TECHNISCHE UNIVERSITÄT WIEN

VODCA2GPPv2 - An updated global model for estimating GPP from microwave satellite observations with enhanced cross-biome consistency

MASTERARBEIT

zur Erlangung des akademischen Grades

Master of Science

im Rahmen des Studiums

Geodäsie und Geoinformation

eingereicht von

Raul Lezameta, BSc
Matrikelnummer 01526601

ausgeführt am Department für Geodäsie und Geoinformation
der Fakultät für Mathematik und Geoinformation
der Technischen Universität Wien

Betreuung
Betreuer*in: Univ.Prof. Dr.rer.nat. Dr. Wouter Dorigo, MSc
Mitwirkung: Univ.Ass.in Dipl.-Ing.in Ruxandra-Maria Zotta

Wien, 22.11.2023

(Unterschrift Verfasser*in) (Unterschrift Betreuer*in)

Declaration

I hereby declare that I have written this thesis independently, that I have completely specified the utilized sources and resources and that I have definitely marked all parts of the work - including tables, maps and figures - which belong to other works or to the internet, literally or extracted, by referencing the source as borrowed.

Vienna, 22th November 2023

Raul Lezameta

Acknowledgements

This thesis would not have been possible without the guidance and support of several people.

First of all, I want to thank my supervisor *Ruxandra-Maria Zotta* for her guidance and support throughout the process. She never failed to emphasise that I should not hesitate to ask for help whenever necessary, and her feedback and helpful advice were invaluable. I am also grateful to *Prof. Wouter Dorigo* for his support and for giving me the opportunity to work on this thesis, as well as for the trust he placed in me by allowing me to work as a student employee in the CLIMERS research group alongside my studies.

I also want to thank my fellow students and colleagues in the GEO department who have made my time at the University so special. Thank you for being part of this journey and showing me how fun science can be within a supportive group of like-minded people.

A very special thank you goes to the many friends I made during my time in Vienna. They are what made my time here truly unforgettable. Thanks for reminding me to put my studies aside every once in a while and enjoy the moment, thanks for the great memories we made along the way, and thanks for friendships that I'm sure will last a lifetime.

Finally, I would like to express my gratitude to my parents. Without them, I would definitely not be where I am today.

Thank you for always being there for me, and thank you for your unconditional support and belief in me. I am eternally grateful for everything you have done for me and the many sacrifices you have made for me. But most of all, I want to thank you for always letting me know that you are proud of me no matter what, and that I could do whatever I wanted in life as long as it made me happy. Because at the end of the day, that is what brought me here.

Thank you,
Grazie,
Gracias,
Danke,
Eskerrik asko!

Thanks also to *Fedrico Marotta* for creating the sensational kaobook LaTeX template I used to write this thesis.
Check out kaobook on GitHub!

Abstract

The monitoring of Gross Primary Production (GPP) on a global scale is essential for understanding the role of terrestrial ecosystems in the carbon cycle. Over the past few decades, significant progress has been made in the ability to globally monitor GPP using process-based models and remote sensing techniques. Despite these advancements, there are still substantial differences between GPP products and large uncertainties in GPP estimates. Recently, Vegetation Optical Depth (VOD) has emerged as a useful indicator for deriving GPP from microwave satellite observations. The carbon-sink driven approach developed by Teubner et al. (2019) utilizes VOD as a proxy for the carbon-sink strength of terrestrial ecosystems to derive GPP. Wild et al. (2022) further adapted this approach, creating a global long-term GPP dataset called VODCA2GPP, based on VOD observations from the Vegetation Optical Depth Climate Archive (VODCA). This approach has shown promising results with good agreements with in-situ GPP observations and independent GPP datasets. However, the model still exhibits limited performance in certain regions and biomes, particularly in arid regions and the tropics, where in-situ data is scarce.

This study builds on the VODCA2GPPv1 model by Wild et al. (2022) and tries to make it more consistent across biomes. This was done by employing a new random forest machine learning model, by merging three different eddy covariance datasets to more than double the training data in comparison with VODCA2GPPv1 and by adding two new predictors: Land Cover and low frequency VOD.

Validation with in-situ GPP observations showed significant improvements in comparison with VODCA2GPPv1. Median correlations increased from 0.67 to 0.78 r, RMSE decreased from 2.81 to 2.25 $gC/m^2/d$, and bias decreased from 0.25 to -0.04 $gC/m^2/d$. Analyzing the cross-validation results based on land cover demonstrated a more consistent performance of the model, making it better suited for diverse regions. Comparisons with the independent FLUXCOM, MODIS and TRENDY GPP datasets revealed good temporal agreement with mean global correlations of 0.56, 0.62 and 0.42 r respectively, which could mostly be improved in comparison to VODCA2GPPv1 (+0.06, -0.02 and +0.03 r). Furthermore, the new model reduced global overestimation with respect to these datasets (bias to FLUXCOM and MODIS could be reduced by 0.44 and 0.45 $gC/m^2/d$ respectively).

However, the new model still has limitations. It still tends to globally overestimate GPP, particularly in tropical regions. Additionally, it exhibits limited performance in arid environments, highlighting the importance of accounting for water limitation in future models.

Overall, the inclusion of new predictors and additional in-situ data has resulted in a model that aligns better with in-situ GPP observations and independent GPP datasets. It also demonstrates improved consistency across different biomes and land cover classes. VODCA2GPPv2 complements existing GPP products and its long temporal availability makes it a valuable tool for studying the carbon cycle over extended time periods.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

1.1.1 What is GPP?

At leaf level, terrestrial plants fix atmospheric carbon-dioxide (CO_2) as organic compounds by net photosynthesis. At ecosystem scale, the gross uptake of CO_2 is known as Gross Primary Production (GPP) (Anav et al. 2015). GPP is defined as the sum of all carbon fixed by primary producers (i.e. autotrophic organisms like plants) through the process of photosynthesis (Beer et al. 2010). It is the largest carbon flux in the carbon cycle (Beer et al. 2010) and is also considered the primary driver of the terrestrial carbon sink responsible for the uptake of approximately 30 % of anthropogenic CO_2 emissions (Friedlingstein et al. 2020).

(Anav et al. 2015): Anav et al. (2015), *Spatiotemporal Patterns of Terrestrial Gross Primary Production: A Review*

(Beer et al. 2010): Beer et al. (2010), *Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate*

(Friedlingstein et al. 2020): Friedlingstein et al. (2020), *Global Carbon Budget 2020*

Given its central role in the global carbon budget and the increasing need to comprehend the role of the terrestrial biosphere in the global carbon cycle, developing a clear understanding of the spatio-temporal patterns of GPP has become crucial (Anav et al. 2015). Consequently, quantifying GPP has become a significant focus in studies of global climate change (Anav et al. 2015) and understanding GPP, and its variability, has become vital in carbon cycle studies (Yang et al. 2022).

(Anav et al. 2015): Anav et al. (2015), *Spatiotemporal Patterns of Terrestrial Gross Primary Production: A Review*

(Yang et al. 2022): Yang et al. (2022), *Divergent Historical GPP Trends among State-of-the-Art Multi-Model Simulations and Satellite-Based Products*

Understanding and quantifying global photosynthesis is also crucial for society, as photosynthesis supports production of food, fiber, wood and fuel for humanity (Ryu et al. 2019). From a technical perspective, GPP has been used to study terrestrial carbon sinks (Cavaleri et al. 2017), predict crop yields (Marshall et al. 2018; Reeves et al. 2005), and investigate the impact of environmental factors such as precipitation (Wang et al. 2020) and soil moisture (Trugman et al. 2018) on carbon sequestration, among many other use cases.

(Ryu et al. 2019): Ryu et al. (2019), *What Is Global Photosynthesis? History, Uncertainties and Opportunities*

1.1.2 How is GPP currently retrieved?

Despite the importance of accurately quantifying GPP, it poses a challenge due to the small spatial scale at which photosynthesis occurs. Locally, GPP can be measured using the eddy-covariance technique, which estimates the net exchange of carbon dioxide (CO2), water vapor, and energy between land ecosystems and the atmosphere (Tramontana et al. 2016). However, this technique is limited to a few hundred so-called FLUXNET sites worldwide (Tramontana et al. 2016) and is therefore not sufficient for comprehensive global GPP monitoring on its own.

(Tramontana et al. 2016): Tramontana et al. (2016), *Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms*

This issue is worsened by the uneven spatial distribution of FLUXNET sites (Tramontana et al. 2016), leading to a scarcity of observations in certain biomes and climates. The vast majority of sites are located in temperate regions and in the Northern Hemisphere, making global GPP monitoring all the more challenging. These limitations make it impossible to directly observe GPP on a global level and only tentative observation-based estimates of global terrestrial GPP have been possible so far (Beer et al. 2010). However, advancements in Terrestrial Biosphere models (TBMs) and remote sensing (RS) techniques have made it feasible to estimate GPP at a global scale (Fisher et al. 2014).

(Tramontana et al. 2016): Tramontana et al. (2016), *Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms*

(Beer et al. 2010): Beer et al. (2010), *Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate*

(Fisher et al. 2014): Fisher et al. (2014), *Modeling the Terrestrial Biosphere*

Dynamic Global Vegetation Models (DGVMs)

In particular, over the past few decades, significant progress has been made in developing Dynamic Global Vegetation Models (DGVMs) that can simulate GPP (Fisher et al. 2014). By integrating biogeography, biogeochemistry, biophysics, and vegetation dynamics (Fisher et al. 2014), DGVMs are capable of simulating terrestrial carbon and biogeochemical cycles (O'Sullivan et al. 2020). As a result, they can effectively model photosynthesis and GPP.

(Fisher et al. 2014): Fisher et al. (2014), *Modeling the Terrestrial Biosphere*

(O'Sullivan et al. 2020): O'Sullivan et al. (2020), *Climate-Driven Variability and Trends in Plant Productivity Over Recent Decades Based on Three Global Products*

The photosynthetic process takes place at cellular and intercellular levels. This makes it impossible to model GPP at a global scale on process level. That is why most DGVMs use a biochemical approach called enzyme kinetics, encapsulated by Farquhar et al. (1980) and commonly referred to as the "Farquhar model". This approach combines carbon, water, and energy through stomatal conductance, bypassing the molecular process and makes it possible to obtain GPP without having to model the individual photosynthetic cells (Fisher et al. 2014). However, due to the complexity of terrestrial ecosystems, all DGVMs make simplifications that result in divergent estimates of GPP (O'Sullivan et al. 2020). These differences arise from variations in equations and parameterization of ecosystem processes such as photosynthesis, leaf phenology, canopy scaling, and nutrient cycling (O'Sullivan et al. 2020). Additionally, the presence of numerous tunable parameters in DGVMs can cause large inter-model spreads in GPP simulations (Yang et al. 2022). Hence, recently many efforts have been made to constrain the global GPP magnitude based on satellite observations (Yang et al. 2022).

(Farquhar et al. 1980): Farquhar et al. (1980), *A Biochemical Model of Photosynthetic CO2 Assimilation in Leaves of C3 Species*

(Fisher et al. 2014): Fisher et al. (2014), *Modeling the Terrestrial Biosphere*

(O'Sullivan et al. 2020): O'Sullivan et al. (2020), *Climate-Driven Variability and Trends in Plant Productivity Over Recent Decades Based on Three Global Products*

(Yang et al. 2022): Yang et al. (2022), *Divergent Historical GPP Trends among State-of-the-Art Multi-Model Simulations and Satellite-Based Products*

Some examples of well-known DGVMs are LPJ, IBIS, ORCHIDEE, CLM, JULES, SDVGM, among others. Many of these models are part of the TRENDY DGVM ensemble run and were used as independent validation data in this thesis (see Subsection 2.3.3).

Remote Sensing (RS) based GPP estimation

In the past two to three decades, remote sensing (RS)-based models have been used to retrieve and quantify spatio-temporal patterns of GPP (Sun et al. 2019). Compared with process-oriented ecosystem models that entail a complex combination of model parameterizations, RS-based approaches are relatively simpler and more efficient for exploring dynamic changes in GPP and their spatio-temporal variations at global scales (Sun et al. 2019).

(Sun et al. 2019): Sun et al. (2019), *Evaluating and Comparing Remote Sensing Terrestrial GPP Models for Their Response to Climate Variability and CO2 Trends*

In the simplest form GPP can be estimated from RS data using simple vegetation index (VI) based models. These models are based on empirical estimations using VIs like the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI) or the Leaf Area Index (LAI) (Sun et al. 2019). They are generally based on the linkage between chlorophyll and the presence of photosynthetic biomass, which is essential for primary production (Sun et al. 2019).

(Sun et al. 2019): Sun et al. (2019), *Evaluating and Comparing Remote Sensing Terrestrial GPP Models for Their Response to Climate Variability and CO2 Trends*

Models, based on the light-use efficiency (LUE) theory (Monteith 1972) are more complex and have a stronger physical foundation compared to VI-based models. They are based on the assumption that GPP is proportional to the absorbed photosynthetically active radiation (APAR). The fraction of APAR (fAPAR) is usually estimated from optical RS data and provides the linkage to GPP (Sun et al. 2019). This approach is powerful and well-constrained at large scales, because fAPAR can be observed globally, consistently and with reasonable accuracy. Nonetheless, how much of the absorbed light gets converted to carbon is highly uncertain (Fisher et al. 2014). MODIS GPP is an example of a widely used LUE-based model (Steven W Running and Zhao 2015), it was also used as an independent validation datasets for this thesis (see Subsection 2.3.1).

(Monteith 1972): Monteith (1972), *Solar Radiation and Productivity in Tropical Ecosystems*

(Sun et al. 2019): Sun et al. (2019), *Evaluating and Comparing Remote Sensing Terrestrial GPP Models for Their Response to Climate Variability and CO2 Trends*

(Fisher et al. 2014): Fisher et al. (2014), *Modeling the Terrestrial Biosphere*

(Steven W Running and Zhao 2015): Steven W Running et al. (2015), *User's Guide: Daily GPP and Annual NPP (MOD17A2/A3) Products NASA Earth Observing System MODIS Land Algorithm*

Recently Solar-Induced Fluorescence (SIF) has also received much attention as a potential indicator for photosynthetic activity (Damm et al. 2010). Unlike light-use efficiency approaches affected by light conversion uncertainty[1], fluorescence is a direct by-product of photosynthesis and has been shown to scale linearly with GPP at global scale (Fisher et al. 2014). Empirical comparisons of SIF and GPP have demonstrated that SIF, even without any model assumptions, exhibits equal or even better predictive skill than traditional VI-based models (Frankenberg et al. 2011). SIF has also already been used in combination with Neuronal Networks to estimate GPP with very promising results (Alemohammad et al. 2017).

(Damm et al. 2010): Damm et al. (2010), *Remote Sensing of Sun-Induced Fluorescence to Improve Modeling of Diurnal Courses of Gross Primary Production (GPP)*

1: e.g., light could be absorbed but not used in photosynthesis

(Fisher et al. 2014): Fisher et al. (2014), *Modeling the Terrestrial Biosphere*

(Frankenberg et al. 2011): Frankenberg et al. (2011), *New Global Observations of the Terrestrial Carbon Cycle from GOSAT: Patterns of Plant Fluorescence with Gross Primary Productivity*

(Alemohammad et al. 2017): Alemohammad et al. (2017), *Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A Statistically Based Estimate of Global Surface Turbulent Fluxes and Gross Primary Productivity Using Solar-Induced Fluorescence*

Lastly, machine learning (ML)-based models have recently been employed to estimate GPP by upscaling eddy covariance flux tower measurements to regional and global scales using remotely sensed ancillary variables. An example is FLUXCOM (Tramontana et al. 2016), a global GPP product that utilizes a machine learning approach by integrating FLUXNET observations with remote sensing data. FLUXCOM was used as a validation dataset in this thesis (see Subsection 2.3.2).

The approach followed in this thesis can also be counted to this last category of ML-based models.

1.1.3 The carbon sink-driven GPP estimation approach

Most RS-based GPP models follow a source-driven (sunlight) approach, i.e. they estimate GPP either based on the amount of absorbed (fAPAR) or re-emitted (SIF) sunlight. In recent years, however, it has been proposed that plant growth may be stronger limited by sink- rather than source-activity (Fatichi et al. 2014; Körner 2015), and that considering sinks of fixed carbon can improve constrains in global vegetation models (Leuzinger et al. 2013).

(Fatichi et al. 2014): Fatichi et al. (2014), *Moving beyond Photosynthesis: From Carbon Source to Sink-Driven Vegetation Modeling*
(Körner 2015): Körner (2015), *Paradigm Shift in Plant Growth Control*

(Leuzinger et al. 2013): Leuzinger et al. (2013), *A Sink-Limited Growth Model Improves Biomass Estimation along Boreal and Alpine Tree Lines*

After assessing the relationship between GPP and microwave-derived Vegetation Optical Depth (VOD) (Teubner et al. 2018), Teubner et al. (2019) proposed a "carbon sink-driven approach to estimate GPP from microwave satellite observations". They used VOD as a proxy for the carbon sink strength of terrestrial ecosystems.

(Teubner et al. 2018): Teubner et al. (2018), *Assessing the Relationship between Microwave Vegetation Optical Depth and Gross Primary Production*

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

1.1.4 What is VOD? - Vegetation Optical Depth as carbon sink proxy

Vegetation Optical Depth (VOD) is a measure of the attenuation of microwave radiation caused by vegetation, which can be derived from passive and active microwave satellite observations. The amount of attenuation (and therefore VOD) depends on various factors, such as the density, type, and water content of vegetation and the wavelength of the sensor (Moesinger et al. 2020). VOD is related to above-ground dry biomass (AGB) (Y. Y. Liu et al. 2015) and its relative water content (RWC) (Momen et al. 2017) and increases with vegetation water content (VWC) (Jackson and Schmugge 1991). Short wavelengths experience a higher attenuation by vegetation, than longer ones (Jackson and Schmugge 1991). This makes short wavelength VOD more sensitive to leaf moisture content, while long wavelength VOD is more sensitive to deeper vegetation layers (e.g. stem biomass) (Chaparro et al. 2019).

(Moesinger et al. 2020): Moesinger et al. (2020), *The Global Long-Term Microwave Vegetation Optical Depth Climate Archive (VODCA)*

(Y. Y. Liu et al. 2015): Y. Y. Liu et al. (2015), *Recent Reversal in Loss of Global Terrestrial Biomass*

(Jackson and Schmugge 1991): Jackson et al. (1991), *Vegetation Effects on the Microwave Emission of Soils*

(Chaparro et al. 2019): Chaparro et al. (2019), *Sensitivity of L-band Vegetation Optical Depth to Carbon Stocks in Tropical Forests: A Comparison to Higher Frequencies and Optical Indices*

Due to its sensitivity to the VWC and AGB, VOD provides the opportunity for studying large-scale vegetation dynamics (Teubner et al. 2021) as well as for different carbon cycle studies. Its applications range from biomass (Y. Y. Liu et al. 2015; Momen et al. 2017) and drought (H. Liu et al. 2018) monitoring to phenology (Jones et al. 2011) analyses and estimating the likelihood of wildfire occurrence (Forkel et al. 2017).

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

Compared to optical vegetation indexes, VOD has distinct advantages for monitoring vegetation. These include higher sensitivity to high biomass (Y. Y. Liu et al. 2015) due to slower saturation and the ability to be retrieved (depending on the wavelength) even under cloud cover (Y. Y. Liu et al. 2011). Such advantages make VOD preferable for monitoring tropical forest areas (Teubner et al. 2019) and therefore specially relevant in high productivity areas.

(Y. Y. Liu et al. 2015): Y. Y. Liu et al. (2015), *Recent Reversal in Loss of Global Terrestrial Biomass*
(Y. Y. Liu et al. 2011): Y. Y. Liu et al. (2011), *Global Long-Term Passive Microwave Satellite-Based Retrievals of Vegetation Optical Depth*

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

VOD is expected to be related to GPP, because of its sensitivity to AGB. Biomass and temporal changes in biomass, relate to Net Primary Production (NPP) and Autotrophic Respiration (R_a) (Teubner et al. 2019), the sum of which constitutes GPP (Bonan 2016). Due to this causal relationship between biomass and GPP, a relationship is expected between VOD and GPP (Teubner et al. 2019).

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*
(Bonan 2016): Bonan (2016), *Ecological Climatology: Concepts and Applications*

Teubner et al. (2018) analyzed the relationship between VOD and GPP and came to the conclusion, that "VOD time series should be used jointly with changes in VOD for the estimation of GPP across biomes". Based on these results, Teubner et al. (2019) proposed a "carbon-sink driven approach to estimate GPP from microwave satellite observations", where they used single frequency VOD as well as its temporal changes to predict GPP. Later, the model was further adapted by Teubner et al. (2021). This new version included 2 m air temperature as an additional predictor to account for the temperature dependency of autotrophic respiration.

(Teubner et al. 2018): Teubner et al. (2018), *Assessing the Relationship between Microwave Vegetation Optical Depth and Gross Primary Production*

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

Finally, Wild et al. (2022) developed the VODCA2GPP model using the VODCA v2 CXKu dataset (Zotta et al. in prep.), which is a long-term multi-sensor and multi-frequency VOD dataset. The model was used to create a

> "new, global, long-term (1988–2020) gross primary production dataset from microwave remote sensing". Wild et al. 2022

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

(Zotta et al. in prep.): Zotta et al. (in prep.), *VODCA v2: A Multi-Sensor and Frequency Vegetation Optical Depth Dataset for Long-Term Canopy Dynamics and Biomass Monitoring, in Preparation*

1.2 Motivation

Despite the importance of global monitoring of GPP and ongoing research, there is currently no consensus on GPP predictability and GPP trends (Dunkl et al. 2023; Yang et al. 2022).

(Dunkl et al. 2023): Dunkl et al. (2023), *Gross Primary Productivity and the Predictability of CO_2: More Uncertainty in What We Predict than How Well We Predict It*

(Yang et al. 2022): Yang et al. (2022), *Divergent Historical GPP Trends among State-of-the-Art Multi-Model Simulations and Satellite-Based Products*

In addressing the need for more research on GPP estimation, Teubner et al. (2019, 2021) and Wild et al. (2022) have demonstrated the potential of their novel approach for estimating GPP using VOD. This method shows promise as an alternative to traditional RS-based approaches and can complement existing GPP products. However, there are still limitations in the model that can be addressed.

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

One major concern is the spatially uneven performance of the model, when compared to in-situ GPP measurements and other independent GPP products. Biases as well as model uncertainties are generally much larger in the Southern Hemisphere, especially in tropical and sub-tropical regions. This is problematic, as these regions are of particular interest for carbon cycle studies. Generally these areas of weak agreement match with areas of low in-situ data availability. This is especially true for the Southern Hemisphere, where in-situ data is sparse. Additionally, the model performance is not consistent across all land cover classes, with larger discrepancies mainly in semi-arid environments (e.g. savannas, open shrublands, grasslands etc.)

1.3 Objective

The aim of this thesis is to explore some limitations of the VODCA2GPP model and discuss and compare different approaches to improve its performance. This includes addressing drawbacks such as insufficient in-situ data and uneven performance across different land cover classes. To achieve this, additional in-situ data is incorporated, and predictors like land cover are included in a pursuit to make the model more suitable for different biomes. The thesis will also compare different modeling approaches, specifically the Generalized Additive Models (GAM) approach used in Teubner et al. (2021) and Wild et al. (2022) and a Random Forest (RF) regressor.

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*
(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

In the course of this investigation, an updated version of the VODCA2GPP model, referred to as VODCA2GPPv2, is developed. Additionally, the following research questions are addressed:

Research Questions

Q1 How does including additional in-situ GPP observations affect the performance of the VODCA2GPPv2 model?

Q2 How does the inclusion of additional predictors affect the performance of the VODCA2GPPv2 model? What roles do the different predictors play?

Q4 How does the RF modeling approach compare to the originally used GAM model?

Q5 How does the VODCA2GPPv2 model compare to in-situ GPP observations? How do bias, correlation and RMSE compare in cross validation?

Q6 How does the new VODCA2GPPv2 model compare to independent GPP products? How do bias, correlation and RMSE compare?

Q7 What are the spatio-temporal patterns of GPP? How does GPP vary over time and space? Are the anomalies of different GPP products comparable?

1.4 Thesis Outline

This thesis starts with this introduction Chapter 1. Afterwards in Chapter 2 "Data" the input data[2] to the VODCA2GPP model as well as independent validation datasets are presented. Chapter 3 "Methods" describes the methodology to derive GPP from VOD and other predictors. It goes into more detail on the carbon sink-driven GPP estimation approach, the VODCA2GPPv1 and the new VODCA2GPPv2 model. Additionally, the means to validate the model are presented. Chapter 4 "Results" presents the results of the model validation and the comparison to independent GPP products. Chapter 5 "Discussion" discusses the results, their implications and future research directions. Finally, Chapter 6 "Conclusions" concludes and summarizes the thesis.

2: Remotely sensed predictors and in-situ GPP

Data 2

VODCA2GPPv2 predicts GPP using VOD, air temperature and land cover as predictors. It is trained on in-situ GPP from three different FLUXNET datasets and is finally compared to three independent validation datasets. The following sections will present and discuss the data used for VODCA2GPPv2 in detail.

Figure 2.1 gives an overview of the input data used for VODCA2GPPv2 and shows its temporal coverage. It contains the temporal coverage of the predictors (green box), the in-situ GPP (orange box) and the independent GPP datasets (purple box) used for model evaluation.

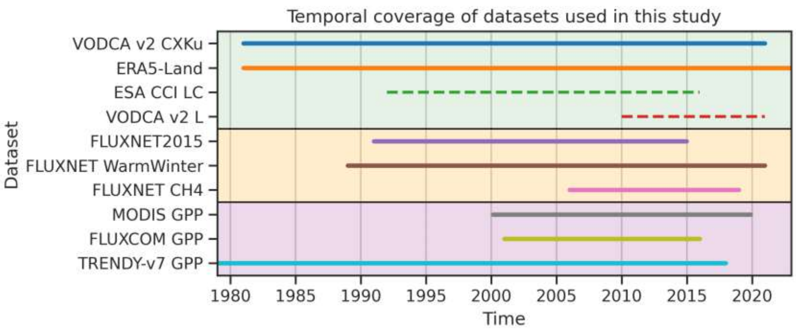


Figure 2.1: Overview of datasets used in this study and their temporal coverage. Predictor datasets are shown with green, in-situ GPP datasets with orange and evaluation GPP datasets with purple background. For static predictors (dashed lines) the mean value was used for all timestamps, even if the predictor value was not available at that time. (e.g. ESA CCI LC, VODCA v2 L)

2.1 Predictors

VODCA2GPPv2 predicts GPP using VOD, air temperature and land cover as predictors. Table 2.1 contains an overview of the predictor variables used for VODCA2GPPv2 and their origin datasets. They will be discussed in detail in the following subsections.

Table 2.1: Datasets and according predictor variables used in the VODCA2GPPv2 model.

