,n) and hydraulic conductivity at saturation (k_s).*

Soil	d	sa	si	cl	Θ_r	Θ_s	α	n	k_s
	$[cm]$	$[\%]$	$[\%]$	$[\%]$	$[\frac{cm^3}{cm^3}]$	$[\frac{cm^3}{cm^3}]$	$[cm^{-1}]$	$[-]$	$[\frac{cm}{d}]$
Acrisol	145								
1	13	58	23	19	0.058	0.389	0.024	1.369	18.9
2	20	54	21	25	0.068	0.395	0.023	1.340	11.9
3	28	48	20	32	0.076	0.406	0.023	1.301	7.6
4	41	43	19	37	0.082	0.419	0.023	1.280	7.1
5	43	44	20	36	0.081	0.416	0.023	1.287	6.8
Cambisol	134								
1	15	36	38	26	0.073	0.421	0.011	1.462	8.2
2	17	32	40	28	0.077	0.431	0.010	1.469	10.5
3	24	33	41	26	0.074	0.426	0.009	1.489	11.3
4	33	35	38	26	0.074	0.425	0.011	1.458	8.3
5	39	37	39	24	0.070	0.417	0.010	1.478	8.8
Chernozem	169								
1	31	16	52	31	0.086	0.458	0.008	1.514	12.6
2	26	16	53	31	0.085	0.461	0.008	1.516	12.2
3	24	16	54	30	0.084	0.458	0.007	1.528	12.2
4	35	17	55	28	0.081	0.452	0.007	1.550	12.3
5	53	20	54	26	0.078	0.443	0.006	1.566	12.7
Podzoluvisol	169								
1	19	40	50	10	0.045	0.405	0.007	1.590	33.3
2	14	40	52	8	0.041	0.410	0.007	1.603	44.5
3	35	40	44	16	0.056	0.401	0.008	1.538	15.6
4	54	43	38	20	0.062	0.406	0.011	1.477	9.2
5	47	49	34	18	0.058	0.398	0.015	1.436	13.6
Rendzina	32								
1	32	48	31	21	0.063	0.400	0.016	1.414	11.6

| Soil | d | sa | si | cl | Θ_r | Θ_s | α | n | k_s |
	$[cm]$	$[\%]$	$[\%]$	$[\%]$	$[\frac{cm^3}{cm^3}]$	$[\frac{cm^3}{cm^3}]$	$[cm^{-1}]$	$[-]$	$[\frac{cm}{d}]$
Ferralsol	165								
1	14	45	19	36	0.080	0.414	0.023	1.282	7.2
2	21	42	18	40	0.084	0.423	0.024	1.260	8.1
3	31	39	18	44	0.087	0.435	0.025	1.242	9.1
4	48	37	17	46	0.089	0.438	0.025	1.232	10.3
5	41	37	17	46	0.089	0.438	0.025	1.232	10.3
Gleysol	122								
1	16	39	32	28	0.076	0.421	0.014	1.400	5.84
2	25	38	29	32	0.080	0.426	0.016	1.358	5.18
3	35	34	31	35	0.083	0.435	0.015	1.362	5.76
4	46	38	27	35	0.082	0.427	0.018	1.334	5.06
Phaeozem	142								
1	22	31	40	29	0.078	0.434	0.010	1.463	10.6
2	20	29	36	35	0.085	0.446	0.013	1.396	8.2
3	25	25	33	42	0.090	0.460	0.015	1.334	9.0
4	31	24	34	42	0.091	0.463	0.014	1.339	9.5
5	44	30	35	35	0.084	0.444	0.013	1.389	7.6
Lithosol	10								
1	10	75	11	14	0.052	0.378	0.032	1.451	42.8
Fluvisol	128								
1	19	38	38	24	0.070	0.415	0.011	1.469	8.0
2	27	35	38	28	0.076	0.429	0.011	1.450	8.1
3	36	36	37	27	0.074	0.423	0.011	1.449	7.5
4	43	41	33	26	0.072	0.414	0.014	1.419	6.7
Kastanozem	122								
1	18	31	46	23	0.070	0.422	0.007	1.543	13.9
2	19	29	45	26	0.075	0.431	0.008	1.519	12.8
3	23	18	40	42	0.094	0.475	0.013	1.370	13.4
4	24	20	46	34	0.087	0.461	0.009	1.464	12.3
5	38	25	49	26	0.076	0.436	0.007	1.544	12.9

Soil	d	sa	si	cl	Θ_r	Θ_s	α	n	k_s
	$[cm]$	$[\%]$	$[\%]$	$[\%]$	$[\frac{cm^3}{cm^3}]$	$[\frac{cm^3}{cm^3}]$	$[cm^{-1}]$	$[-]$	$[\frac{cm}{d}]$
Luvisol	119								
1	16	74	14	12	0.048	0.381	0.034	1.456	48.0
2	23	70	12	18	0.057	0.378	0.030	1.372	26.3
3	25	62	11	27	0.068	0.383	0.027	1.294	13.3
4	45	54	12	34	0.076	0.396	0.027	1.260	11.3
Greyzem	179								
1	23	21	55	24	0.074	0.437	0.006	1.588	13.2
2	22	15	58	27	0.081	0.452	0.006	1.566	12.3
3	42	15	52	33	0.088	0.467	0.009	1.494	12.4
4	33	15	54	32	0.087	0.467	0.008	1.501	12.0
5	59	20	50	30	0.083	0.453	0.008	1.516	12.3
Nitosol	149								
1	14	40	26	35	0.081	0.426	0.019	1.326	5.2
2	19	33	22	45	0.089	0.446	0.021	1.259	8.7
3	27	28	21	51	0.093	0.462	0.022	1.233	14.5
4	39	28	19	53	0.094	0.464	0.023	1.218	14.8
5	50	27	19	54	0.094	0.467	0.023	1.216	15.4
Histosol	150								
1	20	42	21	38	0.083	0.424	0.022	1.282	6.4
2	23	28	38	35	0.085	0.451	0.012	1.408	9.6
3	37	31	30	40	0.087	0.447	0.016	1.326	6.8
4	40	54	24	22	0.064	0.394	0.022	1.366	14.0
5	30	74	13	14	0.051	0.381	0.033	1.435	40.7
Podzol	134								
1	14	77	16	7	0.041	0.385	0.040	1.564	72.8
2	17	79	16	5	0.038	0.387	0.042	1.655	91.1
3	16	80	15	6	0.040	0.387	0.042	1.667	91.5
4	27	81	13	6	0.042	0.384	0.040	1.726	102.8
5	50	85	11	4	0.043	0.384	0.040	2.033	168.0

Soil	d	sa	si	cl	Θ_r	Θ_s	α	n	k_s
	$[cm]$	$[\%]$	$[\%]$	$[\%]$	$[\frac{cm^3}{cm^3}]$	$[\frac{cm^3}{cm^3}]$	$[cm^{-1}]$	$[-]$	$[\frac{cm}{d}]$
Arenosol	170								
1	19	88	7	6	0.050	0.380	0.035	2.203	213.7
2	24	88	7	6	0.050	0.380	0.035	2.203	213.7
3	34	88	6	6	0.051	0.377	0.034	2.236	223.0
4	93	86	8	7	0.049	0.380	0.035	1.992	157.3
Regosol	87								
1	20	41	49	10	0.045	0.403	0.007	1.578	31.1
2	33	42	49	9	0.042	0.404	0.007	1.572	33.6
3	34	46	44	10	0.043	0.397	0.010	1.516	25.7
Solonetz	140								
1	12	52	28	20	0.061	0.395	0.019	1.394	14.6
2	17	47	26	27	0.072	0.406	0.019	1.363	8.4
3	24	42	26	32	0.078	0.417	0.018	1.342	5.8
4	35	39	29	33	0.080	0.427	0.017	1.356	5.2
5	52	42	28	30	0.077	0.416	0.017	1.365	6.1
Andosol	151								
1	23	50	36	14	0.050	0.392	0.014	1.451	18.6
2	25	55	32	12	0.046	0.387	0.019	1.421	27.1
3	29	56	33	12	0.045	0.391	0.020	1.420	27.9
4	30	55	32	13	0.048	0.389	0.019	1.419	25.1
5	42	59	28	13	0.047	0.387	0.024	1.400	28.3
Ranker	35								
1	16	52	25	18	0.057	0.394	0.018	1.409	16.3
2	19	57	24	19	0.058	0.390	0.023	1.374	18.4
Vertisol	127								
1	15	27	25	48	0.092	0.462	0.019	1.263	12.5
2	31	24	20	55	0.095	0.472	0.021	1.219	17.1
3	35	23	20	57	0.096	0.477	0.021	1.212	17.3
4	46	31	20	50	0.092	0.457	0.023	1.231	12.5

Soil	d	sa	si	cl	Θ_r	Θ_s	α	n	k_s
	$[cm]$	$[\%]$	$[\%]$	$[\%]$	$[\frac{cm^3}{cm^3}]$	$[\frac{cm^3}{cm^3}]$	$[cm^{-1}]$	$[-]$	$[\frac{cm}{d}]$
Planosol	103								
1	17	55	28	17	0.055	0.391	0.021	1.397	19.4
2	21	57	24	19	0.058	0.390	0.023	1.374	18.4
3	27	41	25	35	0.081	0.424	0.020	1.320	5.4
4	38	42	26	32	0.078	0.417	0.018	1.342	5.8
Xerosol	122								
1	17	34	45	21	0.066	0.414	0.007	1.543	14.5
2	25	33	42	25	0.072	0.424	0.009	1.501	12.1
3	38	31	43	26	0.074	0.428	0.009	1.504	12.4
4	42	36	41	23	0.069	0.416	0.009	1.500	10.9
Yermosol	131								
1	13	48	41	11	0.044	0.394	0.012	1.487	22.4
2	20	47	41	13	0.005	0.396	0.011	1.491	18.2
3	25	44	36	20	0.062	0.403	0.012	1.458	9.6
4	36	50	32	18	0.058	0.395	0.016	1.424	14.9
5	37	33	49	18	0.061	0.411	0.006	1.589	18.0
Solonchak	123								
1	8	41	23	37	0.083	0.425	0.021	1.298	5.7
2	17	45	18	37	0.081	0.414	0.024	1.272	7.7
3	23	27	35	39	0.088	0.457	0.014	1.362	8.6
4	21	33	32	36	0.084	0.441	0.015	1.361	6.2
5	54	35	27	39	0.086	0.439	0.018	1.313	5.5

Remote Sensing Data

For our modeling approach we used LAI time-series of 10-days data-composites derived using satellite remote sensing data. The LAI describes the phenology of vegetation and thus controls interception. We used LAI time series based on CYCLOPES which can be downloaded from the POSTEL (Pole d'Observation des Surfaces continentales par Teledetection) database. This global dataset is freely available for the period 1999 - 2007 with a spatial $1\,\text{km}^2$ resolution. Since the HWSD soil map is also available on this resolution no spatial interpolation is needed to be applied. However, for each pixel analysis of the LAI time series was conducted to fill data gaps and eliminate outliers, using harmonic analysis (HA), which is based on Bittner (1993). This was needed since our model needs gap-free and continuous time series. HA decomposes a time series into a linear combination of suitable trigonometric functions, i.e. sine and cosine oscillations of particular periodicities. The HA technique corresponds to an approximate deconvolution of the power spectrum by iteratively finding and subtracting the highest peak of the time series power spectrum. This method was adapted for the correction of LAI time series data (Niklaus et al. (2012)).

6.5 Results and discussion

For a regional quality assessment we chose to compare our model results with data taken from the ECMWF ERA-INTERIM re-analysis. The spatial resolution of this dataset is $0.25\,^\circ \times 0.25\,^\circ$. The temporal resolution is a daily time step. Since this dataset is based on the same precipitation dataset we presuppose in a general comparability of the two products. We chose the time period of 2002 to 2007 as observation time and the German territory as simulation area. The November and December 2007 were excluded from the analysis due to massive changes in the ECMWF model approach. The ECMWF numerical weather model subdivides a 289 cm soil core into 4 layers (7 cm, 21 cm, 72 cm and 189 cm), with PWP at $0.171\,\%$ and field capacity (FC) at $0.323\,\%$ water content, which is assumed to be valid global. Infiltration obeys the Darcy Law and is effected by evaporation from the bare soil portion and evapotranspiration from vegetation.

Our model - in contrast - treats each soil type depending on its calculated
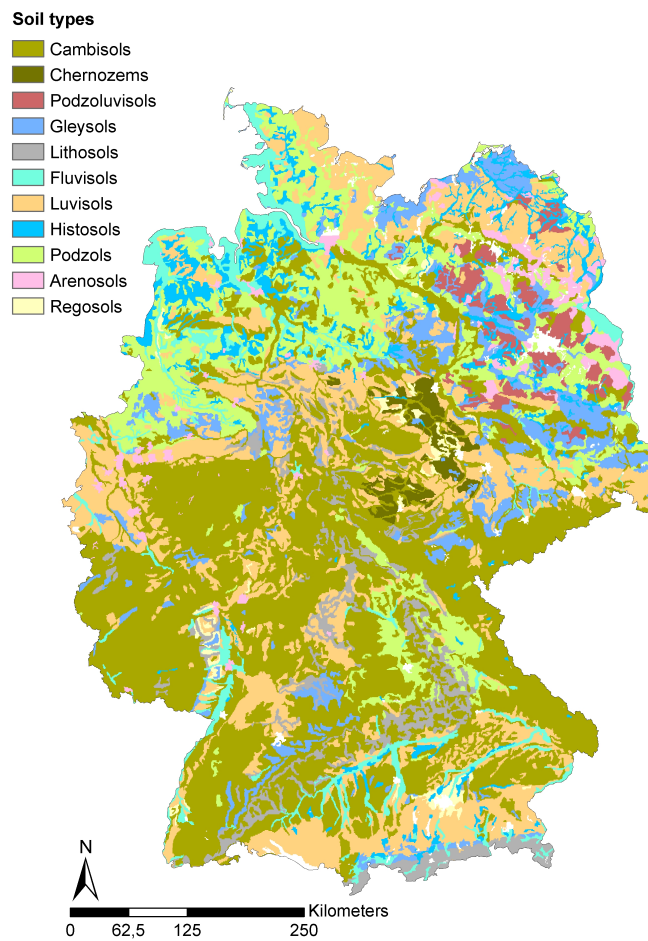
Figure 6.1: *Main soil types for Germany following the FAO '74 systematic (FAO (1974)).*

grain size distribution and layering. In figure 6.1 the distribution of the eleven main FAO soil types for Germany is presented.

From figure 6.1 it is apparent that Germany's soil cover can be characterized using eleven soil types, from which Cambisols and Luvisols can be identified as dominating. However, the spatial distribution is not homogenous. The Northern and Eastern regions of Germany show the highest heterogeneity of soil cover, whereas the middle regions of Germany are more homogeneous and dominated by Cambisols and Luvisols. Mountain ranges as the northern Alps, the "Fraenkische Alp" are covered with Lithosols, a very shallow soil which we

assume to have only 10 cm depth.
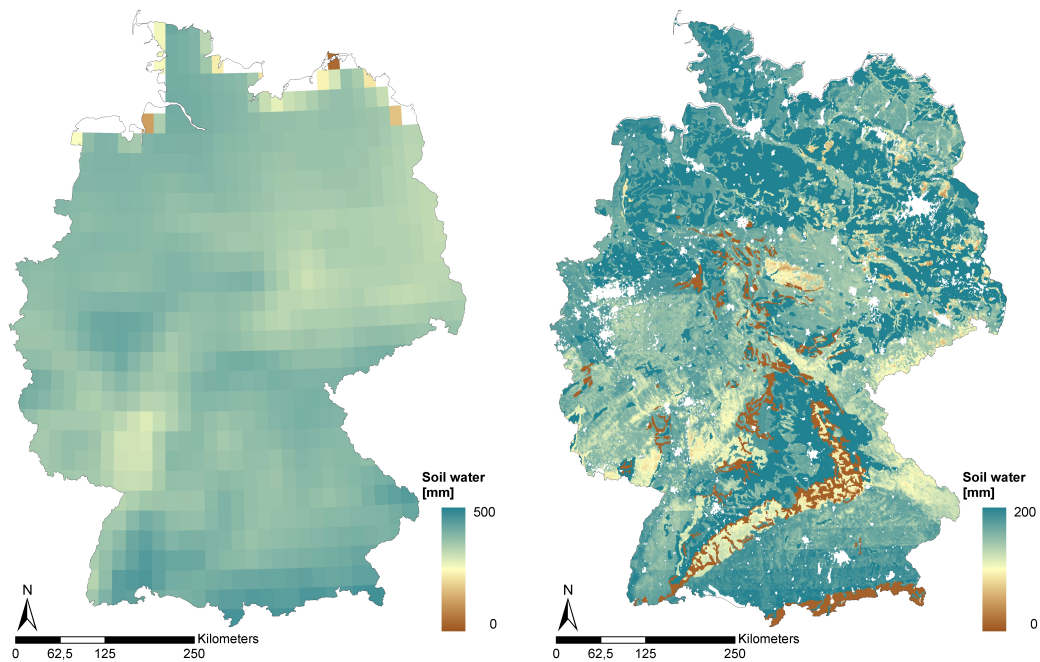
Figure 6.2: *Mean plant available soil water content given in millimetres for the period 2002 to October 2007 for ECMWF (left) and own estimations (right). High values are shown in blue, moderate values in yellow and low values in red. Whit pixels represent no data.*

In order to compare the soil water characteristics we calculated the mean plant available soil water content for the observation time, for both: the ECMWF soil water content product and our own model results (figure 6.2). Analysis revealed that the mean plant available soil water content for ECMWF is 364 ($\pm$ 59) mm. For our own estimates we found a corresponding value of 170 ($\pm$ 83) mm, which is roughly half of the ECMWF, but combined with a higher standard deviation. The lower water content might be explained with the fact that our soil depth ranges soil dependant from 10 cm to 208 cm, but the ECMWF soil is set static to 289 cm. Since the ECMWF data does not show rough transitions, the higher standard deviation can be explained with the fact that we take into account individual soil characteristics, based on the soil distribution as shown in 6.1. Considering the costal zone of the ECMWF result it becomes apparent that the lowest values can be found here. In addition for some areas no data are available, which is due to the global modeling
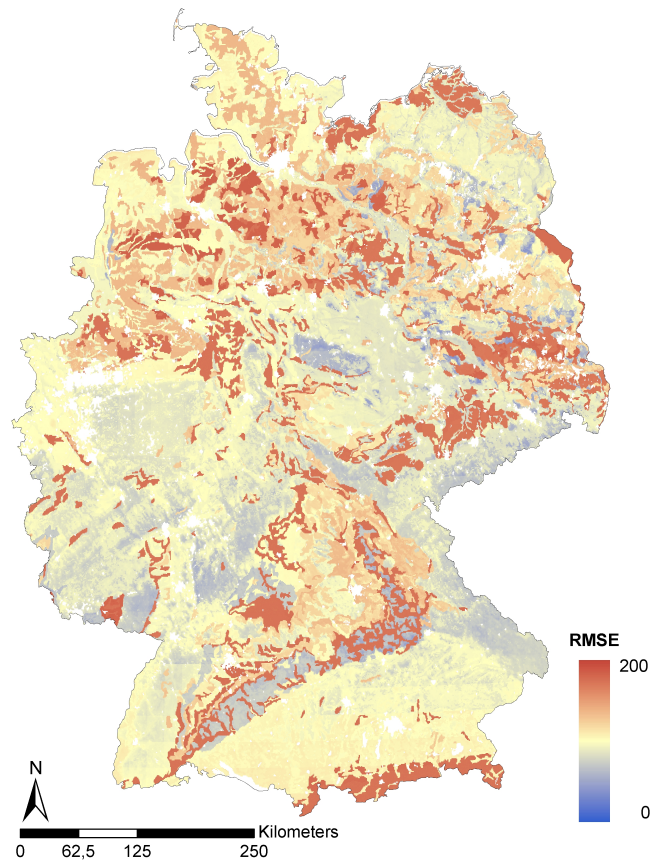
Figure 6.3: *Root Mean Square Error for the ECMWF and our own time series (2002 to October 2007) for Germany. Low RMSEs are shown in blue, moderate in yellow and high in red. White pixels represent urban areas.*

scheme, in which costal zones are sometimes treated as water, although more than 50 percent of the grid cell is covered with land, and vise versa.

