

Functional Biogeography: testing new approaches based on Gross Primary Production

Dissertation

zur Erlangung des akademischen Grades eines

Doctor rerum naturalium (Dr. rer. nat.)



**vorgelegt dem Rat der Fakultät für
Biowissenschaften der
Friedrich-Schiller-Universität Jena**

**von B.Sc. Daniel Ernesto PABON MORENO
geboren am 26. Juli 1993 in Cúcuta, Kolumbien**

November 2022

Gutachter:

1. Prof. Dr. Christine Römermann (Friedrich-Schiller-Universität Jena)
2. Prof. Dr. Miguel D. Mahecha (Universität Leipzig)
3. Prof. Dr. Michael Hauhs (Universität Bayreuth)

Day of the public defence: 13 June 2023

Acknowledgements

I would like to thank my supervisors, Professor Dr. Miguel Mahecha, Dr. Mirco Migliavacca, Professor Dr. Markus Reichstein, and Professor Dr. Christine Römermann, for believing in me to complete this huge task. To the International Max-Planck Research School for Global Biogeochemical cycles (IMPRS-gBGC), especially to Dr. Steffi Rothhardt for all the support during my entire Ph.D. To the Empirical Inference of the Earth System Research group: Dr. Talie Musavi, Dr. Xuanlong Ma, Dr. Nora Linscheid, Dr. Christopher Krich, Dr. Milan Flach, Dr. Guido Kraemer, Fabian Gans, and Dr. Lina Maria Estupinan, thanks for all the ideas and feedback that helped to shape this dissertation. To all the people involved in the TRuStEE (Training on Remote Sensing for Ecosystem modElling) international training network, especially Dr. Micol Rossini, Dr. Cinzia Panigada, Dr. Christian van der tool, Dr. Marco Celesti, and Dr. David Martini for all the support and the good time we spent together during the training courses. I also would like to thank Birgitta Wiehl, Linda Maack, Alice Ratajczak, and Sarah Nuschke for all the outstanding support. To Dr. Carlos Sierra, and the spontaneous chats that motivated me to finish this dissertation. To Professor Dr. Björn Reu, the first person who believed that I could be a scientist. To my family, Elsy, Egidio, Camilo, my wife Andrea, and my son Emilio for all your love. A todos los y las estudiantes y profesores que han ofrendado su vida para que la universidad publica y de calidad siga existiendo en Colombia. Gracias a su sacrificio he podido convertirme en científico. Last but not least, to myself for all the effort and commitment I have put into my work and, most importantly, for keeping me alive in the process.

Special thanks to Lina Estupinan, Danitza Gutierrez, Melissa Ruiz, Andrea Silva, and Diego Ortiz for proof reading the introduction and discussion. As well as to Nora Linscheid and Christian Reimers for their help with the translation of the summary into German.

This research was funded by the European Unions Horizon 2020 Research and Innovation Program via the TRuStEE Project through the Marie Skłodowska-Curie (Grant Number: 721995).

Contents

Summary	5
Zusammenfassung	7
1 Introduction	9
1.1 Terrestrial Gross Primary Production	9
1.2 Functional Biogeography	11
1.3 An approach to Functional Biogeography using Gross Primary Production	12
1.4 Research questions and summary of the manuscripts	13
2 A Functional Classification System based on maximum Gross Primary Production	15
3 Ecosystem physio-phenology revealed using circular statistics	31
4 On the Potential of Sentinel-2 for Estimating Gross Primary Production	49
5 Concluding Discussion	63
A Complementary results: Empirical comparison between GPP_{max} and GPP_{sat}	85
B Supplementary information: Ecosystem physio-phenology revealed using circular statistics	91
C Supplementary information: On the Potential of Sentinel-2 for Estimating Gross Primary Production	141
D Declaration Forms	217
E Curriculum Vitae	221
F Declaration of honor / Ehrenwörtliche Erklärung	229