Dataset	Variable and unit	Sensors	Temporal coverage/resolution	Spatial coverage/resolution	Reference
VODCA v2	CXKu-band Vegetation Optical Depth (VOD) [-]	AMSR-E, AMSR2, SSM/I, TMI, WindSat	1987-2018 / daily	Global / 0.25°	Zotta et al. (in prep.)
	L-band Vegetation Optical Depth (VOD) [-]	SMAP, SMOS	2010-2020 / daily	Global / 0.25°	
ERA5-Land	2m Air Temperature [°C]	- (reanalysis)	1981-2020 / hourly	Global / 9km	Muñoz-Sabater et al. (2021)
ESA CCI Land Cover 2.0.7	Fractional Coverage of Plant Functional Types (PFTs) [-]	AVHRR, PROBA-V, Envisat MERIS, SPOT-VGT	1992-2015 / yearly	Global / 300m	Defourny and ESA Land Cover CCI project team (2017)

2.1.1 VODCA v2 - The Vegetation Optical Depth Climate Archive

The Vegetation Optical Depth Climate Archive (VODCA) (Moesinger et al. 2020) is a VOD dataset, combining VOD retrievals from multiple passive microwave sensors (Table 2.2), derived through the Land Parameter Retrieval Model (LPRM). VODCA harmonizes the retrievals from different satellites and time periods with different measurement configurations[1] to finally provide three VOD products in different spatial bands: Ku-band (period 1987–2017), X-band (1997–2018), and C-band (2002–2018)

(Moesinger et al. 2020): Moesinger et al. (2020), *The Global Long-Term Microwave Vegetation Optical Depth Climate Archive (VODCA)*

1: microwave frequencies, measurement incidence angles, orbit characteristics, radiometric quality, spatial footprint

Table 2.2: List of sensors used in VODCA CXKu.

Sensor	Time period used	AECT	C-band [GHz]	X-band [GHz]	Ku-band [GHz]	Reference
AMSR-E	Jun 2002–Oct 2011	13:30	6.93	10.65	18.70	van der Schalie et al. (2017)
AMSR2	Jul 2012 - Jan 2019	13:30	6.93, 7.30	10.65	18.70	van der Schalie et al. (2017)
SSM/I F08	Jul 1987–Dec 1991	18:15			19.35	Owe et al. (2008)
SSM/I F11	Dec 1991–May 1995	17:00–18:15			19.35	Owe et al. (2008)
SSM/I F13	May 1995–Apr 2009	17:45–18:40			19.35	Owe et al. (2008)
TMI	Dec 1997–Apr 2015	Asynchronous		10.65	19.35	Owe et al. (2008) and van der Schalie et al. (2017)
WindSat	Feb 2003–Jul 2012	18:00	6.80	10.70	18.70	Owe et al. (2008) and van der Schalie et al. (2017)

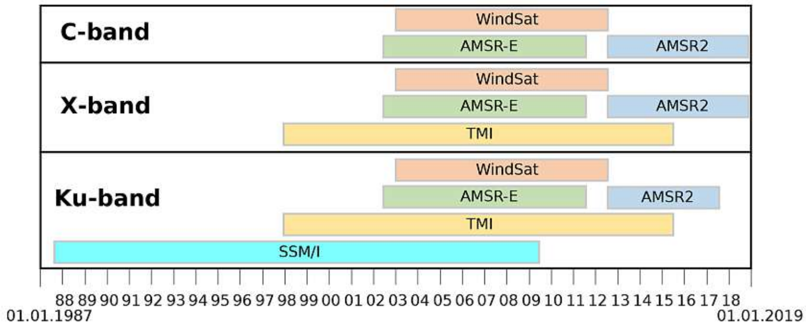


Figure 2.2: Temporal coverage of sensors used in VODCA CXKu. Figure taken from Moesinger et al. (2020).

Here a new improved version of VODCA, VODCA v2 (Zotta et al. in prep.) version was used, which uses observations from the same sensors as VODCA v1 (Table 2.2, Figure 2.2), but merges them into one long-running multi-frequency VOD product, to increase temporal and spatial coverage and reduce random errors.

(Zotta et al. in prep.): Zotta et al. (in prep.), *VODCA v2: A Multi-Sensor and Frequency Vegetation Optical Depth Dataset for Long-Term Canopy Dynamics and Biomass Monitoring, in Preparation*

This multi-frequency product named VODCA v2 CXKu provides a single long-term vegetation metric (1988-2020), exceeding the temporal length of the individual single-frequency products (VODCA v2 C, X and Ku). It is obtained by first rescaling the C- and Ku-band observations to X-band to remove biases between the bands and then computing a weighted average to fuse overlapping observations. The reference frequency for the scaling of the different frequencies is therefore X-band.

VODCA v2 CXKu merged 15 passive VOD datasets[2] retrieved from 7 different sensors using the Land Parameter Retrieval Model (LPRM) (van der Schalie et al. 2017).

2: Scaling of the single-sensor VOD observations was done by applying cumulative distribution function (CDF) matching.

(van der Schalie et al. 2017): van der Schalie et al. (2017), *The Merging of Radiative Transfer Based Surface Soil Moisture Data from SMOS and AMSR-E*

The LPRM is based on radiative transfer theory introduced by (Mo et al. 1982) and uses forward modelling to simulate the top-of-atmosphere brightness temperatures under a wide range of conditions. Although primarily developed for soil moisture, it simultaneously solves for VOD using an analytical solution by Meesters et al. (2005), utilizing the ratio between H- and V-polarized observations (van der Schalie et al. 2017). The LPRM assumes that the temperatures of soil and vegetation are the same, but this assumption may not hold true during the day when solar radiation causes uneven heating. Therefore, VODCA v2 relies solely on nighttime observations, assuming that they reflect thermal equilibrium.

(Mo et al. 1982): Mo et al. (1982), *A Model for Microwave Emission from Vegetation-Covered Fields*

(Meesters et al. 2005): Meesters et al. (2005), *Analytical Derivation of the Vegetation Optical Depth from the Microwave Polarization Difference Index*

(van der Schalie et al. 2017): van der Schalie et al. (2017), *The Merging of Radiative Transfer Based Surface Soil Moisture Data from SMOS and AMSR-E*

Preprocessing of the level-2 LPRM-retieved VOD data includes projecting the data onto a 0.25° x 0.25° grid via nearest-neighbour resampling and selecting the closest nighttime value within a 24h window. Data are masked for radio-frequency interference, negative VOD retrievals and low land surface temperatures (< 0 °C).

3 VOD predictors based on VODCA CXKu were used in this study: the 8-daily mean of VODCA CXKu, the temporal difference of the 8-daily means and a static median VOD predictor (see Table 2.3).

Predictor	Description
VOD	8-daily mean of VODCA CXKu
dVOD	temporal difference of 8-daily VODCA CXKu mean
medVOD*	median of VODCA CXKu (static predictor)

Table 2.3: 3 VODCA CXKu Predictors used in this study.

In addition to VODCA CXKu, a preliminary version of VODCA at L-band was used as a predictor. It is processed like VODCA CXKu, but based on L-band VOD observations from the Soil Moisture Active Passive (SMAP) and Soil Moisture and Ocean Salinity (SMOS) missions. Longer wavelength VOD, like L-band VOD, is less attenuated by vegetation and as a consequence saturates later than shorter wavelength VOD (Jackson and Schmugge 1991). This characteristic enhances its sensitivity to deeper vegetation layers (Chaparro et al. 2019), making it particularly useful for highly productive areas with tall vegetation and high vegetation density, such as the tropics

(Jackson and Schmugge 1991): Jackson et al. (1991), *Vegetation Effects on the Microwave Emission of Soils*

(Chaparro et al. 2019): Chaparro et al. (2019), *Sensitivity of L-band Vegetation Optical Depth to Carbon Stocks in Tropical Forests: A Comparison to Higher Frequencies and Optical Indices*

VODCA L is only available from the years 2010 to 2020, therefore the temporal dynamcis of VODCA L are not used in this study. Instead, the mean of VODCA L is used as a static predictor (see Table 2.4).

Predictor	Description
L-VOD*	mean L-band VOD from 2010 to 2020

Table 2.4: The VODCA L Predictor used in this study.

2.1.2 ERA5-Land - 2m Air Temperature

2 m air temperature, provided by ERA5-Land (Muñoz-Sabater et al. 2021) was used to account for the temperature dependency of autotrophic respiration. ERA5-Land is an enhanced global dataset for the land component of the fifth generation of European ReAnalysis (ERA5) produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5-Land is available hourly at a spatial resolution of 9 km.

(Muñoz-Sabater et al. 2021): Muñoz-Sabater et al. (2021), *ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications*

2 m air temperature was aggregated[3] to 8-daily means to get to the T2m predictor used in this study (see Table 2.5).

3: over day and night

Table 2.5: The 2m Air Temperature Predictor used in this study.

Predictor	Description
T2m	8-daily mean of 2m air Temperature

2.1.3 ESA CCI LC - Plant Functional Types

Land cover information was used to make the VODCA2GPP model more generalizable and to account for the poor global distribution of in-situ GPP measurement stations.[4]

4: In situ data is poorly distributed across the globe, with most sites being located in the Northern Hemisphere and in temperate regions.

The ESA Climate Change Initiative (ESA CCI) provides annual land cover maps which classify the Earth's land surface into 23 level 1 and 14 level 2 (sub) land cover classes following the United Nations Land Cover Classification System (LCCS). Version 2.0.7 (Defourny and ESA Land Cover CCI project team 2017), used in this study, covers all the years from 1992 to 2015 at a spatial resolution of 300 m.

(Defourny and ESA Land Cover CCI project team 2017): Defourny et al. (2017), *ESA Land Cover Climate Change Initiative (Land_Cover_cci): Global Land Cover Maps, Version 2.0.7*

The workflow used to derive the ESA CCI LC maps is made to guarantee a high consistency over time. To achieve this, a unique baseline LC map was generated using data from the MERIS FR and RR archive from 2003 to 2012. Changes in land cover were then detected based on different satellite data[5] from 1992 to 2015. With these changes, the baseline map was then updated to create the annual LC maps from 1992 to 2015.

5: AVHRR time series; SPOT-VGT time series; PROVA-V

For this study the LC maps were first aggregated into a 0.25° spatial resolution grid and the 37 LC classes were then converted into fractional coverages[6] of 11 Plant Functional Types (PFTs) (Table 2.6) using a custom conversion table. These conversions were done using ESA's CCI-LC User Tool (ESA 2014).

6: 0-1; all classes sum up to 1.

(ESA 2014): ESA (2014), *CCI-LC User Tool*

Table 2.6: 11 Plant Functional Type (PFT) Predictors used in this study.

	Vegetation			
Predictor	Leaf type	Phenology	Growth Form	Other
pftTreeBE*	Broadleaved	evergreen	Tree	-
pftTreeBD*	Broadleaved	deciduous	Tree	-
pftTreeNE*	Needle-leaved	evergreen	Tree	-
pftTreeND*	Needle-leaved	deciduous	Tree	-
pftShrubBE*	Broadleaved	evergreen	Shrub	-
pftShrubBD*	Broadleaved	deciduous	Shrub	-
pftShrubNE*	Needle-leaved	evergreen	Shrub	-
pftHerb*	-	-	Herbaceous cover	-
pftCrop*	-	-	Cropland	-
pftBare*	-	-	-	Bare soil
pftNoLand*	-	-	-	No land

PFTs are a key feature of current generation earth system models and represent groupings of plant species that share similar structural, phenological, and physiological traits (Poulter et al. 2015). Individual PFTs combine growth-form (trees, shrubs, herbaceous vegetation, crops) with leaf type (broadleaved, needle-leaved) and phenology (evergreen, deciduous).

(Poulter et al. 2015): Poulter et al. (2015), *Plant Functional Type Classification for Earth System Models: Results from the European Space Agency's Land Cover Climate Change Initiative*

2.2 Target variable - in situ GPP from FLUXNET

GPP is the target variable of the VODCA2GPP model. In-situ GPP from three different FLUXNET datasets was used to train and validate the VODCA2GPPv2 model, namely: FLUXNET2015 (Pastorello et al. 2020) the Fluxnet-CH4 Community Product (Delwiche et al. 2021) and the FLUXNET Warm Winter release (Team and Centre 2022).

FLUXNET refers to a global network of micrometeorological tower sites that use eddy covariance[7] techniques to measure the exchanges of carbon dioxide, water vapor, and energy between terrestrial ecosystems and the atmosphere, across a wide variety of biomes and climates (Baldocchi 2003).

7: atmospheric measurement technique to measure and calculate vertical turbulent fluxes within atmospheric boundary layers

(Baldocchi 2003): Baldocchi (2003), *Assessing the Eddy Covariance Technique for Evaluating Carbon Dioxide Exchange Rates of Ecosystems: Past, Present and Future*

GPP is derived from measured CO_2 fluxes by calculating net ecosystem exchange (NEE) from CO_2 turbulent and storage fluxes and partitioning NEE into its components of ecosystem respiration (RECO) and gross primary production (GPP) (Pastorello et al. 2020).

(Pastorello et al. 2020): Pastorello et al. (2020), *The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data*

2.2.1 FLUXNET datasets

FLUXNET2015

FLUXNET2015 (Pastorello et al. 2020) is the most complete and newest (official) FLUXNET dataset. It provides ecosystem-scale data on CO_2, water, and energy exchange between the biosphere and the atmosphere, and other meteorological and biological measurements, from 212 sites around the globe (up to 2014).

(Pastorello et al. 2020): Pastorello et al. (2020), *The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data*

VODCA2GPPPv2 uses the February 2020 update of FLUXNET2015, whereas the older version by Wild et al. (2022) still used the November 2016 version of FLUXNET2015. This new update of FLUXNET2015 included many sites being changed to a Creative Commons Attribution CC-BY 4.0 license, meaning that a total of 206 sited could now be used (compared to 110 sites in the original version of VODCA2GPP).

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

FLUXNET Warm Winter

The FLUXNET Warm Winter dataset (Team and Centre 2022) is a third-party re-release of most European sites from FLUXNET2015 which now have a longer temporal coverage (up to 2020). Additionally, it adds some totally new sites, not included in FLUXNET2015.

(Team and Centre 2022): Team et al. (2022), *Warm Winter 2020 Ecosystem Eddy Covariance Flux Product for 73 Stations in FLUXNET-Archive Format—Release 2022-1*

Both, the FLUXNET2015 and Warm Winter datasets have been processed using the same pipeline, making them fully compliant and integrable with each other.

FLUXNET-CH4 Community Product

FLUXNET-CH4 (Delwiche et al. 2021) is a community product of eddy-covariance methane and CO_2 flux measurements. The dataset contains 81 sites globally, most of which are not present in FLUXNET2015.

(Delwiche et al. 2021): Delwiche et al. (2021), *FLUXNET-CH_4: A Global, Multi-Ecosystem Dataset and Analysis of Methane Seasonality from Freshwater Wetlands*

2.2.2 Merging of FLUXNET datasets

To mitigate the issue of the small amount of in-situ GPP observations present in VODCA2GPP v1, all three FLUXNET Datasets presented before, were merged to obtain a single dataset containing as many stations as possible. This section will describe the merging procedure in detail.

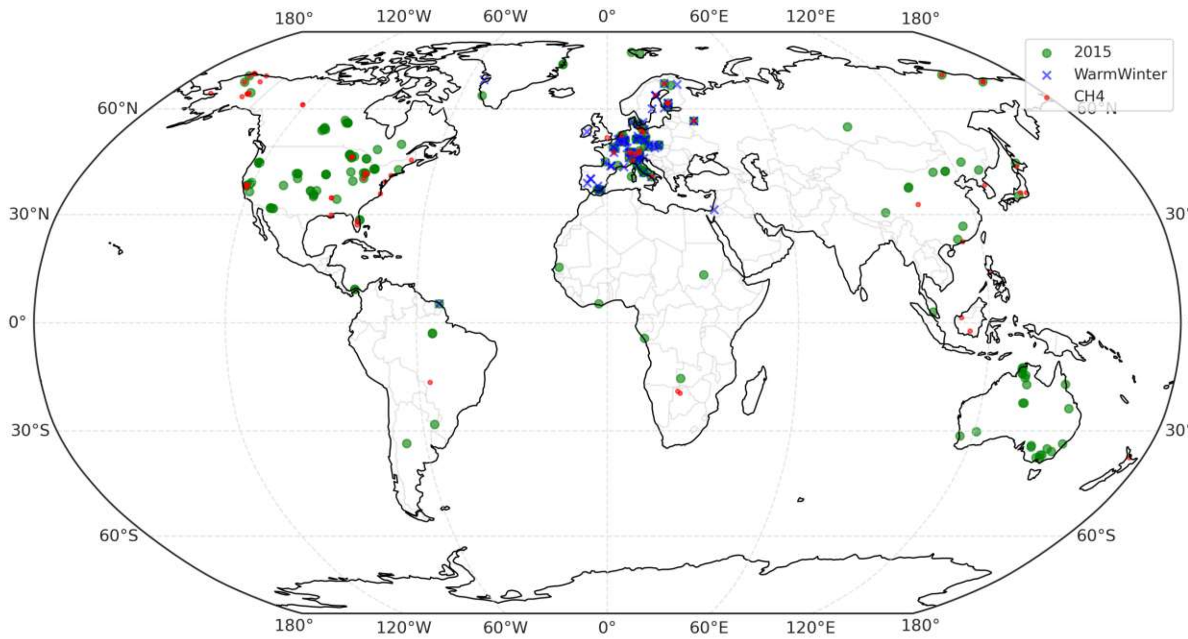


Figure 2.3: Spatial distribution FLUXNET sites from the FLUXNET 2015, Warm Winter, and CH4 datasets. Stations are colored by their origin dataset.

As can be seen in Figure 2.3, showing the spatial distributions of all three FLUXNET datasets, some stations are present in multiple datasets, while others are only present in one dataset. All stations only available in one dataset were used as is, while stations present in multiple datasets were only used once. When deciding which dataset to use for the stations present in multiple datasets, the Warm Winter dataset was prioritized since it had the longest observation time (up to 2020) and nearly always also contained the observations from FLUXNET2015 and FLUXNET-CH4. If not available, the FLUXNET2015 dataset was used over the FLUXNET-CH4 (Delwiche et al. 2021) dataset since the observations in the latter were shorter and less consistent with the other datasets.

(Delwiche et al. 2021): Delwiche et al. (2021), *FLUXNET-CH$_4$: A Global, Multi-Ecosystem Dataset and Analysis of Methane Seasonality from Freshwater Wetlands*

All datasets provide Gross Primary Productivity (GPP) derived using daytime and nighttime partitioning, which were then averaged to obtain a single GPP value for each day, following the suggestions from Teubner et al. (2021) and Wild et al. (2022). Both FLUXNET2015 and Warm Winter datasets also provided the `NEE_VUT_MEAN_QC` quality control flag, which was used to filter out days with a quality flag below 0.5. This means that only days where more than 50 % of the data was considered to be of good quality were utilized.

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

To ensure that the two newly introduced FLUXNET datasets are integrable with FLUXNET2015, a comparison of various metrics was conducted. The datasets were checked for consistency by comparing the mean and standard deviation of the overlapping GPP observations from the different datasets, as well as from stations laying in the same climate or land cover classes. Additionally, the GPP timeseries from overlapping stations were examined and found to be consistent across datasets. Detailed comparative plots illustrating the differences between the datasets can be found in the appendix (see Figures A.2 to A.5).

As for FLUXNET WarmWinter, Team and Centre (2022) themselves state that their data is "fully compliant and integrable with the FLUXNET2015 release".

(Team and Centre 2022): Team et al. (2022), *Warm Winter 2020 Ecosystem Eddy Covariance Flux Product for 73 Stations in FLUXNET-Archive Format—Release 2022-1*

The result of the merging process is a combined dataset containing in-situ GPP observations at 267 Fluxnet stations. Of these, 145 stations originate from the FLUXNET2015 dataset, 70 stations come from the WarmWinter dataset (with 15 of them not present in FLUXNET2015), and 52 stations are from the FLUXNET-CH4 dataset (which are all completely new stations). The resulting dataset will in the following be referred to as FLUXNETmerged. A list, as well as a map (Appendix A), of all stations present in the FLUXNETmerged dataset can be found in Table A.1 in the appendix.

2.3 GPP Evaluation datasets

Three independent GPP datasets were used to evaluate the performance of the VODCA2GPP model. Two remotely sensed GPP products, namely MODIS GPP (S. Running et al. 2015) and FLUXCOM GPP (Tramontana et al. 2016), and with TRENDY GPP also a product derived from an ensemble of DGVM runs.

(S. Running et al. 2015): S. Running et al. (2015), *MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006*

(Tramontana et al. 2016): Tramontana et al. (2016), *Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms*

2.3.1 MODIS GPP

MODIS GPP provides GPP estimates based on the light-use efficiency (LUE) approach by Monteith (1972), which relates plant productivity to the amount of solar radiation absorbed by the vegetation. The MODIS algorithm uses the optically derived fAPAR[8] as a proxy for the absorbed solar energy to derive GPP.

(Monteith 1972): Monteith (1972), *Solar Radiation and Productivity in Tropical Ecosystems*

8: Fraction of Absorbed Photosynthetically Active Radiation

In this study the MOD17A2H Version 6 Data Product (S. Running et al. 2015) was used. It is an 8-day composite product with a spatial resolution of 500 m. For the sake of comparison with the VODCA2GPP model, the MODIS GPP product was aggregated to a spatial resolution of 0.25°.

(S. Running et al. 2015): S. Running et al. (2015), *MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006*

2.3.2 FLUXCOM GPP

FLUXCOM GPP (Tramontana et al. 2016) is a global GPP dataset derived from upscaling of in-situ eddy covairiance measurements using machine learning techniques. The upscaling was carried out using remotely sensed ancillary variables all derived from optical observations from the Moderate Resolution Imaging Spectrometer (MODIS).

(Tramontana et al. 2016): Tramontana et al. (2016), *Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms*

Here FLUXCOM RS was used, one of two FLUXCOM GPP products. While FLUXCOM RS+MET, the other FLUXCOM GPP Product, is based on meteorological data and mean seasonal cycle of remotely sensed variables, FLUXCOM RS is based on remotely sensed variables only.

This remote sensing data includes Land Surface Temperature, several vegetation indices (NDVI, EVI, LAI)[9], the fAPAR as well as the water indices NDWI[10] and LSWI[11].

9: Normalized Differenced Vegetation Index; Enhanced Vegetation Index; Leaf Area Index
10: Normalized Differenced Water Index
11: Land Surface Water Index

FLUXCOM RS GPP has a 10 km spatial resolution and a temporal resolution of 8 days. For this study the FLUXCOM GPP product was aggregated to a spatial resolution of 0.25°.

2.3.3 Trendy-v7 GPP

TRENDY is an ensemble run of DGVMs, which is performed annually to support the Global Carbon Project's (GCP) assessment of the global carbon budget. The specific version used in this study is TRENDY-v7, which is the seventh version of the TRENDY dataset created for the GCP's 2018 global carbon budget assessment (Le Quéré et al. 2018).

(Le Quéré et al. 2018): Le Quéré et al. (2018), *Global Carbon Budget 2018*

TRENDY-v7 includes 16 DGVMs:

- CABLE-POP
- CLASS-CTEM
- CLM5.0
- DLEM
- ISAM
- JSBACH
- JULES
- LPJ
- LPJ-GUESS
- LPX
- OCN
- ORCHIDEE
- ORCHIDEE-CNP
- SDGVM
- SURFEXv8
- VISIT

In this study, the mean GPP of all 16 DGVMs was utilized. The dataset covers the time period from 1901 to 2017 at a spatial resolution of 0.5°.

3 Methods

3.1 The carbon sink-driven GPP estimation approach

This thesis builds on the carbon sink-driven GPP estimation approach introduced by Teubner et al. (2019) and further improved and reworked by Teubner et al. (2021).

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

The biogeochemical basis of their GPP model is the relationship between GPP, ecosystem net uptake of carbon (NPP - Net Primary Production) and autotrophic respiration (R_a) (Bonan 2016):

(Bonan 2016): Bonan (2016), *Ecological Climatology: Concepts and Applications*

$$GPP = \underbrace{R_a}_{R_m + R_g} + NPP \tag{3.1}$$

R_a can further be split up into maintenance (R_m) and growth respiration (R_g), which are proportional to biomass and change in biomass respectively.

The first sink-driven GPP model by Teubner et al. (2019) was based solely on VOD variables. Besides using the VOD time series itself, the model also incorporated two additional VOD-predictors: the temporal changes in VOD (dVOD) and the temporal median of VOD (medVOD).

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

While VOD itself relates to maintenance respiration, the temporal changes in VOD (dVOD) relate to growth respiration and NPP. The temporal median of VOD (medVOD) on the other hand serves as a proxy for vegetation density, it was incorporated to account for larger structural vegetation components and make the resulting model more closely related to biomass changes of smaller structural vegetation components such as leaves.

Teubner et al. (2021) later improved the model by incorporating temperature as an additional predictor variable. This addition accounts for the strong temperature dependence of autotrophic respiration, which is mainly attributed to its maintenance part (Bonan 2016). The improved formulation of the model, considers the temperature dependence of maintenance respiration through a term representing the interaction between temperature (T2m) and VOD (Teubner et al. 2021):

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

(Bonan 2016): Bonan (2016), *Ecological Climatology: Concepts and Applications*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

$$GPP(VOD, T2m) = te(VOD, T2m) + s(\Delta VOD) + s(mdn(VOD)) \tag{3.2}$$

This equation (3.2), represents the model formulation as defined by Teubner et al. (2021) and used for VODCA2GPP by Wild et al. (2022).

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

All of Teubner et al. (2019, 2021)'s versions of the model, as well as VODCA2GPP (Wild et al. 2022) used the Generalized Additive Models (GAM) approach to model the relationship between GPP and the respective predictor variables. The GAM approach is a flexible non-parametric method that allows for the estimation of non-linear relationships between the response and predictor variables. It was chosen over other non-linear regression methods[1] since it is more interpretable and allows for the estimation of the uncertainty of the model parameters.

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

1: such as random forest (RF)

3.2 The updated VODCA2GPPv2 model

In the process of improving the VODCA2GPP model, different model configurations were trained and tested. The different versions of the model are summarized in Table 3.1. They differ in the predictor variables used and the type of regression model used to relate the predictor variables to the response variable (GPP) as well as in the amount of used in-situ GPP data during model training.

The model IDs introduced in Table 3.1 will be used to refer to the different versions of the model throughout this thesis.

Table 3.1: Overview of the different versions of the VODCA2GPP model. The models differ in the predictor variables (static predictors are marked with *), the type of regression model and the training data used. The final version of the model is highlighted in green.