To compare the ECMWF soil water product with our own estimates we calculated the root mean square error (RMSE) based on daily values for the whole observation period and area. The result is presented in figure 6.3.

Areas which are symbolized with blue and yellow colours indicate a good agreement of the model behaviour, whereas red colours represent areas of highly different conformability. From figure 6.3 it can be deduced that the lowest
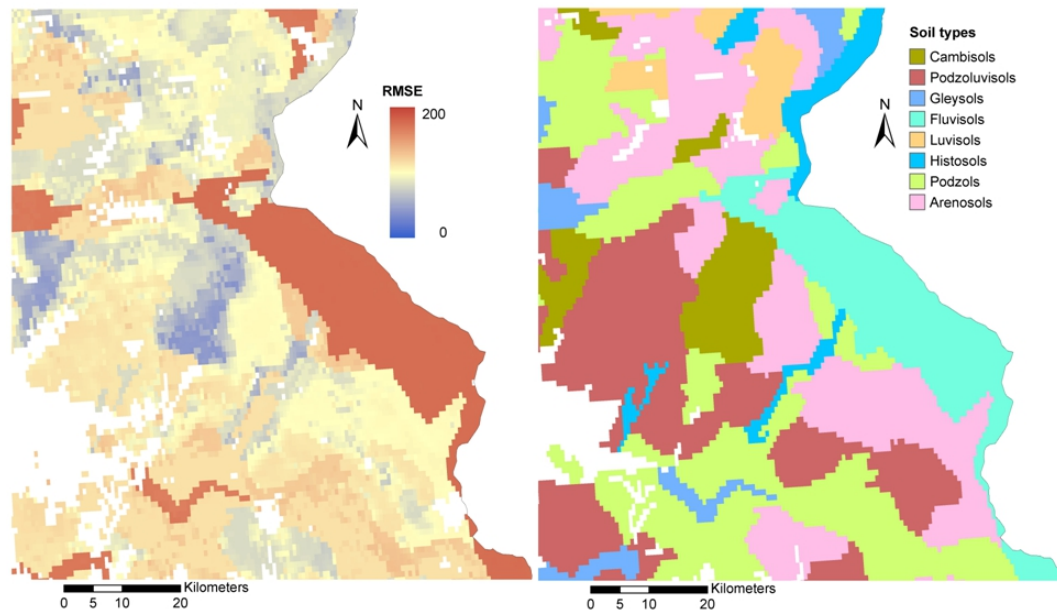
Figure 6.4: *RMSE (left) and main soil types (right) for the Oderbruch-regioin in East Germany. The colour schme follows the scheme of figure 6.1 and figure 6.3.*

RMSEs, and thus the highest degree of agreement, can be found for regions of homogenous soils as e.g. the Cambisol-region in central Germany. The highest RMSE up to 200 can mainly be found for areas which correspond to soils under constant wet conditions, as for instance Gleysols and Fluvisols, and with regions of shallow soils (Lithosols) are reported (see figure 6.1).This is again due to higher detailed soil map which we used and our approach to respect the individual characteristics of soils (see table 6.1). Therefore, our model can represent regional conditions in more detail.

When zooming to the "Oderbruch-region" located at the North-Eastern boarder of Germany this finding can be confirmed (figure 6.4). Here it can be seen that the pixels which are described as Fluvisols and Gleysols show the highest level of disagreement (red). The soil parameters for these heavy soils, also called "minute soils", are highly discrepant to the ECMWF soil properties. Areas with Arenosols, Podzols, and Podzoluvisols show intermediate (yellow) and Cambisols the best agreement (blue). This can be explained with the spatial resolution ($0.25\,° \times 0.25\,°$) of the ECMWF soil water content product. Heterogeneous soil landscapes, as situated here cannot be represented with this coarse resolution. Therefore, the ECMWF product has to be seen as integra-

tive, since these differences in hydrological behaviour and soil properties are not represented.

In order to compare the inter-annual behaviour of the infiltration process of the two models we focused on three soil examples. Since the ECMWF soil bulk and compartments are not directly comparable to our individually calculated soil characteristics, but are based on the same precipitation dataset, we chose to compare the general behaviour for three soil examples (Cambisol, Gleysol, and Lithosol). The results are presented in the figure 6.5, figure 6.6 and figure 6.7. The examples were chosen due to their importance to the global and European soil distribution and because of our finding of high and low agreement, as discussed above. According to the HSWD map around 9% of global soils are Cambisols and 15% are under saturated conditions and 12% are Lithosols. The ECMWF data was taken for latitude: $51.0°$ / longitude: $8.0°$ (Cambisol), latitude: $52.8°$ longitude: $13.0°$ (Gleysol) and latitude: $47.6°$ / longitude: $11.3°$ (Lithosol).
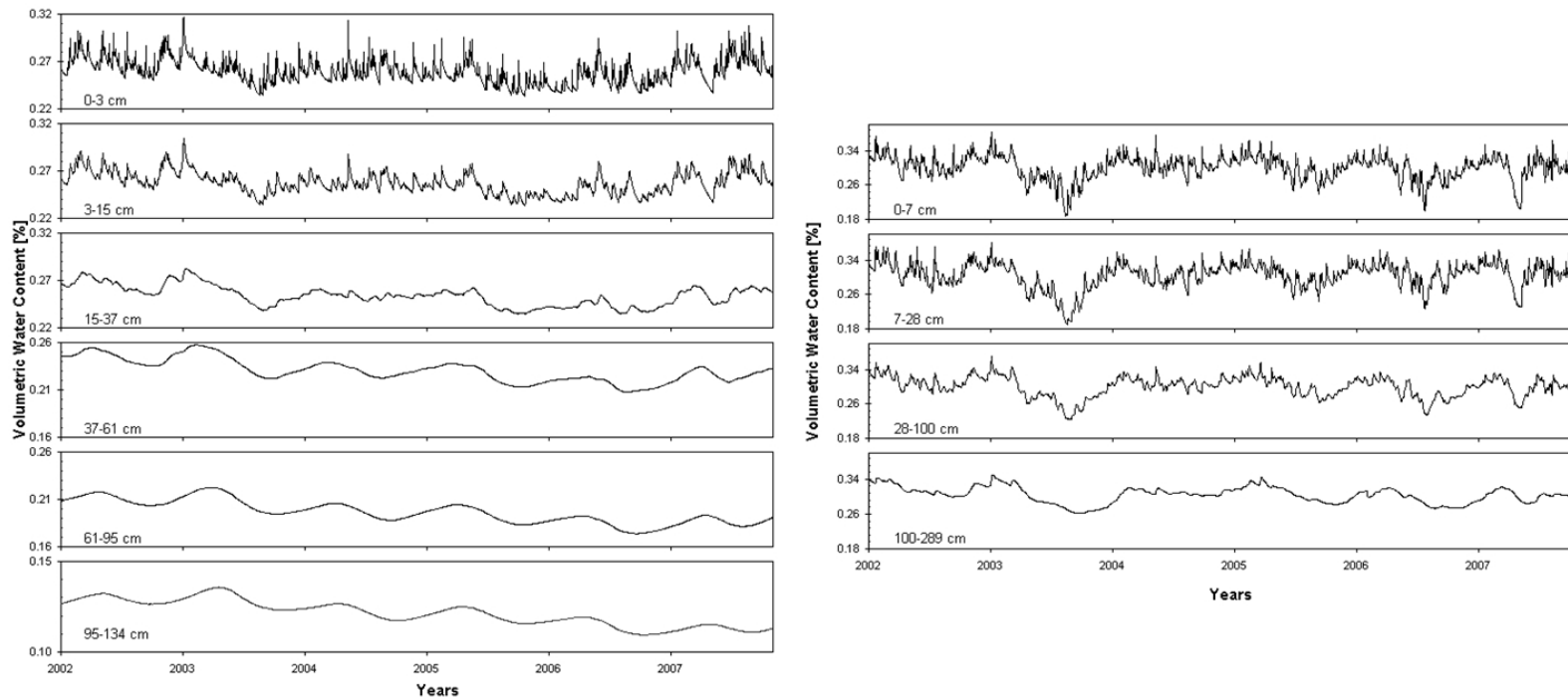
Figure 6.5: *Volumetric water content for a Cambisol site at 8.0° longitude / 51.0° latitude. Left: own model results for six soil layers (0-3 cm, 3-15 cm, 15-37 cm, 37-61 cm, 61-95 cm, 95-134 cm). Right: ECMWF results for four soil layers (0-7 cm, 7-28 cm, 28-100 cm, 100-289 cm). Both given in percent water volume per soil layer on a daily basis from January 2002 to October 2007.*
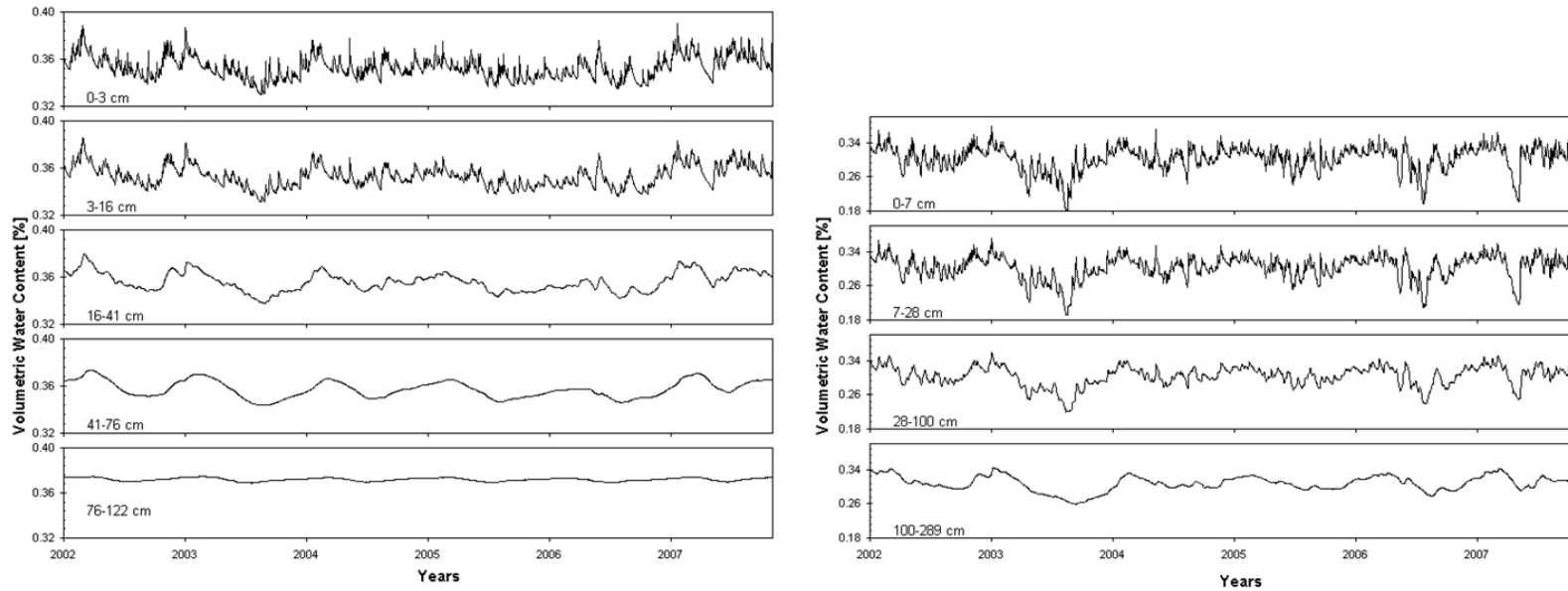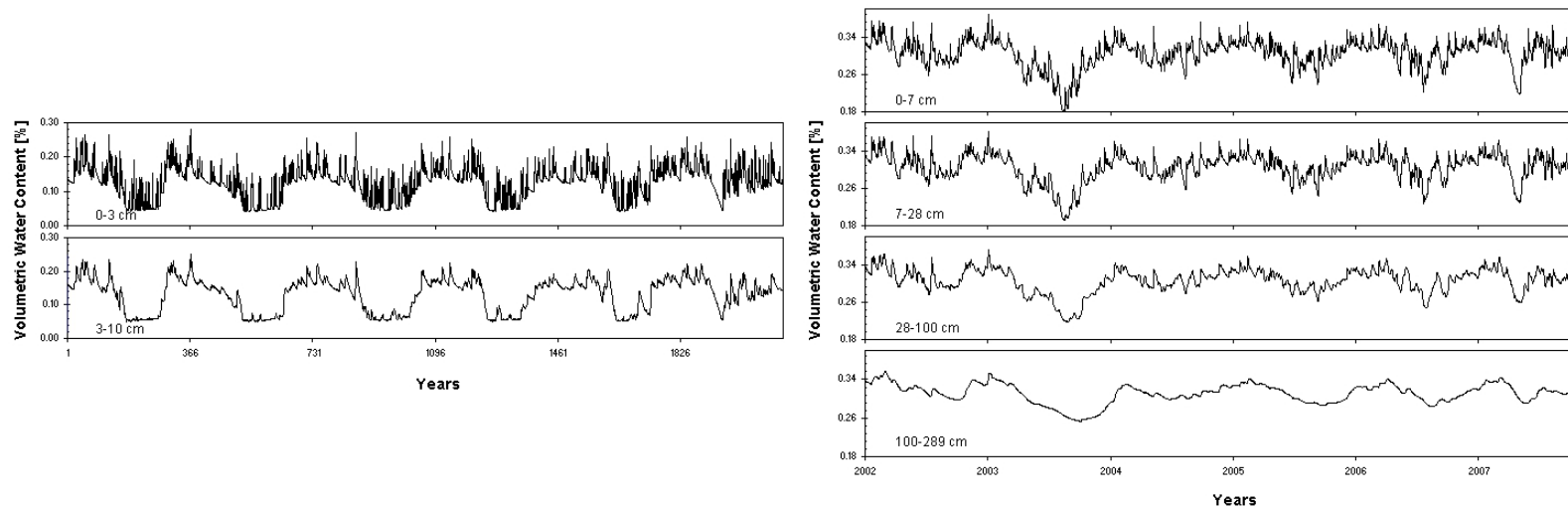
Figure 6.6: *Volumetric water content for a Gleysol site at 13.0° longitude / 52.8° latitude. Left: own model results for five soil layers (0-3 cm, 3-16 cm, 16-41 cm, 41-76 cm, 76-122 cm). Right: ECMWF results for four soil layers (0-7 cm, 7-28 cm, 28-100 cm, 100-289 cm). Both given in percent water volume per soil layer on a daily basis from January 2002 to October 2007.*

Figure 6.7: *Volumetric water content for a Lithosol site at 11.3° longitude / 47.6° latitude. Left: own model results for two soil layers (0-3 cm, 3-10 cm). Right: ECMWF results for four soil layers (0-7 cm, 7-28 cm, 28-100 cm, 100-289 cm). Both given in percent water volume per soil layer on a daily basis from January 2002 to October 2007.*

From figure 6.5 and figure 6.6 it is apparent that the volumetric water content of the two topmost layers of both models are comparable in their general characteristics, but with differences in their absolute values.

The first layer of both models is highly influenced by precipitation, which can explain the high short time variability of water content. This is to some extend also valid for the second layer, which is still highly influenced by daily precipitation events, but already shows little hysteresis effects. ECMWF third layer still shows characteristics which are highly dependant on precipitation, which was not expected, since this layer already represents the soil water content in 28 cm to 100 cm depth. Since we cannot find this behaviour in our soil layers 3 to n this layer cannot be compared with our model results. Our third and ECMWF forth layer can again directly be compared and show hysteresis effects and the start of long term water movement behaviour in soil.

A closer look to figure 6.5 shows a decrease in our volumetric water content from layer 1 to 6. This is due to our model formulation in which we assume a complete dry layer beneath our last model layer, to respect water run off processes, or to be more precise to allow water to leave the system. On the other hand, as can be seen from 6.1, Cambisol has low k_s values for all layers and thus a strong water holding capacity, which results in not too steep negative slopes during summer periods (e.g. 2003 and 2006). To model saturated conditions, as necessary for Gleysols, we assume a saturated layer beneath the last model layer. With this more or less steady state conditions can be achieved, as can be seen in layer 5 of figure 6.6.

In contrast figure 6.7 shows that the soil water availability of the two modelled layers of Lithosols cannot be compared with the characteristics of the corresponding ECMWF plot. Since for Lithosols we assume a soil depth of only 10 cm they are highly dependant on precipitation events and loose water immediately under non precipitation conditions. However, due to the mathematical limitations of our model, a total loss of soil water cannot be modelled. A residue of minimum 3 % water always remains in the soil core.

Generally, long term water characteristics as we can describe with soil layer four to six, saturation conditions or strong water holding capacities cannot be found in the ECMWF model results. This can be expected since this additional information is not taken into account in the ECMWF model.

6.6   Conclusions

We adapted, refined and presented a one-dimensional soil water transport model for regional and global environmental applications, following the van Genuchten approach. It uses remote sensing based time series of the Leaf Area Index to treat interception. In addition we calculated for all 128 FAO '74 soil types the individual soil depth, layering, grain size distribution and van Genuchten parameters. In order to compare our model with other data, we applied our model for the period 2002 to October 2007 for Germany and compared the results with ECMWF soil water content data for the same period. We found good agreements for regions of Cambisols and bad agreements for regions of soils under stagnant moisture (e.g. Gleysols) shallow soils (Lithosols) or regions of heterogeneous soil landscapes of strongly varying soil qualities at small-scale. Reasons for different agreement levels can be seen in the more detailed soil map, which was used in our adapted and refined modelling approach and in the fact that the ECMWF model only comprises one globally generalized soil type.

Modelled characteristics of the plant available soil water in the unsaturated zone are typically used for vegetation models. Therefore, our presented modelling approach could be helpful in the assessment of the soil water at a broader spatial scale. Our model will, additionally, allow assessing the soil water in vegetation models on a global, regional and local scale, if all needed data are available. However, for local applications our approach should be further tested in additional countries and for specific small scale case studies as e.g. lysimeter stations.

Furthermore, this new modelling approach could be seen as useful for applications which calculate the plant available soil water content in scenario models. In particular, more reliable forecasts will be of great consequence for the estimation of the impacts of global climate change upon vegetation species distributions, water availability and thus food security.

6.7 Acknowledgements

Chapter 7

Generation of a global, gap-free SPOT-VGT LAI dataset using spectral analysis techniques

Markus Niklaus, Kurt P. Günther, Markus Tum, Michael Bittner,
International Journal of Remote Sensing, under review

7.1 Abstract

The application of time series analysis of vegetation indices for modeling phenological plant development usually requires continuous, gap-free data. Datasets are often characterised by gaps that are caused by clouds, large solar zenith angles, topology etc. A combination of mean phenology data, derived from the original dataset, and harmonic analysis was used to fill these gaps with reasonable data. From the LAI time series of CYCLOPES and geoland2 derived from SPOT VEGETATION images, we were able to generate global gap-free and outlier-corrected LAI time series for the period 1999 - 2007 and 2010. The comparison of the smoothed LAI values with the original data shows a high degree of agreement on a global scale. The LAI reference maps, elaborated in the VALERI and BigFoot projects, were used to validate the reprocessed data. This validation shows an accuracy of 83 % compared to 82 % of the original data set.

7.2   Introduction

When modelling the development of vegetation information about the seasonal greenness of the earth's surface is essential. This information can be derived both from in-situ measurements and by satellite or airborne remote sensing observations. Information products that are derived primarily from remote sensing data are used at regional, national or global scales. The Normalized Difference Vegetation Index (NDVI) is a well known and frequently used vegetation index. It is directly linked to the photosynthetic capacity of plant canopies (Myneni et al. (1995), Sellers et al. (1992)). Sellers (1985) showed that plant-physiological parameters such as Leaf Area Index (LAI) and the Fraction of Absorbed Photosynthetical Active Radiation (fAPAR) can be derived from vegetation indices. Remote sensing can provide products for all three parameters, such as the AVHRR (Advanced Very High Resolution Radiometer) NDVI (Eidenshink (1992)), MODIS (Moderate resolution Imaging Spectroradiometer) NDVI, LAI and fAPAR (Myneni et al. (2002)), SPOT-VGT (Satellite Pour l'Observation de la Terre - Vegetation) LAI and fAPAR (Baret et al. (2007)) and MERIS (Medium Resolution Imaging Spectrometer) NDVI and fAPAR (Günther and Maier (2007), Gobron et al. (2004)).