Summary

Functional biogeography focuses on understanding the causes and consequences of the spatial distribution of functional traits. One of the functions that plants fulfill in ecosystems is to regulate the exchange of energy and matter through photosynthesis (i.e., Gross Primary Production, GPP). In this direction, previous studies define the maximum photosynthetic capacity (GPP_{max}) as an Ecosystem Functional Property (EFP). It is expected that EFPs will help to better understand how ecosystems are affected by climate change and global warming. In this dissertation, I present how to understand an ecosystem function such as GPP and an EFP as GPP_{max} through three main axes: Magnitude, Time and Space. Specifically, in the first axis, I focus on understanding how environmental variables limit GPP_{max} at a global scale. In the second axis, time, I evaluate how climatic variables affect the timing (i.e. Day of the year, DOY) when GPP_{max} is reached ($DOYGPP_{max}$). Finally, on the last axis, I explore how new satellite missions can help improve GPP prediction. In the first axis, my approach using an optimal ecosystem process as GPP_{max} and its limiting factors allow us to generate a new functional map that relate optimum ecosystem processes with climate information. This new map can be used as input for future global-scale modeling studies. On the second axis, Time, I find that for most of the ecosystems across the globe, an increase in short-wave incoming radiation, temperature, and vapor pressure deficit will produce an earlier $DOYGPP_{max}$ (when compared with the mean $DOYGPP_{max}$) during the growing season. In the last axis, Space, I find that red-edge vegetation indices (estimated from Sentinel-2 images) improve the performance of the prediction of GPP. The exploration of the axes proposed in this dissertation contribute to frame our understanding of vegetation functional biogeography, by combining concepts and techniques from biometeorology, ecosystem physio-phenology, and satellite earth observations.



Zusammenfassung

Die funktionelle Biogeographie konzentriert sich darauf die Ursachen und Folgen der räumlichen Verteilung funktioneller Merkmale zu verstehen. Eine dieser Funktionen, die Pflanzen in Ökosystemen erfüllen, ist die Regulierung des Energie- und Stoffaustauschs durch Photosynthese (d.h. die Bruttoprimärproduktion, GPP). Diesbezüglich definieren frühere Studien die maximale Photosynthesekapazität (GPP_{max}) als eine funktionelle Eigenschaft des Ökosystems (EFP). Die EFPs sollen dazu beitragen, den Einfluss von Klimawandel und globaler Erwärmung auf Ökosysteme besser zu verstehen. In dieser Dissertation präsentiere ich, wie man eine Ökosystemfunktion wie GPP und eine EFP wie GPP_{max} entlang dreier Hauptachsen verstehen kann: Stärke, Zeit und Raum. Für die erste Achse konzentriere ich mich insbesondere auf das Verständnis, wie Umweltvariablen GPP_{max} auf globaler Ebene begrenzen. Für die zweite Achse, der Zeit, untersuche ich, wie klimatische Variablen den Zeitpunkt (d.h. den Tag des Jahres, DOY) an dem GPP_{max} erreicht wird ($DOYGPP_{max}$) beeinflussen. Für die letzte Achse untersuche ich schließlich, wie neue Satellitenmissionen zur Verbesserung der GPP-Vorhersage beitragen können. Für die erste Achse ermöglicht mein Ansatz eine neuen funktionalen Karte zu erstellen, die, unter Verwendung eines optimalen Ökosystemprozesses als GPP_{max} und seiner begrenzenden Faktoren, optimale Ökosystemprozesse mit Klimainformationen verbindet. Diese neue Karte kann als Input für zukünftige globale Modellierungsstudien verwendet werden. Für die zweite Achse, die Zeit, stelle ich fest, dass weltweit für die meisten Ökosysteme ein Anstieg der kurzweiligen einfallenden Strahlung, der Temperatur und des Sättigungsdefizits zu einem früheren $DOYGPP_{max}$ (im Vergleich zum mittleren $DOYGPP_{max}$) während der Wachstumsperiode führt. Für die letzte Achse, der räumlichen Verteilung, stelle ich fest, dass die (aus Sentinel-2-Bildern geschätzten) Vegetationsindizes am roten Rand die Vorhersage von GPP verbessern. Die Erforschung entlang der in dieser Dissertation vorgeschlagenen Achsen trägt dazu bei, unser Verständnis der funktionalen Biogeographie der Vegetation zu erweitern, indem Konzepte und Techniken aus der Biometeorologie, der Ökosystemphysiologie und der satellitengestützten Erdbeobachtung kombiniert werden.