			Predictors					
Model ID	Regressor	FLUXNETmerged?	VOD	dVOD	medVOD*	T2m	LC*	L-VOD*
GAM	GAM		x	x	x	x		
GAM+	GAM	x	x	x	x	x		
RF+	RF	x	x	x	x	x		
RF+_LC	RF	x	x	x	x	x	x	
RF+_LC_LVOD	RF	x	x	x	x	x	x	x

The first Version of the Model (**GAM**) still used GAM and limited in-situ data (Feb 2016 update of FLUXNET 2015; subset of FLUXNET 2015). It is equivalent to the model proposed by Wild et al. (2022) and was trained using the same workflow. All other versions of the model used the new FLUXNETmerged (see Subsection 2.2.2) dataset to have more training data available. The latter versions of the model also use random forest instead of GAM as regressor and subsequently more predictors like land cover (LC) and L-band VOD (L-VOD) were added.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

LC was introduced to account for the uneven spatial distribution of model performance in VODCA2GPPv1 (Wild et al. 2022). This problem was partially addressed with the inclusion of new in-situ GPP data. However, the addition of LC as a predictor variable was expected to further improve the spatial consistency of the model. This is because LC allows the model to account for the changing VOD-GPP relationships of different vegetation types. This is especially important in areas not well represented by the in-situ data, such as in the tropics. With similar goals L-band VOD was introduced as a proxy for vegetation density. However, due to its latter saturation and deeper penetration depth compared to optically derived LC, as well as the higher frequency microwave VOD, it exhibits higher sensitivity to tall and dense vegetation.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

The incentive for switching from GAM to RF is derived from Schmidt et al. (2023), who in assessing the sensitivity of VOD to different vegetation parameters, compared the performance of GAM and RF. Their conclusion was that, in most cases, GAM is insufficient for accurately describing the relation of VOD to vegetation parameters, particularly when including land cover predictors, and predicting across land cover classes.[2] This incentive led to switching to a RF regressor, especially considering the added complexity given by the joint introduction of LC predictors.

(Schmidt et al. 2023): Schmidt et al. (2023), *Assessing the Sensitivity of Multi-Frequency Passive Microwave Vegetation Optical Depth to Vegetation Properties*

2: Schmidt et al. (2023) assessed the sensitivity of VOD to different vegetation parameters by aiming to predict VOD using AGB, Live-Fuel Moisture (LFMC), and LAI. They compared the performance of GAM and RF, and their conclusion was that, in most cases, GAM alone is insufficient for accurately predicting VOD, particularly when including land cover predictors, and predicting across land cover classes. While a simpler additive approach like GAM was adequate for individual land cover types, they discovered that the relationship between VOD and other vegetation properties cannot be easily captured with global linear, monotonic, and bivariate regressions. Instead, it requires accounting for the non-linear interactions among various ecosystem properties.

Independent of the changing model configurations, all models were trained, applied and validated using the same workflow, described in the following Subsections and in the next Section.

Figure 3.1 gives an overview of the timespans on which the model training (orange box), predictions (purple box) and evaluations were performed.

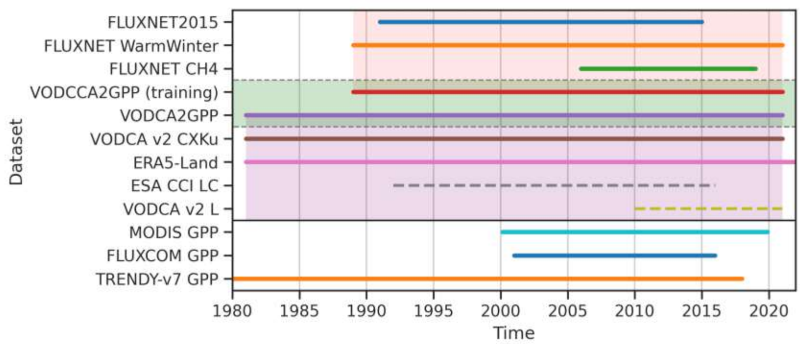


Figure 3.1: Overview of the timespans of model training (orange), application (purple) and evaluation of VODCA2GPPv2.

3.2.1 Preprocessing

Since all input data was already available at (or had previously been converted to) a 0.25° spatial resolution, no spatial resampling was necessary. Temporal resampling, however, was applied primarily to reduce noise and computation times. All input data[3] was resampled to a 8-days temporal resolution. The final VODCA2GPP model prediction therefore also represents the mean of daily GPPs for an 8-day period.[4] Since VODCA v2 already incorporates extensive quality flagging (e.g. for frozen conditions and radio-frequency interference), no additional data processing was necessary.

3: response variable: FLUXNET GPP; predictor variables: VODCA v2 CXKu, ERA5-Land T2m, ESA CCI LC PFTs, VODCA v2 L

4: The 8-day temporal resolution was chosen because the usage of short time intervals (on the order of several days) is crucial in reducing the influence of larger vegetation components (e.g. stems) and makes the model more sensitive to changes in leaf biomass (Wild et al. 2022). Additionally, the validation datasets MODIS and FLUXCOM GPP have the same 8-daily resolution which enhances comparability and facilitates the validation.

3.2.2 Training the model

To train the model, firstly all grid-points where the response variable (in-situ GPP) was available were selected. Depending on the model version, different combinations of predictor variables were chosen for every grid-point and then related to the response variable (GPP) at the respective in-situ stations. As some stations were located in the same grid-point, they had identical predictor values, but different values for the response variable (GPP).

Subsequently, the maximum temporal overlap between predictor and response variables was determined for each station. Additionally, all time points where not all the predictor variables were available were removed from the training data. This was done to ensure that the model was trained on a consistent set of predictor variables for all stations. To increase the robustness of the derivation, VOD and dVOD were smoothed before training the model using a SavitzkyGolay filter with a window size of 11 data points as suggested by Teubner et al. (2021).

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

Finally, this data was used to train a random forest regression model, using the `scikit-learn` (Pedregosa et al. 2011) implementation of the random forest algorithm. To find the optimal hyperparameters for the model, several combinations of hyperparameters were tested. The parameters in Table 3.2 were found to have the best performance in a 10-fold cross-validation and were therefore used to train the final model.

(Pedregosa et al. 2011): Pedregosa et al. (2011), *Scikit-Learn: Machine Learning in Python*

Table 3.2: Final values of the hyperparameters tested for the random forest model. The hyperparameters with the best performance in a 10-fold stratified group cross validation were used to train the final model.

Hyperparameter	Value	Description
n_estimators	1200	The number of trees in the forest.
max_features	5	The number of features to consider when looking for the best split.
min_samples_split	sqrt(max_features)	The minimum number of samples required to split an internal node.
min_samples_leaf	15	The minimum number of samples required to be at a leaf node.
bootstrap	True	Whether bootstrap samples are used when building trees.

For the sake of model-comparisons, some models (see Table 3.1) were also trained following the GAM approach from VODCA2GPPv1 proposed by Wild et al. (2022) (see Equation 3.2).

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

3.2.3 Applying the model

The trained model was then applied to all grid points to get a global dataset of GPP observations at a 0.25° spatial resolution spanning the years from 1988 to 2020.

3.3 Model evaluation

3.3.1 Site-level cross validation

To evaluate the performance of the model, a cross validation was performed at site-level. For this, 10 different versions of the model were trained and validated. In each fold of the cross validation, a different set of stations (roughly 10%) was used for validation, while the remaining stations were used for training. Each station was used for validation exactly once, which is why this approach is called pseudo-random. The different folds were also stratified by land cover class, this means that the distribution of land cover classes in the training and validation sets was set to be as close as possible. This approach can be referred to as 10-fold stratified group cross validation[5].

5: In 10-fold stratified group cross-validation, the data is divided into ten equal-sized parts. During each iteration of model training and evaluation, nine parts of the data are used for training the model (i.e., these parts form the training set), while the remaining one part is used to evaluate the performance of the model (i.e., this part forms the validation set).

In this case, stratified group cross-validation involves first dividing the data into groups based on a grouping variable, which is station ID. These groups are then added either to the training or validation set, such that all observations from a single group (i.e., station) are either used to train or validate the model.

The stratification was done by land cover class, which means that each fold contains approximately the same proportion of each land cover class as the whole dataset. This ensures that the performance of each model is evaluated on a representative sample of the data, rather than being biased towards any particular class.

To evaluate the performances of the different models, different performances metrics were calculated at each fold. These metrics were calculated for every station individually in order to get one value of each metric for every station.

The metrics used are Pearson's correlation coefficient r (Equation 3.3), the root mean squared error RMSE (Equation 3.4) and the bias (Equation 3.5):

$$r(y, \hat{y}) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{3.3}$$

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.4}$$

$$\text{bias}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{3.5}$$

Where:

- y is the vector of observed (in-situ) GPP
- $\hat{y}$ is the vector of predicted GPP
- $\bar{y}$ is the mean of the observations
- $\bar{\hat{y}}$ is the mean of the predictions
- n is the number of observations

3.3.2 Feature Importances of Predictors

To assess the feature importances two measures, namely the mean decrease in impurity (MDI) and the SHAP values, were calculated.

Mean decrease in impurity

The Mean Decrease in Impurity (MDI) is a measure of feature importance in Random Forests that calculates each feature importance as the sum over the number of splits (across all trees) that include the feature, weighted by the number of samples that are affected by the split (Breiman 2001). It is sometimes called "gini importance" and is defined as the total decrease in node impurity (weighted by the probability of reaching that node[6]) averaged over all trees of the ensemble (Pedregosa et al. 2011).

(Breiman 2001): Breiman (2001), *Random Forests*

6: which is approximated by the proportion of samples reaching that node

(Pedregosa et al. 2011): Pedregosa et al. (2011), *Scikit-Learn: Machine Learning in Python*

SHAP values

SHAP values are a model-agnostic measure of feature importance that can be calculated for any machine learning model. They are based on Shapley values, which are a concept from cooperative game theory. Shapley values are a way to fairly distribute the "payout" of a game among the players. In the context of machine learning, the "game" is the prediction task and the "players" are the features. They are based on the idea of assigning each feature an importance score for a particular prediction by computing the contribution of each feature to the difference between the expected model output and the actual model output (Molnar 2022).

(Molnar 2022): Molnar (2022), *Chapter 8.5 Shapley Values*

In contrast to MDI, which is a global measure of feature importance, SHAP values are a local measure of feature importance. This means that MDI measures how much each feature reduces impurity across all trees in the forest while SHAP values measure how much each feature contributes to a specific prediction. This makes SHAP values more interpretable than MDI, since they can be used to explain individual predictions (Loecher 2022).

(Loecher 2022): Loecher (2022), *Debiasing MDI Feature Importance and SHAP Values in Tree Ensembles*

The SHAP values have been calculated using the `shap` python package (Lundberg 2023).

(Lundberg 2023): Lundberg (2023), *Slundberg/Shap - A Game Theoretic Approach to Explain the Output of Any Machine Learning Model.*

3.3.3 Comparison with independent GPP datasets

To assess the performance and validate the predictions of different model versions, as well as of the final model, the predictions were compared to the independent GPP datasets FLUCXOM, MODIS, and TRENDY GPP (see Section 2.3 for details on the datasets).

Maps of temporal correlations and biases between datasets were created. Additionally, latitudinal GPP means were calculated for each dataset and overlaid in a single plot to compare the latitudinal biases between the different datasets. Finally, spatio-temporal GPP anomalies were calculated for each dataset and displayed via hovmöller diagrams to assess the ability of the model to capture interannual variability.

All comparisons were done on the maximum overlapping timespans, which were determined by the availability of the comparison datasets. Figure 3.1 gives an overview of the overlapping timespans between the different datasets.

Results 4

4.1 Comparisons between model versions

4.1.1 Agreement with independent GPP datasets

The use of additional in-situ (**GAM+**) data improved the correlation between the modeled GPP and the independent GPP datasets. Correlation improvements are in the range of 0.02 and 0.01 r, when comparing to the remotely sensed GPP products MODIS and FLUXCOM GPP (see Figure 4.1). However, the magnitudes of regional improvements can be considerably higher. Spatially, there is a significant variation in the magnitude of correlation improvements, with certain regions experiencing substantial changes ($\pm 0.3\ \Delta r$). Furthermore, MODIS and FLUXCOM GPP exhibit similar spatial patterns in terms of correlation improvements. The most notable correlation gains are observed in tropical savannas, as well as semi-arid and temperate regions, whereas high productivity regions such as tropical forests, tropical monsoon regions around the equator and especially arid regions exhibit the biggest reductions.

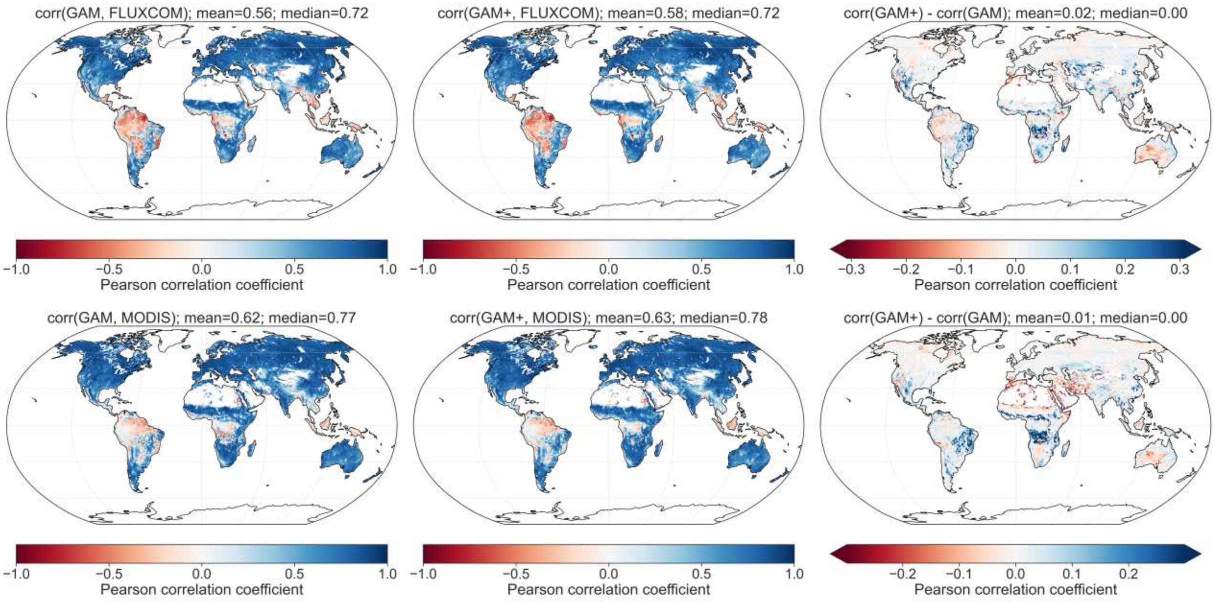


Figure 4.1: Difference in correlation between models with and without added in-situ data and GPP from FLUXCOM (top) and MODIS (bottom). The correlations are based on the common observation period between 2001 and 2016 with a 0.25° spatial and 8d temporal resolution.

When comparing to the GPP from the TRENDY model ensemble (see Figure 4.2) only regional improvements in correlation can be observed. These improvements are particularly notable in the Southern Hemisphere, with the most significant improvement observed in Australia. However, in the global mean, no improvements in correlation could be achieved.

It is important to note that the correlations with TRENDY GPP are generally much lower compared to correlations with the remotely sensed GPP products (0.38 r vs. 0.58-0.63 r).

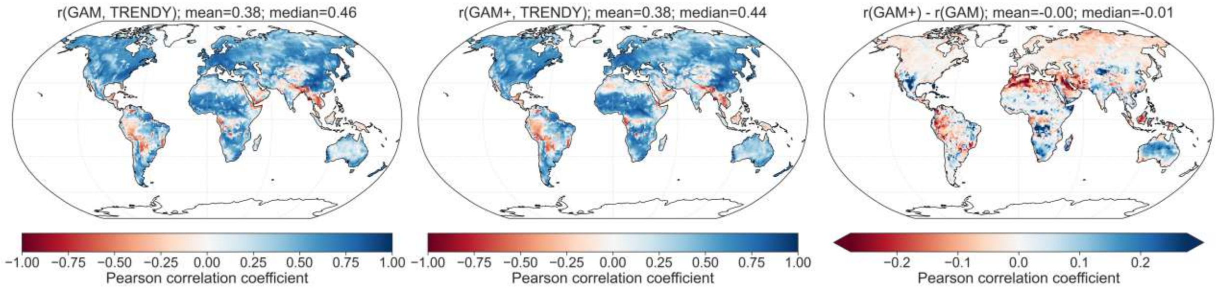


Figure 4.2: Difference in correlation between models with and without added in-situ data and GPP from TRENDY. The correlations are based on the common observation period between 1988 and 2017 with a 0.5° spatial and 1 month temporal resolution.

Switching to a RF regressor and incorporating land cover information (**RF+_LC**) lead to significant improvements in correlation with FLUXCOM GPP (see Figure 4.3; first row). The improvements are substantial at a Δr of 0.05 and are accompanied by significant changes in the spatial patterns of correlation magnitudes. The most notable improvements are observed in tropical regions, particularly in areas with a tropical rainforest and tropical monsoon climate, as well as to a lesser extent in tropical savanna. Correlations in these regions completely shift from strongly negative to clearly positive. On the other hand, temperate regions only showed slight improvements[1], while desert and arid regions (and to a smaller extent, semi-arid regions) exhibited significant decreases in correlation.

1: Temperate regions had already been improved in **GAM+** through the inclusion of more in-situ stations, most of which were located in temperate regions.

However, the results for MODIS (Figure 4.3; second row) were not as favorable, as more areas demonstrated decreasing correlation. Overall, the mean correlation with MODIS decreased by 0.03, primarily due to large decreases in arid regions.[2]

2: Even though the patterns of correlation increases and decreases were similar to those in FLUXCOM, MODIS covers more arid regions where the correlations decreased the most. This leads to reduced correlations in MODIS but not in FLUXCOM. Additionally, the tropical regions where correlations increased are arguably more interesting for GPP.

Additionally, the inclusion of land cover significantly increased the correlations with TRENDY GPP as shown in Figure 4.4. The average increase is 0.05 r globally.

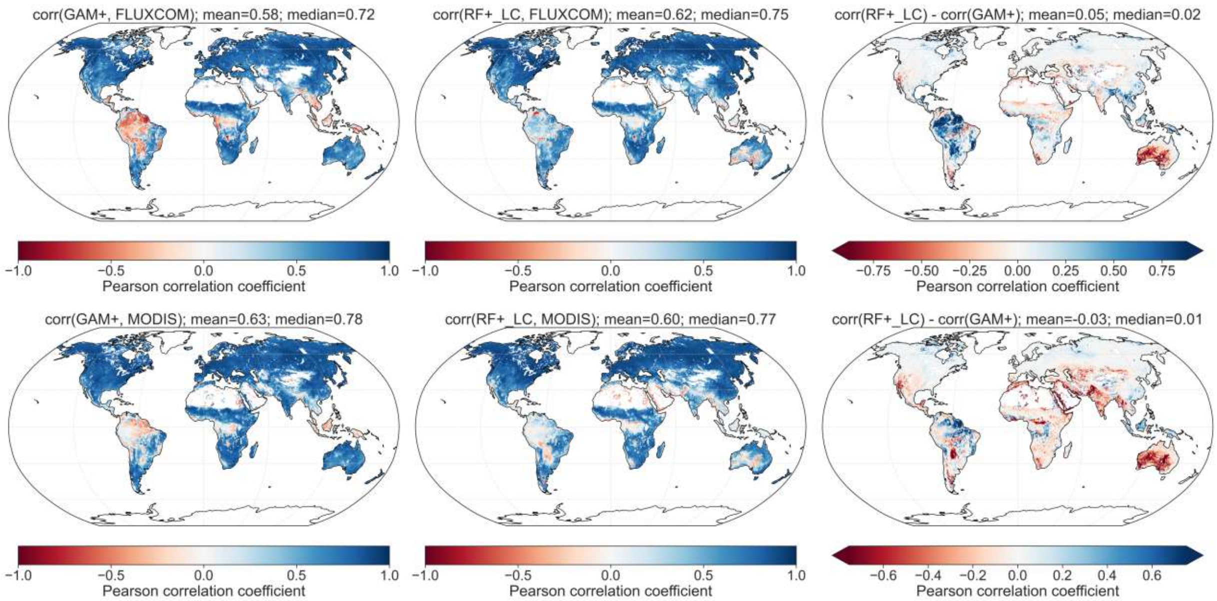


Figure 4.3: Difference in correlation between models with and without land cover data and GPP from FLUXCOM (top) and MODIS (bottom). The correlations are based on the common observation period between 2001 and 2016 with a 0.25° spatial and 8d temporal resolution.

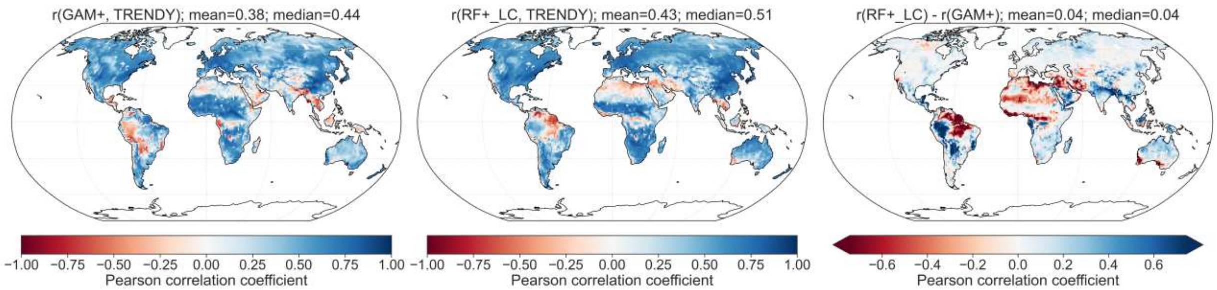


Figure 4.4: Difference in correlation between models with and without land cover data and reference GPP from TRENDY. The correlations are based on the common observation period between 1988 and 2017 with a 0.5° spatial and a monthly temporal resolution

Finally, the inclusion of L-band VOD (**RF+_LC_LVOD**) as a predictor does only yield minor improvements in correlation with the independent GPP datasets. The improvements are minimal, with negligible gains of 0.01 r only for MODIS (see Figure 4.5). For FLUXCOM, the changes are overall insignificant, with no discernible spatial patterns following climate or land cover classes. In contrast, the improvements for MODIS are more pronounced, bringing the correlations with MODIS and FLUXCOM to a similar level. The most noticeable improvements are observed in the Australian desert, although this pattern does not hold true for all desert regions, as decreases in correlation are observed in the Sahara, Arabian desert, and Middle East.

Similar to the remotely sensed GPP products, the improvements in correlation with TRENDY GPP obtained by adding L-band VOD are also negligible (see Figure 4.6).

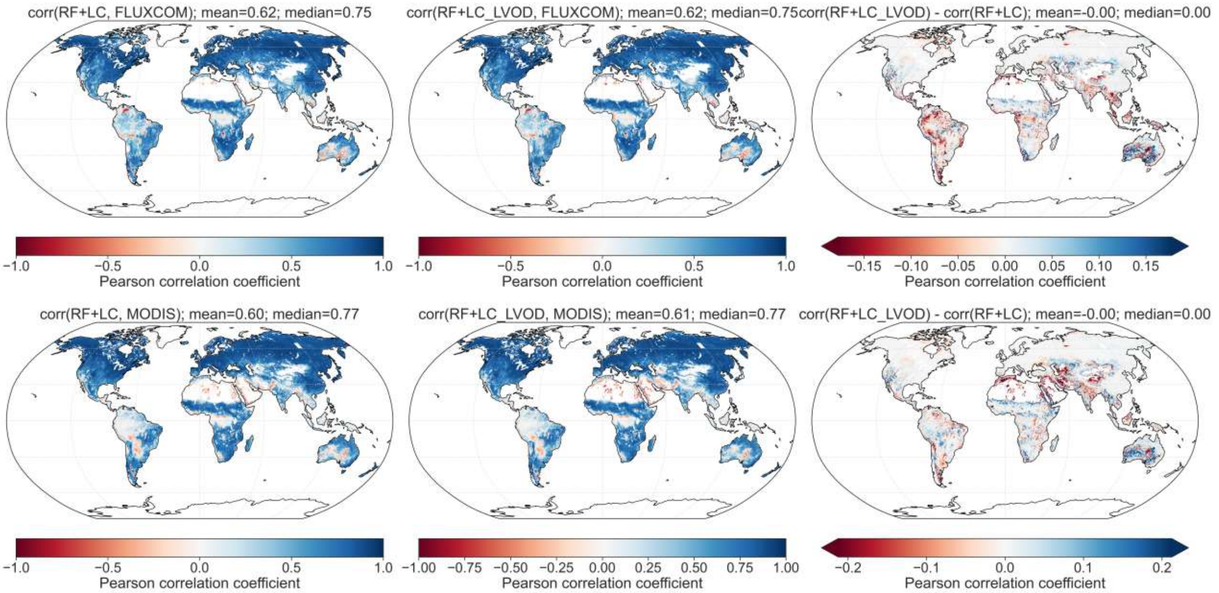


Figure 4.5: Difference in correlation between models with and without L-band VOD and GPP from FLUXCOM (top) and MODIS (bottom). The correlations are based on the common observation period between 2001 and 2016 with a 0.25° spatial and 8d temporal resolution.

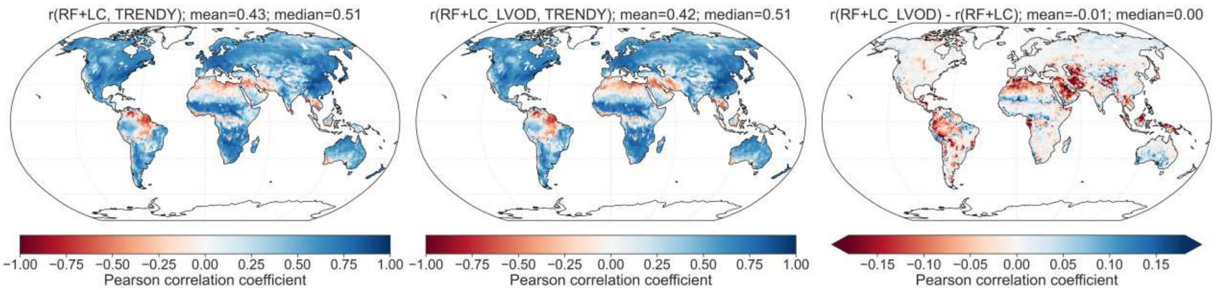


Figure 4.6: Difference in correlation between models with and without L-band VOD and GPP from reference GPP from TRENDY. The correlations are based on the common observation period between 1988 and 2017 with a 0.5° spatial and a monthly temporal resolution.

Overall, the correlation improvements from the first model (**GAM**), as used by (Wild et al. 2022), to the final random forest model with land cover and L-band VOD (**RF+_LC_LVOD**) are highly significant when comparing to the remotely sensed and DGVM-based GPP products. The total correlation improvements from the original model version for MODIS and FLUXCOM, as well as for TRENDY, are shown in the Appendix in Figures B.1 and B.2, respectively.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

4.1.2 Cross Validation

Figure 4.7, contains the cross validation metrics (r Equation 3.3, RMSE Equation 3.4, bias Equation 3.5) for the different models. Each value represents the respective performance metric calculated for one specific site during cross validation. The box-plots show the distribution of the performance metrics across all sites.