For mechanistic models such as the regional biosphere model (RBM) or C-Fix (Richters (2005), Veroustraete et al. (2002)) is information about the development of vegetation one of the major inputs. The uptake of CO_2, referred to as Net Primary Productivity (NPP), is usually computed as a function of the light use efficiency (LUE) and a vegetation parameter such as LAI or fAPAR. For dynamic vegetation models such as the Lund-Potsdam-Jena Dynamic Global Vegetation Model (LPJ) (Sitch et al. (2003)), the Biome-BioGeochemical Cycles model (Biome-BGC) or the Biosphere Energy Transfer Hydrology Model BETHY/DLR (Knorr (1997), Knorr and Heimann (2001), Wißkirchen (2005), Thornton (1998)) such input information is essential to also estimate the carbon balance between biosphere and atmosphere using a parameterization of the photosynthetic activity.

For both model types, it is important to have information about the plant's development as a spatially and temporally continuous, gap-free dataset. However, this requirement is not often achieved. Clouds are the most common

reason for large data gaps. In winter and particularly in the Northern Hemisphere solar zenith angles exceeding 60 ° often restrict the processing of remote sensing data, or even prevent it. This is because the low solar altitudes above the horizon result in increased atmospheric disturbances due to the non-linear increase of the atmospheric path. There are also larger shadows and a generally lower illumination thus increasing the signal-to-noise ratio. Topological effects caused by steep terrain, for example, can lead to physiologically unrealistic observations. Therefore, such data points need to be filtered, and gaps need to be filled. The amount of applications using such processed data has generated a large number of approaches to create them.

Jakubauskas et al. (2002) carried out Fourier time series analysis of AVHRR NDVI data to quantify seasonal and inter annual land use / land cover changes. Azzali and Menenti (2000) correlated the amplitudes of the Fourier Transform with the aridity and the vegetation types in southern Africa. Dilmaghani et al. (2007) applied the Lomb-Scargle periodogram (Lomb (1976), Scargle (1989)) to time series of air quality and water quality. They used datasets with irregular sampling periods and missing data, and derived continuous and gap free time series. The TIMESAT program, developed by Jönsson and Eklundh (2002), Jönsson and Eklundh (2004), was first adapted to AVHRR data, providing smoothing functions using the methods of the asymmetric Gaussian, double logistic or adaptive Savitzky-Golay method (Chen et al. (2004)). Furthermore, it was used by Gao et al. (2008) to analyze MODIS-LAI time series applying the asymmetric Gaussian smoothing function. Yuan et al. (2011) used a combined approach of the modified Temporal Spatial Filter (mTSF) method for filling data gaps and the Savitzky-Golay filter of the TIMESAT algorithm to reprocess the MODIS LAI product.

Colditz et al. (2008) introduced the Time-Series Generator (TiSeG), which checks the MODIS Quality Assurance (QA) flags in a first step. These are determined by aerosol quantity, atmospheric correction conditions, cloud cover, shadow, and sun-target-viewing geometry (Myneni et al. (2002), Huete et al. (1999)). In a second step, the user compares several QA flags in order to mask or interpolate data gaps using linear, polynomial or cubic spline functions in space or time. The approaches shown here each have their advantages over others, especially when considering the variety of applications the processed

data are used for. This is discussed in several publications (Hird and McDermid (2009), Musial et al. (2011), Moffat et al. (2007)). But only few (Yuan et al. (2011)) are used to produce gap free datasets of LAI on a global scale. The presented work in this paper is based on the Harmonic Analysis (HA) approach (Bittner et al. (1994)). This technique has been used, for example, to operationally process satellite based ozone column measurements (Bittner et al. (1998), Meisner et al. (1999), Ebertseder et al. (2006)). We adapted the HA in order to generate global LAI time series. With this scheme the SPOT-VGT LAI time series is examined to receive a gap-free and outlier corrected global data set. In this work the changes to the algorithm are described, the quality of this approach is discussed and finally the global LAI data set is presented, validated and discussed.

7.3   Data and Methods

As described above, LAI time series are often used as input for biophysical process models. These models need equidistant gridded, continuous time series (e.g. 8 days, 10 days or monthly composites). Since LAI time series contain data gaps, a method is needed to generate continuous sequences of input data.

Data Description

For this study LAI time series of the CYCLOPES (Carbon cycle and change in land observational products from an ensemble of satellites) archive for the period 1999 - 2007 and of the geoland2 project for the year 2010 are used. The geoland2 data were produced also using the CYCLOPES processor. The products can be downloaded, as 10 day composites, from the POSTEL database (http://postel.mediasfrance.org) and the geoland2-Biopar platform (http://www.geoland2.eu) respectively. The spatial resolution is about $1\,\mathrm{km}^2$ and the files are available as $10\,^\circ \times 10\,^\circ$ tiles (labelled H[0-35]V[0-17]). We applied the HA method for each pixel to eliminate data gaps and outliers.

Harmonic Analysis

The seasonal variability in LAI can be described as a superposition of oscillations with different periods (see figure 7.1). The first fundamental harmonic can for example reflect for temperate vegetation the annual cycle of the mean temperature or for sub-tropical vegetation the rain seasons. The HA follows the concept to represent a time series as a linear combination of n sinusoids.
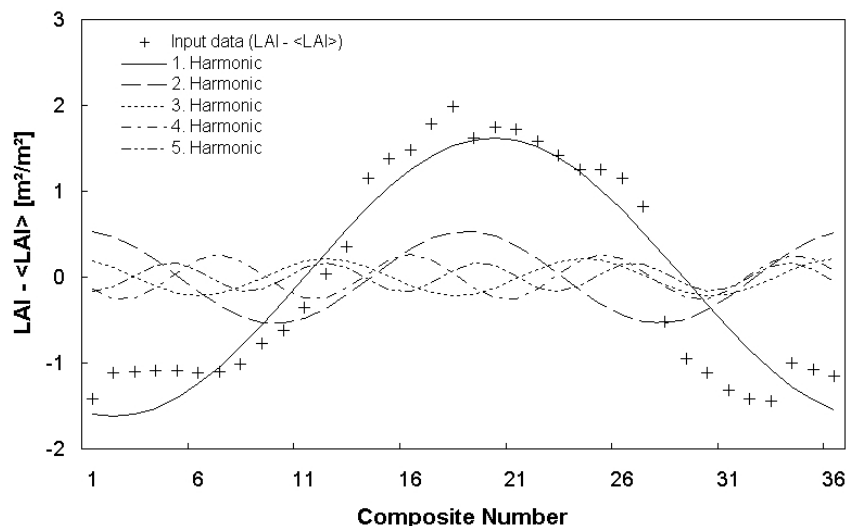

Figure 7.1: *Decomposition of LAI time series for one year into five harmonics using the technique of the HA. The LAI values are normalized to zero by subtracting the mean LAI value of the time series.*

Each oscillation is defined by its amplitude A_i, phase φ_i and frequency f_i, where i ranges from 1 to n. The technique corresponds to an approximate deconvolution of the power spectrum by successively subtracting the highest peak, then computing a new spectrum, and so on. The resulting linear combination of sinusoids, however, turns out not to be unimodal when fitting the data. To be more precise, the variance of the time series can be greatly reduced if all parameters of the actual and former sinusoids are varied simultaneously. For example, when fitting the second sine function, the period, amplitude and phase of the first sine function are free parameters, and are therefore iterated. The fitting is done by pixel based identification of dominant harmonic oscillations as e.g. changes within the seasonal cycles. Here the major challenge is to discriminate natural and disturbance induced dominant spectral structures

(Bittner et al. (1994)). It has to be taken into account that outliers within a time series can strongly affect the model result.

To solve the linear system of equations, the Newton-Raphson (Ortega and Rheinbold (1970)) and Cholesky (Golub and Van Loan (1996)) methods are used. A set of several sinusoids with corresponding amplitudes, phases and frequencies are computed with their sum representing a fit to the original data:

$$y\left(t\right) = \sum_{i=1}^{n} A_i \times sin\left(\frac{2\pi}{\tau_i}t + \varphi_i\right) \tag{7.1}$$

$y(t)$ represents the LAI value, $A_i(t)$ the amplitude, $\tau_i(t)$ the period and $\varphi_i(t)$ the phase for the ith oscillation. This procedure of HA is applied for each time step of the time series. Therefore a sequence of wave lengths which are oriented in the length of the time series is distributed in equidistant sections. Within this array the optimum amplitude and phase are determined to adapt the oscillations. This results in a periodogram from which the variance of dominant spectral characteristics can be deduced. The wavelength which describes the data variability of the whole time series the best is selected and optimized in addition of the measured values for each pixel, concerning amplitude and phase. This is repeated for further oscillation detections. With each step the previously found oscillations are optimized. The amount of oscillations to describe a time series with maximum accuracy reaches saturation at a certain point (Bittner (1997)).

Modifications to the program

The HA method was originally designed to process time series of ozone information in the stratosphere (Bittner (1997)). In order to use the routine of HA for LAI time series, an adaption had to be performed. This becomes apparent when applying the original routine to the LAI data, as shown in figure 7.2. The fitted function (solid line) is off the measured data over the whole period, showing unrealistic plant development, especially during the dormant phase of vegetation. This is mainly due to lack of data at the beginning and at the end of the year. But also the vegetative phase is not fitted successfully by this setup of the HA.
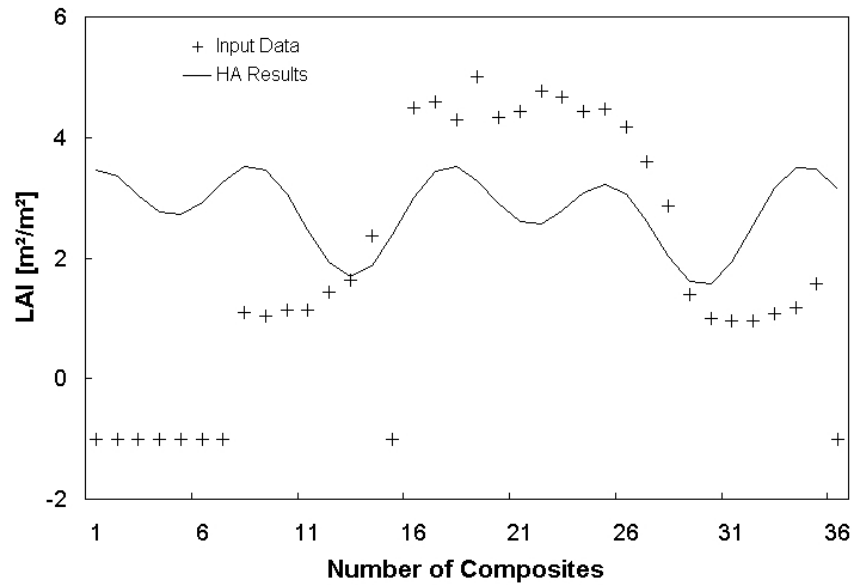
Figure 7.2: *Example for applying the original (solid line) HA to the CYCLOPES data (crosses) for one pixel in Massachusetts, USA, for the year 2003. A value of -1 represents a data gap.*

Pre-processing of the data

The situation shown in figure 7.2 requires a method to fill large data gaps, since the original setup of the method was not suitable for extrapolating larger periods. As already mentioned, data gaps are mainly due to cloud contamination of the pixel or due to high sun zenith angles. Most satellite data providers limit the data processing to sun zenith angles of less than $60\,°$. Therefore, data gaps are more frequent during the wintertime in the Northern Hemisphere. In equatorial and tropical regions permanent cloud cover cause periods of missing data over nearly the whole year. The number of missing data or the length of data gaps thus depends on latitude and season. It turned out, that data gaps of five and more missing composites result in unacceptable uncertainties within the HA, whereas smaller gaps have no significant impact. Hence, the criterion was included that large gaps of five or more missing composites should be filled with LAI values derived from a mean phenology for the land cover type under investigation. Therefore, for each $10\,° \times 10\,°$ tile mean phenologies were calculated for each land cover type by averaging all corresponding LAI values for each compositing period independently. For the identification of the

different land cover types the Global Land Cover GLC2000 (Bartholome et al. (2002), Fritz et al. (2003)) was used, differentiating 24 land cover types. The GLC2000 is included as an additional layer in the CYCLOPES dataset.

Filling gaps with mean values of phenologies is based on the assumption that the phenology of each land cover type in a $10° \times 10°$ tile is similar for all pixels. This mean phenology is shown in figure 7.3 for a tree-cover class of the tile 'H10V4' covering the northeast of the US and south-eastern parts of Canada (upper left corner: 80 W 50 N; lower right corner: 70 W 40 N) as it would be used for the gaps in the time series shown in figure 7.2.
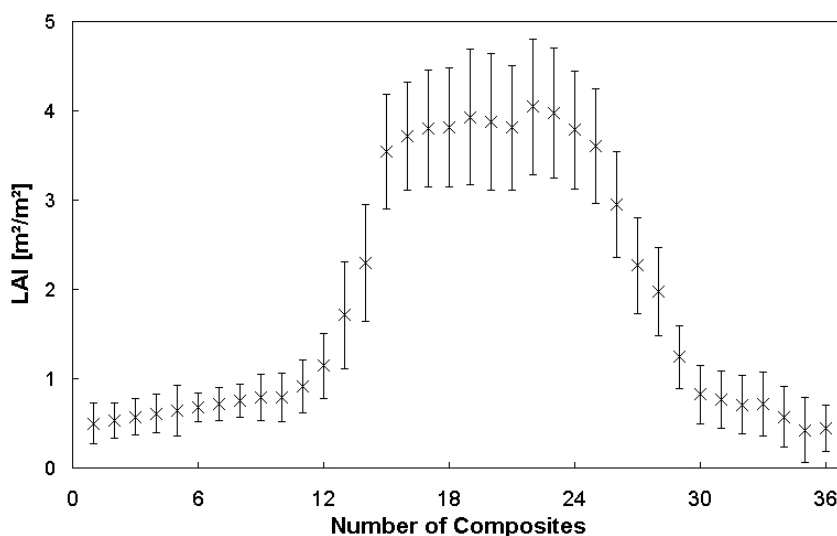
Figure 7.3: *Mean phenology and standard deviation of tree-cover class for tile H10V4.*

A second step to improve the HA is to detect outliers within the LAI time series. In order to understand the occurrence of outliers in LAI time series, especially from SPOT-VEGETATION (SPOT-VGT), one has to be aware of the different pre-processing steps that are performed in order to train and feed the neural net for deriving LAI (Baret et al. (2007)). After cloud screening and atmospheric correction the Roujean BRDF (Bidirectional Reflectance Distribution Function) model (Roujean et al. (1992)) is applied to normalize the top of canopy reflectances of the three VEGETATION bands (B2, B3 and MIR) to nadir viewing conditions. For the BRDF correction cloud-free observations collected over a time window of ± 15 days are needed. In an iterative

process outliers in the reflectances due to cloud contamination or improper atmospheric correction are detected and eliminated. Finally, the medians are calculated for all three BRDF corrected bands as well as the median of the sun zenith angle. These normalized reflectances are used as input for the neural network. Nevertheless, outliers in the LAI time series are observed. One explanation might be, that the median input variables for the neural net do not fit the input range which was used during the training of the neural net. This deviation might occur when clouds, haze or shadows are not detected in a reliable manner. Also, the atmospheric correction might fail when high aerosol loads occurred which are not included in the aerosol climatology. In general, outliers in the LAI time series as high-frequency oscillations can affect the HA by modifying amplitudes and phases and thus resulting in major errors. Outliers may be detected using a simple stochastic threshold. For example, outliers can be defined as values deviating from the range of the mean phenology $\pm 3\sigma$ (standard deviation). The 3σ-range includes 99.7% of all values as confident and the rest as outliers. Tests confirmed that with this criterion not all outliers can be detected.

A more sophisticated two step based approach is proposed and tested. In a first step each gap is filled with its corresponding value from the mean phenology. In a second step, a comparison of three consecutive data points of the time series is done. Outliers are characterized when the gradients from the first to the second value and from the second to the third value are opposite and one gradient exceeds a value of $2\,\mathrm{m^2\,m^{-2}\,10d^{-1}}$. A continuous, steep increase or decrease of the LAI does not result in outlier identification. Investigations have shown, and other works confirm (Verger et al. (2011)), that the change of sign of gradients combined with an in- or decrease of the LAI of more than $2\,\mathrm{m^2\,m^{-2}\,10d^{-1}}$ represents unreasonable fast changes in vegetation physiology. Steeper slopes are marked as outliers and, henceforth, are excluded from the HA. Since vegetation reacts more dynamically to its environment this method is regarded as more suitable to detect outliers, which are due to errors in measurements.

When all data gaps and outliers are detected, the data set is doubled and placed one after another to get periodical boundary conditions (figure 7.4). Hence, the missing data at the two ends can be analyzed in the middle of the
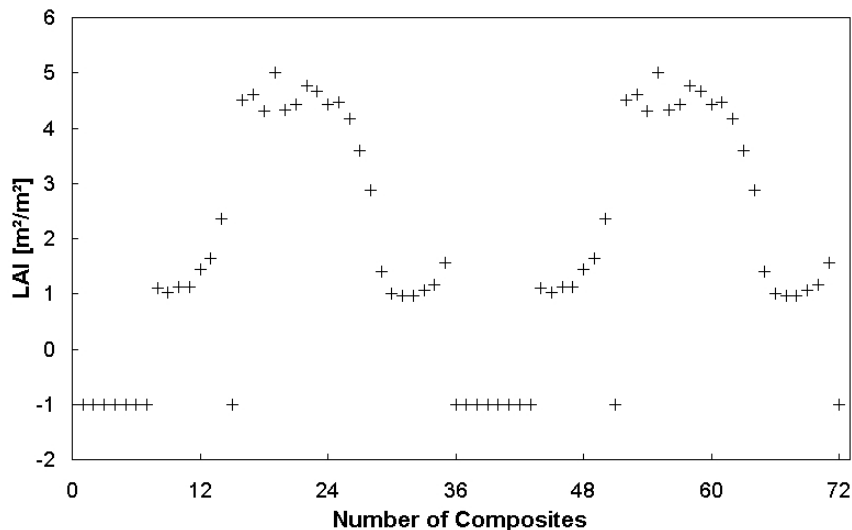
Figure 7.4: *Doubled data set of the time series shown in figure 7.2. A value of -1 represents a data gap.*

doubled time series. Thus, the major data gap for the time series shown in figure 7.2 covers, after doubling, eight data points (figure 7.4), consisting of the one missing data point at the end of the year and seven missing data points at the beginning of the year.

Method modifications

The third improvement of the HA is the selection of the number of oscillations, n, and of their periods, τ. In the original HA version an upper and lower period (τ_{up}, τ_{lo}) are to be selected in order to define the interval of allowed periods. From these, the best fit curve is to be derived together with the number of oscillations to be used for the linear combination of the selected sinusoids (see Eqn. 1). The start values of the periods for the n oscillations to be selected are distributed equally between the upper and lower period. Thus the ith period is defined as (τ_{up} - τ_{lo})$\times i/n$. With this distribution of periods commonly only two periods were identified before the exit conditions are met. However, two periods are usually insufficient to describe the characteristics of the underlying LAI time series. To solve this shortcoming a combination of two modifications was used. First, the exit conditions for finding the harmonic oscillation have been changed. To abort the iterations of computing the best fit, the residua of

the actual calculation and of the previous step are compared. The difference of the two variances of the residua is controlled by a quality factor ϵ. This factor was set to a constant value of 0.5 in the original setup and is now changed to 0.001 at the beginning of the iterations. If this threshold is not met after four iterations ϵ is doubled respectively.