CHAPTER 1

Introduction

Since 1790, when the industrial revolution began, the concentration of carbon dioxide (CO₂) in the atmosphere has been increasing (Andres et al., 2012; Eyring et al., 2021). Humans learned to burn fossil fuels to increase the number of goods and create new services, a practice that remains until now (Friedlingstein et al., 2019). A negative effect of fuel combustion is that large amounts of CO₂ are released into the atmosphere. Since CO₂ is one of the main greenhouse gases, its increase in the atmosphere has led to an increase in global air temperature in the last decades (i.e., global warming). The increase in temperature and deforestation caused by the industrial revolution has changed the climate regimes across the globe (i.e., climate change Gulev et al., 2021). Circularly, climate change has drastically impacted the ecosystem's dynamics and composition, affecting ecosystem processes such as carbon uptake, water regulation, and cloud formation (Canadell et al., 2021).

The transformation of energy and matter in terrestrial ecosystems is mainly constrained by the capacity of plants to absorb and transform CO₂ and water into carbohydrates, lipids, proteins, and oxygen (O₂) (i.e., photosynthesis, Chen and Blankenship, 2021). Photosynthesis products are used by other organisms through the trophic network (Chapin et al., 2011b) and constitute the main building blocks of other forms of life in the biosphere. To understand how different ecosystems are being affected by climate change and global warming, it is necessary to measure the exchange of energy and matter between the ecosystems and the atmosphere (Baldocchi, 2008). Having reliable knowledge of the photosynthetic activity of plants (Gross Primary Production, GPP) can give us a better understanding of how ecosystems respond in terms of energy and matter fluxes to the increase in atmospheric CO₂ and temperature (Campbell et al., 2017).

1.1 Terrestrial Gross Primary Production

Terrestrial GPP is the measurement of CO₂ uptake by ecosystems through plant photosynthesis (Dokulil, 2019; Chapin et al., 2011a). In the late 1940s and early 1950s, Montgomery (1948), and Swinbank (1951) developed the theoretical basis of the eddy covariance method (EC). The EC method estimates the exchange of

matter and energy between the atmosphere and biosphere using the covariance of vertical wind velocity fluctuations and the chemical component fluctuation to be measured (e.g., CO₂, H₂O) (Aubinet et al., 2012). Nevertheless, it was not until the 1980s that scientists started registering the continuous exchange of CO₂ between the atmosphere and plant's canopy (Verma et al., 1986). The continuous registers were possible thanks to the technological development of gas analyzers and sonic anemometers (Ohtaki and Matsui, 1982; Ohtaki, 1984). Later, in the 1990s, several regional networks of EC towers started (Aubinet et al., 2012). At the beginning of the 2000s, the FLUXNET initiative (Baldocchi et al., 2001), a meta-network of numerous regional networks with EC data, was created (For a complete review of the history of the eddy covariance method, see: Baldocchi, 2003). FLUXNET aimed to create a single data repository of standardized EC information accessible to any scientist. Today, the latest FLUXNET dataset holds 206 EC-towers (1532 site-years) around the world, covering several vegetation types and climate classes (Pastorello et al., 2020). Thanks to FLUXNET, it has been possible to understand how ecosystems respond to climate change and global warming in terms of energy and matter fluxes (Yu et al., 2019). Understanding the relationship between the biotic and abiotic components of the Earth is more relevant than ever; as CO₂ levels continue to rise, climate change and global warming have an increasing impact on human societies around the world (Pörtner et al., 2022).

The EC technique has, however, some constraints. On the one hand, fluxes estimation is only possible under stable atmospheric conditions on flat terrain and where the canopy is homogeneous (Schmid and Lloyd, 1999); on the other hand, the spatial scale of fluxes' footprint ranges from hundreds of meters to a few kilometers (Schmid, 1994). These limits represent a challenge for regional and global estimations of GPP. Furthermore, to estimate the state of vegetation on a global scale. Fortunately, the eddy covariance technique has not been the only tool scientists have developed to study vegetation productivity. The space race prompted the development of satellite missions and sensors to study and track changes on the land surface (Belward and Skøien, 2015). The information provided by satellite missions combined with the knowledge of the optical properties of plants (Tucker, 1979) has made it possible to monitor vegetation dynamics at regional (Goward et al., 1985, 1991), and global scale (Ryu et al., 2019). Later, the combination of remote sensing information and model-based process understanding led to the development of the first continuous GPP global product (Running et al., 2004). In the last decades, with the development of new mathematical algorithms as machine learning techniques (Breiman, 2001), new estimates of GPP at a global scale were possible by combining satellite images and estimates of GPP from EC towers (Jung et al., 2009, 2020, 2019; Tramontana et al., 2016). As technology improves, new sensors have made it possible to increase the temporal, spatial, and spectral resolution at which we monitor vegetation (Thépaut et al., 2018). Examples of these new sensors are the Sentinel satellites from the European Space Agency. Nevertheless, evaluating whether the new Sentinel satellites allow us to better predict the fluxes between ecosystems and the atmosphere is a question that remains open.