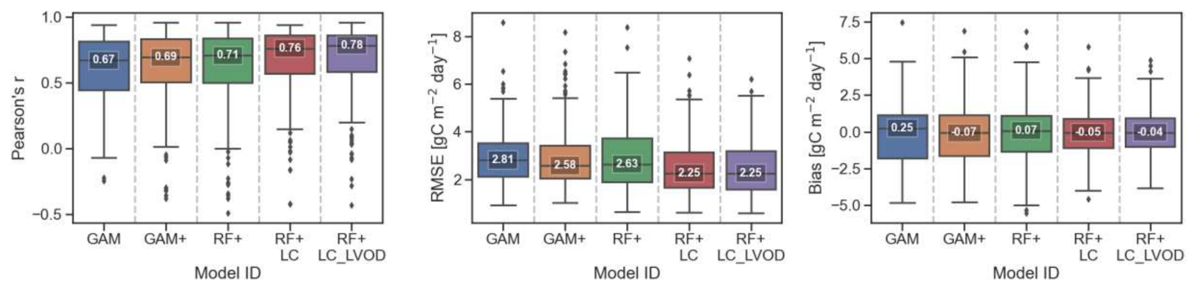


Figure 4.7: Box-plots of cross validation (CV) performance metrics (r Equation 3.3, RMSE Equation 3.4, bias Equation 3.5) for different models. Each value represents the respective performance metric calculated for one specific site during CV. The center line as well as the annotation represent the median while box extents represent the 25th and 75th percentiles. The maximum length of the whiskers is 1.5 times the interquartile range, and outliers are shown as single points.

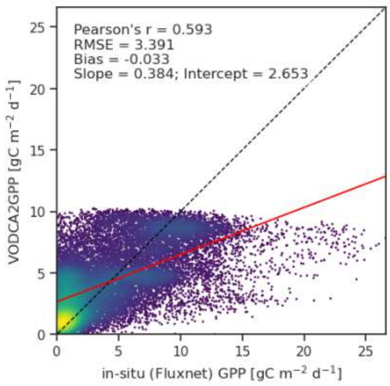
The site based cross validation results (Figure 4.7), show clear improvements for every new iteration of the model. The step from the simple GAM model to adding more in-situ data (blue to orange) is clearly beneficial to the model, as can be seen in the higher correlations (r), the reduced RMSE and the slightly lower bias[3]. Differences between GAM and RF (orange to green) are not as clear, with correlations staying similar and RMSE getting slightly larger while bias marginally improves. Biggest increases are achieved when adding new predictors (green to red and purple). Especially, with the inclusion of LC (red) bringing significant improvements across all metrics. Overall, the final model is significantly better than the first model in all of the chosen metrics.

3: Although it is hard to say if those are really model improvements or just come from the pure fact that more stations (267 compared to 110) are available now.
Many of the new stations are in the Northern Hemisphere and in temperate regions (where the model is known to work better).
The apparently better values could also come from the fact, that the ratio changed towards more "easy to predict" stations. This would show here without necessarily improving the model.

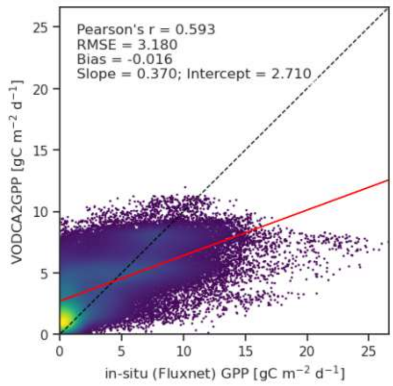
Continuous improvements can also be observed in Figure 4.8, containing plots of in-situ GPP vs. predicted[4] GPP for the different models. The diagonal line represents the 1:1 line, where predicted GPP equals in-situ GPP. The closer the points are to the diagonal line, the better the model performs.

4: during cross validation

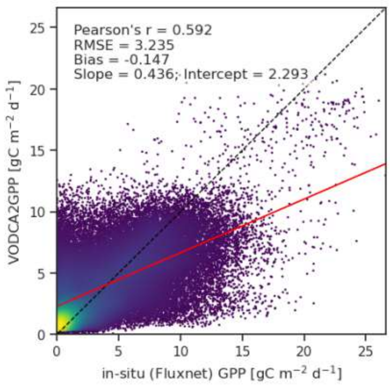
(a) GAM - simple GAM model



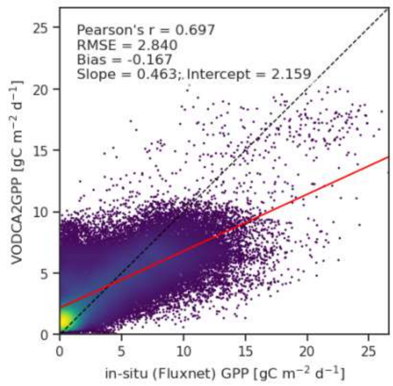
(b) GAM+ - GAM with added in-situ GPP



(c) RF+ - RF with added in-situ data



(d) RF+_LC - RF with added in-situ data and LC



(e) Final VODCA2GPPv2 model: **RF+_LC_-LVOD** - RF with added in-situ data, LC and L-band VOD

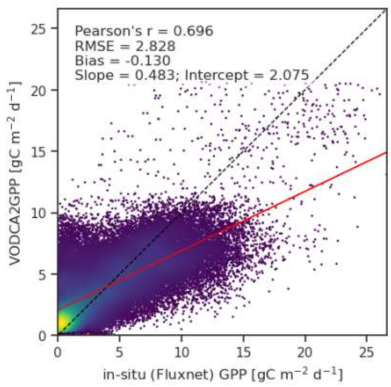


Figure 4.8: Scatter plots of predicted (during CV) GPP against in-situ GPP for the models: **(a)** GAM, **(b)** GAM+, **(c)** RF+, **(d)** RF+_LC, and **(e)** RF+_LC_LVOD. The dashed line represents the 1:1 line, the red line represents the linear regression fit.

The two GAM models (Fig. 4.8a, b) exhibit nearly identical distribution patterns. This is expected since they both rely on the same model. The only difference between them is the higher amount of in-situ data used for GAM+ (Fig. 4.8b).[5] However, it appears that this additional in-situ data does not noticeably impact the performance of the model at site level. Both models reach saturation early, resulting in underestimation of high GPP values. In fact, no GPP values above 10 $gCm^{-2}day^{-1}$ (and almost none for **GAM+**) are predicted.

5: The two GAM figures also indicate that there has been minimal addition of new high GPP in-situ data. This is to be expected, given that most new sites are located in temperate regions in the Northern Hemisphere.

The scattering patterns exhibit changes when switching to a RF regressor (Fig. 4.8c-e). Notably, the inclusion of the new land cover (4.8d) and, to a lesser extent, the L-band VOD predictors (4.8e) leads to further changes in the patterns. All RF models are capable of predicting GPP values above 10 $gCm^{-2}day^{-1}$ to some extent. However, the distribution of points still remains fairly saturated. Especially for the **RF+** model, which is arguably not better than the GAM version.

The addition of land cover in the **RF+_LC** and **RF+_LC_LVOD** models (Figures 4.8d,e) results in a notable improvement in the point distribution, with a higher density of points aligning close to the 1:1 line. As a result, there is an overall better agreement between predicted and in-situ GPP. The linear regression fit (red line) also shows a better fit to the 1:1 line.

Improvements from adding L-band VOD (Fig. 4.8e) are minor again.[6]

6: L-band VOD has other advantages though, such as the better agreement with GPP from MODIS at pixel level.

4.1.3 Latitudinal GPP bias

Figure 4.9 shows the latitudinal GPP patterns of the different models, as well as MODIS and FLUXCOM GPP. It contains the latitudinal GPP means calculated on the common observation period of 2001 to 2016 on pixels present in all datasets.

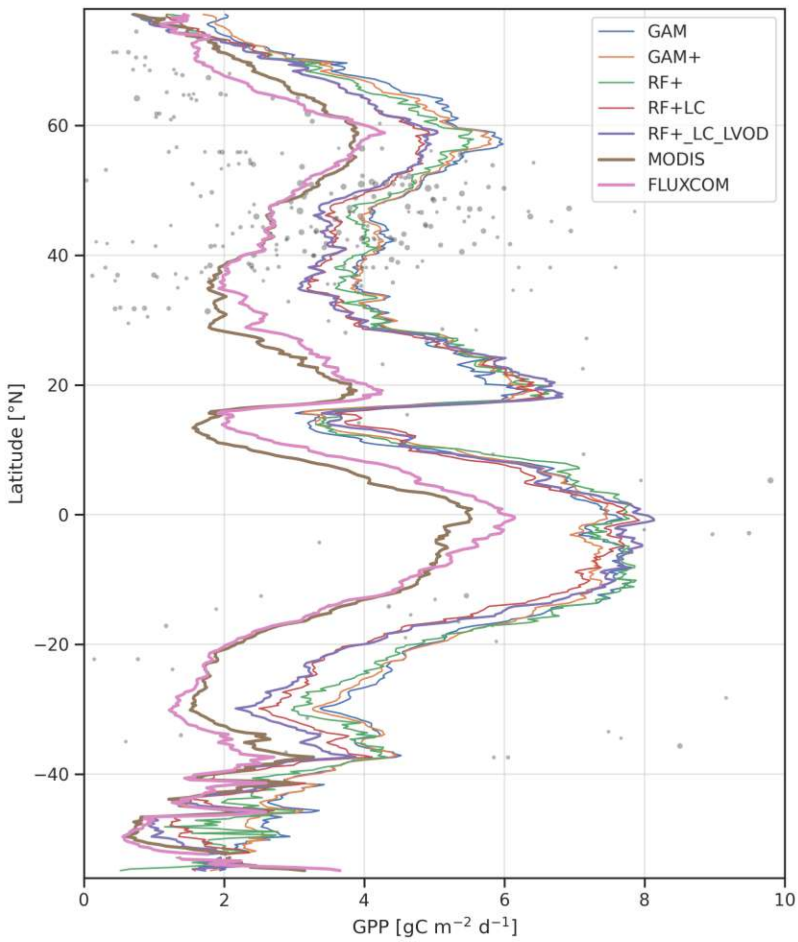


Figure 4.9: Latitudinal GPP mean for different VODCA2GPP models, as well as MODIS and FLUXCOM GPP. The mean was calculated for the common observation period of 2001 to 2016, considering only pixels present in all datasets. The gray dots represent the mean GPP observed at FLUXNET sites, with the marker size representing the number of observations at each site.

As depicted in Figure 4.9, the MODIS and FLUXCOM models exhibit a high level of agreement in predicting GPP. Both models demonstrate similar latitudinal patterns and minimal biases across all latitudes.

Similarly, the various VODCA2GPP models also display comparable latitudinal GPP patterns, although they significantly overpredict GPP compared to MODIS/FLUXCOM.[7] However, this bias is reduced by including new predictors and using a RF regressor.

7: Note that this overprediction is observed in relation to MODIS/FLUXCOM but not necessarily in comparison with in-situ GP (see below).

While including additional in-situ data only slightly reduces the bias (shown by the orange line; **GAM+**), switching to a RF regressor noticeably reduces the bias, especially in the subtropics and temperate zone. The inclusion of land cover and, to a lesser extent, L-VOD further reduces the bias significantly, further increasing the agreement between the VODCA2GPP models and MODIS/FLUXCOM. Unfortunately, there are not many improvements in reducing the bias in the tropics. Notably, it seems that other models perform better in those regions.[8]

8: Interestingly, the final model with L-VOD, which successfully mitigated bias in many latitudes, seems to underperform in large parts of the tropics, particularly in terms of bias towards MODIS/FLUXCOM. It should be noted, however, that predicting GPP in the tropics is notoriously challenging, and there is currently no consensus on GPP magnitudes in those regions. GPP from optical observations, which tend to saturate earlier than microwave data, can certainly not be considered a reliable "true reference" for GPP in the tropics.

Moreover, upon examination of the gray dots representing in-situ GPP observations in Figure 4.9, it becomes apparent that MODIS and FLUXCOM do not necessarily exhibit better agreement with in-situ observation than the VODCA2GPP models. In fact, the VODCA2GPP models seem to more closely resemble in-situ GPP, as their latitudinal means align closer with what is the center of the value range observed in the in-situ GPP data.

This raises the question of whether the VODCA2GPP models offer a superior fit, implying that the accuracy of MODIS/FLUXCOM may be comparatively lower. Overall, in this simple visual comparison, there appears to be better agreement between the VODCA2GPP models and in-situ GPP than with MODIS/FLUXCOM. Although it is challenging to make a definitive assessment, this outcome is certainly not surprising given the simple machine learning approach used in the VODCA2GPP models and more importantly by the fact that the VODCA2GPP models in contrast to the other products are trained on this exact in-situ GPP data shown in the plot.[9] Consequently, it is expected that the VODCA2GPP models would closely align with the patterns observed in the shown in-situ GPP data.

9: MODIS does not employ any in-situ GPP for training, while FLUXCOM does but not the exact subset utilized by VODCA2GPP and displayed here.

The confidence in the correctness of the VODCA2GPP models is further strengthened by comparing its latitudinal means to those of TRENDY GPP (Figure 4.10). Notably, TRENDY GPP exhibits a significantly smaller bias towards VODCA2GPP compared to MODIS and FLUXCOM (as shown in Figure 4.9). In general, the different VODCA2GPP models, especially the final model with L-VOD (purple), demonstrate strong agreement with TRENDY GPP and exhibit highly similar latitudinal patterns.[10]

10: However, it should be noted that Figures Figures 4.9 and 4.10 cannot be directly compared because the latitudinal means were calculated for different time periods. Moreover, the spatial availability of the data sets also differs, with TRENDY GPP including more regions than MODIS/FLUXCOM. As a result, the latitudinal means were calculated using different latitudinal pixel subsets, making them not directly comparable. Additionally, TRENDY GPP has a lower spatial resolution compared to MODIS-/FLUXCOM, which may further affect the latitudinal means.

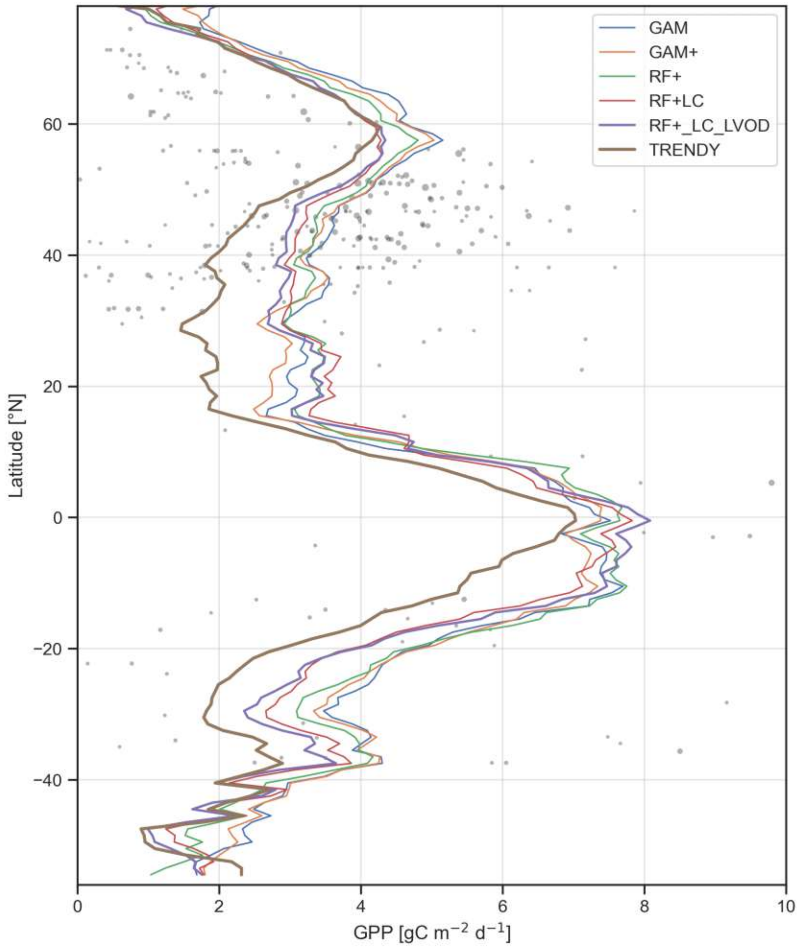


Figure 4.10: Latitudinal GPP mean for different VODCA2GPP models, as well as TRENDY GPP. The mean was calculated for the common observation period of 1988 to 2017, considering only pixels present in all datasets. The gray dots represent the mean GPP observed at FLUXNET sites, with the marker size representing the number of observations at each site.

4.2 The final VODCA2GPPv2 model

The previously presented results show the **RF+_LC_LVOD** model to be the best performing model. Since it was chosen as the final model it will be referred to as **VODCA2GPPv2** from now on. The following sections will focus on the results of this model.

4.2.1 Bias to independent GPP datasets

Figure 4.11 shows the mean GPP for, and biases between, the original GAM model (**bb**), the updated VODCA2GPPv2 model (**aa**), and reference GPP from MODIS (**cc**) and FLUXCOM (**dd**). The maps are arranged in a matrix like layout, with the main diagonal showing the mean GPP for each model and the off-diagonal elements showing the difference in mean GPP (bias) between the models. The GPP means as well as biases are all calculated for the common observation period of 2001 to 2016.

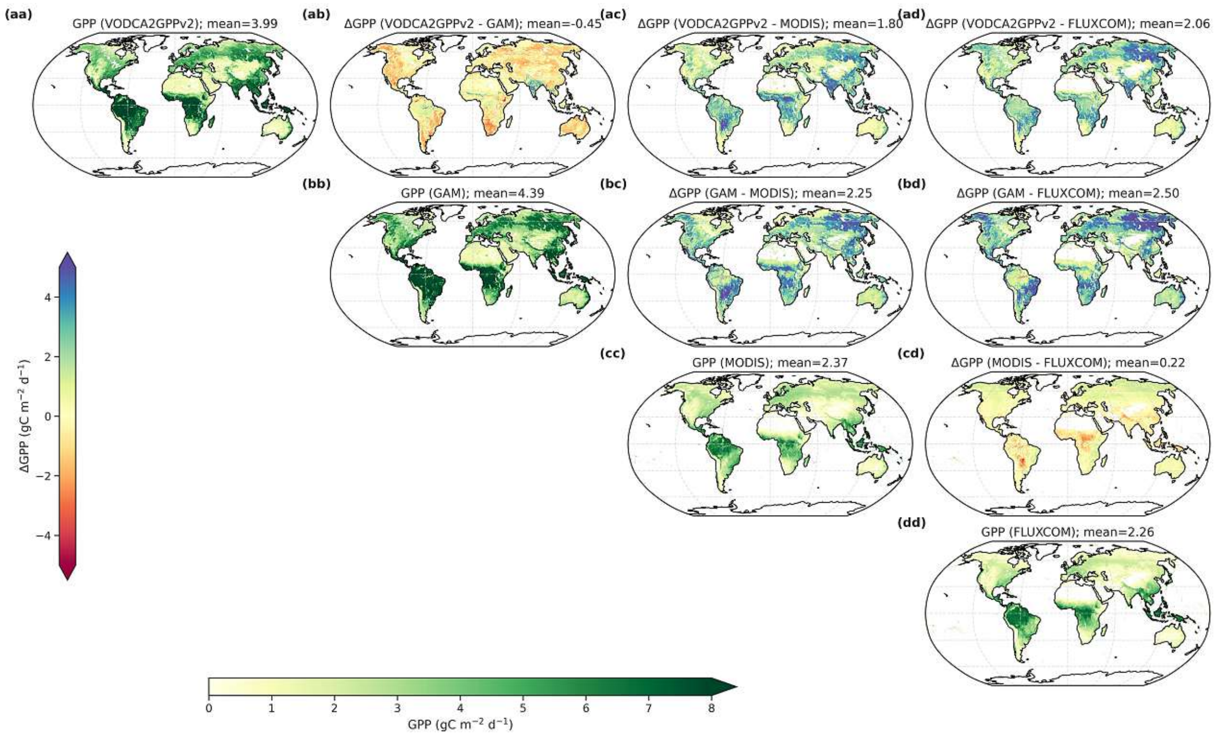


Figure 4.11: Mean GPP and bias for different models and MODIS and FLUXCOM GPP, arranged in a matrix like layout. The main diagonal shows the mean GPP for each model and the off-diagonal elements show the difference in mean GPP (bias) between the models. Means as well as biases are calculated for the common observation period between 2001 and 2016 with a 0.25° spatial resolution.

One thing immediately apparent from Figure 4.11, is that VODCA2GPP versions overestimate GPP across nearly all regions, often with significant discrepancies.

This behavior, while being very pronounced for the original GAM version of the model is slightly mitigated in the new VODCA2GPPv2. This can be observed by comparing the biases between the two versions of VODCA2GPP and the reference datasets[11]. Generally, the new VODCA2GPPv2 model yields lower GPP predictions compared to GAM, which is also reflected in reduced biases with the reference GPP. Across the map, biases with respect to the reference datasets have improved (with the mean bias decreasing by 0.45 and 0.44 $gCm^{-2}d^{-1}$ for MODIS and FLUXCOM), except for India where the bias increased. These findings are consistent with the observations from the latitudinal bias plot (Figure 4.9) discussed in Subsection 4.1.3, which also provides evidence of the overprediction of final VODCA2GPP compared to the original version at approximately 20° north[12].

Overall, Figure 4.11 demonstrates a significant improvement in the reduction of biases with the independent GPP datasets for VODCA2GPP. However, it should be noted that the biases, although mitigated to some extent, still remain relatively large. This is particularly evident when comparing with the bias between the two comparison datasets MODIS and FLUXCOM (Figure 4.11**cd**), which show a much higher level of agreement.

11: The Bias between the GAM model and the validation datasets (**bc**, **bd**) is consistently larger than the bias of VODCA2GPPv2 version with the same datasets (**ac**, **ad**).

12: In the latitudinal plot, VODCA2GPPv2's prediction is noticeably larger than the one from the original version within a latitude range of about 15° centered around 20° north. This observation aligns with the region of increased bias in India whereas in almost all other regions, the bias maps as well as latitudinal bias plots show decreased bias.

4.2.2 Cross Validation Results - Site based comparison with in-situ GPP

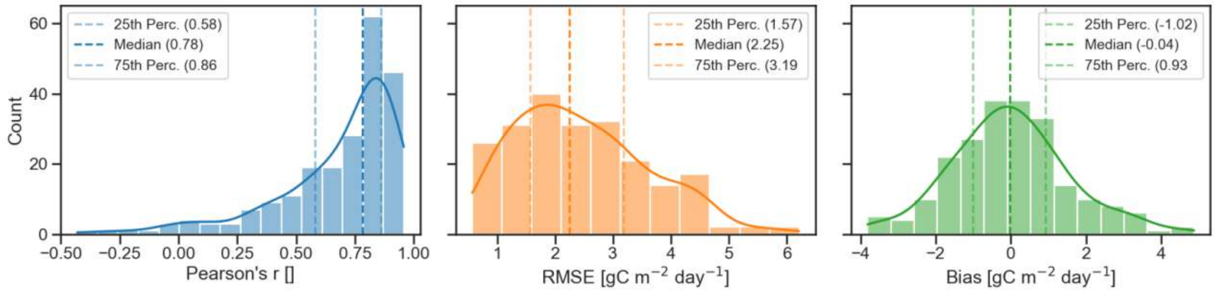


Figure 4.12: Histograms of site-based cross validation performance metrics (r eq. 3.3; RMSE eq. 3.4; bias eq. 3.5) for the final VODCA2GPPv2 model.

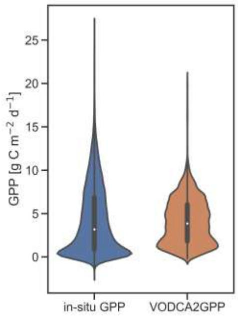


Figure 4.13: Violin plots of in-situ GPP and VODCA2GPPv2 GPP at all sites.

The cross-validation metrics (r eq. 3.3, RMSE eq. 3.4, bias eq. 3.5) for the VODCA2GPPv2 model are displayed in Figure 4.12. The histograms illustrate the performance metric distribution across all sites. Correlations between the model and in-situ GPP are particularly strong, with over 75% of the sites exhibiting higher correlations than 0.58 r. Additionally, the median correlation is remarkably high at 0.78 r. The bias follows a fairly normal distribution centered around 0, indicating that the model does not have a tendency to consistently over- or underpredict GPP. Roughly 50% of the sites have a bias of smaller magnitude than 1. However, it's worth noting that both bias and RMSE can become considerably large, especially when considering the total range of GPP values (see Figure 4.13).

Displaying the correlation coefficients on a map (Figure 4.14) reveals that the model performs best in Europe and North America. Generally, the Northern Hemisphere outperforms the Southern Hemisphere significantly. The model's performance aligns well with the distribution of sites, as most sites are located in temperate regions in the Northern Hemisphere. This indicates that the uneven distribution of sites is inherited by the model, resulting in its superior performance in the Northern Hemisphere and temperate regions.

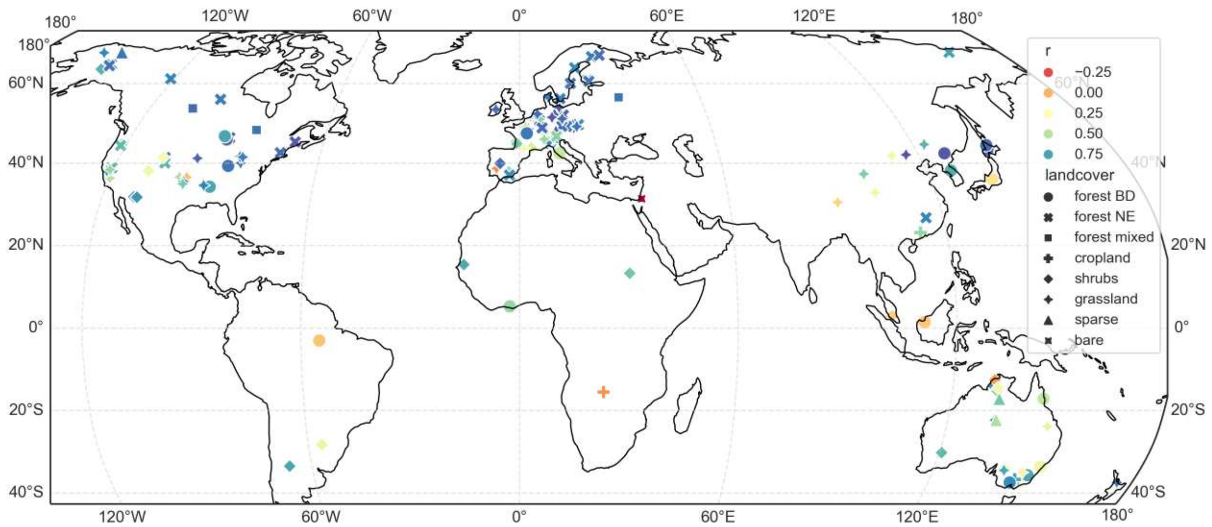


Figure 4.14: Map of Pearson correlation coefficients (r Equation 3.3) between predicted (during CV) and reference (in-situ) GPP at FLUXNET sites. Sites are colored by correlation coefficient (continuous color map), the marker type represents the LC class.