The second step was rearranging the start values following the principles of the Bessel-formula (Schönwiese (1985)). The new start values for the periods are defined by selecting the smallest period and the number of oscillations. Thus the distribution of the periods is no longer equidistant and the ith period is defined as τ_{lo}/i. In the iteration to find the appropriate harmonic to the residuals of the function, here oscillations with shorter periods are proven before the exit conditions are met and thus considered for the description of the time series. This procedure has the additional benefit to reduce the computational time, since the modified HA will take significantly longer to find a fit to the data, especially due to the doubling of the data set and the decreased quality flag.

The number of computed harmonics is set to a minimum of two oscillations and controlled by the restriction of a minimum period of 60 days, or six data points. Technically the method allows for a total of n oscillations, but due to the period limitation this is never met.

For checking the quality of the results of the modified HA the Root Mean Squared Error (RMSE) is computed for each pixel and year.

$$RMSE = \sqrt{\frac{sum\left(LAI_{input} - LAI_{HA}\right)^2}{n}} \qquad (7.2)$$

with LAI_{input} being the CYCLOPES or geoland2 input data, LAI_{HA} the analysis results and n the number of good values (meaning without data gaps) taken for the HA. To get a globally comparative value this RMSE is normalized to the mean LAI value of the pixel resulting in a Normalized RMSE value (NRMSE):

$$NRMSE = RMSE/\overline{LAI} \qquad (7.3)$$

The NRMSE compares the results of the modified HA directly with the mea-

sured LAI input data by calculating the sum of the squared differences between the estimated LAI and the measured LAI. The normalization of the RMSE to the mean LAI allows a better comparison of the error on a continental or global scale. As an example, suppose a grassland pixel with an average LAI of 1 and a deciduous tree pixel with an average LAI of 3 have the same RMSE of 0.1. The NRMSE for the time series of the deciduous tree pixel is better (NRMSE = 0.033) than for the grassland pixel (NRMSE = 0.1) indicating that the fit for the deciduous tree pixel is more reliable.

7.4   Results

To analyze the impact of the changes to the HA technique the algorithm was applied to the LAI time series shown previously in figure 7.2. The pixel under investigation represents the temporal development of the LAI of a mixed forest area in the Harvard Forest, Massachusetts, USA. Figure 7.5 shows the results of the improved HA, where values of the mean phenology are denoted as triangles. Only the first seven and the last value are used to fill the data gaps of the input time series. It was found that small gaps of less than five data points can be extrapolated by the modified version of the HA without filling with mean phenology data. By inserting the mean phenology the plant's dormant phase during winter time is represented in a phenological more reasonable way. The results of the modified HA fit the measured and inserted LAIs with good accuracy, having a NRMSE of 0.10.

Figure 7.6 shows another example of the temporal evolution of the LAI for a pixel in tile H19V4, covering Italy, the Alpine and the Balkan region (upper right corner: 10 E 50 N; lower right corner: 20 E 40 N) for the year 2001. Following the scheme of GLC2000, this pixel represents the vegetation type of temperate, broadleaved, deciduous trees (class 6 of the GLC2000 classification). This time series includes two data gaps. One at the beginning of 2001 (twelve LAI composites = 120 days) and one at the end of the year (three LAI composites = 30 days), where no LAI values were measured. The latter can be handled by the HA and does not need to be filled with values from the mean phenology. However, the bigger first gap was filled with values of the mean phenology for the vegetation class of the pixel (triangles). The solid
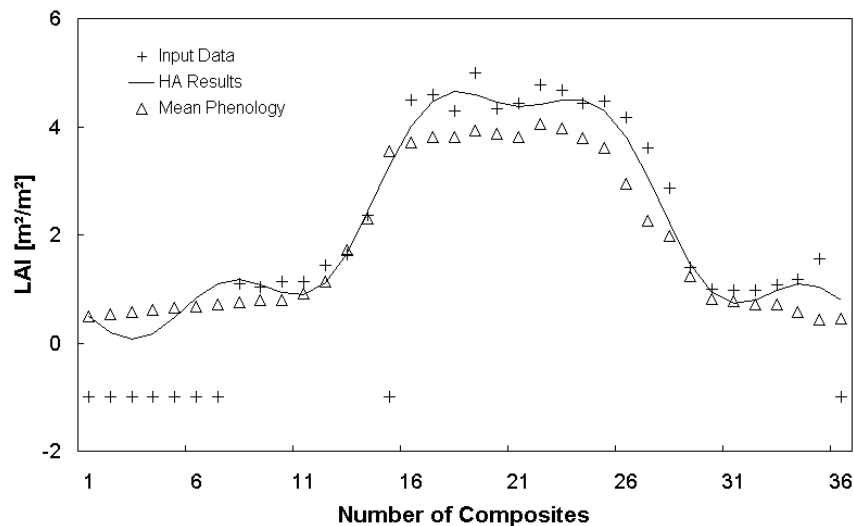
Figure 7.5: *Results of the modified HA based processor (solid line) applied to the input data (crosses) previously shown in figure 7.2. Triangles show the mean phenology of a mixed forest area for this pixel. A value of -1 represents a data gap.*

curve shows the result of the modified HA-processor for the year 2001. With a total of 15 missing data points the NRMSE for this pixel is about 0.09.

The effect of applying the modified HA to a representative set of full tiles of SPOT-VGT LAI data can be seen in figures 7.7 and 7.8. The tiles shown are H7V5 (North America), H19V4 (Europe), H12V8 (South America), H18V7 (Africa), H29V12 (Australia) and H29V4 (Asia). The examples are shown for different years and different time steps, comparing the tiles before (SPOT-VGT) and after (HA) applying the analysis. For land masked pixel data gaps are due to cloud cover or snow, according to Baret et al. (2007). Looking to the results, it is obvious that not all data gaps over land are fitted by the HA (shown as white pixel). In tile H19V4 (figure 7.7, middle column) the remaining data gaps represent the Alps, which are masked out as bare areas by GLC2000 classification, or are covered by snow most of the year. Gaps in the results for tile H12V8 (figure 7.7, right column) stem from the amount of less than ten available data points over the year, where no analysis is performed. Although the algorithm allows for n harmonic oscillations, depending on the number of data points, generally a maximum of five oscillations is sufficient, by means of the criteria described in the methods chapter, to fit the LAI time series by a superposition of harmonic functions. For the tile H19V4, about 5 %
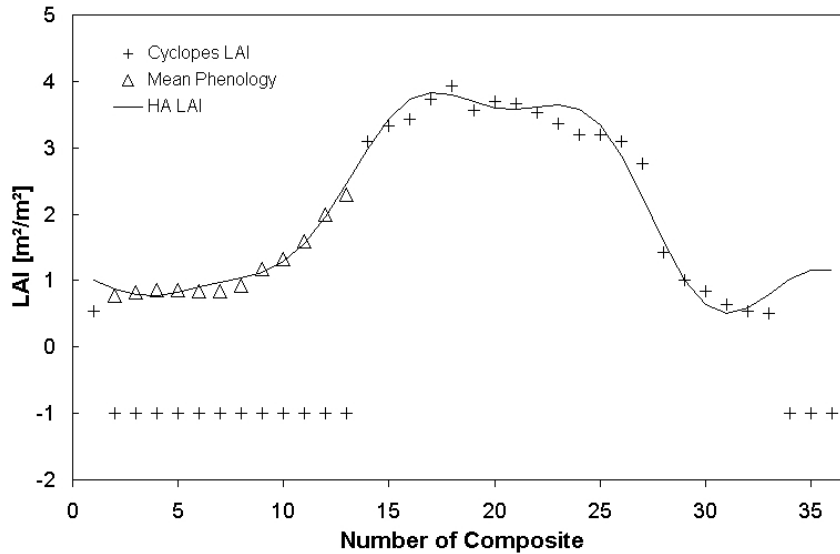
Figure 7.6: *Time series of CYCLOPES (crosses) and HA (solid line) data for one pixel in northern Italy for the year 2001. Triangles show the filled values of the mean phenology for the GLC2000 vegetation class temperate, broadleaved, deciduous trees. A value of -1 represents a data gap.*

of the processed pixel needed the minimum amount of two oscillations, about 25 % were fitted with three, about 46 % with four and about 24 % with five oscillations.

The modified HA was applied to the full, global CYCLOPES LAI archive and the LAI product derived by VITO for geoland2, covering the time period from 1999 to 2007 and the year 2010, respectively. Hence, ten years of LAI data have been processed, resulting in a global, spatio-temporal continuous dataset with a spatial resolution of $1\,\text{km}^2$. Figure 7.9a shows a global map of the LAI for the 18th dataset, representative for end of June 2001. Here the vegetative phase is at its maximum in the Northern Hemisphere and at the minimum in the Southern Hemisphere. LAI values from 0 to $7.9\,\text{m}^2\,\text{m}^{-2}$ are estimated by the modified HA-processor. The effect of having less than ten data points available for computation, as described earlier, can globally be found for Equatorial Guinea, Cameroon and Gabon in Africa and for the Amazon Basin in Brazil, for Guyana, the west coast of Colombia and Ecuador. In these regions huge gaps in the LAI time series result from cloud contamination over most time of the year.
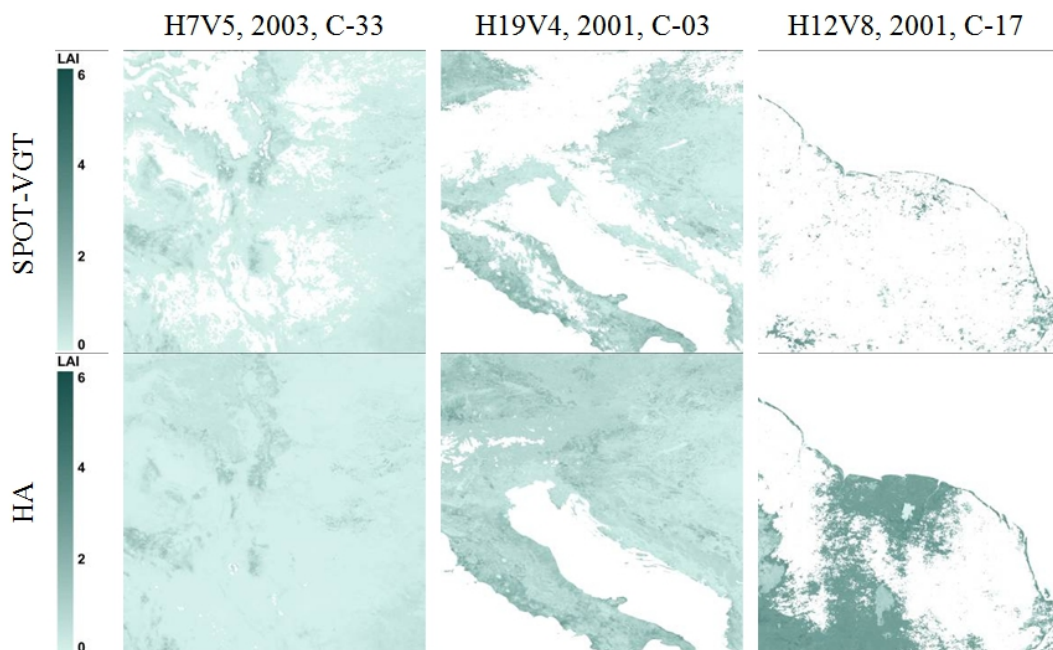
Figure 7.7: *Comparison of representative SPOT-VGT tiles H7V5 (left, North America), H19V4 (middle, Europe) and H12V8 (right, South America) before (first row, SPOT-VGT) and after (second row, HA) applying the analysis to the data. The different tiles and time steps are described above the figure, where C-xx stands for the composite of the year. White spots represent pixel where no LAI data was recorded. Sea masked pixel are also marked as no data but of no interest for LAI investigations.*

In figure 7.9b the global NRMSE is shown for the year 2001. It turns out that the HA-processor produces good results for wide areas, comparable with the original data. The amount of NRMSE values higher than 0.5 for this year is about 1.1 %, with a global mean error of 0.12. For the period 1999 - 2007 and the year 2010 the highest global mean NRMSE is found for 2010 (NRMSE=0.14) with the fraction of values higher than 0.5 of 1.3 %.

To directly assess the quality of our product independently, we compared it to a global set of ground measurements (see table 7.1). These validation points are taken from the datasets of the VALERI and BigFoot projects, available at http://w3.avignon.inra.fr/valeri and http://www.daac.ornl.gov, respectively. The reference maps were produced by extending local ground measurements of true LAI to a wider area, using high spatial resolution satellite images and a specific transfer function (Morisette et al. (2006)). VALERI sites are $9\,km^2$
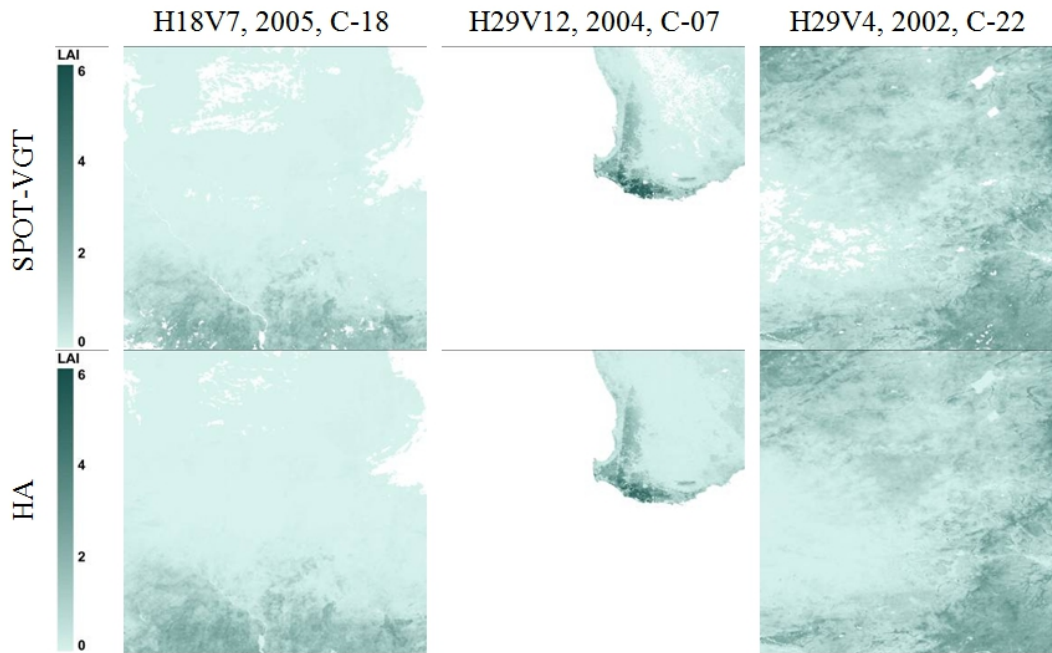
Figure 7.8: *Comparison of representative SPOT-VGT tiles H18V7 (left, North America), H29V12 (middle, Europe) and H29V4 (right, South America) before (first row, SPOT-VGT) and after (second row, HA) applying the analysis to the data. The different tiles and time steps are described above the figure, where C-xx stands for the composite of the year. White spots represent pixel where no LAI data was recorded. Sea masked pixel are also marked as no data but of no interest for LAI investigations.*

$(3\,\mathrm{km} \times 3\,\mathrm{km})$ in size, BigFoot sites cover $25\,\mathrm{km}^2$ ($5\,\mathrm{km} \times 5\,\mathrm{km}$), each site surrounding an eddy flux tower (Cohen et al. (2006)).

Following the guidelines of Morisette et al. (2006) the LAI-Map values were reprojected to the WGS-84 datum of the SPOT-VGT LAI products. Projections from LAI-Map pixels fitting more than 2/3 of the corresponding SPOT-VGT pixel were taken for the quality assessment. Taking into account errors from reprojecting the LAI reference map to the SPOT-VGT projection, the mean values of all pixels were calculated, referred to as Map, CYC and HA in table 7.1, for the reference map values, the SPOT-VGT values and the reprocessed SPOT-VGT values, respectively.
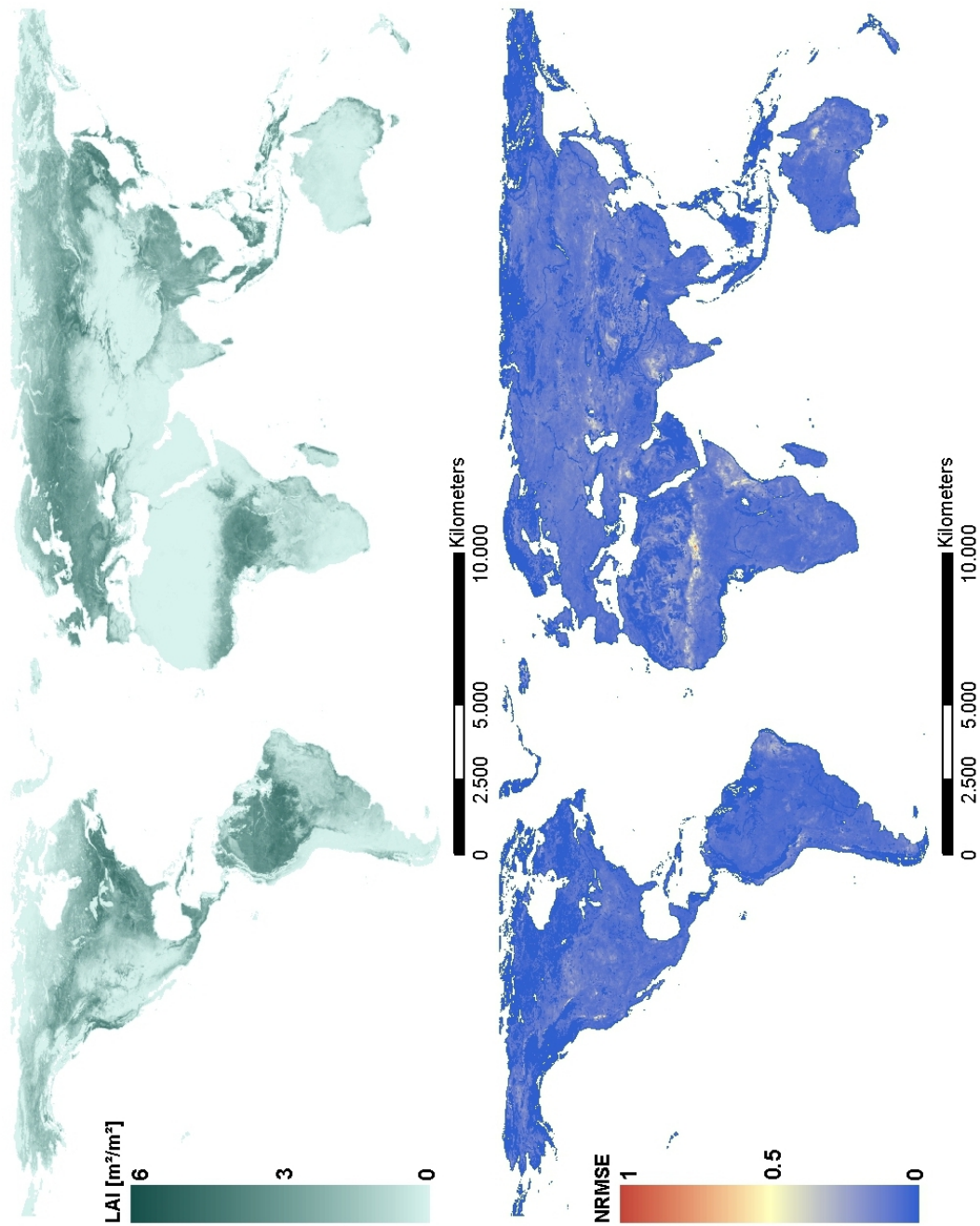
Figure 7.9: *Global maps for (a) the LAI of the 18th composite and (b) the corresponding NRMSE for 2001. White pixels represent no data values.*

Table 7.1: *Characteristics of the validation sites (total 24) and associated reference maps (total 43). Abbreviations in the header are "Lat" for latitude, "Lon" for longitude, "LC" for land cover. "Map" are the mean values of the LAI reference map data, projected to the corresponding SPOT-VGT pixel. "CYC" are the mean values of the SPOT-VGT pixels and "HA" are the mean values of the reprocessed SPOT-VGT pixels. Land cover classes are "EBF" for evergreen broadleaf forest, "MF" for mixed forest and "ENF" for evergreen needleleaf forest.*