1.2 Functional Biogeography

Functional Biogeography combines ecology, geography, and environmental sciences to understand what organisms and ecosystems do, rather than the elements and interactions that constitute organisms and ecosystems (Violle et al., 2014; Malaterre et al., 2019). From an epistemic perspective, classic ecology focus on the interaction of the species with the environment. In functional biogeography, the species concept is not seen as a unit of evolution (Duckworth et al., 2000). Instead, it focuses on the understanding of the functions of the organisms (For a complete discussion between classic ecology and functional biogeography, see: Violle et al., 2014). For example, plants absorb CO₂ and water to synthesize sugars, lipids, and proteins and release oxygen and water vapor during the same process. The synthesized compounds then work as support for other organisms (Chase et al., 2000; Warne et al., 2010, e.g.). In this sense, the definition of a function can be delimited in a series of metabolic processes where the Earth's biotic elements transform energy and matter from different sources to produce compounds with higher chemical potential energy. Following this idea, Reichstein et al. (2014) proposed that ecosystem functional properties (EFPs) are measures of optimal biogeochemical processes that can be quantified at the ecosystem level, for example, using the eddy covariance technique. It is expected that EFPs help to characterize and understand ecosystem dynamics and how ecosystem properties change into the current climate change and global warming scenario (Reichstein et al., 2014).

Timing as a Plant Trait and an Ecosystem Functional Property

The original formulation of EFPs was developed to link plant traits with ecosystem fluxes reducing the confounding effect of climate regimes (Figure 1.1). Nevertheless, an essential element in this formulation was not considered: the timing of optimal ecosystem processes. In the last decades, the consequences of plant phenology shifts on the global carbon cycle have been observed (Richardson et al., 2013, 2010; Buitener et al., 2015). For this reason, Wolkovich et al. (2014) notes that while several studies in the last decade have focused on understanding plant phenology as a functional trait, a multidisciplinary approach that includes ecology, biometeorology, and phylogenetics is still necessary. In this dissertation, I argue that EFPs can also be interpreted as optimum ecosystem states derived from fluxes without removing the climate effect (For a complete comparison between GPP_{max} and GPP_{sat} see the Appendix A). Then, the timing of optimal ecosystem processes as the maximum Gross Primary Production (GPP_{max}) should also be included as part of the functional biogeography in general.