Interestingly, the correlations in Australia are not as strong as those in North America and Europe, despite the presence of numerous sites. This discrepancy may be attributed to the fact that Australia is a region with limited water availability, where temperature is not such a strong constraint on GPP as in other regions. The poor performance in the tropics, on the other hand, can likely be attributed to the limited availability of sites and consequently training data in these areas.

Model performance does not only vary spatially, but also depends on the land cover class of the site. This can be observed by looking at the different markers representing each land cover class in the correlation map (Figure 4.14). However, it becomes even more clear when the sites are grouped into land cover classes and box plots are used to visualize the performance metrics of each group. This is demonstrated in the following Figure 4.15.

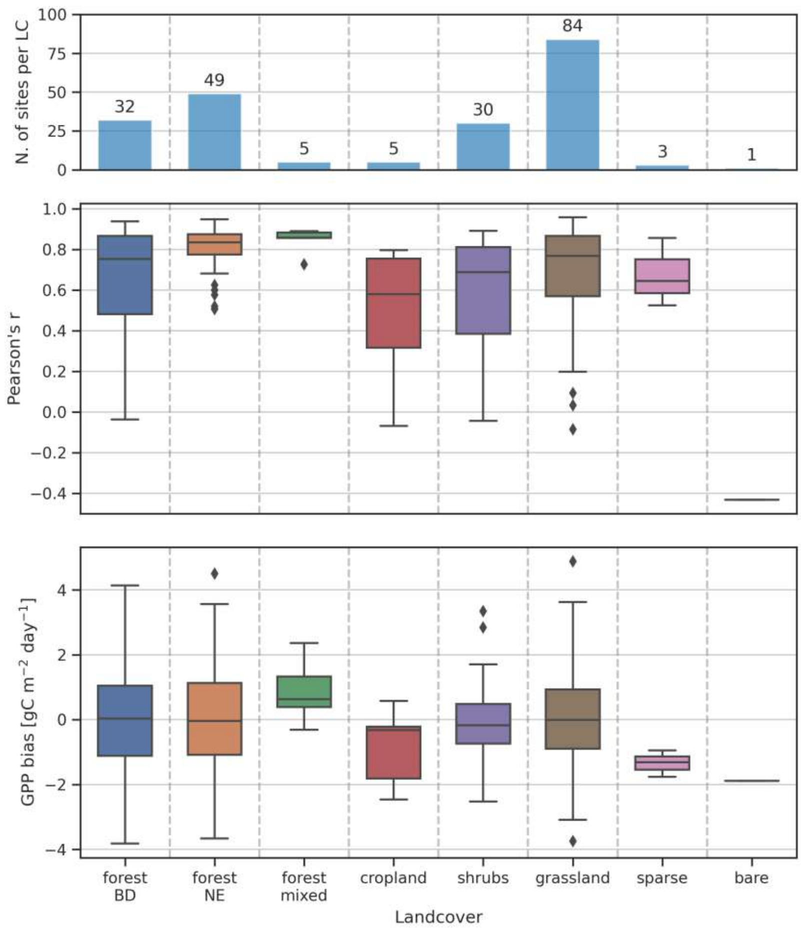


Figure 4.15: Box-plots of Pearson correlation coefficients (mid) and biases (bottom) calculated during CV, grouped by land cover classes. Together with histogram showing amount of sites per LC class.
Each value represents the respective performance metric calculated for one specific site during cross validation. Box extents represent the 25th and 75th percentiles. The maximum length of the whiskers is 1.5 times the interquartile range, outliers points are shown as single dots.

As seen in Figure 4.15, correlations between predicted and in-situ GPP vary considerably between land cover classes. The land cover classes with the highest number of sites[13] tend to exhibit the best model performance. The superior model performance for these classes can be attributed to the larger sample size of training data, and the model therefore being trained to work better for these classes. Correlations are also generally higher on forested sites, especially on mixed forests. On the other hand, low and sparse vegetation sites[14] generally have slightly lower correlations, except for grasslands which perform quite well. In forested sites, NE stations consistently display exceptionally high correlations across almost all sites,

13: grassland, needle-leaved evergreen forest, broadleaved deciduous forest, shrubs

14: such as croplands, shrubs, and sparse vegetation while BD stations exhibit a wider range of correlations. This difference may be attributed to the fact that NE sites are exclusively located in the Northern Hemisphere, whereas BD sites are more geographically dispersed. Additionally, the temporal dynamics of evergreen vegetation in needle-leaved evergreen forests may be easier to predict compared to broadleaved deciduous forests.

Similar patterns observed for correlations are also evident for the bias (bottom of Figure 4.15). The most represented classes, such as grasslands, needle-leaved evergreen (NE) and broadleaved deciduous (BD) forests, and shrubs, consistently demonstrate the best performance, with biases centered around 0. The effects of uneven representation in the training data are, however, even more pronounced for biases than for correlations. The underrepresented classes notably exhibit worse performance, showing clear systematic biases with a clear tendency to overpredict GPP (positive bias) in mixed forests and underpredict GPP (negative bias) in croplands[15].

15: This negative bias for croplands is particularly interesting as it could show the impacts of irrigation practices.

Overall, Figure 4.15 demonstrates that the model performs best for the most represented land cover classes. This is expected since the model is trained to excel on these classes. However, there is a noticeable difference in performance for underrepresented classes. On correlations the model still performs relatively well for these classes. On the other hand, the performance on bias is notably worse, for the underrepresented classes. This indicates that the generalizability of the model, particularly in terms of bias, is not yet ideal.

Note

Figure B.3, in the Appendix, contains a similar Figure to 4.15, with the difference that it also includes the correlations and bias for the GAM+ model. This allows to illustrate the improvements on model generalizability though the inclusion of LC and LVOD across biomes and will be referred to in Section 5.4 in the discussion.

4.2.3 Feature Importances

The Mean Decrease in Impurity (MDI) and SHAP values, described in Subsection 3.3.2, were utilized to assess the importance of various predictors in VODCA2GPPv2. These values are visualized in Figure 4.16 (bottom row: 4.16c and 4.16d), alongside the feature importances for the simple model (top row: 4.16a and 4.16b) which only uses the original VODCA2GPP model's previous predictors (VOD, dVOD, medVOD, T2M).

Both MDI (4.16a and 4.16c) and SHAP values (4.16b and 4.16d) highlight temperature as the most significant predictor, for the basic as well as for the full-feature model. This is logical considering the strong temperature dependence of R_a. Additionally, the simple model assigns high importance to median VOD. It is important to note that median VOD serves solely as a static predictor without providing temporal dynamics. As a result, in the simple model the temporal dynamics are primarily driven by temperature.

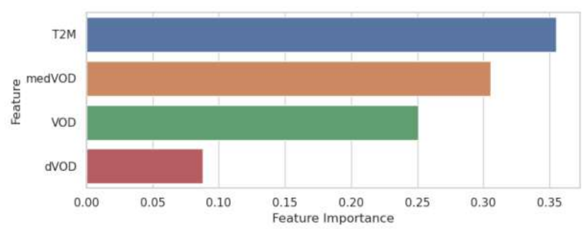
In the full-feature model, although temperature remains the most important predictor, the VOD timeseries becomes the dominant VOD predictor. This aligns well with the expected behavior of a VOD-based model, where VOD should primarily be utilized for capturing temporal dynamics. Consequently, the importance of median VOD decreases as its role is assumed by land cover data in the form of fractional coverages of PFTs[16].

16: It is worth mentioning that median VOD, suggested as a predictor by Teubner et al. (2019) to serve as a proxy for vegetation density, may no longer be necessary, as its role is mostly fulfilled by land cover data in the form of fractional coverages of PFTs.

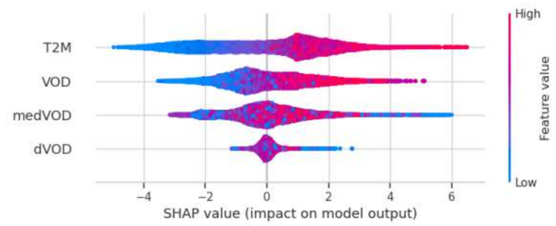
Many of the land cover predictors exhibit high feature importances, with Broadleaved Deciduous (BD) and Bare Soil (Bare) being the most important ones.[17]

17: It is interesting to observe the inverse relationship between BD and Bare with the predicted GPP. A high feature value of BD increases the predicted GPP, while a high feature value of Bare is associated with lower GPP value.

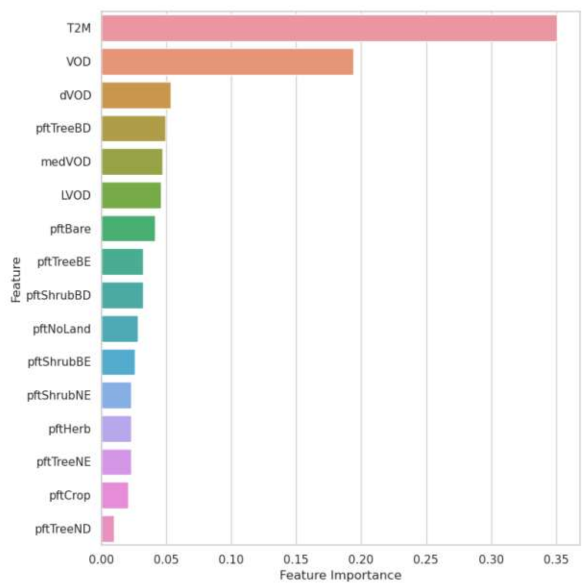
(a) Mean Decrease in Impurity (MDI) - **RF+**



(b) SHAP values - **RF+**



(c) Mean Decrease in Impurity (MDI) - **RF+_LC_LVOD**



(d) SHAP values - **RF+_LC_LVOD**

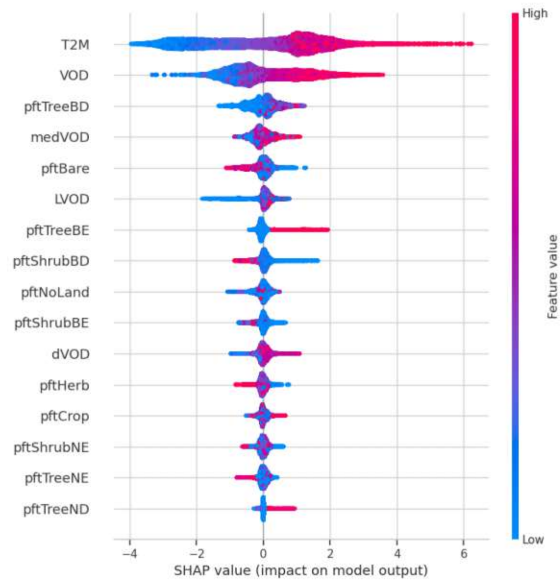


Figure 4.16: MDI and SHAP values for the random forest models trained on the simple (VOD, T2M) and extended (VOD, T2M, LC, LVOD) feature sets.

4.2.4 Spatio-temporal GPP patterns - GPP anomalies

To compare anomaly pattern in space and time between the different datasets, the GPP anomalies were calculated for each dataset by subtracting the mean GPP for each month from the respective monthly GPP values. The anomalies were averaged on a latitudinal basis and are shown in Figure 4.17. For comparisons Figure 4.17 also contains the GAM model, which is equivalent to the VODCA2GPPv1 model by Wild et al. (2022). The anomalies are calculated for the common observation period of 2001 to 2016.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

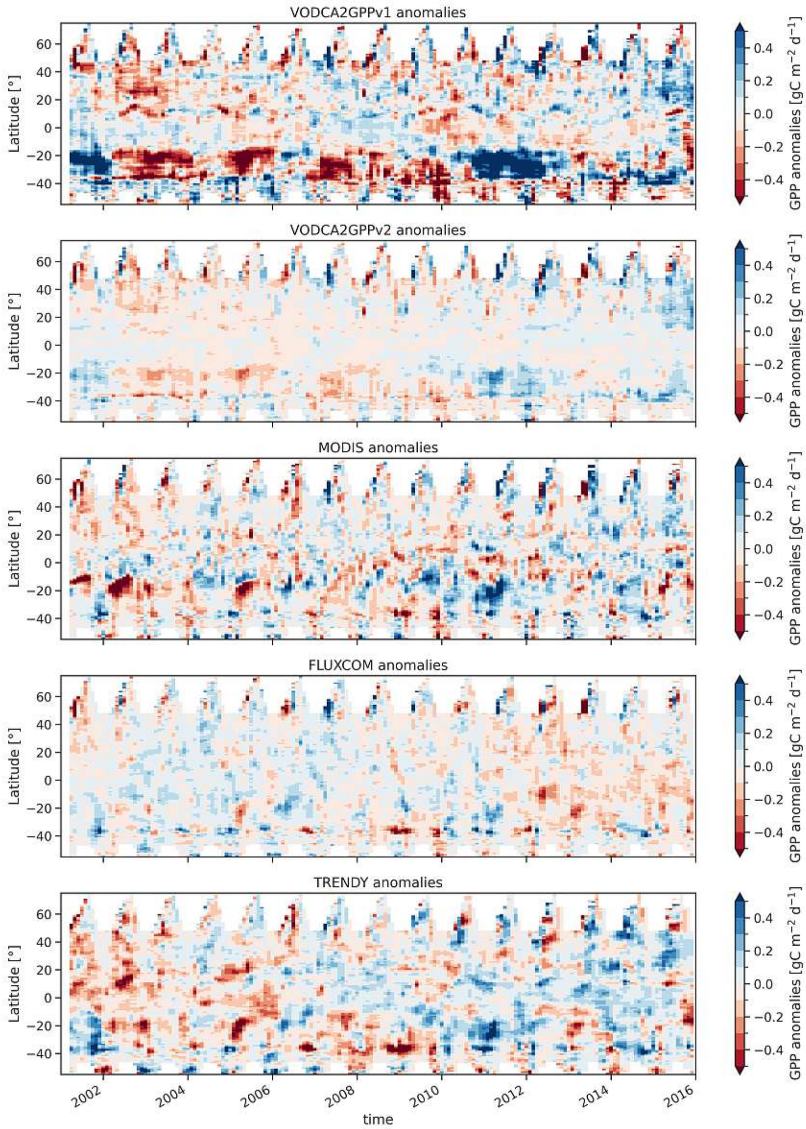


Figure 4.17: Hovmöller plots of monthly mean GPP and GPP anomalies.
The anomalies are calculated by subtracting the mean GPP for each month from the respective monthly GPP values on a latitudinal basis.

Visually anomalies from VODCA2GPPv2 seem to match MODIS anomalies best, followed by TRENDY, while FLUXCOM anomalies match worst. Several of the extreme events captured in VODCA2GPPv2 are also detected in at least one of the other datasets. The most prominent anomalies examples present in VODCA2GPPv1 and highlighted by Wild et al. (2022) are also detected in VODCA2GPPv2. These include the pronounced positive anomalies centered at 20-30 °S from 2010 to 2012, likely resulting from record-breaking rainfall in Australia during that period (Wardle et al. 2013). VODCA2GPPv2 also captures the prominent negative anomalies

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

(Wardle et al. 2013): Wardle et al. (2013), *Greening of Arid Australia: New Insights from Extreme Years* around 20°C in 2002/2003 and early 2005 discussed by Wild et al. (2022). These anomalies can be attributed to severe drought events occurring in those years (Bureau of Meteorology 2003, 2005), which are often associated with El Niño events (Taschetto and England 2009). Furthermore, a distinct decline in GPP during 2015/2016, within similar latitudes, may be linked to El Niño-related drought events (Zhai et al. 2016).

Although Wild et al. (2022) reported that extreme events were more evident in VODCA2GPPv1 compared to their comparison datasets[18], this is not the case for VODCA2GPPv2. VODCA2GPPv2 is clearly less influenced by extreme events than it's predecessor. This distinction is clearly depicted in Figure 4.17, where GAM and VODCA2GPPv2 exhibit very similar patterns but with a substantial difference in magnitude. While part of this difference can be explained by the overall shorter range of GPP values in VODCA2GPPv2 (VODCA2GPPPv2 on average predicts 0.45 $gC/m^2/d$ less than GAM; refer to Figures 4.9 and 4.11), the disparity in magnitude remains significant. One possible explanation for this discrepancy is the importance of the new land cover predictors in VODCA2GPPv2 (see Figure 4.16), which, due to their static nature, may reduce temporal dynamics in the prediction and consequently diminish the magnitude of the anomalies.

(Bureau of Meteorology 2003): Bureau of Meteorology (2003), *Annual Climate Report 2003*

(Bureau of Meteorology 2005): Bureau of Meteorology (2005), *Annual Climate Report 2005*

(Taschetto and England 2009): Taschetto et al. (2009), *El Niño Modoki Impacts on Australian Rainfall*

(Zhai et al. 2016): Zhai et al. (2016), *The Strong El Niño of 2015/16 and Its Dominant Impacts on Global and China's Climate*

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

18: They used the same comparison datasets: FLUXCOM, MODIS, and TRENDY GPP

Discussion 5

5.1 Observed bias between VODCA2GPP and independent GPP products

There is a minimal bias observed between VODCA2GPP and in-situ GPP measurements (Figures 4.7 and 4.15). However, a substantial bias exists between VODCA2GPP and other RS based GPP products (Figures 4.9 and 4.11).

In tropical regions, where bias between VODCA2GPP and FLUXCOM and MODIS is the largest, this bias can be partly explained by a reported and observed tendency of FLUXCOM and MODIS to underestimate GPP in these regions. For instance, the FLUXCOM RS setup used in this study, has been reported to yield lower global estimates compared to the FLUXCOM RS+METEO setup or GPP estimates from vegetation models (Jung et al. 2020). Similarly, MODIS has been found to underestimate GPP in tropical regions (Turner et al. 2006). The need for improved constraints on GPP estimates, particularly in the tropics, is widely acknowledged (MacBean et al. 2018), and various studies have addressed this issue (MacBean et al. 2018; Wu et al. 2020). However, the low availability of in-situ estimates often hampers these efforts.

(Jung et al. 2020): Jung et al. (2020), *Scaling Carbon Fluxes from Eddy Covariance Sites to Globe: Synthesis and Evaluation of the FLUXCOM Approach*

(Turner et al. 2006): Turner et al. (2006), *Evaluation of MODIS NPP and GPP Products across Multiple Biomes.*

(MacBean et al. 2018): MacBean et al. (2018), *Strong Constraint on Modelled Global Carbon Uptake Using Solar-Induced Chlorophyll Fluorescence Data*

(Wu et al. 2020): Wu et al. (2020), *Using SMOS Soil Moisture Data Combining CO2 Flask Samples to Constrain Carbon Fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS)*

Outside the tropics, there are still discrepancies in absolute GPP, although they are significantly less pronounced. One possible explanation, already discussed by Wild et al. (2022), for this behavior in regions with pronounced seasonality is the presence of high VOD during winter months, where little to no primary productivity is expected. This overestimation can be attributed to the water content in vegetation that remains present even during dormant periods. The sensitivity of microwaves to this water content leads to non-zero VOD and, consequently, non-zero GPP (Teubner et al. 2021). This bias affects regions with strong seasonality and a pronounced dormant period, which could possibly explain (part of) the bias observed in temperate and continental climates.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

In arid regions, the bias is possibly affected by the effect of isohydricity. This term refers to the water regulation adaptation of plants in cases of low water availability. In drought-prone regions, plants often reduce transpiration by limiting stomatal conductance in order to maintain a constant water potential even during times of extreme water scarcity (Sade et al. 2012). This isohydric behavior of vegetation could partly explain the relatively high VOD and consequently the overestimated GPP in those regions (Teubner et al. 2021).

(Sade et al. 2012): Sade et al. (2012), *Risk-Taking Plants*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

Furthermore, surface water contamination has been observed in some VODCA pixels, which partially contain water bodies such as lakes and rivers. These pixels consistently exhibit lower VOD values compared to neighboring pixels without water bodies. This discrepancy has two implications. Firstly, it leads to underestimation in the VODCA2GPP model for pixels that contain surface water. Secondly, it affects the model training process. If a station falls within a water-contaminated pixel, the

VOD is underestimated at the 0.25° pixel scale, while in-situ GPP remains largely unaffected. Consequently, this leads to a slight but systematic global overestimation. Although masking for water-contaminated pixels seems promising, Wild et al. (2022) showed it would significantly reduce the already limited data available for training, and was thus deemed not worth the trade-off.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

Finally, it must be acknowledged that there is currently no consensus among GPP datasets, and especially the productivity magnitudes can vary greatly between datasets. GPP from DGVMs seems to match the magnitudes of VODCA2GPP more closely, as was tested with TRENDY GPP, but it is important to note that these models are not without their own limitations. On the other hand, the optical nature of many RS datasets makes them more susceptible to saturation, which undoubtedly impacts their relatively lower predictions. Hence, VODCA2GPP, but also the other GPP datasets for that matter, should not be regarded as an absolute reference but rather as supplementary data aiding in a more comprehensive understanding of global GPP and its role in the carbon cycle.

5.2 Limited availability of in-situ GPP and independence of validation datasets

A common challenge in the process of upscaling GPP measurements to derive global GPP estimates is the limited availability of in-situ observations. In the context of VODCA2GPPv1, this issue was particularly pronounced, but was partially addressed in this thesis via the inclusion of two new FLUXNET datasets. As a result, the number of sites could be more than doubled. However, although this expansion of training data was important and beneficial, the problem of insufficient global coverage remains unresolved as most sites overlap with regions already covered by the previous training data. Achieving equal coverage across all regions still remains a distant goal, as shown in Figure 2.3 and Appendix A.

The scarcity of in-situ GPP measurements, especially their uneven distribution, not only hinders achieving spatially consistent upscaling performance but also impedes fair evaluation and validation at the global scale.

Alternatively, VODCA2GPP (and global GPP products in general) can be evaluated by comparing them with independent global GPP products. However, this approach poses a different challenge, as the question of whether different RS GPPs are truly independent has to be addressed. In the absence of alternative high-accuracy GPP observations, FLUXNET GPP is extensively used in deriving most (if not all) global RS GPP products. For example, both FLUXNET and MODIS rely on in-situ GPP measurements from FLUXNET to some extent. FLUXCOM was trained against FLUXNET GPP[1] (Jung et al. 2020; Tramontana et al. 2016), and MODIS GPP has been partly calibrated using data from select FLUXNET stations (Steven W. Running et al. 2004). Consequently, they cannot be considered fully independent from VODCA2GPP. However, at present, there are no other alternatives for constraining global GPP estimates besides utilizing FLUXNET measurements (Teubner et al. 2021).

1: although with a different subset of stations

(Jung et al. 2020): Jung et al. (2020), *Scaling Carbon Fluxes from Eddy Covariance Sites to Globe: Synthesis and Evaluation of the FLUXCOM Approach*

(Tramontana et al. 2016): Tramontana et al. (2016), *Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms*

(Steven W. Running et al. 2004): Steven W. Running et al. (2004), *A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production*

In contrast, process-driven GPP estimations such as TRENDY GPP from DGVMs can be largely considered independent from VODCA2GPP.

5.3 The random forest regressor

The random forest algorithm was chosen over other machine learning algorithms, and specially over the GAM approach of VODCA2GPPv1 (Wild et al. 2022) because of its ability to handle complex interactions between predictors and its robustness against overfitting. This is particularly important in the context of VODCA2GPPv2, where the number of predictors is relatively high compared to the number of observations. Even when comparing simpler models with the original feature set of VODCA2GPPv1, Random Forest performed slightly better than GAM in cross-validation, especially in terms of correlation and bias. It also reduced latitudinal bias to MODIS, FLUXCOM, and TRENDY GPP datasets (Figures 4.9 and 4.10). While the difference in models with limited predictors was noticeable, it was not substantial. The real improvements came from the model with land cover predictors. Although GAM was not tested with new predictors, the added complexity of the model makes RF a better choice. This aligns with the suggestion by Schmidt et al. (2023), who proposed that GAM may not be sufficient to accurately model the complex relationship between VOD and vegetation properties, particularly when including land cover predictors.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

(Schmidt et al. 2023): Schmidt et al. (2023), *Assessing the Sensitivity of Multi-Frequency Passive Microwave Vegetation Optical Depth to Vegetation Properties*

5.4 Land cover and improved generalizability of the model

As demonstrated in previous figures, the performance of the model in relation to in-situ measurements and bias with independent GPP is inconsistent across different regions of the world (Figure 4.11), latitudes (Figures 4.9 and 4.10) or LC classes (Figure 4.15). Some reasons for the large discrepancies between GPP products and the implications of limited and unevenly distributed in-situ GPP, have already been discussed in Sections 5.1 and 5.2. However, the question remains whether the model is capable of generalizing across different biomes and LC classes, and whether the inclusion of LC as a predictor has been successful in this regard.

The inclusion of LC information has significantly contributed to reducing latitudinal bias to the FLUXCOM, MODIS, and TRENDY datasets. In terms of correlations, it has reversed the negative agreement observed in a lot of the highly densely vegetated areas like the Amazon rainforest, resulting in improved consistency across different biomes. Generally, there have been substantial improvements in the correlation of the model with independent GPP in regions where in-situ data is limited. This suggests that the model can effectively use LC information for generalization, partially compensating for the lack of in-situ data.

However, it is important to note that the model still heavily relies on in-situ data. This dependence becomes evident when considering CV performance across different LC classes. As depicted in Figure 4.15 and discussed in Subsection 4.2.2, the model performs better in LC classes that are well-represented in the training data. Notably, this behavior is more prominent in the version of VODCA2GPP without LC (refer to Figure B.3 in the appendix for a comparison of the 2 model versions). Specifically, when using LC predictors, correlations exhibit greater consistency across LC classes, and biases are closer to zero with a significant reduction in systematic over- and underestimation based on LC class.

Despite these improvements, there is still room for further enhancement, as performance continues to vary across LC classes even after incorporating LC information. This variability can partly be attributed to the microwave-based approach, which may exhibit varying sensitivities to different vegetation types. Moreover, the lack of sufficient in-situ data for certain LC classes also contributes to this variability. Finally, while LC data can contribute to generalization across different LC classes to some extent, it remains imperative to have adequate in-situ data in order to accurately train the model for all the different LC classes.

5.5 L band VOD and the wavelength dependency of the VOD - GPP relationship

L-band VOD was introduced for its ability to penetrate deeper vegetation layers, making it more sensitive to areas with high biomass and vegetation density. However, assessing the results is challenging, as there are improvements in reducing latitudinal biases (Figures 4.9 and 4.10) and in correlation with MODIS (Figure 4.5), but also an increased positive bias towards independent GPP datasets in the tropics. This outcome can be attributed to L-band VOD being less saturated in tropical regions, leading to higher predictions. However, it contradicts the objective of reducing latitudinal bias.