Site	Country	Lat	Lon	LC	Date	Map	CYC	HA
AekLoba	Sumatra	2.63	99.58	EBF	2001.06.01	3.3	3.1	3.4
Alpilles	France	43.81	4.71	Crop	2002.07.20	1.4	1.2	1.2
Camerons	Australia	32.6	116.25	EBF	2004.04.06	2.2	1.8	1.7
Counami	French Guiana	5.34	-53.24	EBF	2001.10.18	4.8	2.7	3.1
Demmin	Germany	53.89	13.21	Crop	2004.07.23	4.5	3.3	3.1
Donga	Benin	9.77	1.75	Grass	2005.06.20	1.8	1.4	1.7
Fundulea	Romania	44.41	26.59	Crop	2001.03.17	1.0	0.5	0.4
Fundulea	Romania	44.41	26.59	Crop	2001.05.02	3.3	1.8	1.9
Fundulea	Romania	44.41	26.59	Crop	2002.06.09	1.2	1.5	1.3
Gilching	Germany	48.08	11.32	Crop	2002.07.08	5.5	2.7	2.8
Gnangara	Australia	-31.53	115.88	EBF	2004.03.03	1.0	0.6	0.6
Järvselja	Estonia	58.3	27.26	MF	2000.08.26	3.0	3.0	3.0
Järvselja	Estonia	58.3	27.26	MF	2002.07.13	4.2	2.9	3.2
Nezer	France	44.57	-1.05	ENF	2001.06.20	3.4	2.9	2.9
Nezer	France	44.57	-1.05	ENF	2002.04.21	2.4	1.2	1.5

Site	Country	Lat	Lon	LC	Date	Map	CYC	HA
PlandeDieu	France	44.2	4.95	Crop	2004.06.29	1.3	0.5	0.6
Puechabon	France	43.72	3.65	MF	2001.06.12	3.0	1.9	1.9
Romilly	France	48.45	3.8	Crop	2000.06.05	3.8	3.8	3.8
Sonian	Belgium	50.77	4.41	MF	2004.07.28	5.6	2.6	2.6
Sud-Oest	France	43.51	1.24	Crop	2002.07.20	1.7	2.1	1.8
Turco	Bolivia	-18.24	-68.19	Grass	2001.07.31	0.3	0.1	0.1
Turco	Bolivia	-18.24	-68.19	Grass	2002.08.29	0.1	0.0	0.1
Turco	Bolivia	-18.24	-68.19	Grass	2003.04.25	0.1	0.1	0.2
Wankama	Niger	13.65	2.64	Grass	2005.06.22	0.1	0.1	0.0
Zhang Bei	China	41.28	114.69	Grass	2002.08.23	1.2	1.2	1.3
AGRO	USA	40.01	-88.29	Crop	2000.07.07	2.6	3.0	3.1
AGRO	USA	40.01	-88.29	Crop	2000.08.11	3.2	3.0	3.0
HARV	USA	42.53	-72.17	MF	2002.08.24	4.3	3.8	4.0
KONZ	USA	39.09	-96.57	Grass	2000.06.06	2.2	2.1	2.1
KONZ	USA	39.09	-96.57	Grass	2000.08.26	2.0	1.3	1.3
KONZ	USA	39.09	-96.57	Grass	2001.06.18	2.8	2.4	2.3
KONZ	USA	39.09	-96.57	Grass	2001.08.16	2.6	1.8	1.8
NOBS	Canada	55.89	-98.48	ENF	2000.07.14	3.5	2.4	2.2
NOBS	Canada	55.89	-98.48	ENF	2001.07.14	3.5	2.4	2.4
NOBS	Canada	55.89	-98.48	ENF	2002.07.14	3.2	2.3	2.2

Site	Country	Lat	Lon	LC	Date	Map	CYC	HA
SEVI	USA	34.35	-106.7	Grass	2002.07.26	0.1	0.1	0.1
SEVI	USA	34.35	-106.7	Grass	2002.08.22	0.3	0.1	0.2
SEVI	USA	34.35	-106.7	Grass	2002.09.09	0.4	0.1	0.2
SEVI	USA	34.35	-106.7	Grass	2002.11.15	0.3	0.1	0.1
SEVI	USA	34.35	-106.7	Grass	2003.06.23	0.1	0.1	0.1
SEVI	USA	34.35	-106.7	Grass	2003.07.28	0.1	0.1	0.1
SEVI	USA	34.35	-106.7	Grass	2003.09.15	0.1	0.1	0.1
SEVI	USA	34.35	-106.7	Grass	2003.11.21	0.1	0.0	0.1

For this direct validation a total of 79 LAI values from 40 stations world-wide were analyzed. Differences in land cover classification were considered as false classification in the GLC2000 product, and the corresponding site was excluded from the direct validation. Hence, there were 43 LAI values of 24 stations left to be used for the validation. The land cover types of these stations are evergreen broadleaf forest (EBF), evergreen needleleaf forest (ENF), mixed forest (MF), crops and grassland. The scatter plot of the validation data is shown in figure 7.10. From comparing the LAI data of the original and the reprocessed SPOT-VGT data with the reference map values, it is obvious that there is no significant change in the quality of the data. The R^2 value increases slightly from 0.82 to 0.83 and the calculated RMSE decreases from 0.92 to 0.88. So the good quality of the SPOT-VGT LAI product could be preserved.
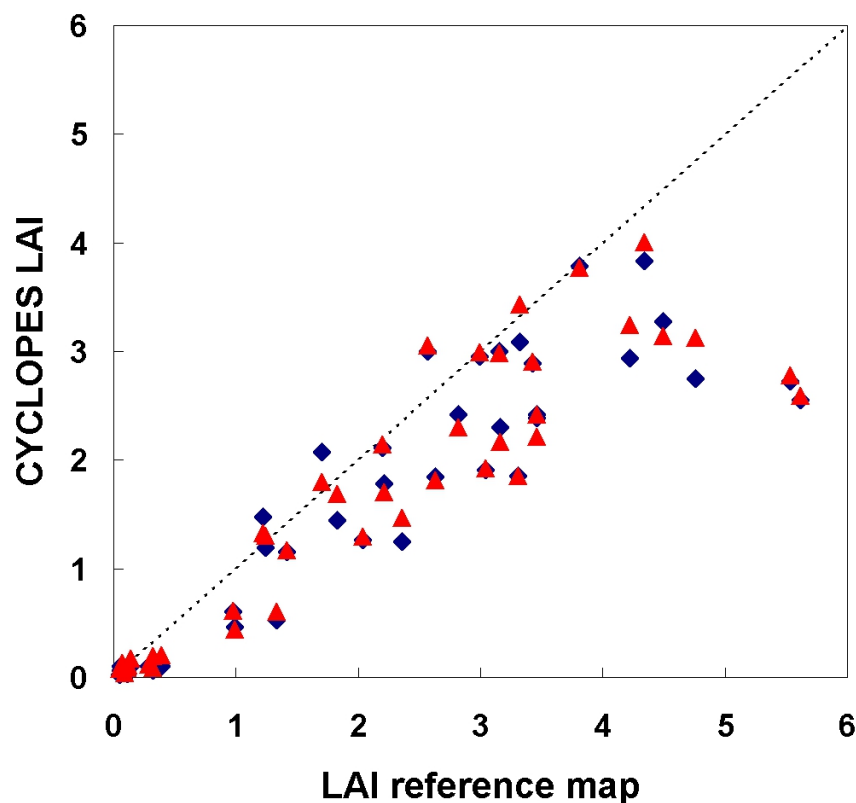
Figure 7.10: *Scatter plot of the direct validation of original (blue diamonds) and reprocessed (red triangles) SPOT-VGT LAI values with results of the LAI reference map.*

To assess the temporal fit of the reprocessed data to the original time series, the development over the nine year period from 1999 to 2007 and for 2010 is shown in figure 7.11, as examples for the main land cover type. The reprocessed time series provides a smooth fit to the SPOT-VGT input data with good agreement with the measurements of the LAI reference maps. However, for the Fundulea site this only applies for the observed value in 2002. The SPOT-VGT as well as the HA values for 2001 deviate from the observed data by a factor of about 2. In Yuan et al. (2011) this is also observed when comparing the reference map data with original and improved MODIS LAI data.
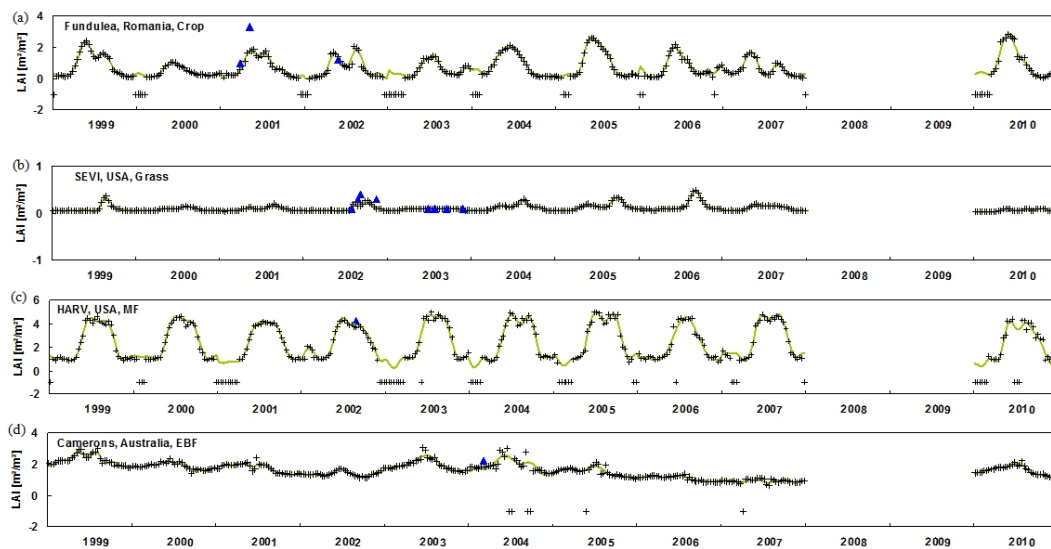
Figure 7.11: *Temporal development of the original (black crosses) and the reprocessed (green solid line) SPOT-VGT data for crop sites (a), grassland sites (b), mixed forest sites (c) and evergreen forest sites (d). Blue triangles mark the reference map LAI values of each site. The site and country names are labelled at each plot.*

Some original data, as the Camerons and Nezer series show short term fluctuations where the fit strongly deviates from the sensor data. Higher peaks with slopes greater than $2\,\mathrm{m^2\,m^{-2}\,10d^{-1}}$ are marked as outliers, but still there are smaller fluctuations left, which seem not to be representative for the natural development of evergreen trees. Here, the HA provides smoothed LAI curves, which, however, show features of these fluctuations.

7.5 Discussion

The modification of the HA and the associated improvement will be discussed in the following. Figure 7.12 compares the results of the original HA-algorithm and the results of the modified algorithm for a pixel in the Harz, Germany, representing a needle forest. Three outliers have been detected marked as filled circles. The large gaps at the beginning and at the end of the year are filled with the phenological mean of a needle forest before the HA is applied. The error bars of the results from the modified HA show the NRMSE of 0.22 of this pixel, calculated from the available input data and applied also to the results, where no input data was present.
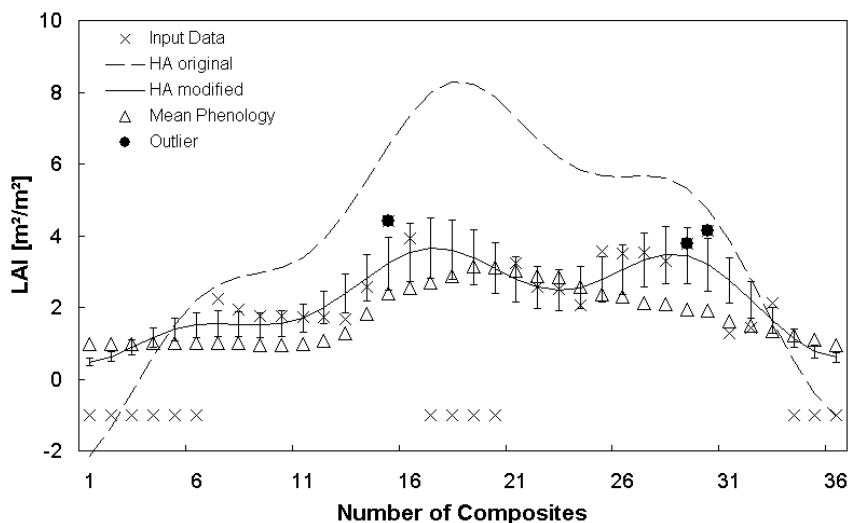
Figure 7.12: *Comparison of applying the original (solid line) and modified (dashed line) HA to the input data (crosses) for a pixel in the Harz, Germany, for 2001. Error bars at the solid line show the NRMSE of the results of the modified HA for this pixel and this year. Triangles show the mean phenology of a needle forest for this pixel, filled circles are detected outliers of the input data. A value of -1 represents a data gap.*

To further discuss this example the amplitudes, phases and periods of the three computed harmonics for the original and the modified HA are listed in table 7.2.

The three oscillations are drawn in figure 7.13, normalized to zero by subtracting the mean LAI value. The most obvious difference relates to the first

Table 7.2: *Amplitudes, phases and periods of the three computed harmonics by the original and the modified HA, for the time series shown in figure 7.12.*

	A_1	φ_1	τ_1	A_2	φ_2	τ_2	A_3	φ_3	τ_3
Original HA	4.82	1.11	7.45	1.01	1.49	1.90	0.05	-1.24	3.81
Modified HA	1.19	2.08	5.85	0.59	1.21	1.89	0.35	2.72	2.87

harmonic, where the original HA fits an amplitude four times higher than the modified HA.
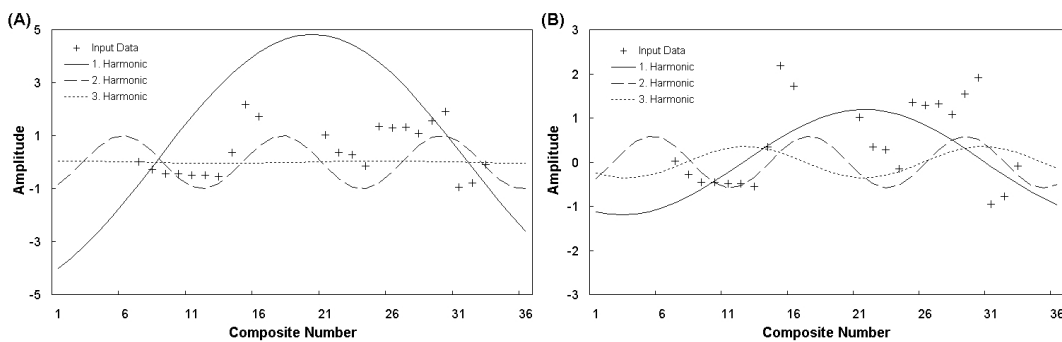
Figure 7.13: *The three harmonics of the fitted oscillation computed by the original (a) and the modified (b) HA, for the time series shown in figure 7.12.*

The period of the first oscillation is also extended by 1.6 composites (i.e. 16 days). The second harmonics have again a difference of about 40 % with the higher amplitude found by the original HA, but correlate in respect to phase and period. The third harmonic of the original setup, with amplitude of 0.05 (figure 7.13a), does not contribute much to the resulting oscillation. However, from the modified setup the third oscillation was found being out of phase with the first two harmonics damping the final curve in the maximum vegetative phase. The comparison of the original and the modified HA shows that large data gaps must be filled with reasonable background values. In our case, the background values for each time step are determined for each land cover class as the mean value calculated for the whole tile. This approach is similar to the "per class mean" introduced by Yuan et al. (2011) for processing MODIS LAI time series. But our approach for background filling can be applied for time series even without the availability of quality control information.

Considering the outliers and filling the bigger data gap with mean phenology data (figure 7.12) we achieve a suitable fitting of the LAI time series over the year. With a NRMSE of 0.22 a higher accuracy is obtained by the modified HA compared to the results obtained with the original HA (dashed line), revealing a NRMSE of 0.92.

This increase in quality is accompanied by an increase of computational demand. By doubling the time series in order to improve stationary and the consequent gap filling method resulted in an increased computing time by a factor of 7 compared to the original HA.

Looking again to the set of representative tiles shown in the last section (see figures. 7 and 8) we computed the absolute difference (SPOT-VGT - HA) of the original and the reprocessed LAI data. The deviation of the reprocessed data was calculated for available values in the SPOT-VGT product and can be seen in the first row of figures. 14 and 15. Most of the pixels do not deviate more than $0.5\,\mathrm{m^2\,m^{-2}}$ with maximum differences of about $\pm\,1\,\mathrm{m^2\,m^{-2}}$. Higher values are observed at coastal lines that could stem from projection errors during the reprocessing. The highest differences can be seen for evergreen broadleaf forest at the Australian south-western coast (tile H29V12 in figure 7.15). It becomes apparent that the fluctuations, described earlier for this land cover class, are not limited spatially. In contrast, the NRMSE for this region again shows low values, describing the smoothing of the time series by the HA. In general, the NRMSE shows good agreement of the reprocessed product with the original SPOT-VGT data over wide areas. High errors are faced in the mountainous regions, as in the Alps, or areas with heterogeneous meteorological conditions, as in the dry areas of Nigeria, which is part of the Sahel. Looking at the global NRMSE map (figure 7.9b) this becomes apparent for the whole transition zone of the Sahel desert. Here precipitation is strongly coupled to monsoonal rain seasons in the south and trade winds in the north (Tetzlaff and Peters (1986)) accompanied with short rainy seasons (3 to 4 months), irregular rain events and long droughts during the rest of the year (Kandij et al. (2006)). Hence, the temporal signature of the measured LAI shows low values (< 0.1) during the dry period with a sharp increase to nearly two orders of magnitude higher values (> 2.5) when the rainy season starts, as can be seen in figure 7.16. With our criterion for the detection of

outliers this sharp increase followed by a decrease is marked as an outlier and thus set to missing value. During the next processing step, the same temporal behaviour is observed for the next LAI values and therefore marked as second outlier. The result of the modified HA shows the very short vegetation period but with reduced maximum LAI of about 1.6 instead of about 3.0.

The fractions of filled gaps for the several tiles and the particular years are shown in the third row of figures. 14 and 15. As an example, the gap filling in tile H19V4 occurred for about 24 % of the pixels. 57 % of these pixels have filled gaps of 5 to 7 data points and 10 % of more than 18 data points, which mainly represent mountainous regions (orange and red areas).