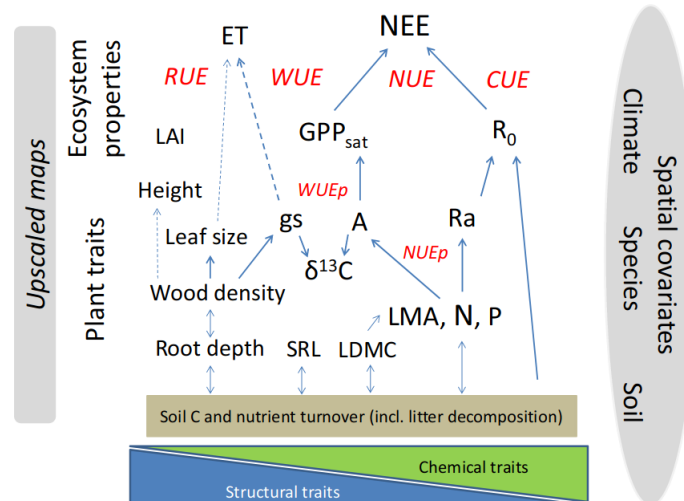


Figure 1.1: Major relationships between plant structural and chemical/physiological traits and ecosystem functional properties related to carbon and water fluxes, embedded in an upscaling framework considering spatial covariates. Tradeoffs related to water-use (*WUE*), nitrogen-use (*NUE*), radiation-use (*RUE*), and carbon-use (*CUE*) efficiencies at plant (*p*) and ecosystem scale are printed in italics in red. Note that soil C and nutrient turnover processes are also important ecosystem properties, affecting carbon fluxes directly via heterotrophic soil respiration and indirectly via effects of nutrient availability on plant functional traits and ecosystem structure. *A*, photosynthetic capacity; $\delta^{13}\text{C}$, stable carbon isotope ratio; *ET*, evapotranspiration; GPP_{sat} , gross primary productivity at saturating light; *gs*, maximum stomatal conductance; *LAI*, leaf area index; *LDMC*, leaf dry matter content; *LMA*, leaf mass per area; *N*, tissue nitrogen concentration; *NEE*, net ecosystem exchange of CO_2 ; *P*, tissue phosphorus concentration; R_0 , ecosystem respiration at reference temperature; *Ra*, plant respiration; *SRL*, specific root length. Figure and legend reproduced from Reichstein et al. (2014).

1.3 An approach to Functional Biogeography using Gross Primary Production

Although plant functions are regulated by different metabolic processes (Figure 1.1). In this dissertation, I will focus on one function: the terrestrial gross primary production, and one EFP: the optimum photosynthetic capacity (GPP_{sat} or GPP_{max}) of plants per unit area per second. To have a complete understanding of GPP , and GPP_{max} in the context of biogeography, it is necessary to develop three axes: (1) Magnitude (here it refers to the value *per-se*), Timing, and Space (Figure 1.2). These reference axes are present throughout the entire dissertation at different levels.

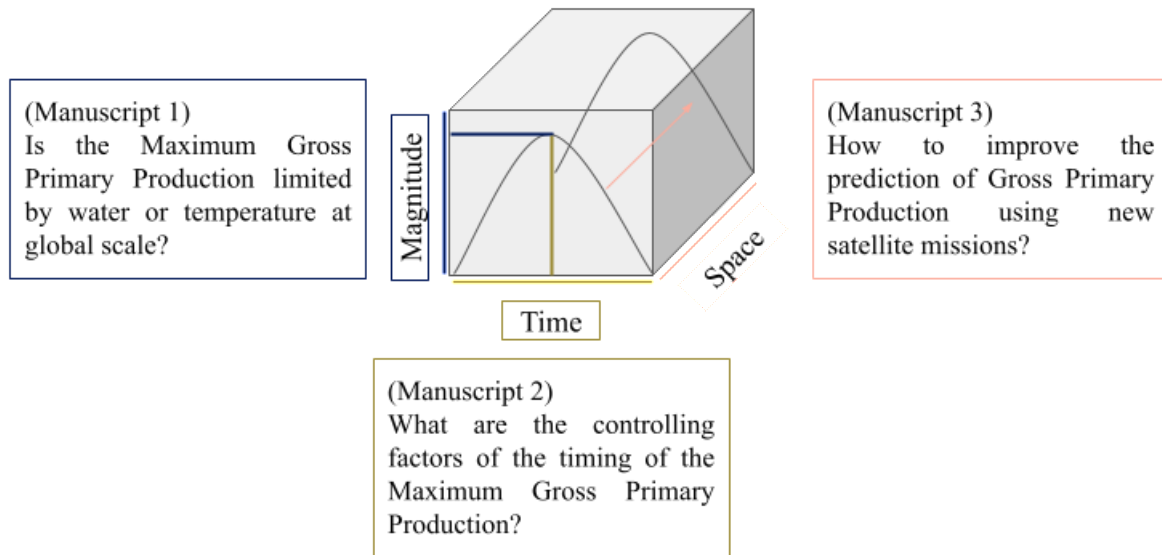


Figure 1.2: Conceptual representation of the three axes of study on this dissertation. The gray line represents the gross primary production variation during the growing season. The first axis **Magnitude** refers to the value of gross primary production (GPP) and GPP_{max} *per-se*. The second axis **Time** is the timing when GPP_{max} is reached. The third axis, **Space**, represents the prediction/upscaling of GPP using new remote sensing information. The arrow represents the increase in spatial resolution. From local eddy covariance Towers to regional and global predictions of GPP

1.4 Research questions and summary of the manuscripts

Research Question 1.

Is the Maximum Gross Primary Production limited by water or temperature at a global scale?

In **Chapter 2 / Manuscript 1** I investigate in what ecosystems GPP_{max} is limited either by temperature or precipitation. Based on these results, I proposed new functional classes and analyzed how these classes are related to previous vegetation classification systems. This study allows us to better understand the relationship between climate variables and GPP. The resulting categories, from the relationship between GPP limitation by water or temperature, may be useful for future exercises to predict the impact of climate change on ecosystems.