L-band VOD provides more robust AGB estimates compared to lower frequency VOD. Nevertheless, the impact of potential saturation with biomass on GPP estimation is less straightforward, especially in densely vegetated areas like the tropics (Teubner et al. 2021). Teubner et al. (2019, 2018) demonstrated that X-band VOD has a stronger relationship with GPP than L-band VOD. This finding may appear unexpected considering the higher sensitivity of L-band VOD to AGB. However, AGB is largely composed by woody structural components. In contrast, X-band VOD is more sensitive to metabolically active plant parts like leaves and fine roots, making it a suitable estimator for GPP (Teubner et al. 2021). The use of VOD as a proxy for aboveground metabolically active parts, related to GPP, is supported by its sensitivity to water content in metabolically active cells.

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

(Teubner et al. 2018): Teubner et al. (2018), *Assessing the Relationship between Microwave Vegetation Optical Depth and Gross Primary Production*

(Teubner et al. 2021): Teubner et al. (2021), *Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production*

While L band VOD might not be particularly suited as the primary VOD input for predicting the temporal dynamics of GPP, it can still be used as an additional predictor. Although its temporal dynamics may not be particularly useful for predicting GPP, its ability to saturate less and its sensitivity to deeper vegetation layers might be useful when considering it as a proxy for vegetation density, similar to how medianVOD[2] is used. Additionally, using LVOD only as a static predictor, without considering its temporal dynamics, has the significant advantage of not reducing

2: high frequency VOD.

the training data. The low availability period from 2010 to 2020 would otherwise severely reduce the training data and the timespan of possible VODCA2GPP predictions.

Ultimately it is difficult to make a final assessment as many of the insecurities regarding global GPP estimations also play a role when evaluating the performance increases of this new predictor. However, especially the cross-validation results, which are independent of the global GPP datasets, suggest that L-band VOD can be a useful supplementary predictor for GPP.

5.6 Future research

The new predictors included in this study are not based on GPP drivers or environmental conditions that influence GPP. Instead, the aim was to stay true to the sink-driven approach by Teubner et al. (2019) and try to adapt it to work better on a global scale. Thus, the predictors aimed to enhance generalizability and consistency across different biomes. This objective was successfully achieved through the inclusion of LC data, in the form of fractional coverages of PFTs, and to a lesser extent, through a static predictor based on L-band VOD.

(Teubner et al. 2019): Teubner et al. (2019), *A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations*

The importance of LC is significant because it characterizes the vegetation type within each grid cell, and GPP exhibits substantial variation among different vegetation types. Furthermore, the fractional coverage aspect of the predictor provides information on vegetation density. Similarly, L-band VOD serves as a proxy for vegetation density, with additional sensitivity to vegetation height. Its deeper penetration allows for higher sensitivity to vegetation height when compared to median (high frequency) VOD and PFTs. However, its results are less conclusive than those of LC as was discussed in Section 5.5.

This raises the question of whether the median VOD, fractional coverages of PFTs, and the static L band VOD predictor all provide a unique contribution in the context of capturing vegetation density. Notably, the feature importance of median VOD decreases when the new predictors are included, resulting in similar levels of importance for both L-band VOD and median VOD. LC should definitely be retained as it is the only predictor that directly provides information on vegetation type[3]. However, further tests should be conducted to determine if both L-band VOD and median VOD are necessary.

3: medianVOD and L-band VOD provide indirect information on vegetation type through their sensitivity to vegetation density

Additionally, exploring the incorporation of predictors that more directly represent the drivers of GPP, such as water availability, solar radiation, or atmospheric CO_2 levels, provides material for future research. Specifically, water availability could have significant implications, as evidenced by decreased performance in the case of Australia despite ample training data availability. Australia is known to be a water-limited region, suggesting the importance of water availability for predicting GPP. Soil moisture or the Standardized Precipitation Evapotranspiration Index (SPEI) are indicators of water availability that could be considered as potential predictors.[4]

4: Microwave derived soil moisture is particularly interesting as it is not only a good indicator for water availability but also aligns with the microwave-based approach of VODCA2GPP.

Furthermore, other potential predictors to consider are solar radiation and CO_2 concentrations. While these predictors hold promise, it is crucial to assess whether their incorporation would still align with the sink-driven approach. Solar radiation, for instance, is clearly a source-driven predictor. It is essential to determine whether the potential improvements of new predictors outweigh the deviation from this approach or if it is more worthwhile to continue pursuing a strictly sink-driven approach to retain the unique insights into GPP dynamics specific to it.

6 Conclusions

In this thesis the VODCA2GPPv2 model, a new updated Gross Primary Production model from microwave-derived VOD observations was developed. Building on the first version of the VODCA2GPP model by Wild et al. (2022), its uneven spatial performance was addressed in an aim to make it more consistent across biomes. Several enhancements to the original model were made to achieve this goal: three different in-situ GPP observation datasets were merged to more than double the amount of available training locations, a new machine learning algorithm was employed, namely a random forest regressor and two new predictors were introduced, LC and (low frequency) L-band VOD.

(Wild et al. 2022): Wild et al. (2022), *VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing*

Comparisons with the independent GPP records from MODIS, FLUXCOM, and TRENDY revealed that this new model is capable of capturing temporal GPP patterns more effectively than VODCA2GPPv1 in many regions of the world. Additionally, the amount of overestimation (in comparison to independent GPP) could be significantly reduced. Comparisons with local GPP measurements demonstrated that the new model is more consistent across different biomes and land cover types and exhibits improved performance during cross-validation, with higher correlations and reduced bias and RMSE. Furthermore, the model reaches saturation at a slower rate than VODCA2GPPv1 and is capable of predicting higher GPP values.

These findings imply that the changes were successful in developing a new model that is more generalizable. The new model is less dependent on in-situ data distribution and density and able to better capture the spatial patterns of GPP across biomes, land cover type and latitudes.

However, it should be noted that there is still a tendency for the model to perform better in regions and land cover classes with high in-situ data density and that its dependence on in-situ data could only be mitigated but not eliminated. Furthermore, while overestimation of GPP at the global scale could be reduced, it still remains high, especially in the tropics where GPP predictability is known to be notoriously difficult.

To address some of these limitations, future research could consider incorporating soil moisture as a predictor to account for water availability. This may help address the performance decrease observed in regions where photosynthetic activity is hampered by limited water availability.

Furthermore, it may be beneficial to include additional drivers of GPP, including direct ones such as radiation. Although this possible deviation from the carbon sink-driven approach needs to be assessed carefully.

In conclusion, the improvements implemented in this thesis have resulted in an enhanced GPP model that demonstrates closer agreement with independent GPP datasets and in-situ observations. These results further increase the confidence in the carbon sink-driven GPP estimation approach. Moreover, the unique approach makes the resulting VODCA2GPPv2 dataset a valuable complementary dataset, which, if used jointly with traditional RS-based models, can aid in a more comprehensive understanding of the dynamics of GPP and its role in the global carbon cycle.

Bibliography

Alemohammad, Seyed Hamed, Bin Fang, Alexandra G. Konings, Filipe Aires, Julia K. Green, Jana Kolassa, Diego Miralles, Catherine Prigent, and Pierre Gentine. 'Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A Statistically Based Estimate of Global Surface Turbulent Fluxes and Gross Primary Productivity Using Solar-Induced Fluorescence'. In: *Biogeosciences* 14.18 (Sept. 2017), pp. 4101–4124. DOI: 10.5194/bg-14-4101-2017. (Visited on 09/05/2023) (cited on page 5).

Anav, Alessandro, Pierre Friedlingstein, Christian Beer, Philippe Ciais, Anna Harper, Chris Jones, Guillermo Murray-Tortarolo, Dario Papale, Nicholas C. Parazoo, Philippe Peylin, Shilong Piao, Stephen Sitch, Nicolas Viovy, Andy Wiltshire, and Maosheng Zhao. 'Spatiotemporal Patterns of Terrestrial Gross Primary Production: A Review'. In: *Reviews of Geophysics* 53.3 (2015), pp. 785–818. DOI: 10.1002/2015RG000483. (Visited on 06/15/2023) (cited on page 3).

Baldocchi, Dennis D. 'Assessing the Eddy Covariance Technique for Evaluating Carbon Dioxide Exchange Rates of Ecosystems: Past, Present and Future'. In: *Global Change Biology* 9.4 (2003), pp. 479–492. DOI: 10.1046/j.1365-2486.2003.00629.x. (Visited on 05/24/2023) (cited on page 15).

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, M. Altaf Arain, Dennis Baldocchi, Gordon B. Bonan, Alberte Bondeau, Alessandro Cescatti, Gitta Lasslop, Anders Lindroth, Mark Lomas, Sebastiaan Luyssaert, Hank Margolis, Keith W. Oleson, Olivier Roupsard, Elmar Veenendaal, Nicolas Viovy, Christopher Williams, F. Ian Woodward, and Dario Papale. 'Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate'. In: *Science* 329.5993 (Aug. 2010), pp. 834–838. DOI: 10.1126/science.1184984. (Visited on 06/15/2023) (cited on pages 3, 4).

Bonan, Gordon B. *Ecological Climatology: Concepts and Applications*. Third edition. New York, NY, USA: Cambridge University Press, 2016 (cited on pages 6, 21).

Breiman, Leo. 'Random Forests'. In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. DOI: 10.1023/A:1010933404324. (Visited on 09/19/2023) (cited on page 26).

Bureau of Meteorology. *Annual Climate Report 2003*. Tech. rep. 2003 (cited on page 45).

– *Annual Climate Report 2005*. Tech. rep. 2005 (cited on page 45).

Cavaleri, Molly A., Adam P. Coble, Michael G. Ryan, William L. Bauerle, Henry W. Loescher, and Steven F. Oberbauer. 'Tropical Rainforest Carbon Sink Declines during El Niño as a Result of Reduced Photosynthesis and Increased Respiration Rates'. In: *New Phytologist* 216.1 (2017), pp. 136–149. DOI: 10.1111/nph.14724. (Visited on 09/06/2023) (cited on page 3).

Chaparro, David, Grégory Duveiller, Maria Piles, Alessandro Cescatti, Mercè Vall-llossera, Adriano Camps, and Dara Entekhabi. 'Sensitivity of L-band Vegetation Optical Depth to Carbon Stocks in Tropical Forests: A Comparison to Higher Frequencies and Optical Indices'. In: *Remote Sensing of Environment* 232 (Oct. 2019), p. 111303. DOI: 10.1016/j.rse.2019.111303. (Visited on 09/06/2023) (cited on pages 6, 13).

Damm, Alexander, Jan Elbers, André Erler, Beniamino Gioli, Karim Hamdi, Ronald Hutjes, Martina Kosvancova, Michele Meroni, Franco Miglietta, André Moersch, Jose Moreno, Anke Schickling, Ruth Sonnenschein, Thomas Udelhoven, Sebastian Van Der LINDEN, Patrick Hostert, and Uwe Rascher. 'Remote Sensing of Sun-Induced Fluorescence to Improve Modeling of Diurnal Courses of Gross Primary Production (GPP)'. In: *Global Change Biology* 16.1 (2010), pp. 171–186. DOI: 10.1111/j.1365-2486.2009.01908.x. (Visited on 09/18/2023) (cited on page 5).

Defourny, Pierre and ESA Land Cover CCI project team. *ESA Land Cover Climate Change Initiative (Land_Cover_cci): Global Land Cover Maps, Version 2.0.7.* 2017 (cited on pages 11, 14).

Delwiche, Kyle B. et al. 'FLUXNET-CH_4: A Global, Multi-Ecosystem Dataset and Analysis of Methane Seasonality from Freshwater Wetlands'. In: *Earth System Science Data* 13.7 (July 2021), pp. 3607–3689. DOI: 10.5194/essd-13-3607-2021. (Visited on 10/17/2022) (cited on pages 15, 16).

Dunkl, István, Nicole Lovenduski, Alessio Collalti, Vivek K. Arora, Tatiana Ilyina, and Victor Brovkin. 'Gross Primary Productivity and the Predictability of CO_2: More Uncertainty in What We Predict than How Well We Predict It'. In: *Biogeosciences* 20.16 (Aug. 2023), pp. 3523–3538. DOI: 10.5194/bg-20-3523-2023. (Visited on 09/11/2023) (cited on page 7).

ESA. *CCI-LC User Tool*. 2014. (Visited on 09/20/2023) (cited on page 14).

Farquhar, G. D., S. von Caemmerer, and J. A. Berry. 'A Biochemical Model of Photosynthetic CO2 Assimilation in Leaves of C3 Species'. In: *Planta* 149.1 (June 1980), pp. 78–90. DOI: 10.1007/BF00386231. (Visited on 09/01/2023) (cited on page 4).

Fatichi, Simone, Sebastian Leuzinger, and Christian Körner. 'Moving beyond Photosynthesis: From Carbon Source to Sink-Driven Vegetation Modeling'. In: *The New Phytologist* 201.4 (2014), pp. 1086–1095. (Visited on 09/19/2023) (cited on page 6).

Fisher, Joshua B., Deborah N. Huntzinger, Christopher R. Schwalm, and Stephen Sitch. 'Modeling the Terrestrial Biosphere'. In: *Annual Review of Environment and Resources* 39.1 (Oct. 2014), pp. 91–123. DOI: 10.1146/annurev-environ-012913-093456. (Visited on 09/01/2023) (cited on pages 4, 5).

Forkel, Matthias, Wouter Dorigo, Gitta Lasslop, Irene Teubner, Emilio Chuvieco, and Kirsten Thonicke. 'A Data-Driven Approach to Identify Controls on Global Fire Activity from Satellite and Climate Observations (SOFIA V1)'. In: *Geoscientific Model Development* 10.12 (Dec. 2017), pp. 4443–4476. DOI: 10.5194/gmd-10-4443-2017. (Visited on 09/06/2023) (cited on page 6).

Frankenberg, Christian, Joshua B. Fisher, John Worden, Grayson Badgley, Sassan S. Saatchi, Jung-Eun Lee, Geoffrey C. Toon, André Butz, Martin Jung, Akihiko Kuze, and Tatsuya Yokota. 'New Global Observations of the Terrestrial Carbon Cycle from GOSAT: Patterns of Plant Fluorescence with Gross Primary Productivity'. In: *Geophysical Research Letters* 38.17 (2011). DOI: 10.1029/2011GL048738. (Visited on 09/04/2023) (cited on page 5).

Friedlingstein, Pierre et al. 'Global Carbon Budget 2020'. In: *Earth System Science Data* 12.4 (Dec. 2020), pp. 3269–3340. DOI: 10.5194/essd-12-3269-2020. (Visited on 09/06/2023) (cited on page 3).

Jackson, T.J. and T.J. Schmugge. 'Vegetation Effects on the Microwave Emission of Soils'. In: *Remote Sensing of Environment* 36.3 (June 1991), pp. 203–212. DOI: 10.1016/0034-4257(91)90057-D. (Visited on 06/21/2023) (cited on pages 6, 13).

Jones, Matthew O., Lucas A. Jones, John S. Kimball, and Kyle C. McDonald. 'Satellite Passive Microwave Remote Sensing for Monitoring Global Land Surface Phenology'. In: *Remote Sensing of Environment* 115.4 (Apr. 2011), pp. 1102–1114. DOI: 10.1016/j.rse.2010.12.015. (Visited on 09/06/2023) (cited on page 6).

Jung, Martin et al. 'Scaling Carbon Fluxes from Eddy Covariance Sites to Globe: Synthesis and Evaluation of the FLUXCOM Approach'. In: *Biogeosciences* 17.5 (Mar. 2020), pp. 1343–1365. DOI: 10.5194/bg-17-1343-2020. (Visited on 09/26/2023) (cited on pages 47, 48).

Körner, Christian. 'Paradigm Shift in Plant Growth Control'. In: *Current Opinion in Plant Biology* 25 (June 2015), pp. 107–114. DOI: 10.1016/j.pbi.2015.05.003. (Visited on 09/19/2023) (cited on page 6).

Le Quéré, Corinne et al. 'Global Carbon Budget 2018'. In: *Earth System Science Data* 10.4 (Dec. 2018), pp. 2141–2194. DOI: 10.5194/essd-10-2141-2018. (Visited on 09/01/2023) (cited on page 18).

Leuzinger, Sebastian, Corina Manusch, Harald Bugmann, and Annett Wolf. 'A Sink-Limited Growth Model Improves Biomass Estimation along Boreal and Alpine Tree Lines'. In: *Global Ecology and Biogeography* 22.8 (2013), pp. 924–932. DOI: 10.1111/geb.12047. (Visited on 09/19/2023) (cited on page 6).

Liu, Huimin, Qingming Zhan, Chen Yang, and Jiong Wang. 'Characterizing the Spatio-Temporal Pattern of Land Surface Temperature through Time Series Clustering: Based on the Latent Pattern and Morphology'. In: *Remote Sensing* 10.4 (Apr. 2018), p. 654. DOI: 10.3390/rs10040654. (Visited on 09/06/2023) (cited on page 6).

Liu, Yi Y., Richard A. M. de Jeu, Matthew F. McCabe, Jason P. Evans, and Albert I. J. M. van Dijk. 'Global Long-Term Passive Microwave Satellite-Based Retrievals of Vegetation Optical Depth'. In: *Geophysical Research Letters* 38.18 (2011). DOI: 10.1029/2011GL048684. (Visited on 09/06/2023) (cited on page 6).

Liu, Yi Y., Albert I. J. M. van Dijk, Richard A. M. de Jeu, Josep G. Canadell, Matthew F. McCabe, Jason P. Evans, and Guojie Wang. 'Recent Reversal in Loss of Global Terrestrial Biomass'. In: *Nature Climate Change* 5.5 (May 2015), pp. 470–474. DOI: 10.1038/nclimate2581. (Visited on 09/06/2023) (cited on page 6).

Loecher, Markus. 'Debiasing MDI Feature Importance and SHAP Values in Tree Ensembles'. In: *Machine Learning and Knowledge Extraction*. Ed. by Andreas Holzinger, Peter Kieseberg, A. Min Tjoa, and Edgar Weippl. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 114–129. DOI: 10.1007/978-3-031-14463-9_8 (cited on page 26).

Lundberg, Scott. *Slundberg/Shap - A Game Theoretic Approach to Explain the Output of Any Machine Learning Model.* Apr. 2023. (Visited on 04/11/2023) (cited on page 26).

MacBean, Natasha, Fabienne Maignan, Cédric Bacour, Philip Lewis, Philippe Peylin, Luis Guanter, Philipp Köhler, Jose Gómez-Dans, and Mathias Disney. 'Strong Constraint on Modelled Global Carbon Uptake Using Solar-Induced Chlorophyll Fluorescence Data'. In: *Scientific Reports* 8.1 (Jan. 2018), p. 1973. DOI: 10.1038/s41598-018-20024-w. (Visited on 09/26/2023) (cited on page 47).

Marshall, Michael, Kevin Tu, and Jesslyn Brown. 'Optimizing a Remote Sensing Production Efficiency Model for Macro-Scale GPP and Yield Estimation in Agroecosystems'. In: *Remote Sensing of Environment* 217 (Nov. 2018), pp. 258–271. DOI: 10.1016/j.rse.2018.08.001. (Visited on 09/06/2023) (cited on page 3).

Meesters, A.G.C.A., R.A.M. De Jeu, and M. Owe. 'Analytical Derivation of the Vegetation Optical Depth from the Microwave Polarization Difference Index'. In: *IEEE Geoscience and Remote Sensing Letters* 2.2 (Apr. 2005), pp. 121–123. DOI: 10.1109/LGRS.2005.843983 (cited on page 13).

Mo, T., B. J. Choudhury, T. J. Schmugge, J. R. Wang, and T. J. Jackson. 'A Model for Microwave Emission from Vegetation-Covered Fields'. In: *Journal of Geophysical Research: Oceans* 87.C13 (1982), pp. 11229–11237. DOI: 10.1029/JC087iC13p11229. (Visited on 09/08/2023) (cited on page 13).

Moesinger, Leander, Wouter Dorigo, Richard de Jeu, Robin van der Schalie, Tracy Scanlon, Irene Teubner, and Matthias Forkel. 'The Global Long-Term Microwave Vegetation Optical Depth Climate Archive (VODCA)'. In: *Earth System Science Data* 12.1 (Jan. 2020), pp. 177–196. DOI: 10.5194/essd-12-177-2020. (Visited on 02/23/2023) (cited on pages 6, 12).

Molnar, Christoph. 'Chapter 8.5 Shapley Values'. In: *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable*. Munich, Germany: Independently published, Feb. 2022 (cited on page 26).

Momen, Mostafa, Jeffrey D. Wood, Kimberly A. Novick, Robert Pangle, William T. Pockman, Nate G. McDowell, and Alexandra G. Konings. 'Interacting Effects of Leaf Water Potential and Biomass on Vegetation Optical Depth'. In: *Journal of Geophysical Research: Biogeosciences* 122.11 (2017), pp. 3031–3046. DOI: 10.1002/2017JG004145. (Visited on 09/06/2023) (cited on page 6).

Monteith, J. L. 'Solar Radiation and Productivity in Tropical Ecosystems'. In: *The Journal of Applied Ecology* 9.3 (Dec. 1972), p. 747. DOI: 10.2307/2401901. (Visited on 05/24/2023) (cited on pages 5, 17).

Muñoz-Sabater, Joaquín, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, Brecht Martens, Diego G. Miralles, María Piles, Nemesio J. Rodríguez-Fernández, Ervin Zsoter, Carlo Buontempo, and Jean-Noël Thépaut. 'ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications'. In: *Earth System Science Data* 13.9 (Sept. 2021), pp. 4349–4383. DOI: 10.5194/essd-13-4349-2021. (Visited on 02/11/2023) (cited on pages 11, 13).

O'Sullivan, Michael, William K. Smith, Stephen Sitch, Pierre Friedlingstein, Vivek K. Arora, Vanessa Haverd, Atul K. Jain, Etsushi Kato, Markus Kautz, Danica Lombardozzi, Julia E. M. S. Nabel, Hanqin Tian, Nicolas Vuichard, Andy Wiltshire, Dan Zhu, and Wolfgang Buermann. 'Climate-Driven Variability and Trends in Plant Productivity Over Recent Decades Based on Three Global Products'. In: *Global Biogeochemical Cycles* 34.12 (Dec. 2020). DOI: 10.1029/2020GB006613. (Visited on 09/03/2022) (cited on page 4).

Owe, Manfred, Richard de Jeu, and Thomas Holmes. 'Multisensor Historical Climatology of Satellite-Derived Global Land Surface Moisture'. In: *Journal of Geophysical Research: Earth Surface* 113.F1 (2008). DOI: 10.1029/2007JF000769. (Visited on 02/23/2023) (cited on page 12).

Pastorello, Gilberto et al. 'The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data'. In: *Scientific Data* 7.1 (July 2020), p. 225. DOI: 10.1038/s41597-020-0534-3. (Visited on 10/05/2022) (cited on page 15).

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 'Scikit-Learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830. (Visited on 04/07/2023) (cited on pages 24, 26).

Poulter, B., N. MacBean, A. Hartley, I. Khlystova, O. Arino, R. Betts, S. Bontemps, M. Boettcher, C. Brockmann, P. Defourny, S. Hagemann, M. Herold, G. Kirches, C. Lamarche, D. Lederer, C. Ottlé, M. Peters, and P. Peylin. 'Plant Functional Type Classification for Earth System Models: Results from the European Space Agency's Land Cover Climate Change Initiative'. In: *Geoscientific Model Development* 8.7 (July 2015), pp. 2315–2328. DOI: 10.5194/gmd-8-2315-2015. (Visited on 05/23/2023) (cited on page 14).

Reeves, M. C., M. Zhao, and S. W. Running. 'Usefulness and Limits of MODIS GPP for Estimating Wheat Yield'. In: *International Journal of Remote Sensing* 26.7 (Apr. 2005), pp. 1403–1421. DOI: 10.1080/01431160512331326567. (Visited on 09/06/2023) (cited on page 3).

Running, Steve, Qiaozhen Mu, and Maosheng Zhao. *MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006*. 2015. DOI: 10.5067/MODIS/MOD17A2H.006. (Visited on 05/24/2023) (cited on page 17).

Running, Steven W and Maosheng Zhao. *User's Guide: Daily GPP and Annual NPP (MOD17A2/A3) Products NASA Earth Observing System MODIS Land Algorithm*. Oct. 2015 (cited on page 5).

Running, Steven W., Ramakrishna R. Nemani, Faith Ann Heinsch, Maosheng Zhao, Matt Reeves, and Hirofumi Hashimoto. 'A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production'. In: *BioScience* 54.6 (June 2004), pp. 547–560. DOI: 10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2. (Visited on 09/28/2023) (cited on page 48).

Ryu, Youngryel, Joseph A. Berry, and Dennis D. Baldocchi. 'What Is Global Photosynthesis? History, Uncertainties and Opportunities'. In: *Remote Sensing of Environment* 223 (Mar. 2019), pp. 95–114. DOI: 10.1016/j.rse.2019.01.016. (Visited on 09/06/2023) (cited on page 3).

Sade, Nir, Alem Gebremedhin, and Menachem Moshelion. 'Risk-Taking Plants'. In: *Plant Signaling & Behavior* 7.7 (July 2012), pp. 767–770. DOI: 10.4161/psb.20505. (Visited on 09/26/2023) (cited on page 47).

Schmidt, Luisa, Matthias Forkel, Ruxandra-Maria Zotta, Samuel Scherrer, Wouter A. Dorigo, Alexander Kuhn-Régnier, Robin Van Der Schalie, and Marta Yebra. 'Assessing the Sensitivity of Multi-Frequency Passive Microwave Vegetation Optical Depth to Vegetation Properties'. In: *Biogeosciences* 20.5 (Mar. 2023), pp. 1027–1046. DOI: 10.5194/bg-20-1027-2023. (Visited on 06/06/2023) (cited on pages 23, 49).

Sun, Zhongyi, Xiufeng Wang, Xirui Zhang, Hiroshi Tani, Enliang Guo, Shuai Yin, and Tianyou Zhang. 'Evaluating and Comparing Remote Sensing Terrestrial GPP Models for Their Response to Climate Variability and CO2 Trends'. In: *Science of The Total Environment* 668 (June 2019), pp. 696–713. DOI: 10.1016/j.scitotenv.2019.03.025. (Visited on 06/16/2023) (cited on page 5).

Taschetto, Andréa S. and Matthew H. England. 'El Niño Modoki Impacts on Australian Rainfall'. In: *Journal of Climate* 22.11 (June 2009), pp. 3167–3174. DOI: 10.1175/2008JCLI2589.1. (Visited on 10/17/2023) (cited on page 45).

Team, Warm Winter 2020 and ICOS Ecosystem Thematic Centre. 'Warm Winter 2020 Ecosystem Eddy Covariance Flux Product for 73 Stations in FLUXNET-Archive Format—Release 2022-1'. In: (Feb. 2022). DOI: 10.18160/2G60-ZHAK. (Visited on 10/05/2022) (cited on pages 15, 17).