When looking closer to the upper left of the NRMSE map of this tile (Alpine region) it can be seen that for some pixels the modified HA-processor yields unreasonable results. For the tile H19V4 such results can be found mainly in the Alpine region but also for some lakes (Lake Garda in the west and Lake Balaton in the east). A detailed analysis showed a slight shift of about ± 1 pixel between the land water mask of the Land Cover GLC2000 and of the LAI data itself. This shift becomes apparent in figure 7.17, where Lake Balaton and Lake Neusiedl are shown. Overlaying the highly precise land-water mask derived from the Shuttle Radar Topography Mission (SRTM) to the original CYCLOPES data highlights an offset of the land cover map (figure 7.17a) to the north-west. The SRTM Water Body Data (SWBD) includes water bodies of a size less than 100 m with a spatial accuracy of 20 m in horizontal resolution and 16 m in vertical resolution. The observed mismatch between the GLC2000 and the LAI data can lead to low LAI values at the south-eastern shore, obviously measured over water (figure 7.17b). Our approach leads to high discontinuities in the time series. This high variation leads to oscillations with higher amplitude even for the lower original input data, and the HA analysis would produce negative values. This is caught by setting these values to zero (see figure 7.16), leading to an error of 100 % even for very low values causing high NRMSE values at the edges of lakes. Hence, the modified HA tends to produce unreliable results due to mismatch of the land-water mask and the GLC2000 (figure 7.17c). This shift is present for the CYCLOPES products in the period 1999-2003, and obviously has been corrected from 2004 on. According to Carmona-Moreno (2000) and Sylvander et al. (2000) the
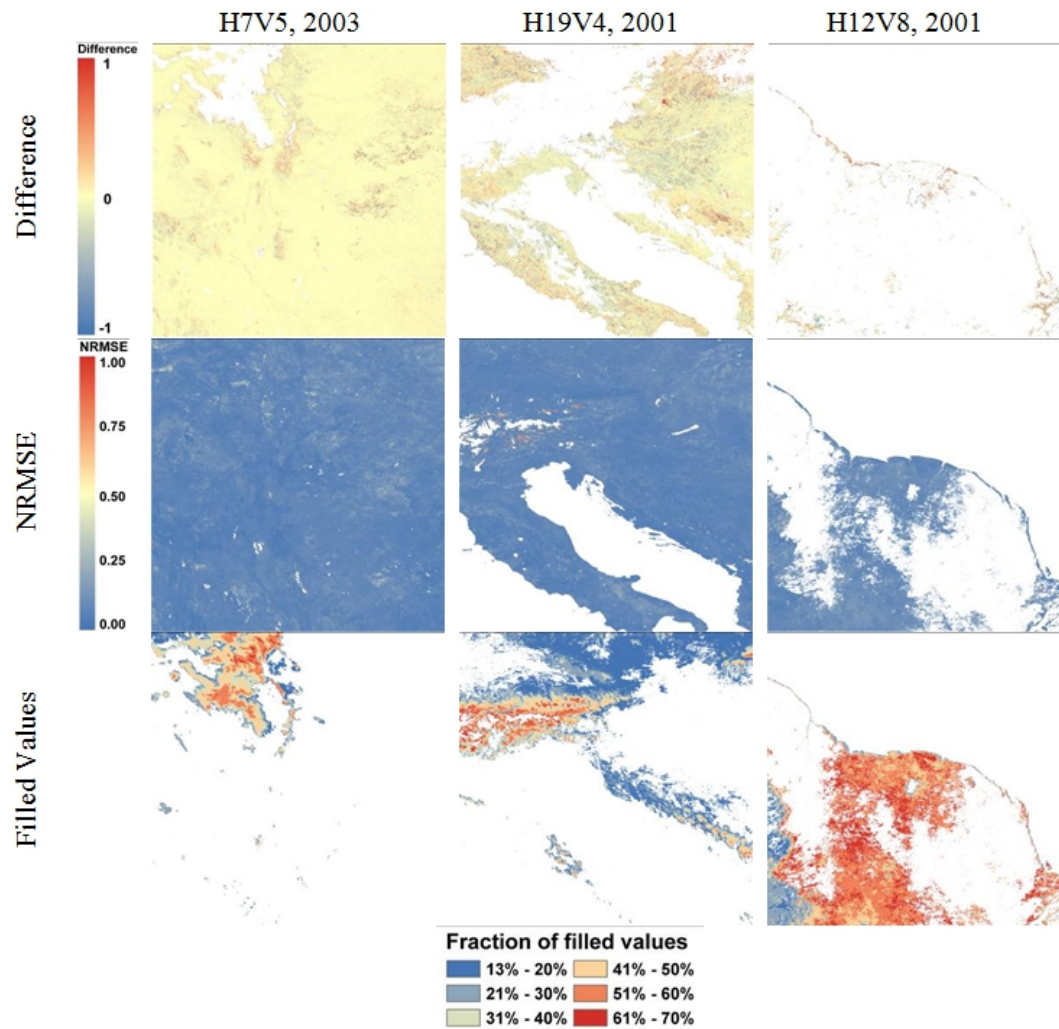
Figure 7.14: *Absolute difference (SPOT-VGT - HA), NRMSE and fraction of filled values for the particular year of the tiles shown in figure 7.7. White pixels show areas where no values have been filled. The different tiles are described above the figure.*
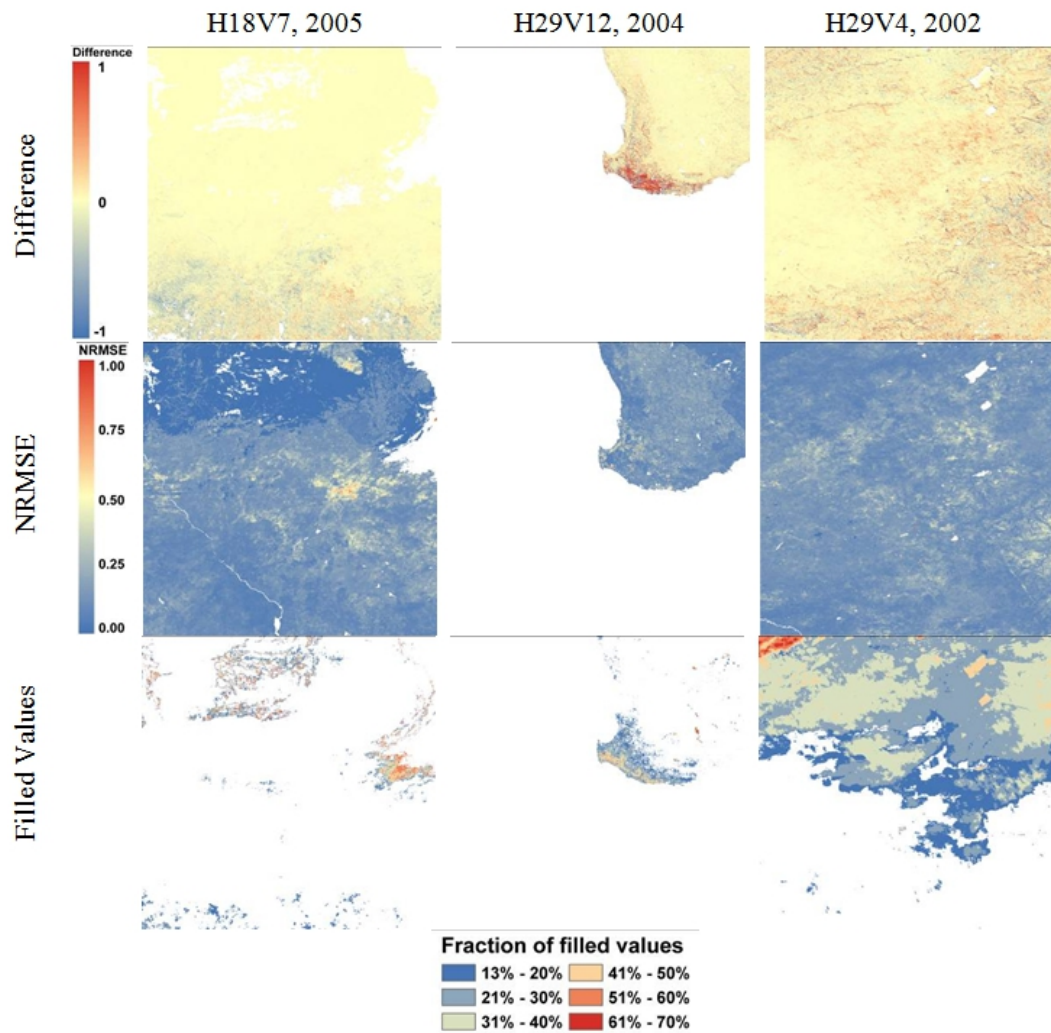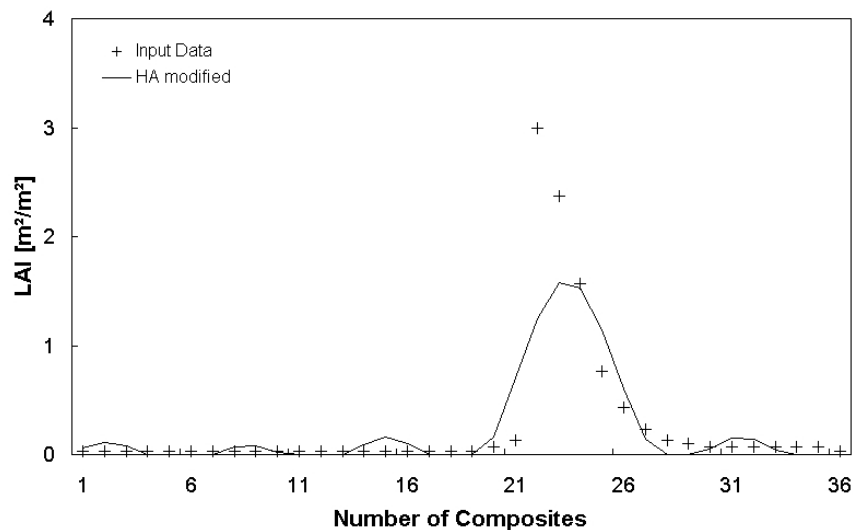
Figure 7.15: *Absolute difference (SPOT-VGT - HA), NRMSE and fraction of filled values for the particular year of the tiles shown in figure 7.8. White pixels show areas where no values have been filled. The different tiles are described above the figure.*

Figure 7.16: *LAI time series for a pixel in the transfer zone of the Sahel desert.*

absolute geolocation accuracy for SPOT-VGT data is estimated to be in the order of 300 m while the multi-temporal registration accuracy is better than 325 m. Sylvander et al. (2000) reported a maximum, multi-temporal registration inaccuracy of about 700 m for 95 % of all distance measurements. Thus a shift of about ± 1 pixel between the LAI data and the GLC2000 is in agreement with the published geolocation accuracy taking into account resampling to rectangular grid map. In addition, the low performance of the modified HA in the Alps can be linked especially with steep terrain and geolocation accuracy.

Focusing on agricultural areas, one encounters two different harvesting methods in tile H19V4. In the temperate northern part agricultural areas are generally harvested once a year, whereas agricultural areas in southern Italy might show two growth cycles for crops. Since the mean phenology is calculated from all agricultural pixels in the entire tile this will lead to a mean phenology which neither reflects one nor two growth cycles. Thus our gap filling procedure will result in higher uncertainties for agricultural areas when different management practices are performed. An improvement of the gap filling technique and thus the modified HA processor could be a regionalization of the mean phenologies for the vegetation classes with respect to e.g. climatic zones instead of calculating them for a full tile.
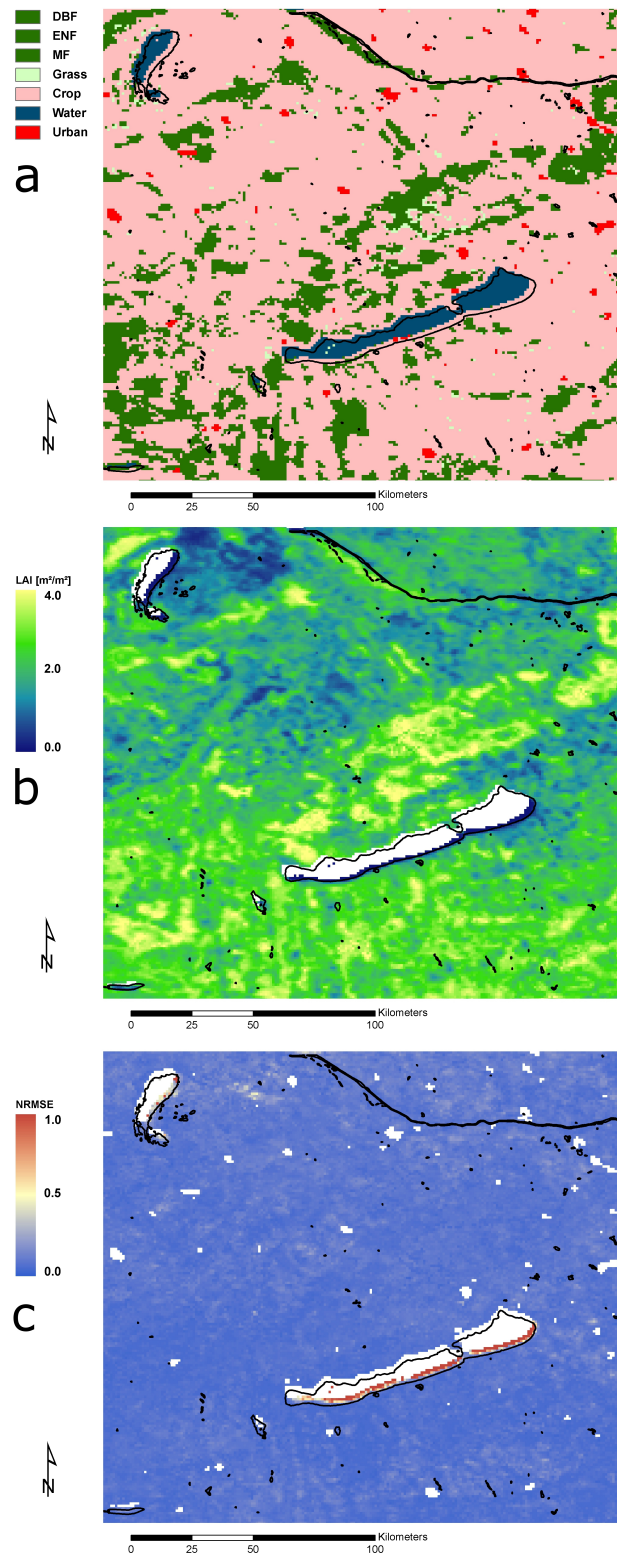
Figure 7.17: *Original land cover (a), LAI (b) and calculated NRMSE (c) for Lake Ballaton (Hungary, centre) and Lake Neusidl (Austria, upper left) with SRTM water mask (black lines). Abbreviations of the land cover types (a) are: DBF - Deciduous Broadleaved Forest; ENF - Evergreen Needleleaved Forest; MF - Mixed Forest.*

In our example, the tile "H19V4" covers several climatic zones according to the Koeppen-Geiger climate classification (Kottek et al. (2006)). The northern and eastern part of the tile (southern Germany, Austria, Hungary) is classified as warm summer continental or hemiboreal climate while Italy, Croatia and Albania are classified as a dry summer subtropical or Mediterranean climate. Nevertheless, gap-free time series of the LAI from SPOT-VGT for greater areas can be achieved by the current setup of the modified HA-processor.

7.6   Conclusions

Cloud contamination, topographic effects and restriction to the solar zenith angle lead to gaps or outliers in the SPOT-VGT LAI time series. To make this data usable for vegetation modelling or other applications requiring continuous data, we applied the technique of the HA to improve this dataset on a global scale by filling and extrapolating the data gaps with reliable values.

For this study we first had to adjust the HA to the characteristics of the biophysical parameter LAI, since the original setup was built for processing time series of ozone information in the stratosphere. The main adaption is the use of a mean phenology to fill greater data gaps, since the method of the HA is not suitable for extrapolating. Then the time series is cleaned from additional outliers that could strongly affect the results of the HA. Modifications to the algorithm itself are the doubling of the input time series to get periodical boundary conditions and the rearrangement of the start values of the periods for saving computing time. Finally the quality factor ϵ as exit condition of the fitting iteration has been improved to get higher accuracy for the resulting fit. With this modified technique we globally computed continuous, gap free time series for the period of 1999 - 2007 and the year 2010. This is, with the reprocessed MODIS LAI product introduced by Yuan et al. (2011), the second of globally available gap free datasets.

The direct validation with the LAI reference map of the BigFoot and VALERI databases shows high accuracy of the product ($R^2 = 0.83$). The comparison of the available input data with the corresponding results of the HA, using the NRMSE, shows good agreement on a global scale, with regional constraints (global mean for 1999-2007 and 2010: NRMSE $= 0.1$). Sudden peaks from

very low values can lead to negative values in the computation of the fitting function. These are set to zero, leading to high values for the NRMSE.

The advantage of the technique of the HA is operational applicability. Furthermore small data gaps of less than five missing data points can be closed without using any additional data for filling. Values for filling bigger data gaps are automatically estimated from the mean phenology of the associated vegetation class using the GLC2000. This dependency to the land cover classification on the other hand makes this approach vulnerable to false classifications or land cover changes.Despite this are methods of time series analysis in general dependent on the quality of the georeferencing of the analyzed data.

Having filled the data gaps in the CYCLOPES and geoland2 SPOT-LAI data, we now can provide a spatiotemporal continuous data set, applicable as phenological input data for global or regional vegetation models. The data is available at http://wdc.dlr.de/.

7.7   Acknowledgments

Chapter 8

Summary and Conclusions

In this thesis project the SVAT model BETHY/DLR was improved and used to estimate sustainable theoretical bio-energy potentials for agricultural and forest areas in Germany and Austria. To assess the quality of the modelled data, two comprehensive approaches were developed to validate the model output with empirical data on land use and biomass increase.

For quality assessment two independent studies were performed, one for agriculture and one for forestry. The years modelled were 2000 and 2001. For agricultural areas, crop yield estimates derived from the national statistics of Germany and Austria were used. It was found that BETHY/DLR slightly overestimates (8 %) the NPP for Austrian districts, while for Germany an underestimation of 17 % was found. A high R^2 (up to 0.79) was observed for both countries, indicating a strong correlation between modelled and empirical yields. The reason for the different results for Germany and Austria was traced to the spatial resolution used ($1 \, km^2$), which does not properly represent the small-scale farm structure in Central Europe, as well as to misclassifications in the land cover dataset used (GLC2000). The validation strategy also included a model comparison with an already validated model (EPIC) for a small area (Marchfeld, Austria). This study revealed that BETHY/DLR responds comparably to EPIC when alternative input data are used, but tends to model more NPP (up to 32 %) in that area than EPIC.

Quality assessment for forest areas in Germany was performed by using mean annual increment data taken from national forest inventories and calculating current annual increments from the modelled NPP. With this method a rea-

171

sonably good agreement was found at the state level (R^2 up to 0.95), but combined with underestimations of up to 34 %. A major reason for this can be seen in the age structure of Germany's forest, which is younger than the global average forest age. BETHY/DLR does not take into account the tree age distribution, but uses photosynthesis-limiting parameters which are valid for the global mean age of untouched forests. The mean age of Canada's forests, for instance, is $\approx$100 years; for the rain forest it is even higher (200 - 300 years). For this reason, the carbon accumulation of younger tree communities could be underestimated. The validation approach was additionally cross-compared with eddy covariance measurements, which supported the finding of underestimation, but also showed good agreement to the annual pattern of GPP (R^2: 0.79).

To estimate sustainable bio-energy potentials, comprehensive approaches were developed and applied. Based on the modelled and validated NPP, sustainable theoretical energy potentials were calculated for both agricultural and forest areas at a $1\,km^2$ resolution. It was assumed that grain and beet yields are not directed towards bio-energy uses; of agricultural crops, only the corn side product straw was modelled as producing bio-energy. Use competitions for straw, such as for animal housing and fertilizer, were taken into account. Following a sustainable approach, for forest areas it was assumed that only the amount of above-ground forest which grew during a year could be used for energy production. Since the end uses of woody biomass are highly diverse, it was chosen to calculate only theoretical potentials, which need further expertise on forest management when they are used. For agriculture, annual energy potentials of $156\,PJ$ to $217\,PJ$ were calculated, which agrees with comparable studies.

During the validation exercises one important model process, the soil water balance, was identified as having potential for improvement. The simple "bucket model" formulation which was used to describe the soil water balance only allowed one soil layer per grid cell. This approach is widely used in vegetation models, but is discussed controversial because of its potential to predict unrealistic water availability. To improve on this, a new one-dimensional water transport model, based on the van Genuchten pedotransfer formulation, was applied. For all 128 FAO soil types the needed van Genuchten parameters were estimated. Furthermore, the characteristics of individual soil layers, such as

their depths and grain size distributions, were calculated based on more than 10,000 soil core measurements. With this modelling approach more realistic soil water predictions will be possible in a future BETHY/DLR version.

The techniques employed and the results achieved in this study will help to increase the level of confidence that decision-makers can place in model results. As models gain in accuracy, and are validated more thoroughly, reliable simulations of bio-geo-physical processes will gain in economic and political importance. In particular, more reliable forecasts will be of great consequence for the estimation of the impacts of global climate change upon vegetation species distributions, food security, and the availability of renewable energy sources.

Bibliography

Anderson, J. M. (1992). Responses of soils to climate change. In Woodward, F. I., editor, *The Ecological Consequences of Global Climate Change*, volume 22, pages 163–211.

Anderson, M. C., Norman, J. M., Diak, G. R., Kustas, W. P., and Mecikalski, J. R. (1997). A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sensing of Environment*, 60:195–216.