Research Question 2.

What are the controlling factors of the timing of the Maximum Gross Primary Production?

In **Chapter 3 / Manuscript 2**, I explore the potential of circular-linear regressions (circular statistics) as a method to correlate changes in the timing of GPP_{max} with the cumulative effect of climate variables (ecosystem memory). These results allow a better understanding of ecosystems' response to climate change not only from a phenological perspective but from a more holistic ecosystem physio-phenology perspective.

Research Question 3.

How to improve the prediction of Gross Primary Production using new satellite missions?

In **Chapter 4 / Manuscript 3**, I present how new satellite missions (i.e. Sentinel-2) from the European Space Agency (ESA) improve the prediction/upscaling of GPP using red-edge vegetation indices that could not be computed previously. I also explore the bias produced by the imbalanced representation of observations by vegetation type (different number of observations) and whether balancing techniques can help to better represent vegetation types and low-frequency observations (rare observations). This work provides a base for future global estimates of GPP and GPP_{\max} based on the new satellite missions from ESA.

CHAPTER 2

A Functional Classification System based on maximum Gross Primary Production

Manuscript Nr. 1

Title of the Manuscript: A Functional Classification System based on maximum Gross Primary Production

Authors: Pabon-Moreno, Daniel E., Migliavacca, Mirco., Reichstein, Markus., Jung, Martin., Weber, Ulrich., Mahecha, Miguel D.

The candidate is: First author

Status: in preparation

Authors' contribution in percentage

Author	Conceptual	Data analysis	Experimental (Numerical)	Writing the manuscript	Provision of material
Daniel E. Pabon-Moreno	40	79	89	83	0
Dr. Mirco Migliavacca	15	8	3	10	0
Prof. Dr. Markus Reichstein	5	5	5	2	0
Prof. Dr. Miguel D. Mahecha	40	8	3	5	0
Dr. Martin Jung	0	0	0	0	50
Ulrich Weber	0	0	0	0	50
Total:	100	100	100	100	100

A Functional Classification System based on maximum Gross Primary Production

Daniel E. Pabon-Moreno¹, Mirco Migliavacca², Markus Reichstein^{1,3}, Marting Jung¹, Ulrich Weber¹, and Miguel D. Mahecha^{4,5}

¹Biogeochemical Integration Department, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany

²European Commission, Joint Research Centre (JRC), Via E. Fermi, 2749, 21027 Ispra VA

³Michael-Stifel-Center Jena for Data-driven and Simulation Science, Friedrich-Schiller-University Jena, 07743 Jena, Germany

⁴Remote Sensing Center For Earth System Research, Leipzig University, 04103 Leipzig, Germany

⁵Helmholtz Centre for Environmental Research, 04318 Leipzig, Germany

Correspondence: Daniel E. Pabon-Moreno (dpabon@bgc-jena.mpg.de)

Abstract.

The classification of ecosystems by their eco-physiological characteristics or response to climate conditions has proven to be of high utility of multiple applications. Evaluating climate change effects on ecosystems, for instance, often is performed by vegetation class. The most widely used classical classification system is based on climate data. Satellite remote sensing observations are used to classify land cover, and modeled species distribution maps lead to accurate ecosystem classifications. However, none of the existing classification products so far considers ecosystem processes as a key dimension of interest in the classification. One of the ecosystem processes that is highly relevant is the gross primary productivity (GPP) i.e. the total amount of CO₂ that is absorbed by plants through photosynthesis. Today, it is possible to monitor GPP dynamics at global scale thanks to the combination of remote sensing and machine learning techniques. In this study, we explore how the maximum annual GPP (GPPmax), a key feature of ecosystem functioning, can be used to classify the vegetated Earth surface, and how this new classification system compares to previous vegetation classification systems as Köppen-Geiger climate classes and the biomes proposed by Olsen. We find that most of the land surface correspond to the class where GPPmax is mainly limited by temperature and precipitation. While the second class in geographical extension is when GPPmax is only limited by temperature. Our classification system shows a low percentage of agreement with Köppen-geiger classes and Olson biomes. Nevertheless, we found that many transitions on köppen-geiger classes and Olson biomes are also presented in our classification system. Considering optimum ecosystem processes and its temporal and climatic component opens new ways to classify the vegetated surface, and integrate the dynamics between biotic and abiotic components of ecosystems at global scale.