Teubner, Irene E., Matthias Forkel, Gustau Camps-Valls, Martin Jung, Diego G. Miralles, Gianluca Tramontana, Robin van der Schalie, Mariette Vreugdenhil, Leander Mösinger, and Wouter A. Dorigo. 'A Carbon Sink-Driven Approach to Estimate Gross Primary Production from Microwave Satellite Observations'. In: *Remote Sensing of Environment* 229 (Aug. 2019), pp. 100–113. DOI: 10.1016/j.rse.2019.04.022. (Visited on 07/16/2022) (cited on pages v, 6, 7, 21, 22, 43, 50, 51).

Teubner, Irene E., Matthias Forkel, Martin Jung, Yi Y. Liu, Diego G. Miralles, Robert Parinussa, Robin van der Schalie, Mariette Vreugdenhil, Christopher R. Schwalm, Gianluca Tramontana, Gustau Camps-Valls, and Wouter A. Dorigo. 'Assessing the Relationship between Microwave Vegetation Optical Depth and Gross Primary Production'. In: *International Journal of Applied Earth Observation and Geoinformation* 65 (Mar. 2018), pp. 79–91. DOI: 10.1016/j.jag.2017.10.006. (Visited on 07/16/2022) (cited on pages 6, 7, 50).

Teubner, Irene E., Matthias Forkel, Benjamin Wild, Leander Mösinger, and Wouter Dorigo. 'Impact of Temperature and Water Availability on Microwave-Derived Gross Primary Production'. In: *Biogeosciences* 18.11 (June 2021), pp. 3285–3308. DOI: 10.5194/bg-18-3285-2021. (Visited on 07/18/2022) (cited on pages 6–8, 16, 21, 22, 24, 47, 48, 50).

Tramontana, Gianluca, Martin Jung, Christopher R. Schwalm, Kazuhito Ichii, Gustau Camps-Valls, Botond Ráduly, Markus Reichstein, M. Altaf Arain, Alessandro Cescatti, Gerard Kiely, Lutz Merbold, Penelope Serrano-Ortiz, Sven Sickert, Sebastian Wolf, and Dario Papale. 'Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms'. In: *Biogeosciences* 13.14 (July 2016), pp. 4291–4313. DOI: 10.5194/bg-13-4291-2016. (Visited on 05/24/2023) (cited on pages 4, 5, 17, 48).

Trugman, A. T., D. Medvigy, J. S. Mankin, and W. R. L. Anderegg. 'Soil Moisture Stress as a Major Driver of Carbon Cycle Uncertainty'. In: *Geophysical Research Letters* 45.13 (2018), pp. 6495–6503. DOI: 10.1029/2018GL078131. (Visited on 09/06/2023) (cited on page 3).

Turner, David P., William D. Ritts, Warren B. Cohen, Stith T. Gower, Steve W. Running, Maosheng Zhao, Marcos H. Costa, Al A. Kirschbaum, Jay M. Ham, Scott R. Saleska, and Douglas E. Ahl. 'Evaluation of MODIS NPP and GPP Products across Multiple Biomes.' In: *Remote Sensing of Environment. 102: 282-292* (2006). (Visited on 09/26/2023) (cited on page 47).

van der Schalie, R., R. A. M. de Jeu, Y. H. Kerr, J. P. Wigneron, N. J. Rodríguez-Fernández, A. Al-Yaari, R. M. Parinussa, S. Mecklenburg, and M. Drusch. 'The Merging of Radiative Transfer Based Surface Soil Moisture Data from SMOS and AMSR-E'. In: *Remote Sensing of Environment* 189 (Feb. 2017), pp. 180–193. DOI: 10.1016/j.rse.2016.11.026. (Visited on 02/23/2023) (cited on pages 12, 13).

Wang, Zhipeng, Yongtao He, Ben Niu, Jianshuang Wu, Xianzhou Zhang, Jiaxing Zu, Ke Huang, Meng Li, Yanan Cao, Yanjie Zhang, Ning Chen, Shuai Yang, and Xiangtao Wang. 'Sensitivity of Terrestrial Carbon Cycle to Changes in Precipitation Regimes'. In: *Ecological Indicators* 113 (June 2020), p. 106223. DOI: 10.1016/j.ecolind.2020.106223. (Visited on 09/06/2023) (cited on page 3).

Wardle, Glenda M., Chris R. Pavey, and Chris R. Dickman. 'Greening of Arid Australia: New Insights from Extreme Years'. In: *Austral Ecology* 38.7 (2013), pp. 731–740. DOI: 10.1111/aec.12073. (Visited on 10/13/2023) (cited on page 44).

Wild, Benjamin, Irene Teubner, Leander Moesinger, Ruxandra-Maria Zotta, Matthias Forkel, Robin van der Schalie, Stephen Sitch, and Wouter Dorigo. 'VODCA2GPP – a New, Global, Long-Term (1988–2020) Gross Primary Production Dataset from Microwave Remote Sensing'. In: *Earth System Science Data* 14.3 (Mar. 2022), pp. 1063–1085. DOI: 10.5194/essd-14-1063-2022. (Visited on 07/18/2022) (cited on pages v, 7, 8, 15, 16, 22–24, 32, 44, 45, 47–49, 55, 67).

Wu, Mousong, Marko Scholze, Thomas Kaminski, Michael Voßbeck, and Torbern Tagesson. 'Using SMOS Soil Moisture Data Combining CO2 Flask Samples to Constrain Carbon Fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS)'. In: *Remote Sensing of Environment* 240 (Apr. 2020), p. 111719. DOI: 10.1016/j.rse.2020.111719. (Visited on 09/26/2023) (cited on page 47).

Yang, Ruqi, Jun Wang, Ning Zeng, Stephen Sitch, Wenhan Tang, Matthew Joseph McGrath, Qixiang Cai, Di Liu, Danica Lombardozzi, Hanqin Tian, Atul K. Jain, and Pengfei Han. 'Divergent Historical GPP Trends among State-of-the-Art Multi-Model Simulations and Satellite-Based Products'. In: *Earth System Dynamics* 13.2 (Apr. 2022), pp. 833–849. DOI: 10.5194/esd-13-833-2022. (Visited on 08/23/2023) (cited on pages 3, 4, 7).

Zhai, Panmao, Rong Yu, Yanjun Guo, Qingxiang Li, Xuejuan Ren, Yaqiang Wang, Wenhui Xu, Yanju Liu, and Yihui Ding. 'The Strong El Niño of 2015/16 and Its Dominant Impacts on Global and China's Climate'. In: *Journal of Meteorological Research* 30.3 (June 2016), pp. 283–297. DOI: 10.1007/s13351-016-6101-3. (Visited on 10/17/2023) (cited on page 45).

Zotta, Ruxandra-Maria, Leander Moesinger, Robin Van Der Schalie, Mariette Vreugdenhil, Wolfgang Preimesberger, and Wouter A. Dorigo. *VODCA v2: A Multi-Sensor and Frequency Vegetation Optical Depth Dataset for Long-Term Canopy Dynamics and Biomass Monitoring, in Preparation*. in prep. (Cited on pages 7, 11, 12).

Appendix

A

Supplementary Materials

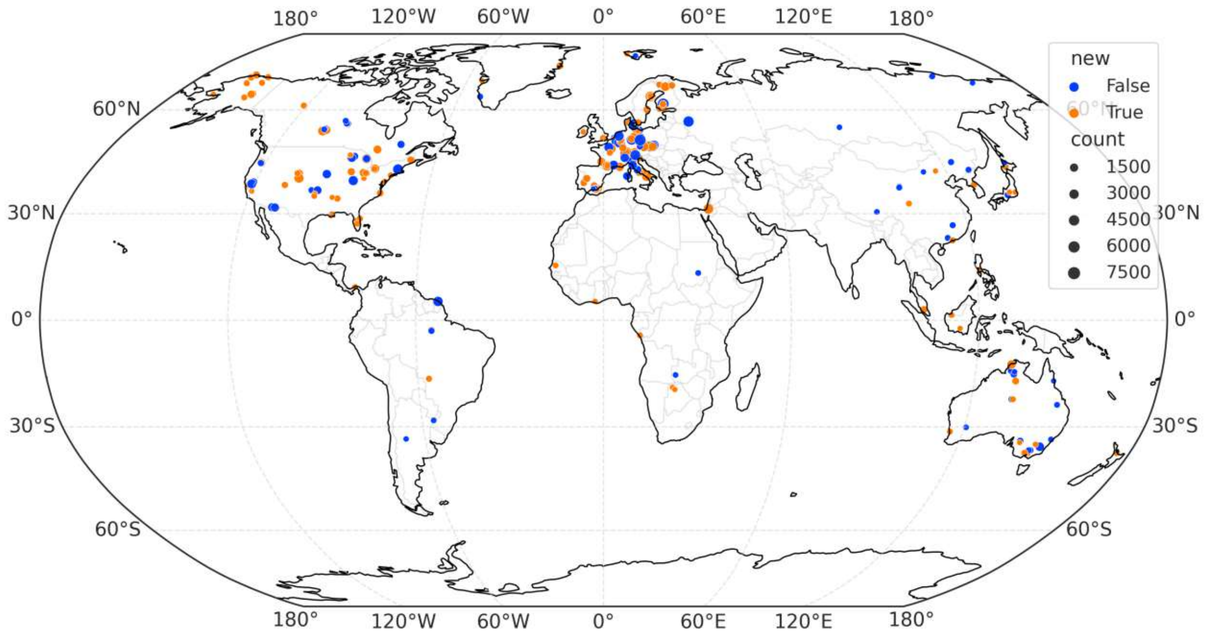


Figure A.1: Spatial distribution of FLUXNET sites used in VODCA2GPPv1 (Wild et al. 2022) (blue), and newly added sites for VODCA2GPP v2 (orange). Marker size is scaled to the number of days of observations.

Table A.1: FLUXNET sites used in this study

FLUXNET ID	Name	Lat [°N]	Lon [°E]	Origin dataset	No. of obs. [days]
AR-SLu	San Luis	-33.46	-66.46	2015	467
AR-Vir	Virasoro	-28.24	-56.19	2015	741
AT-Neu	Neustift	47.12	11.32	2015	3924
AU-ASM	Alice Springs	-22.28	133.25	2015	1557
AU-Ade	Adelaide River	-13.08	131.12	2015	558
AU-Cpr	Calperum	-34.00	140.59	2015	1501
AU-Cum	Cumberland Plain	-33.62	150.72	2015	788
AU-DaP	Daly River Savanna	-14.06	131.32	2015	1933
AU-DaS	Daly River Cleared	-14.16	131.39	2015	2399
AU-Dry	Dry River	-15.26	132.37	2015	1699
AU-Emr	Emerald	-23.86	148.47	2015	863
AU-Fog	Fogg Dam	-12.55	131.31	2015	928
AU-GWW	Great Western Woodlands, Western Australia, Aus...	-30.19	120.65	2015	696
AU-Gin	Gingin	-31.38	115.71	2015	1016
AU-How	Howard Springs	-12.49	131.15	2015	4173
AU-Lox	Loxton	-34.47	140.66	2015	294
AU-RDF	Red Dirt Melon Farm, Northern Territory	-14.56	132.48	2015	611
AU-Rig	Riggs Creek	-36.65	145.58	2015	1326
AU-Rob	Robson Creek, Queensland, Australia	-17.12	145.63	2015	362
AU-Stp	Sturt Plains	-17.15	133.35	2015	2092
AU-TTE	Ti Tree East	-22.29	133.64	2015	888
AU-Tum	Tumbarumba	-35.66	148.15	2015	4616
AU-Wac	Wallaby Creek	-37.43	145.19	2015	1029
AU-Whr	Whroo	-36.67	145.03	2015	1125

Continued on next page

FLUXNET ID	Name	Lat [°N]	Lon [°E]	Origin dataset	No. of obs. [days]
AU-Wom	Wombat	-37.42	144.09	2015	1707
AU-Ync	Jaxa	-34.99	146.29	2015	555
BE-Bra	Brasschaat	51.31	4.52	WarmWinter	6834
BE-Dor	Dorinne	50.31	4.97	WarmWinter	3472
BE-Lcr	Lochristi	51.11	3.85	WarmWinter	675
BE-Lon	Lonzee	50.55	4.75	WarmWinter	5980
BE-Maa	Maasmechelen	50.98	5.63	WarmWinter	1659
BE-Vie	Vielsalm	50.30	6.00	WarmWinter	8451
BR-Npw	Northern Pantanal Wetland	-16.50	-56.41	CH4	1122
BR-Sa1	Santarem-Km67-Primary Forest	-2.86	-54.96	2015	2364
BR-Sa3	Santarem-Km83-Logged Forest	-3.02	-54.97	2015	1221
BW-Gum	Guma	-18.96	22.37	CH4	365
BW-Nxr	Nxaraga	-19.55	23.18	CH4	365
CA-Gro	Ontario - Groundhog River, Boreal Mixedwood Forest	48.22	-82.16	2015	3617
CA-Man	Manitoba - Northern Old Black Spruce (former BO...	55.88	-98.48	2015	3626
CA-NS1	UCI-1850 burn site	55.88	-98.48	2015	1136
CA-NS2	UCI-1930 burn site	55.91	-98.52	2015	1191
CA-NS3	UCI-1964 burn site	55.91	-98.38	2015	1499
CA-NS4	UCI-1964 burn site wet	55.91	-98.38	2015	810
CA-NS5	UCI-1981 burn site	55.86	-98.48	2015	1319
CA-NS6	UCI-1989 burn site	55.92	-98.96	2015	1423
CA-NS7	UCI-1998 burn site	56.64	-99.95	2015	1137
CA-Oas	Saskatchewan - Western Boreal, Mature Aspen	53.63	-106.20	2015	5293
CA-Obs	Saskatchewan - Western Boreal, Mature Black Spruce	53.99	-105.12	2015	4160
CA-Qfo	Quebec - Eastern Boreal, Mature Black Spruce	49.69	-74.34	2015	2547
CA-SCB	Scotty Creek Bog	61.31	-121.30	CH4	1417
CA-SCC	Scotty Creek Landscape	61.31	-121.30	CH4	1338
CA-SF1	Saskatchewan - Western Boreal, forest burned in...	54.49	-105.82	2015	558
CA-SF2	Saskatchewan - Western Boreal, forest burned in...	54.25	-105.88	2015	724
CA-SF3	Saskatchewan - Western Boreal, forest burned in...	54.09	-106.01	2015	949
CA-TP1	Ontario - Turkey Point 2002 Plantation White Pine	42.66	-80.56	2015	2858
CA-TP2	Ontario - Turkey Point 1989 Plantation White Pine	42.77	-80.46	2015	810
CA-TP3	Ontario - Turkey Point 1974 Plantation White Pine	42.71	-80.35	2015	3246
CA-TP4	Ontario - Turkey Point 1939 Plantation White Pine	42.71	-80.36	2015	4531
CA-TPD	Ontario - Turkey Point Mature Deciduous	42.64	-80.56	2015	1058
CG-Tch	Tchizalamou	-4.29	11.66	2015	960
CH-Aws	Alp Weissenstein	46.58	9.79	WarmWinter	3391
CH-Cha	Chamau grassland	47.21	8.41	WarmWinter	5267
CH-Dav	Davos	46.82	9.86	WarmWinter	8619
CH-Fru	Fruebuel grassland	47.12	8.54	WarmWinter	5229
CH-Lae	Laegern	47.48	8.37	WarmWinter	5956
CH-Oe1	Oensingen grassland	47.29	7.73	2015	2388
CH-Oe2	Oensingen crop	47.29	7.73	WarmWinter	6072
CN-Cha	Changbaishan	42.40	128.10	2015	1033
CN-Cng	Changling	44.59	123.51	2015	1199
CN-Dan	Dangxiong	30.50	91.07	2015	709
CN-Din	Dinghushan	23.17	112.54	2015	978
CN-Du2	Duolun_grassland (D01)	42.05	116.28	2015	705
CN-Du3	Duolun Degraded Meadow	42.06	116.28	2015	280
CN-Ha2	Haibei Shrubland	37.61	101.33	2015	1081
CN-HaM	Haibei Alpine Tibet site	37.37	101.18	2015	1062
CN-Hgu	Hongyuan	32.85	102.59	CH4	960
CN-Qia	Qianyanzhou	26.74	115.06	2015	1092
CN-Sw2	Siziwang Grazed (SZWG)	41.79	111.90	2015	413
CZ-BK1	Bily Kriz forest	49.50	18.54	WarmWinter	5477
CZ-BK2	Bily Kriz grassland	49.49	18.54	2015	2011
CZ-KrP	Kresin u Pacova	49.57	15.08	WarmWinter	2544
CZ-Lnz	Lanzhot	48.68	16.95	WarmWinter	2126
CZ-RAJ	Rajec	49.44	16.70	WarmWinter	3196
CZ-Stn	Stitna	49.04	17.97	WarmWinter	3862
CZ-wet	Trebon	49.02	14.77	WarmWinter	5307
DE-Akm	Anklam	53.87	13.68	WarmWinter	2293
DE-Dgw	Dagowsee	53.15	13.05	CH4	1461
DE-Geb	Gebesee	51.10	10.91	WarmWinter	7190
DE-Gri	Grillenburg	50.95	13.51	WarmWinter	5890
DE-Hai	Hainich	51.08	10.45	WarmWinter	7172
DE-HoH	Hohes Holz	52.09	11.22	WarmWinter	2176
DE-Hte	Huetelmoor	54.21	12.18	CH4	2922
DE-Hzd	Hetzdorf	50.96	13.49	WarmWinter	2713
DE-Kli	Klingenberg	50.89	13.52	WarmWinter	5540
DE-Lkb	Lackenberg	49.10	13.30	2015	1253
DE-Lnf	Leinefelde	51.33	10.37	2015	2753
DE-Obe	OberbÃƒÂ¤renburg	50.79	13.72	WarmWinter	4468
DE-RuR	Rollesbroich	50.62	6.30	WarmWinter	3477
DE-RuS	Selhausen Juelich	50.87	6.45	WarmWinter	3091
DE-RuW	Wustebach	50.50	6.33	WarmWinter	2528

Continued on next page

FLUXNET ID	Name	Lat [°N]	Lon [°E]	Origin dataset	No. of obs. [days]
DE-Seh	Selhausen	50.87	6.45	2015	1200
DE-SfN	Schechenfilz Nord	47.81	11.33	2015	817
DE-Spw	Spreewald	51.89	14.03	2015	1464
DE-Tha	Tharandt	50.96	13.57	WarmWinter	8733
DE-Zrk	Zarnekow	53.88	12.89	2015	582
DK-Eng	Enghave	55.69	12.19	2015	1102
DK-Fou	Foulum	56.48	9.59	2015	243
DK-Gds	Gludsted Plantage	56.07	9.33	WarmWinter	161
DK-Sor	Soroe	55.49	11.64	WarmWinter	8470
ES-Abr	Albuera	38.70	-6.79	WarmWinter	1839
ES-Agu	Aguamarga	36.94	-2.03	WarmWinter	4157
ES-Amo	Amoladeras	36.83	-2.25	2015	1630
ES-Cnd	Conde	37.91	-3.23	WarmWinter	1828
ES-LJu	Llano de los Juanes	36.93	-2.75	2015	3030
ES-LM1	Majadas del Tietar North	39.94	-5.78	WarmWinter	2469
ES-LM2	Majadas del Tietar South	39.93	-5.78	WarmWinter	2448
ES-LgS	Laguna Seca	37.10	-2.97	2015	893
ES-Ln2	Lanjaron-Salvage logging	36.97	-3.48	2015	190
FI-Hyy	Hyytiala	61.85	24.29	WarmWinter	8440
FI-Jok	Jokioinen	60.90	23.51	2015	1114
FI-Ken	Kenttarova	67.99	24.24	WarmWinter	969
FI-Let	Lettosuo	60.64	23.96	WarmWinter	3650
FI-Lom	Lompolojankka	68.00	24.21	2015	1069
FI-Qvd	Qvidja	60.30	22.39	WarmWinter	881
FI-Si2	Siikaneva-2 Bog	61.84	24.20	CH4	1827
FI-Sii	Siikaneva	61.83	24.19	WarmWinter	1715
FI-Sod	Sodankyla	67.36	26.64	2015	4803
FI-Var	Varrio	67.75	29.61	WarmWinter	1805
FR-Aur	Aurade	43.55	1.11	WarmWinter	5641
FR-Bil	Bilos	44.49	-0.96	WarmWinter	2311
FR-FBn	Font-Blanche	43.24	5.68	WarmWinter	4541
FR-Fon	Fontainebleau-Barbeau	48.48	2.78	WarmWinter	5465
FR-Gri	Grignon	48.84	1.95	WarmWinter	5238
FR-Hes	Hesse	48.67	7.06	WarmWinter	2522
FR-LBr	Le Bray	44.72	-0.77	2015	3787
FR-LGt	La Guette	47.32	2.28	WarmWinter	1275
FR-Lam	Lamasquere	43.50	1.24	WarmWinter	5384
FR-Pue	Puechabon	43.74	3.60	2015	5059
FR-Tou	Toulouse	43.57	1.37	WarmWinter	1088
GF-Guy	Guyaflux	5.28	-52.92	WarmWinter	6044
GH-Ank	Ankasa	5.27	-2.69	2015	694
GL-Dsk	Disko	69.25	-53.51	WarmWinter	364
GL-NuF	Nuuk Fen	64.13	-51.39	2015	1000
GL-ZaF	Zackenberg Fen	74.48	-20.55	2015	511
GL-ZaH	Zackenberg Heath	74.47	-20.55	2015	1994
HK-MPM	Mai Po Mangrove	22.50	114.03	CH4	1096
ID-Pag	Palangkaraya undrained forest	-2.32	113.90	CH4	365
IE-Cra	Clara	53.32	-7.64	WarmWinter	366
IL-Yat	Yatir	31.35	35.05	WarmWinter	6627
IT-BCi	Borgo Cioffi	40.52	14.96	WarmWinter	5396
IT-BFt	Bosco Fontana	45.20	10.74	WarmWinter	731
IT-CA1	Castel d'Asso1	42.38	12.03	2015	1144
IT-CA2	Castel d'Asso2	42.38	12.03	2015	1155
IT-CA3	Castel d'Asso3	42.38	12.02	2015	977
IT-Cas	Castellaro	45.07	8.72	CH4	730
IT-Col	Collelongo	41.85	13.59	2015	4363
IT-Cp2	Castelporziano2	41.70	12.36	WarmWinter	2112
IT-Cpz	Castelporziano	41.71	12.38	2015	2760
IT-Isp	Ispra ABC-IS	45.81	8.63	2015	679
IT-La2	Lavarone2	45.95	11.29	2015	555
IT-Lav	Lavarone	45.96	11.28	WarmWinter	6167
IT-Lsn	Lison	45.74	12.75	WarmWinter	1724
IT-MBo	Monte Bondone	46.01	11.05	WarmWinter	6172
IT-Noe	Arca di Noe - Le Prigionette	40.61	8.15	2015	3345
IT-PT1	Parco Ticino forest	45.20	9.06	2015	936
IT-Ren	Renon	46.59	11.43	WarmWinter	6758
IT-Ro1	Roccarespampani 1	42.41	11.93	2015	2790
IT-Ro2	Roccarespampani 2	42.39	11.92	2015	3283
IT-SR2	San Rossore 2	43.73	10.29	WarmWinter	2777
IT-SRo	San Rossore	43.73	10.28	2015	4479
IT-Tor	Torgnon	45.84	7.58	WarmWinter	4299
JP-BBY	Bibai bog	43.32	141.81	CH4	1461
JP-MBF	Moshiri Birch Forest Site	44.39	142.32	2015	560
JP-Mse	Mase rice paddy field	36.05	140.03	CH4	366
JP-SMF	Seto Mixed Forest Site	35.26	137.08	2015	1411
JP-SwL	Suwa Lake	36.05	138.11	CH4	366

Continued on next page

FLUXNET ID	Name	Lat [°N]	Lon [°E]	Origin dataset	No. of obs. [days]
KR-CRK	Cheorwon Rice paddy	38.20	127.25	CH4	1461
MY-MLM	Maludam National Park	1.45	111.15	CH4	730
MY-PSO	Pasoh Forest Reserve (PSO)	2.97	102.31	2015	2247
NL-Hor	Horstermeer	52.24	5.07	2015	2406
NL-Loo	Loobos	52.17	5.74	2015	6227
NZ-Kop	Kopuatai	-37.39	175.55	CH4	1461
PA-SPn	Sardinilla Plantation	9.32	-79.63	2015	827
PA-SPs	Sardinilla-Pasture	9.31	-79.63	2015	980
PH-RiF	Philippines Rice Institute flooded	14.14	121.27	CH4	1096
RU-Ch2	Chersky reference	68.62	161.35	CH4	1096
RU-Che	Cherski	68.61	161.34	2015	584
RU-Cok	Chokurdakh	70.83	147.49	2015	1109
RU-Fy2	Fyodorovskoye dry spruce stand	56.45	32.90	WarmWinter	2019
RU-Fyo	Fyodorovskoye	56.46	32.92	WarmWinter	7553
RU-Ha1	Hakasia steppe	54.73	90.00	2015	601
SD-Dem	Demokeya	13.28	30.48	2015	820
SE-Deg	Degero	64.18	19.56	WarmWinter	6769
SE-Htm	Hyltemossa	56.10	13.42	WarmWinter	2156
SE-Nor	Norunda	60.09	17.48	WarmWinter	2525
SE-Ros	Rosinedal-3	64.17	19.74	WarmWinter	2293
SE-Svb	Svartberget	64.26	19.77	WarmWinter	1939
SJ-Adv	Adventdalen	78.19	15.92	2015	377
SJ-Blv	Bayelva, Spitsbergen	78.92	11.83	2015	361
SN-Dhr	Dahra	15.40	-15.43	2015	710
UK-LBT	London_BT	51.52	-0.14	CH4	1461
US-A03	ARM-AMF3-Oliktok	70.50	-149.88	CH4	1461
US-A10	ARM-NSA-Barrow	71.32	-156.61	CH4	2557
US-AR1	ARM USDA UNL OSU Woodward Switchgrass 1	36.43	-99.42	2015	1226
US-AR2	ARM USDA UNL OSU Woodward Switchgrass 2	36.64	-99.60	2015	1125
US-ARM	ARM Southern Great Plains site- Lamont	36.61	-97.49	2015	3455
US-ARb	ARM Southern Great Plains burn site- Lamont	35.55	-98.04	2015	594
US-ARc	ARM Southern Great Plains control site- Lamont	35.55	-98.04	2015	613
US-Atq	Atqasuk	70.47	-157.41	2015	1678
US-BZB	Bonanza Creek Thermokarst Bog	64.70	-148.32	CH4	1096
US-BZF	Bonanza Creek Rich Fen	64.70	-148.31	CH4	1096
US-BZS	Bonanza Creek Black Spruce	64.70	-148.32	CH4	731
US-Beo	Barrow Environmental Observatory (BEO) tower	71.28	-156.61	CH4	730
US-Bes	Barrow-Bes (Biocomplexity Experiment South tower)	71.28	-156.60	CH4	1095
US-Bi1	Bouldin Island Alfalfa	38.10	-121.50	CH4	1096
US-Bi2	Bouldin Island corn	38.11	-121.53	CH4	730
US-Blo	Blodgett Forest	38.90	-120.63	2015	3021
US-CRT	Curtice Walter-Berger cropland	41.63	-83.35	2015	1081
US-Cop	Corral Pocket	38.09	-109.39	2015	1404
US-DPW	Disney Wilderness Preserve Wetland	28.05	-81.44	CH4	1465
US-EDN	Eden Landing Ecological Reserve	37.62	-122.11	CH4	273
US-EML	Eight Mile Lake Permafrost thaw gradient, Healy...	63.88	-149.25	CH4	1080
US-GBT	GLEES Brooklyn Tower	41.37	-106.24	2015	655
US-GLE	GLEES	41.37	-106.24	2015	3281
US-Goo	Goodwin Creek	34.25	-89.87	2015	1400
US-HRA	Humnoke Farm Rice Field Ã¢â‚¬â€œ Field A	34.59	-91.75	CH4	238
US-HRC	Humnoke Farm Rice Field Ã¢â‚¬â€œ Field C	34.59	-91.75	CH4	238
US-Ha1	Harvard Forest EMS Tower (HFR1)	42.54	-72.17	2015	6872
US-Ho1	Howland Forest (main tower)	45.20	-68.74	CH4	2373
US-IB2	Fermi National Accelerator Laboratory- Batavia ...	41.84	-88.24	2015	2577
US-ICs	Imnavait Creek Watershed Wet Sedge Tundra	68.61	-149.31	CH4	1095
US-Ivo	Ivotuk	68.49	-155.75	2015	1131
US-KS1	Kennedy Space Center (slash pine)	28.46	-80.67	2015	277
US-KS2	Kennedy Space Center (scrub oak)	28.61	-80.67	2015	1336
US-LA1	Pointe-aux-Chenes Brackish Marsh	29.50	-90.44	CH4	705
US-LA2	Salvador WMA Freshwater Marsh	29.86	-90.29	CH4	704
US-LWW	Little Washita Watershed	34.96	-97.98	2015	675
US-Lin	Lindcove Orange Orchard	36.36	-119.84	2015	368
US-Los	Lost Creek	46.08	-89.98	2015	2941
US-MAC	MacArthur Agro-Ecology	27.16	-81.19	CH4	944
US-MMS	Morgan Monroe State Forest	39.32	-86.41	2015	5588
US-MRM	Marsh Resource Meadowlands Mitigation Bank	40.82	-74.04	CH4	731
US-Me1	Metolius - Eyerly burn	44.58	-121.50	2015	301
US-Me2	Metolius mature ponderosa pine	44.45	-121.56	2015	3809
US-Me3	Metolius-second young aged pine	44.32	-121.61	2015	1876
US-Me4	Metolius-old aged ponderosa pine	44.50	-121.62	2015	874
US-Me5	Metolius-first young aged pine	44.44	-121.57	2015	934
US-Me6	Metolius Young Pine Burn	44.32	-121.61	2015	1374
US-Myb	Mayberry Wetland	38.05	-121.77	2015	1322
US-NC4	NC_AlligatorRiver	35.79	-75.90	CH4	1827
US-NGB	NGEE Arctic Barrow	71.28	-156.61	CH4	2557
US-NGC	NGEE Arctic Council	64.86	-163.70	CH4	457