Anlauf, R. K., Kersebaum, C., Ping, L. Y., Nuske-Schüler, A., Richter, J., Springob, G., Syring, K. M., and Uterman, J. (1987). *Modelle für Prozesse im Boden: Programme und Übungen*. Ferdinand Enke Verlag Stuttgart.

Azzali, S. and Menenti, M. (2000). Mapping vegetation-soil-climate complexes in southern africa using temporal fourier analysis of NOAA-AVHRR NDVI data. *International Journal of Remote Sensing*, 21:973–996.

Baldocchi, D. (1997). FLUX footprints within and over forest canopies. *Boundary-Layer Meteorology*, 2:273–293.

Baldocchi, D., Falge, E., Gu, L. H., Olsen, R., Hollinger, D., Running, S. W., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechtel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S. (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.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Nino, F., Weiss, M., Samain, O., Roujean, J. L., and Leroy, M. (2007). LAI, fAPAR and FCOVER CYCLOPES global products derived from vegetation. part 1: Principles of the algorithm. *Remote Sensing of Environment*, 110:275–286.

Bartholome, E. and Belward, A. S. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, 26:1959–1977.

Bartholome, E., Belward, A. S., and Achard, F. (2002). GLC2000: Global land cover mapping for the year 2000. *EUR 20524 EN, European Commission, Luxemburg.*

Batjes, N. H. (2009). Harmonized soil profile data for applications at global and continental scales: updates to the wise database. *Soil Use and Management*, 25:124–127.

Battaglia, M., Sands, P. J., White, D., and Mummery, D. (2004). CABALA: a linked carbon, water and nitrogen model of forest growth for silvicultural decision support. *Forest Ecological and Management*, 193:251–282.

Beheydt, D., Boeckx, P., Sleutel, S., Li, C., and van Cleemput, O. (2007). Validation of DNDC for 22 long-term n_2o field emission measurements. *Atmospheric Environment*, 41:6196–6211.

Bittner, M. (1993). *Langperiodische Temperaturoszillationen in der unteren und mittleren Atmosphäre (0 bis 100 km) während der DYNA-Kampagne.* PhD thesis, Bergische Universität; Gesamthochschule Wuppertal, Fachbereich 8; Physik.

Bittner, M. (1997). Value adding Produkte bei GOME. Technical report, DLR: Oberpfaffenhofen.

Bittner, M., Dech, S., and Balzer, W. (1998). Ozonkartierung per Satellit. *Spektrum der Wissenschaft*, 9:54–61.

Bittner, M., Offermann, D., Buaeva, I., Kokin, G., Koshelkov, J., Krivolutsky, A., Tarasenko, D., Gil-Ojeda, M., Hauchecorne, A., Lübken, F. J., De la

Morena, B., Mourier, A., Nakane, H., Oyama, K., Schmidlin, F., Soule, I., Thomas, L., and Tsuda, T. (1994). Long period/large scale oscillations of temperature during the DYANA campaign. *Journal of Atmospheric and Terrestrial Physics*, 56:1675–1700.

BMELV (2004). Die zweite Bundeswaldinventur (BWI2) Das wichtigste in Kürze. *Bonn.*

BMU (2011). *Erneuerbare Energien in Zahlen - Nationale und internationale Entwicklung.* http://www.erneuerbare-energien.de/inhalt/2720.

BMU (2012). Verordnung über die Erzeugung von Strom aus Biomasse (Biomasseverordnung - BiomasseV). *http://www.erneuerbare-energien.de/inhalt/2671/.*

BMVBS (2010). Globale und regionale Verteilung von Biomassepotentialen. *BMVBS-online publication 27/10.*

Bolinder, M. A., Angers, D. A., and Dubuc, J. (1997). Estimating shoot to root ratios and annual carbon inputs in soil for cereal crops. *Agriculture and Ecosystem Environment*, 63:61–66.

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., and Smith, B. (2007). Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, 13:679–706.

Borresen, T. (1999). The effect of straw management and reduced tillage on soil properties and crop yields of spring-sown cereals on two loam soils in norway. *Soil and Tillage Research*, 51:91–102.

Bossard, M., Feranec, J., and Otahel, J. (2000). Corine land cover technical guide: Addendum 2000. *European Enviroment Agency, Copenhagen.*

Boulet, G., Chehbouni, A., Braud, I., Vauclin, M., Haverkamp, R., and Zammit, C. (2000). A simple water and energy balance model designed for regionalization and remote sensing data utilization. *Agricultural and Forest Meteorology*, 105:117–132.

Braden, H. (1985). Ein Energiehaushalts- und Verdunstungsmodell für Wasser- und Stoffhaushaltsuntersuchungen landwirtschaftlich genutzter Einzugsgebiete. *Mitteilungen Deutsche Bodenkundliche Gesellschaft*, 22:294–299.

Braden, H. (1995). *The Model AMBETI: A Detailed Description of a Soil-Plant-Atmosphere Model*. Deutscher Wetterdienst.

Brooks, R. H. and Corey, A. T. (1964). Hydraulic properties of porous media. hydrology paper no. 3. *Colorado State University, Fort Collins, CO*.

Bruijnzeel, L. A. (1990). Hydrology of moist tropical forests and effects of conversion. a state-of-knowledge review. *UNESCO International Hydrological Programme, Paris*.

Burridge, D. M. and Gadd, A. J. (1974). The meteorological office operational 10-level numerical weather prediction model. *British Meteorological Office Tech. Notes Nos. 12 and 48*.

Cai, Z., Sawamoto, T., Li, C., Kang, G., Boonjawat, J., Mosier, A., Wassmann, R., and Tsuruta, H. (2003). Field validation of the DNDC model for greenhouse gas emissions in east asian cropping systems. *Global Biogeochemical Cycles*, 17:1–10.

Calder, I. (1986). A stochastic model of rainfall interception. *Journal of Hydrology*, 89:65–71.

Campbell, G. S. (1974). A simple method for determining unsaturated hydraulic conductivity from moisture retention data. *Soil Science*, 117:311–314.

Carmona-Moreno, C. (2000). Space-VEGETATION software: the software for processing VEGETATION L-Band images. In *Proc. of VEGETATION 2000 Symposium*, pages 295–300.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., and Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, 91:332–344.

Cohen, W. B., Maiersperger, T. K., and Pflugmacher, D. (2006). BigFoot Leaf Area Index surfaces for North and South American sites, 2000-2003. *Oak Ridge National Laboratory, Tennessee, U.S.A.*

Colditz, R. R., Conrad, C., Wehrmann, T., Schmidt, M., and Dech, S. (2008). TiSeG: A flexible software tool for time-series generation of MODIS data utilizing the quality assessment science data set. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(10):3296–3308.

Collatz, G. J., Ribas-Carbo, M., and Berry, J. A. (1992). Coupled photosynthesis- a stomatal conductance model for leaves of C4 plants. *Australian Journal of Plant Physiology*, 19:519–538.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J. (1999). The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics*, 15:183–203.

de Rosnay, P. and Polcher, J. (1998). Modelling root water uptake in a complex land surface scheme coupled to a GCM. *Hyrology and Earth System Sciences*, 2:239–255.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., and Loveland, T. R. (2000). A new global 1-km dataset of percentage tree cover derived from remote sensing. *Gobal Change Biology*, 6:247–254.

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Hartman, M. D., Bremmer, J., Ojima, D. S., and Schimel, D. S. (2001). Simulated interaction of carbon dynamics and nitrogen trace gas fluxes using the DAYCENT model. In Schaffer, M., Ma, L., and Hansen, L., editors, *Modeling carbon and nitrogen dynamics for soil management*, pages 303–332. Boca Raton, Florida: CRC Press.

Dieter, M. and Englert, H. (2001). Abschätzung des Rohholzpotenzials für die energetische Nutzung in der Bundesrepublik Deutschland. *Arbeitsbericht des Instituts für Ökonomie 2001/11, Hamburg.*

DiGregorio, A. and Jansen, L. J. M. (2001). Land cover classification system (LCCS): classification concepts and user manual for software version 1.0. *Rome: United Nations Food and Agricultural Organization.*

Dilmaghani, S., Henry, I. C., Soonthornnonda, P., Christensen, E. R., and Henry, R. C. (2007). Harmonic analysis of environmental time series with missing data or irregular sample spacing. *Environmental Science and Technology*, 41(20):7030–7038.

Dyck, S., Becker, A., Flemming, G., Glugla, G., Golf, W., Grünewald, U., Gurtz, J., Kluge, C., and Peschke, G. (1980). *Angewandte Hydrologie. Teil 1: Berechnung und Regelung des Durchflusses der Flüsse. Teil 2: Der Wasserhaushalt der Flußgebiete.* Verlag für Bauwesen, Berlin.

Eagleson, P. S. (1978). Climate, soil, and vegetation 1-6. *Water Resource Research*, 5:705–764.

Ebertseder, T., Eyring, V., Bittner, M., Dameris, M., and Grewe, V. (2006). Hemispheric ozone variability indices derived from satellite observations and as diagnostics for coupled chemistry-climate models. *Atmospheric Chemistry and Physics*, 6:5105–5120.

EEA (2006). The thematic accuracy of corine land cover 2000: Assessment using LUCAS (land use/cover area frame statistical survey). *EEA Technical Report*, 7:1–90.

Eidenshink, J. C. (1992). The 1990 conterminous U.S. AVHRR data set. *Photogrammetric Engineering and Remote Sensing*, 58:809–813.

Erteld, W. (1963). *Ertragstafelauszüge, für den Gebrauch in der Praxis 2nd Edition.* Neumann Verlag Radebeul.

Evans, S. P. and Trevisan, M. (1995). A soil water-balance 'bucket' model for paleoclimatic purposes 1. model structure and validation. *Ecological Modelling*, 82:109–129.

FAO (1974). United Nations Educational, Scientific and Cultural Organization (FAO-UNESCO), Soil Map of the World, Volume I: Legend. *UNESCO, Paris, France.*

FAO (1988). FAO-UNESCO soil map of the world, revised legend, with corrections and updates. *World Soil Resources Report 60, FAO, Rome.*

FAO/IIASA (2009). Harmonized soil database (version 1.1). *FAO, Rome, Italy and IIASA, Laxenburg, Austria.*

Farquhar, G. D., Caemmerer, S., and Berry, J. A. (1980). A biochemical model of photosynthetic CO_2; assimilation in leaves of C3 species. *Planta*, 149:78–90.

Federer, C. A. (1979). A soil-plant-atmosphere model for transpiration and availability of soil water. *Water Resource Research*, 15:555–562.

Fritsche, U. R., Dehoust, G., Jenseit, W., Hünecke, K., Rausch, L., Schüler, D., Wiegmann, K., Heinz, A., Hiebel, M., Ising, M., Kabasci, S., Unger, C., Thrän, D., Fröhlich, N., Scholwin, F., Reinhardt, G., Gärtner, S., Patyk, A., Baur, F., Bemmann, U., Groß, B., Heib, M., Ziegler, C., Flake, M., Schmehl, M., and Simon, S. (2004). Stoffstromanalyse zur nachhaltigen energetischen Nutzung von Biomasse: Endbericht. *BMU.*

Fritz, S., Bartholome, E., Belward, A., Hartley, A., Stibig, H., Eva, H., Mayaux, P., Bartalev, S., Latifovic, R., Kolmert, S., Roy, P., Agrawal, S., Bingfang, W., Wenting, X., Ledwith, M., Pekel, F., Giri, C., Mücher, S., de Badts, E., Tateishi, R., Champeaux, J. L., and Defourny, P. (2003). Harmonisation, mosaicing and production of the Global Land Cover 2000 database (beta version). *EUR 20849 EN, European Commission, Directorate General, Joint Research Centre (JRC).*

Gao, F., Morisette, J. T., Wolfe, R. E., Ederer, G., Pedelty, J., Masuoka, E., Myneni, R., Tan, B., and Nightingale, J. (2008). An algorithm to produce temporally and spatially continuous MODIS-LAI time series. *Geoscience and Remote Sensing Letters, IEEE*, 5(1):60–64.

Gardner, W. R. (1958). Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table. *Soil Science*, 85:228–232.

Garrigues, S., Lacaze, R., Baret, F., Morisette, J. T., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., Knyazikhin, Y., and Yang, W. (2008). Validation and intercomparison of

global Leaf Area Index products derived from remote sensing data. *Journal of Geophysical Research (Biogeosciences)*, 113:G02028.

Gash, J. (1979). An analytical model of rainfall interception by forest. *Quarterly Journal if the Royal Meteorological Society*, 105:43–55.

Gauder, M., Graeff-Hönninger, S., and Claupein, W. (2011). Identifying the regional straw potential for energetic use on the basis of statistical information. *Biomass and Bioenergy*, 35:1646–1654.

Günther, K. P. and Maier, S. W. (2007). AVHRR compatible vegetation index derived from MERIS data. *International Journal of Remote Sensing*, 28:693–708.

Gobron, N., Aussedat, O., Pinty, B., Taberner, M., and Verstraete, M. M. (2004). Medium Resolution Imaging Spectrometer (MERIS) - an optimized fAPAR algorithm - theoretical basis document. *European Commission - DG Joint Research Centre, Institute for Environment and Sustainability*.

Goetz, S. J., Prentice, S. D., Goward, S. N., Thawley, M. M., and Small, J. (1999). Satellite remote sensing of primary production: an improved production efficiency modelling approach. *Ecological Modelling*, 122:239–255.

Golub, G. and Van Loan, C. (1996). *Matrix computations*. Baltimore: Johns Hopkins University Press.

Grammel, R. H. (1989). *Forstbenutzung*. Paul Parey, Hamburg, Berlin.

Guetter, P. J. and Kutzbach, J. E. (1990). A modified Köppen classification applied to model simulations of glacial and interglacial climates. *Climate Change*, 16:193–215.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Caroll, M., Dimiceli, C., and Sohlberg, R. A. (2003). Global pecent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm,. *Earth Interactions*, 7:1–15.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R. A., Dimiceli, C., and Caroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, 83:303–319.

Haxeltine, A. and Prentice, I. (1996). BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochemical Cycles*, 10(4):693–709.

Heinsch, F. A., Zhao, M., Running, S. W., Kimball, J. S., Nemani, R. R., Davis, K. J., Bolstad, P. V., Cook, B. D., Desai, A. R., Ricciuto, D. M., Law, B. E., Oechel, W. C., Kwon, H., Luo, H., Wofsy, S. C., Dunn, A. L., Munger, J. W., Baldocchi, D. D., Xu, L., Hollinger, D. Y., Richardson, A. D., Stoy, P. C., Siqueira, M. B. S., Monson, R. K., Burns, S. P., and Flanagan, L. B. (2006). Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. *IEEE Transactions on Geoscience and Remote Sensing*, 44:1908–1926.

Hird, J. N. and McDermid, G. J. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113:248–258.

Hofreither, M., Eder, M., Feichtinger, F., Kniepert, M., Liebhard, P., Salhofer, K., Schmid, E., Sinabell, F., and Streicher, G. (2000). Modellanalyse von ökonomischen Instrumenten zum Grundwasserschutz im Zusammenhang mit dem ÖPUL-Programm. Endbericht zum Forschungsprojekt Nr. 1133 im Auftrag des BMLF und BMUJF, Wien. *Institute for Sustainable Economic Development, University of Natural Resources and Life Sciences*.

Houghton, J. T., Filho, L. G. M., Lim, B., Treanton, K., Mamaty, I., Bonduki, Y., Griggs, D. J., and Callander, B. A. (1997). Revised 1996 intergovernmental panel on climate change guidelines for national greenhouse inventories. *IPCC/OECD/IEA*.

Houghton, R. A., Butman, D., Bunn, A. G., Krankina, O. N., Schlesinger, P., and Stone, T. A. (2007). Mapping russian forest biomass with data from satellites and forest inventories. *Environmental Resarch Letters*, 2:7–13.

Huete, A., Justice, C. O., and van Leewen, W. J. D. (1999). MODIS vegetation index (MOD 13). *Algorithm Theoretical Basis Document (ATBD), Version 3.0.*

Hutson, J. L. and Cass, A. (1987). A retentivity function for use in soil-water simulation models. *Journal of Sil Science*, 38:105–113.

IPCC (2007). *Climate Change 2007: Impacts, Adaptions and Vulnerability. Summay for Policymakers.* Intergovernmental Panel on Climate Change.

Izaurralde, R. C., Williams, J. R., McGill, W. B., Rosenberg, N. J., and Quiroga, M. C. (2006). Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modelling*, 192:362–384.

Jackson, R. B., Canadell, J., Ehleringer, J. R., Mooney, H. A., Sala, O. E., and Schulze, E. D. (1996). A global analysis of root distributions for terrestrial biomes. *Oecologia*, 108:389–411.

Jakeman, A. J., Letcher, R. A., and Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, 21:602–614.

Jakubauskas, M., Peterson, D., and Legates, D. (2002). Fourier decomposition of an AVHRR NDVI time series for seasonal and interannual land cover change detection. In Buzzone, L. and Smits, P., editors, *Analysis of Multitemporal Remote Sensing Images, Series in Remote Sensing*, pages 387–394. Singapore, World Scientific.

Jarvis, N. J., Zavattaro, L., Rajkai, K., Reynolds, W. D., Olsen, P. A., McGechan, M., Mecke, M., Mohanty, B., Leeds-Harison, P. B., and Jacques, D. (2000). Indirect estimation of near-saturated hydraulic conductivity from readily available soil information. *Geoderma*, 108:1–17.

Jönsson, P. and Eklundh, L. (2002). Seasonality extraction by function fitting to time series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8):1824–1832.

Jönsson, P. and Eklundh, L. (2004). TIMESAT: a program for analyzing time series of satellite sensor data. *Computers and Geosciences*, 30:833–845.

Jones, C. A., Dyke, P. T., Williams, J. R., Kiniry, J. R., Benson, V. W., and Griggs, R. H. (1991). EPIC: An operational model for evaluation of agricultural sustainability. *Agricultural Systems*, 37:341–350.

Jones, C. A. and Kiniry, J. R. (1986). *CERES-Maize: a simulation model of maize growth and development.* Texas A & M University Press.

JRC (2006). *Monitoring Agriculture with Remote Sensing - Bulletin. Review of: 2005-2006 Campaign and Situation: 1 September to 10 November 2006*, volume 14. Joint Research Centre of the European Commission.

JRC (2007). *Monitoring Agriculture with Remote Sensing - Crop Monitoring in Europe. Review of: 2006-2007 Campaign and Situation: 11st September to 10th November 2007*, volume 15. Joint Research Centre of the European Commission.

Jung, M. (2008). *Uncertainties of terrestrial carbon cycle modelling: Studies on gross carbon uptake of Europe.* PhD thesis, University Hamburg.

Kaltschmitt, M. and Hartmann, H., editors (2001). *Energie aus Biomasse - Grundlagen, Techniken und Verfahren.* Berlin, Heidelberg: Springer-Verlag.

Kandij, S. T., Verchot, L., and Mackensen, J. (2006). Climate change and variability in the sahel region: impacts and adaptation strategies in the agricultural sector. *UNEP and ICRAF*.

Köhler, B. and Kolbe, H. (2007). *Nährstoffgehalte der Fruchtarten im ökologischen Landbau.* Sächsische Landesanstalt für Landwirtschaft, Fachbereich Pflanzliche Entwicklung, Leipzig.

Knorr, W. (1997). *Satellite Remote Sensing and Modelling of the Global CO_2 Exchange of Land Vegetation: A Synthesis Study.* PhD thesis, Faculty of Earth Sciences of the University of Hamburg, Hamburg.

Knorr, W. and Heimann, M. (2001). Uncertainties in global terrestrial biosphere modeling, part i: A comprehensive sensitivity analysis with a new

photosynthesis and energy balance scheme. *Global Biogeochemical Cycles*, 15(1):207–225.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the Köppen Geiger climate classification updated. *Meteorologische Zeitschrift*, 15:259–263.

Köppen, W. (1936). Das Geographische System der Klimate. In Köppen, W. and Geiger, R., editors, *Handbuch der Klimatologie, Vol. I, part C.* Gebrüder Borntreger, Berlin.

Kramer, H. and Krüger, H. H. (1981). Vorrat und Nutzungsmöglichkeiten forstlicher Biomasse in der Bundesrepublik Deutschland. *Der Forst- und Holzwirt*, 36:33–37.

Krinner, G., Viovy, N., de Noblet-Decoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 19:1–33.

KTBL (2005). *Faustzahlen für die Landwirtschaft*. Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V., Darmstadt.

Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., and Monod, H. (2009). Multivariate global sensitivity analysis for dynamic crop models. *Field Crop Research*, 113:312–320.

Lang, R. H., Vine, D. M., and Utku, C. (2004). Estimation of forest biomass from ESTAR image data. *XXth ISPRS Congress Istanbul 2004, Proceedings of Commission VII*.

Larocque, G. R., Bhatti, J. S., Gordon, A. M., Luckai, N., Wattenbach, M., Liu, J., Peng, C., Arp, P. A., Liu, S., Zhang, C. F., Komarov, A., Grabarnik, P., Sun, J., and T.White (2008). Uncertainty and sensitivity issues in process-based models of carbon and nitrogen cycles in terrestrial ecosystems. *Developments in Integrated Environmental Assessment*, 3:307–327.

Li, C., Frolking, S., and Frolking, T. A. (1992). A model of nitrous oxide evolution from soil driven by rainfall events: 1. model structure and sensitivity. *Journal of Geophysical Research*, 97:9769–9776.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462.

Luyssaert, S., Ciais, P., Piao, S. L., Schulze, E. D., Jung, M., Zaehle, S., Schelehaas, M. J., Reichstein, M., Churkina, G., Papale, D., Abril, G., Beer, C., Grace, J., Loustau, D., Matteucci, G., Magnani, F., Nabuurs, G. J., Verbeeck, H., Sulkava, M., van der Werf, G. R., and Janssens, I. A. (2010). The european carbon balance. part 3: forests. *Global Change Biology*, 16:1429–1450.

Massman, W. (1983). The derivation and validation of a new model fort he interception of rainfall by forest. *Agricultural Meteorology*, 28:261–286.

Mayaux, P., Hugh, E., Gallego, J., Strahler, A. H., Herold, M., Agrawal, S., Naumov, S., Miranda, E. E. D., Bella, C. M. D., Ordoyne, C., Kopin, Y., and Roy, P. S. (2006). Validation of the Global Land Cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing*, 44:1728–1739.

Mayer, T. and Jarvis, N. J. (1999). Pedotransfer functions to estimate soil water retention parameters for modified brooks-corey type model. *Geoderma*, 91:1–9.

McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., and Nilsson, S. (2010). Comparison of four global fAPAR datasets over Northern Eurasia for the year 2000. *Remote Sensing of Environment*, 114(5):941–949.

McMaster, G. S. and Wilhelm, W. W. (1997). Growing degree-days: one equation, two interpretations. *Agriculture and Forest Meteorology*, 87:291–300.

Meisner, R. E., Bittner, M., and Dech, S. W. (1999). Computer animation of remote sensing based time series data sets. *IEEE Transactions on Geoscience and Remote Sensing*, 37:1100–1106.

Minasny, B. and McBratney, A. B. (2000). Hydraulic conductivity pedotransfer functions for australian soil. *Australian Journal of Soil Research*, 38:905–926.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hui, D., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J. (2007). Comprehensive comparison of gap filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology*, 147(3-4):209–232.

Monsi, M. and Saeki, T. (1953). Über den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung für die Stoffproduktion. *Japanese Journal of Botany*, 14:22–52.

Monteith, J. L. (1965). Evaporation and environment in the state and movement of water in living organisms. In Fogy, G., editor, *Symposia of the society for experimental biology*, volume 19, pages 205–34.

Morisette, J. T., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J. E., Garrigues, S., Shabanov, N. V., Weiss, M., Fernandes, R. A., Leblanc, S. G., Kalacska, M., Sanchez-Azofeifa, G. A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A. N., Lewis, T. E., Iiames, J. S., Colombo, R., Meroni, M., Busetto, L., Cohen, W. B., Turner, D. P., Warner, E. D., Petersen, G. W., Seufert, G., and Cook, R. (2006). Validation of global moderate resolution LAI products: a framework proposed within the CEOS land product validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1804–1817.

Musial, J. P., Verstraete, M. M., and Gobron, N. (2011). Technical note: Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. *Atmospheric Chemistry and Physics*, 11(15):7905–7923.

Muzylo, A., Llorens, P., Valente, F., Keizer, J. J., Domingo, F., and Gash, J. H. C. (2009). A review of rainfall interception modelling. *Journal of Hydrology*, 370:191–206.

Myneni, R. B., Dong, J., Tucker, C. J., Kaufmann, R. K., Kauppiu, P. E., Liski, J., Zhou, L., Alexeyev, V., and Hughes, M. (2001). A large carbon sink in the woody biomass of northern forests. *Proceedings of the National Academy of Science*, 98:14784–14789.

Myneni, R. B., Hall, F. G., Sellers, P. J., and Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on, Geoscience and Remote Sensing*, 33(2):481–486.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83(1-2):214–231.

Niklaus, M., Günther, K. P., Tum, M., and Bittner, M. (2012). Generation of a global, gap-free spot-vgt lai dataset using spectral analysis techniques. *International Journal of Remote Sensing*, under review.

Ortega, J. and Rheinbold, W. (1970). Iterative solution of nonlinear equations in several variables. *New York, Academic Press*.

Pacan, B. and Dröge, S. (2010). Abschlussbericht zum Forschungsprojekt: Stoffliche und energetische Nutzung von Stroh. *PFI, Pirmasens*.

Pachepsky, Y. A. and Rawls, W. J. (2005). *Development of Pedotransfer Functions in Soil Hydrology*. Elsevier Science, Amsterdam.

Parry, M. (1992). The potential effect of climate changes on agriculture and land use. In Woodward, F. I., editor, *The Ecological Consequences of Global Climate Change*, volume 22, pages 63–92.

Parton, W. J., Hartman, M. D., Ojima, D. S., and Schimel, D. S. (1998). DAYCENT: its land surface submodel: description and testing. *Global Planetary Change*, 19:35–48.

Parton, W. J., Ojima, D. S., Cole, C. V., and Schimel, D. S. (1994). A general model for soil organic matter dynamics: sensitivity to litter chemistry,

texture and management. In Bryants, R. and Arnold, R., editors, *Quantitative modeling of soil forming processes*, pages 147–167. Madison, WI: Soil Science Society of America.

Philip, J. R. (1957). Evaporation, and moisture and heat fields in the soil. *Journal of Meteorology*, 14:139–158.

Pistorius, T. and Zell, J. (2005). Die Dynamik der Kohlenstoffvorräte in Baden-Württemberg zwischen 1987 und 2002: Veränderungen in der Waldbiomasse und Modellierung der Holzproduktspeicher. *Allgemeine Forst und Jagdzeitung*, 176:111–120.

Potter, C., Klooster, S., Myneni, R., Genovese, V., Tan, P. N., and Kumar, V. (2003). Continental-scale comparisons of terrestrial carbon sinks estimated from satellite data and ecosystem modeling 1982 to 1998. *Global Planetary Change*, 39:201–213.

Potter, C. S., Bubler, J., Crill, P., and Lafleur, P. (2001). Ecosystem modeling of methane and carbon dioxide fluxes for boreal forest sites. *Canadian Journal of Forest Research*, 31:208–223.

Potter, C. S., Randerson, J. T., and Field, C. B. (1993). Terrestrial ecosystem production: a process model based on global satellite and surface data. *Global Biogeochemical Cycles*, 7:811–841.

Prentice, I. C., Cramer, W., Harrison, S. P., and Leemans, R. (1992). A global biome model based in plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, 19:117–134.

Recknagel, F. (1984). A comprehensive sensitivity analysis for an ecological simulation model. *Ecological Modelling*, 26:77–96.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Hohila, A., Lousau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Seufert, J. T. G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R. (2005). On the separation of

net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, 11:1424–1439.

Remson, I., Hornberger, G. M., and Molz, F. J. (1971). *Numerical methods in subsurface hydrology.* Wiley, New York.

Richters, J. (2005). Biomass changes in north western Namibia: First results from a remote sensing modelling approach. *EARSeL eProceedings*, 4:157–170.

Ritchie, J. and Otter, S. (1985). *Description of and performance of CERESWheat: a user-oriented wheat yield model.* Department of agriculture, agricultural research service. Washington, DC: ARS-38.

Ritchie, J. T. (1972). Model for predicting evaporation from a row crop with incomplete cover. *Water Resources Research*, 8:1204–1213.

Roujean, J. L., Leroy, M., and Deschanps, P. Y. (1992). A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, 97:20455–20468.

Running, S. W., Thornton, P. E., Nemani, R., and Glassy, J. M. (2000). Global terrestrial gross and net primary productivity from the earth obeservation system. In Sala, O., Jackson, R., Mooney, H., and Howarth, R., editors, *Methods in Ecosystem Science*, pages 44–57. New York: Springer-Verlag.

Russo, D. (1988). Determining soil hydraulic properties by parameter estimation: On the selection of a model for the hydraulic properties. *Water Resources Research*, 24:453–459.

Rutter, A., Kershaw, K., Robins, P., and Morton, A. (1971). A predictive model of rainfall interception in forest. i. derivation of the model from observation in a plantation of corsican pine. *Agricultural Meteorology*, 9:367–384.

Ryan, M. G., Glower, S. T., Hubbard, R. M., Waring, R. H., Gholz, H. L., Cropper Jr., W. P., and Running, S. W. (1995). Woody tissue maintenance respiration of four conifers in contrasting climates. *Oecologia*, 101:133–140.

Sala, O. E. and Austin, A. T. (2000). Methods of estimating aboveground net primary production. *Methods in Ecosystem Science*, pages 31–43.

Sands, P. J., Battaglia, M., and Mummery, D. (2000). Application of process-based models to forest management: experience with promod, a simple plantation productivity model. *Tree Physiology*, 20:383–392.

Scargle, J. D. (1989). Studies in astronomical time series analysis. iii: Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *The Astrophysical Journal*, 343:874–887.

Schaap, M. G., Leij, F. J., and van Genuchten, M. T. (2001). Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, 251:915–921.

Schmid, E., Sinabell, F., and Hofreither, M. F. (2007). Sustainability in practice: a case study on the reorientation of the common agricultural policy in austria. In Schubert, U. and Strömer, E., editors, *Sustainable Development in Europe: Concepts, Evaluation and Application.* Edward Elgar, Cheltenham, UK and Northhampton.

Schmid, E., Sinabell, F., and Liebhard, P. (2004). Effects of reduced tillage systems and cover crops on sugar beet yield and quality, ground water recharge and nitrogen leaching in the pannonic region marchfeld, austria. *Pflanzenbauwissenschaften*, 8:1–9.

Schönwiese, C. (1985). *Praktische Statistik für Meteorologen und Geowissenschaftler (4th ed.).* Gebrüder Borntraeger: Stuttgart.

Sellers, P., Berry, J., Collatz, G., Field, C., and Hall, F. (1992). Canopy reflectance, photosynthesis, and transpiration. iii. a reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sensing of Environment*, 42(3):187–216.

Sellers, P. J. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, 6(8):1335–1372.

Sellers, P. J., Los, S. O., Tucker, C. J., Justice, C. O., Dazlich, D. A., Collatz, G., and Randall, D. A. (1996). A revised land surface parameterization (SiB2) for atmospheric GCMs. part ii: The generation of global fields of terrestrial biophysical parameters from satellite data. *Journal of Climate*, 9:706–737.

Simrad, M., Pinto, N., Fisher, J. B., and Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research*, 116:G04021.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 9:161–185.

Stange, F., Butterbach-Bahl, K., Papen, H., Zechmeister-Boltenstern, S., Li, C., and Aber, J. (2000). A process-oriented model of N_2O and NO emissions from forest soils 2. sensitivity analysis and validation. *Journal of Geophysical Research*, 105:4385–4398.

StartClim (2004). Teilprojekt von StartClim2004: Analysen von Hitze und Trockenheit und deren Auswirkungen in Österreich. Analyse der Auswirkungen der Trockenheit 2003 in der Landwirtschaft Österreichs: Vergleich verschiedener Methoden. *http://www.austroclim.at/startclim/*.

Stewart, J. B. (1977). Evaporation from the wet canopy of a pine forest. *Water Resources Research*, 13:915–921.

Stockle, C. O., Williams, J. R., Rosenberg, N. J., and Jones, C. A. (1992). A method for estimating the direct and climatic effects of rising atmospheric carbon dioxide on growth and yield of crops: Part i - modification of the EPIC model for climate change analysis. *Agricultural Systems*, 38:225–238.

Strauss, F., Formeyer, H., and Schmid, E. (2012). High resolution climate data for Austria in the period from 2008 to 2040 from a statistical climate change model. *International Journal of Climatology*, accepted for publication.

Svoray, T. and Shoshany, M. (2002). SAR-based estimation of areal aboveground biomass (AAB) of herbaceous vegetation in the semi-arid zone: a modification of the water-cloud model. *International Journal of Remote Sensing*, 23:4089–4100.

194

Sylvander, S., Henry, P., Bastien-Thyri, C., Meunier, F., and Fuster, D. (2000). Vegetation geometrical image quality. In *Proceedings of VEGETATION 2000 Symposium*, pages 33–44.

Syring, K. M. and Kersebaum, K. C. (1988). Simulation des 1-dimensionalen Wassertransportes. In Richter, J., editor, *Modelle für Prozesse im Boden*, pages 32–49.

Tangki, H. and Chapalle, N. A. (2008). Biomass varaiation across selectively logged forest within a 225-km^2 region of borneo and its prediction by LANDSAT TM. *Forest Ecology and Management*, 256:1960–1970.

Tetzlaff, G. and Peters, M. (1986). The atmospheric transport potential for water vapour and dust in the sahel region. *Geo Journal*, 12:387–398.

Thornton, P. E. (1998). *Description of a numerical simulation model for predicting the dynamics of energy, water, carbon, and nitrogen in a terrestrial ecosystem.* PhD thesis, University of Montana, USA.

Thrän, D., Bunzel, K., Seyfert, U., Zeller, V., Buchhorn, M., Müller, K., Matzdorf, B., Gaasch, N., Klöckner, K., Möller, I., Starick, A., Brandes, J., Günther, K. P., Tum, M., Zeddies, J., Schönleber, N., Gamer, W., Schweinle, J., and Weimar, H. (2010). Regionale und globale räumliche Verteilung von Biomassepotenzialen. *BBSR*.

Tomasella, J., Pachepsky, Y., Crestana, S., and Rawls, W. J. (2003). Comparison of two techniques to develop pedotransfer functions for water retention. *Soil Science of America Journal*, 67:1085–1092.

Treter, U. (2000). Rolle der borealen Waldländer im globalen CO_2-Haushalt. *Geographische Rundschau*, 52:4–11.

Tum, M., Buchhorn, M., Günther, K. P., and Haller, B. C. (2011). Validation of modelled forest biomass in germany using BETHY/DLR. *Geoscientific Model Development*, 4:1019–1034.

Tum, M. and Günther, K. P. (2011). Validating modelled NPP using statistical yield data. *Biomass and Bioenergy*, 35:4665–4674.

Valentini, R. (2003). *Fluxes of carbon, water and energy of European forests, Ecological studies.* Springer, Berlin.

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science of America*, 44:892–898.

van Oijen, M. (2002). On the use of specific publication criteria for papers on process-based modelling in plant science. *Field Crops Research*, 74:197–205.

Vancley, J. K. and Skovsgaard, J. P. (1997). Evaluating forest growth models. *Ecological Modelling*, 98:1–12.

Vereecken, H., Maes, J., Feyen, J., and Darius, P. (1989). Estimating the soil moisture retention characteristics from texture, bulk density and carbon content. *Soil Science Society of America Journal*, 6:389–403.

Verger, A., Baret, F., and Weiss, M. (2011). A multisensor fusion approach to improve LAI time series. *Remote Sensing of Environment*, 115(10):2460–2470.

Veroustraete, F., Patyn, J., and Myneni, R. B. (1994). Forcing of a simple ecosystem model with fAPAR and climate data to estimate regional scale photosynthetic assimilation. In Veroustraete, F. and Ceulemans, R., editors, *VGT, Modelling and Climate Change Effects*, pages 151–177. Academic Publishing, The Hague, the Netherlands.

Veroustraete, F., Sabbe, H., and Eerens, H. (2002). Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. *Remote Sensing of Environment*, 83(3):376–399.

Verstraeten, W. W., Veroustraete, F., and Feyen, J. (2006). On temperature and water limitation of ecosystem productivity: Implementation in the C-Fix model. *Ecological Modelling*, 1999:4–22.

Wessolek, G., Bohne, K., Duijnisveld, W., and Trinks, S. (2011). Development of hydro-pedotransfer functions to predict capillary rise and actual evapotranspiration for grassland sites. *Journal of Hydrology*, 400:4229–437.

Weynants, M., Vereecken, H., and Javaux, M. (2009). Revisiting vereecken pedotransfer functions: Introducing a closed-form hydraulic model. *Vadose Zone Journal*, 8:86–95.

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

Wigmosta, M. S., Vail, L., and Lettenmaier, D. P. (1994). A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, 30:1665–1679.

Wißkirchen, K. (2005). *Modellierung der regionalen CO_2-Aufnahme durch Vegetation*. PhD thesis, Dissertation - Mathematisch-Naturwissenschaftliche Fakultät der Rheinischen Friedrich-Wilhelms Universität Bonn.

Williams, J. R. (1995). The EPIC model. In Singh, V., editor, *Computer models of watershed hydrology*, pages 909–1000. Highlands Ranch, CO: Water Resource Publications.

Williams, J. R., Jones, C. A., and Dyke, P. T. (1984). A modeling approach to determining the relationship between erosion and soil productivity. *Transactions of the ASAE*, 27:129–144.

Wosten, J. H. M. (1997). Pedotransfer functions to evaluate soil quality. in: Gregorich. *Soil Science*, 25:221–245.

Xiao, Q., McPerson, E., Ustin, S., and Grismer, M. (2000). A new approach to modelling tree rainfall interception. *Journal of Geophysical Research*, 105:173–188.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W. (2011). Reprocessing the MODIS leaf area index products for land surface and climate modelling. *Remote Sensing of Environment*, 115(5):1171–1187.

Zeller, V., Weiser, C., Hennenberg, K., Reinicke, F., Schaubach, K., Thrän, D., Vetter, A., and Wagner, B. (2011). Basisinformationen für eine nachhaltige Nutzung landwirtschaftlicher Reststoffe zur Bioenergiebereitstellung.

Schriftreihe des BMU- Förderprogramms "Energetische Biomassenutzung"
Band 2.

Zähle, S., Sitch, S., Prentice, I. C., Liski, J., Cramer, W., Erhard, M., Hickler, T., and Smith, B. (2006). The importance of age-related decline in forest NPP for modeling regional carbon balances. *Ecological Applications*, 16:1555–1574.

Zhou, X., Peng, C., and Dang, Q. L. (2006). Formulating and parameterizing the allocation of net primary productivity for modelling overmature stands in boreal forest ecosystems. *Ecological Modelling*, 195:264–272.

Curriculum Vitae

Markus Tum

Geburtstag und Geburtsort: 09.06.1983 in Göttingen, Deutschland

Nationalität: Deutsch

Familienstand: ledig

<u>Promotionsstudium</u>

2009   Seit 01.05.2009 Doktorand und wissenschaftlicher Mitarbeiter in der Abteilung Landoberfläche des Deutschen Fernerkundungsdatenzentrums des Deutschen Zentrums für Luft- und Raumfahrt, Oberpfaffenhofen

<u>Grund- und Hauptstudium</u>

2009   Diplom in Geographie - Titel der Arbeit: Variabilitätsuntersuchungen der NPP von Energiepflanzen für das Gebiet Deutschland – Österreich

2006   Vordiplom in Geographie

2003 – 2009   Studium der Geographie an der Georg-August-Universität Göttingen

2002 – 2003   Zivildienst beim Arbeiter Samariter Bund in Hann. Münden

<u>Schulbildung</u>

1995 – 2002   Gymnasium: Grotefend Gymnasium in Hann. Münden, Abschluss: Abitur

1993 – 1995   Orientierungsstufe in Hann. Münden

1989 – 1993   Grundschule in Hann. Münden