Continued on next page

FLUXNET ID	Name	Lat [°N]	Lon [°E]	Origin dataset	No. of obs. [days]
US-NR1	Niwot Ridge Forest (LTER NWT1)	40.03	-105.55	2015	5603
US-Ne1	Mead - irrigated continuous maize site	41.17	-96.48	2015	4383
US-Ne2	Mead - irrigated maize-soybean rotation site	41.16	-96.47	2015	4149
US-Ne3	Mead - rainfed maize-soybean rotation site	41.18	-96.44	2015	4309
US-ORv	Olentangy River Wetland Research Park	40.02	-83.02	2015	335
US-OWC	Old Woman Creek	41.38	-82.51	CH4	669
US-Oho	Oak Openings	41.55	-83.84	2015	3361
US-PFa	Park Falls/WLEF	45.95	-90.27	2015	6357
US-Prr	Poker Flat Research Range Black Spruce Forest	65.12	-147.49	2015	1290
US-SRC	Santa Rita Creosote	31.91	-110.84	2015	1887
US-SRG	Santa Rita Grassland	31.79	-110.83	2015	2466
US-SRM	Santa Rita Mesquite	31.82	-110.87	2015	3974
US-Snd	Sherman Island	38.04	-121.75	CH4	1952
US-Sne	Sherman Island Restored Wetland	38.04	-121.75	CH4	1096
US-Srr	Suisun marsh - Rush Ranch	38.20	-122.03	CH4	1371
US-StJ	St Jones Reserve	39.09	-75.44	CH4	365
US-Sta	Saratoga	41.40	-106.80	2015	1015
US-Syv	Sylvania Wilderness Area	46.24	-89.35	2015	2475
US-Ton	Tonzi Ranch	38.43	-120.97	2015	4809
US-Tw1	Twitchell Wetland West Pond	38.11	-121.65	2015	751
US-Tw2	Twitchell Corn	38.10	-121.64	2015	348
US-Tw3	Twitchell Alfalfa	38.12	-121.65	2015	509
US-Tw4	Twitchell East End Wetland	38.10	-121.64	2015	393
US-Tw5	East Pond Wetland	38.11	-121.64	CH4	365
US-Twt	Twitchell Island	38.11	-121.65	2015	1880
US-UMB	Univ. of Mich. Biological Station	45.56	-84.71	2015	5427
US-UMd	UMBS Disturbance	45.56	-84.70	2015	2679
US-Uaf	University of Alaska, Fairbanks	64.87	-147.86	CH4	2922
US-Var	Vaira Ranch- Ione	38.41	-120.95	2015	5103
US-WCr	Willow Creek	45.81	-90.08	2015	3948
US-WPT	Winous Point North Marsh	41.46	-83.00	2015	1085
US-Whs	Walnut Gulch Lucky Hills Shrub	31.74	-110.05	2015	2737
US-Wi0	Young red pine (YRP)	46.62	-91.08	2015	223
US-Wi1	Intermediate hardwood (IHW)	46.73	-91.23	2015	160
US-Wi2	Intermediate red pine (IRP)	46.69	-91.15	2015	144
US-Wi3	Mature hardwood (MHW)	46.63	-91.10	2015	440
US-Wi4	Mature red pine (MRP)	46.74	-91.17	2015	715
US-Wi5	Mixed young jack pine (MYJP)	46.65	-91.09	2015	234
US-Wi6	Pine barrens #1 (PB1)	46.62	-91.30	2015	250
US-Wi7	Red pine clearcut (RPCC)	46.65	-91.07	2015	170
US-Wi8	Young hardwood clearcut (YHW)	46.72	-91.25	2015	182
US-Wi9	Young Jack pine (YJP)	46.62	-91.08	2015	317
US-Wkg	Walnut Gulch Kendall Grasslands	31.74	-109.94	2015	3888
ZM-Mon	Mongu	-15.44	23.25	2015	677

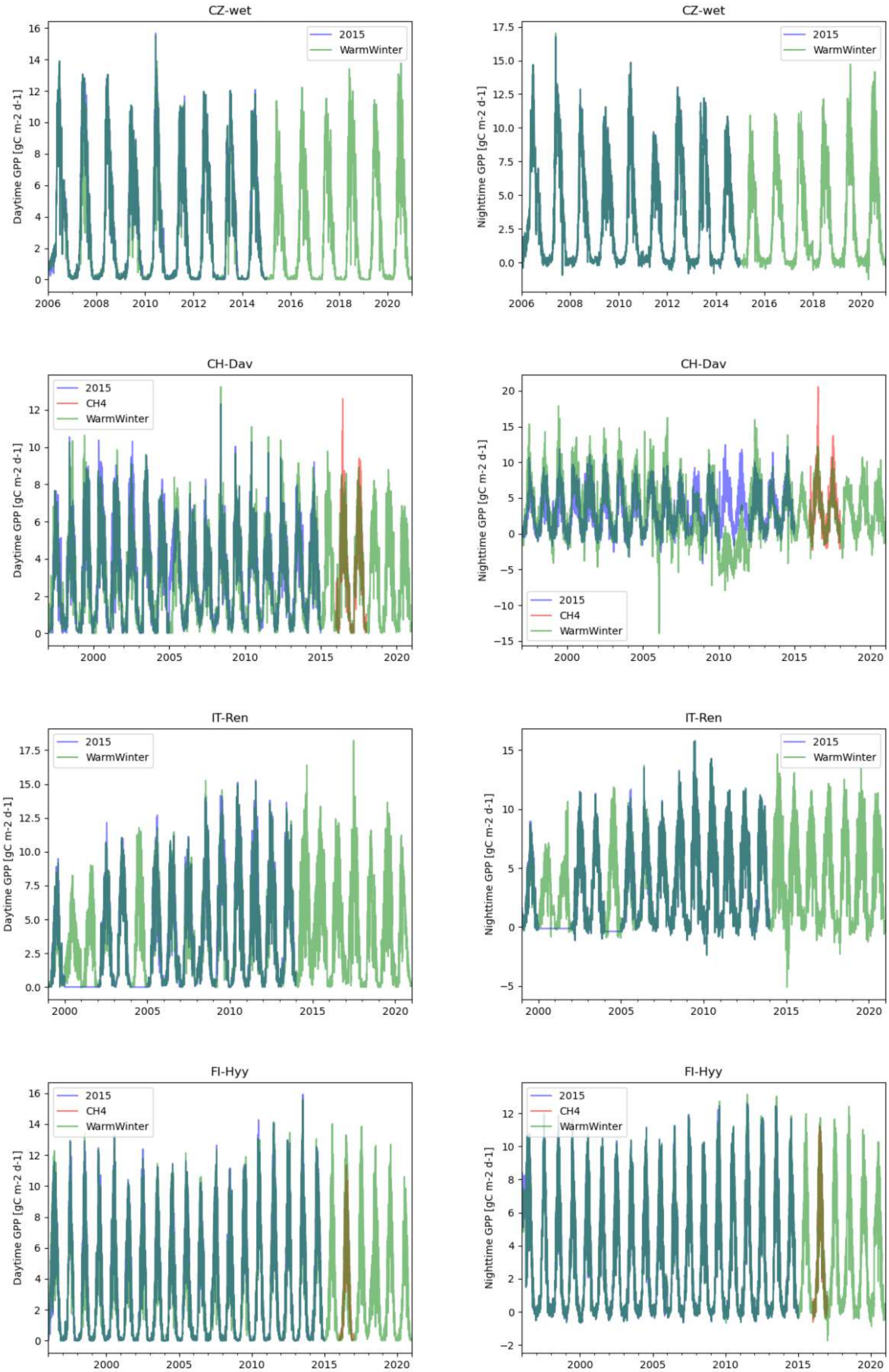


Figure A.2: Comparison of in-situ GPP time series for different FLUXNET datasets for the stations CZ-wet (Czech Republic - Třeboň), CH-dav (Switzerland - Davos), IT-Ren (Italy - Renon) and FI-Hyy (Finnland - Hyytiälä). First column shows GPP derived from daytime partitioning, second column shows GPP derived from nighttime partitioning.

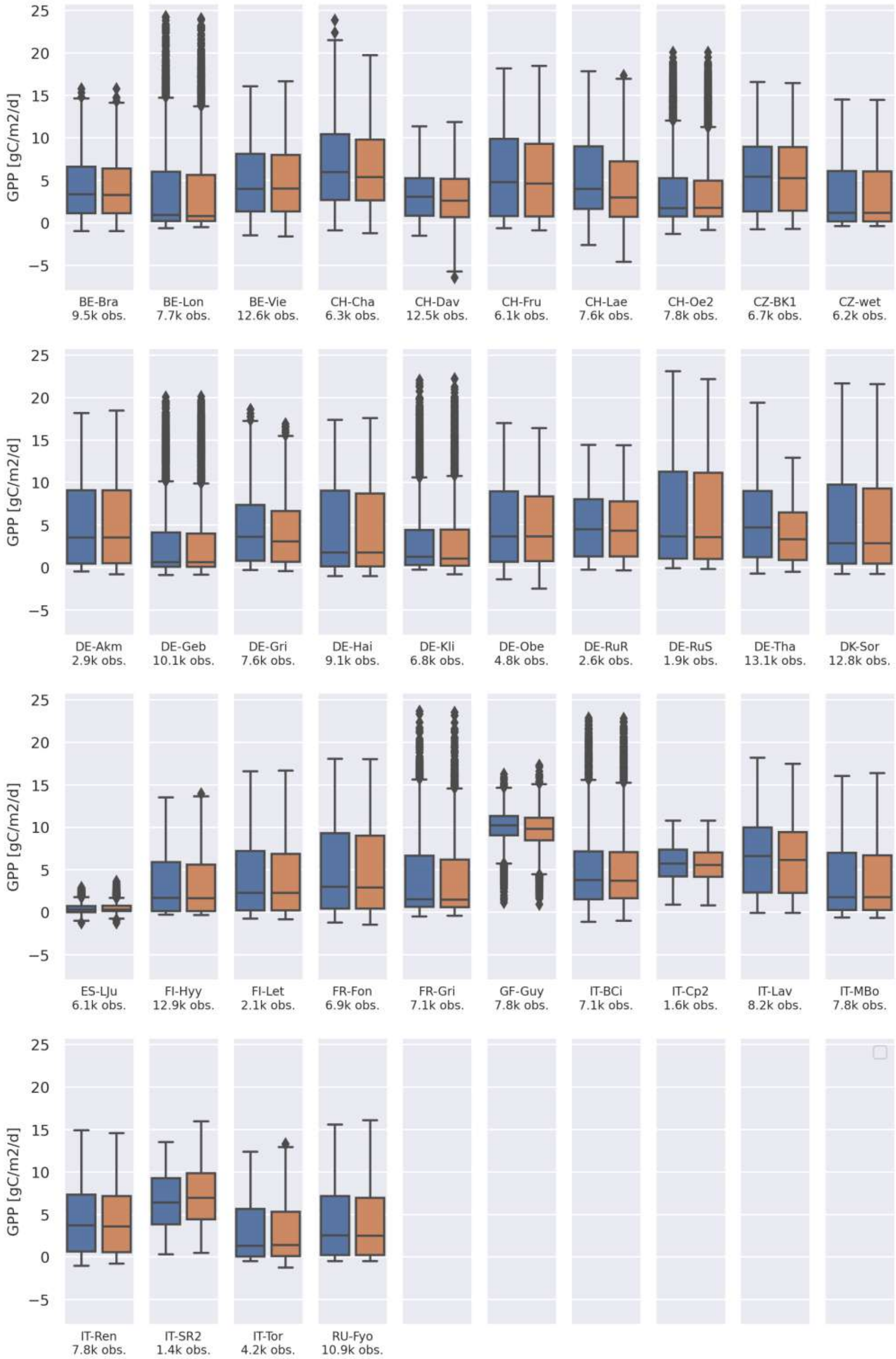


Figure A.3: Box plots of overlapping in situ GPP observations from the FLUXNET2015 (blue) and the FLUXNET Warm Winter (orange) datasets.

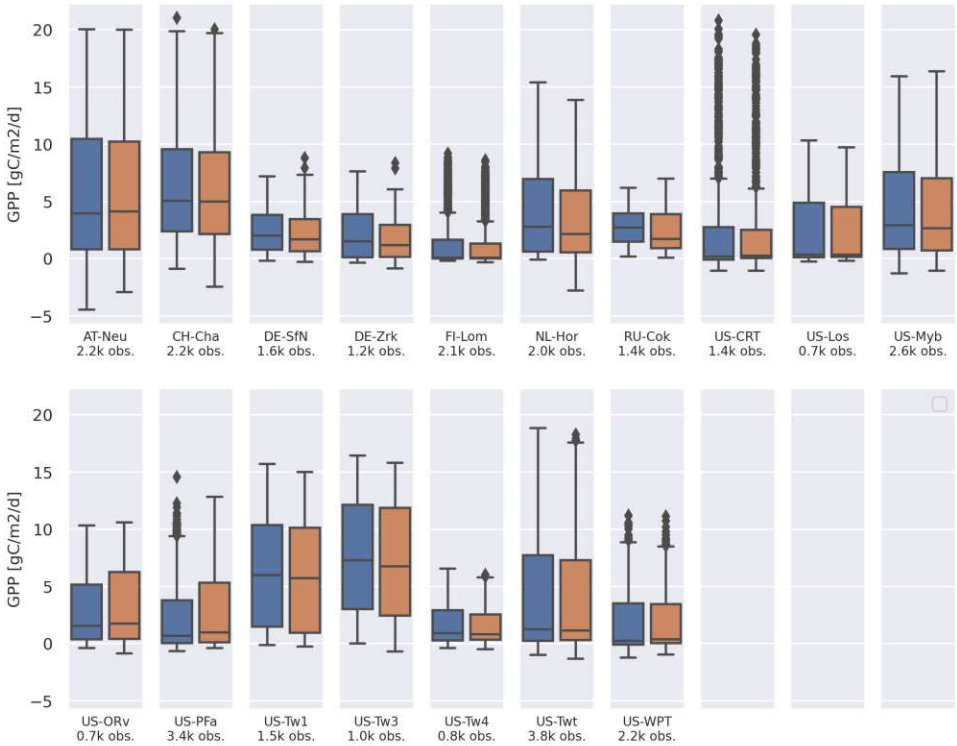


Figure A.4: Box plots of overlapping in situ GPP observations from the FLUXNET2015 (blue) and the FLUXNET CH4 (orange) datasets.

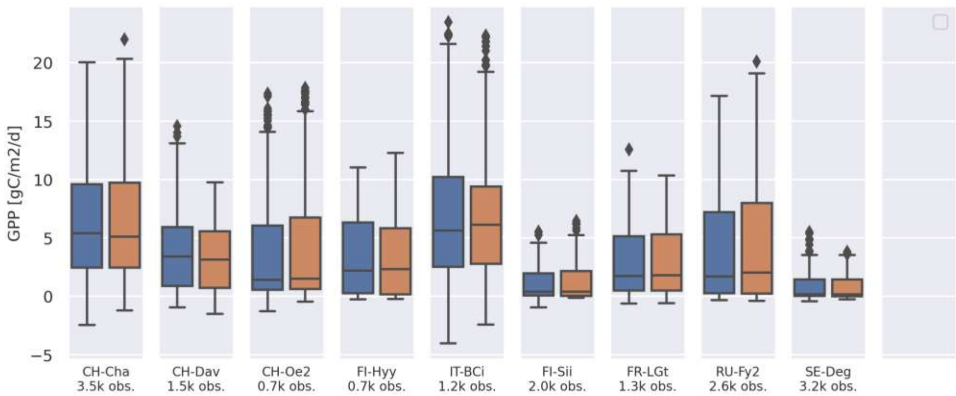


Figure A.5: Box plots of overlapping in situ GPP observations from the FLUXNET CH4 (blue) and the FLUXNET Warm Winter (orange) datasets.

B

Supplementary Results

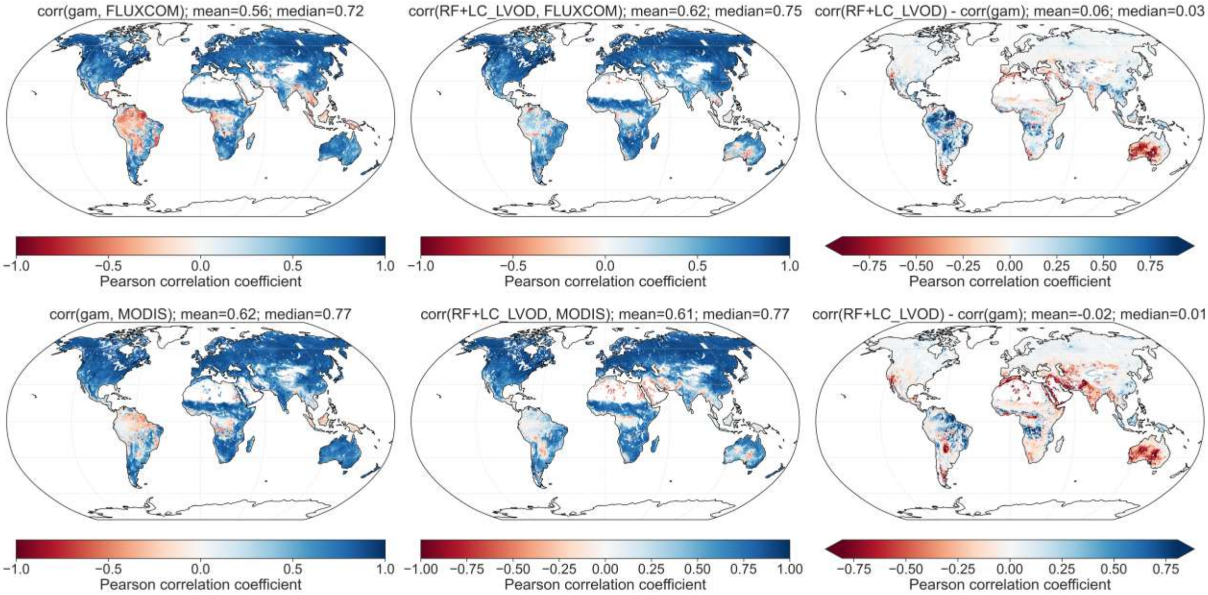


Figure B.1: Difference in correlation between first GAM model and RF model with LC and L-band VOD with GPP from FLUXCOM (top) and MODIS (bottom). The correlations are based on the common observation period between 2001 and 2016 with a 0.25° spatial and 8d temporal resolution.

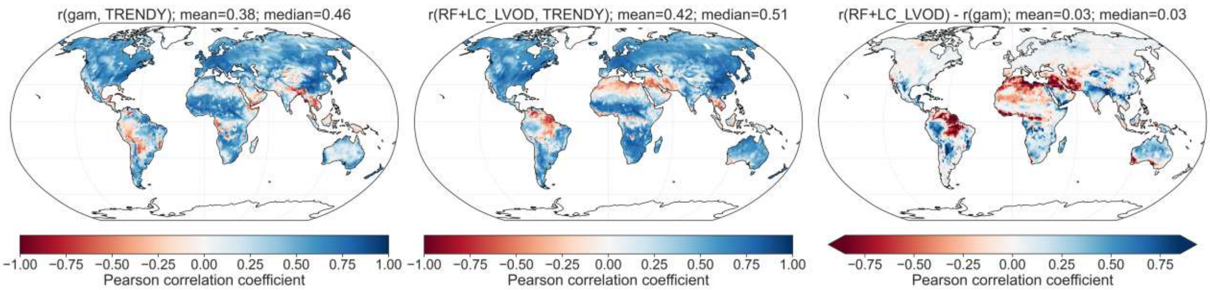


Figure B.2: Difference in correlation between original and final models with GPP from TRENDY. The correlations are based on the common observation period between 1988 and 2017 with a 0.5° spatial and 1 month temporal resolution.

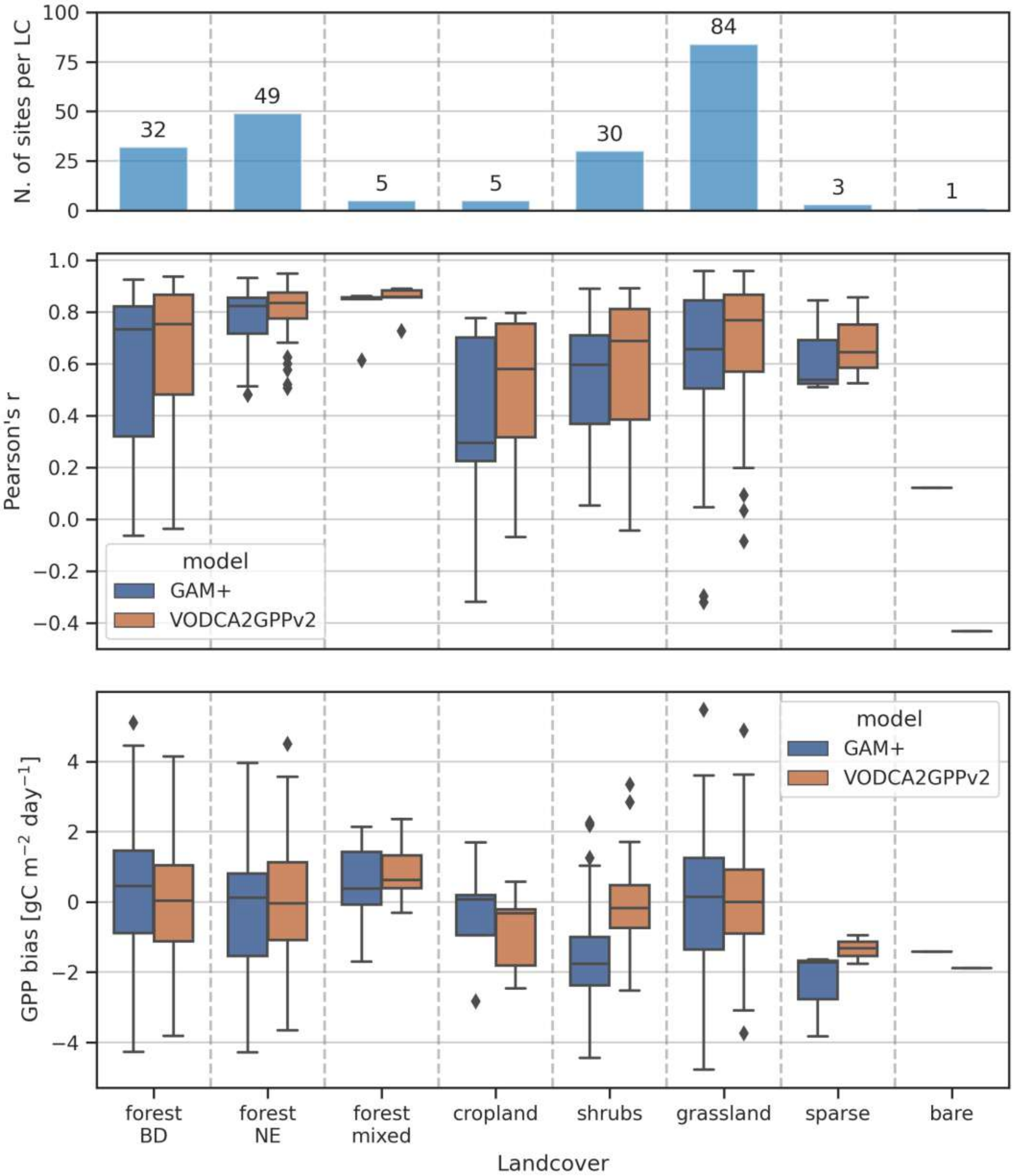


Figure B.3: Box-plots of Pearson correlation coefficients (mid) and biases (bottom) for the VODCA2GPPv2 and GAM+ models, grouped by Land Cover classes. Together with histogram showing amount of sites per LC class.
Each value represents the respective performance metric calculated for one specific site during cross validation. Box extents represent the 25th and 75th percentiles. The maximum length of the whiskers is 1.5 times the interquartile range, outliers points are shown as single dots.