1 Introduction

1. *What is functional classification /classification of the land surface (concept definition, history and utility, for internal use only)*

Functional classification systems are based on what organisms do instead of what organisms are composed (Malaterre et al., 2019). The idea behind of functional classification is that independently of the taxonomic relationship, different organisms can share the same function into the ecosystem (e.g. different plant species share the same role in terms of primary productivity). Examples of functional classification systems can be tracked in time through different centuries until Aristotles who classified animals based on their diet (e.g. carnivorous, herbivorous) (Malaterre et al., 2019); later in the XIX Century, Humboldt (1806) proposed a classification system based on plant characteristics as the growth form settling the bases for the modern functional classification systems of plants. Functional classifications had been, not only applied to living organisms, in the last decades of the XIX Century, Wladdimir Köppen developed a climate classification system based on data from several weather stations around the world and the distribution of the vegetation (Köppen, 1884) generating one of the first world climate classification systems. Even if Köppen-Geiger system is considered as a climate classification, Köppen built his model based on the relationship between plants and the climate conditions. In this sense, we can consider Köppen-Geiger classes as the first prototype of a functional classification of the vegetated land surface. Nevertheless, modern functional classification systems started in the 80th, when the concept of plant functional types (PFT) was developed (Gillison, 1981; Box, 1981) at the same time as the International Geosphere Biosphere Program was starting. PFTs were defined based on structural, and physiological characteristics of the vegetation and its interaction with climatic conditions (Box, 1995). These works were the based for later, when satellite products where available, to build land cover maps at global scale. The PFTs not only allow the scientist and modelers to represent the enormous diversity of plants at global scale in a lower number of classes that interact with the environment in similar ways, but also to understand the geographical distribution of vegetation in previous geological times, and to try to understand how vegetation is responding to climate change (Lavorel et al., 2007).

2. *New Information, new techniques, new frontiers.*

Today, the development of satellite missions opened new opportunities to understand the dynamics between biotic and abiotic components of the land surface. Thanks to this, we were able to have a record of the changes and processes occurring on the Earth's surface during the last decades (Belward and Skøien, 2015). This is illustrated by the images captured by Landsat and MODIS missions (Friedl et al., 2002; Markham and Helder, 2012), and more recently the first world land cover map at 10 meters resolution, developed using radar and optical information from Sentinel-1 and Sentinel-2 satellites (Zanaga et al., 2021). But we have not only seen progress in the area of land/functional classification. Today several global products of biophysical variables as leaf area index (Fang et al., 2019), fraction of the Absorbed photosynthetic active radiation (FAPAR), and global products of ecosystem processes as Gross Primary production (Tramontana et al., 2016; Bodesheim et al., 2018; Jung et al., 2020) among others are available thanks to the combination of earth observations from satellite missions, process-based model and machine learning techniques. In this sense, as new data-analytic tools have been widely became available, classical classification schemes have been challenged. For instance, Netzel and Stepinski (2016); Zscheischler et al. (2012) used clustering methods to generated and evaluated

new climate classes, opening new ways to interpret the different climatic regions and the relationship between vegetation distribution and climate conditions.

55
3. *From plant functional types to optimality principles (Background, current status, and aims of our study).*

Wullschleger et al. (2014) proposed that a radical approach to overcome PFTs is to abandon the concept, and focus in plant/ecosystem optimality principles as shown by Pavlick et al. (2013) in the JeDi global dynamic vegetation model. In their approach, PFTs are replaced by a series of ecophysiological traits parameters, as carbon allocation, turnover, and senescence among others, allowing to better represent the functional diversity of plants at global scale. In this direction and from a data-driven perspective, Reichstein et al. (2014) proposed the ecosystem functional properties' concept, that are quantities that allow to correlate ecosystem processes and functions in an integrated manner with plant functional traits. An example of this approach is shown by Musavi et al. (2016) linking the potential photosynthetic capacity with leaves nitrogen content and Leaf Area Index. More recently, Migliavacca et al. (2021) showed that leaf optimality can be extrapolated to ecosystems in a three-dimensional space, where the first dimension is mainly governed by the maximum CO₂ uptake. Now that new global products, techniques and theories have been developed to understand the interaction of biotic and abiotic components of ecosystems at global scale, we can revisit plant functional types from a new perspective that integrate ecosystem processes, climate conditions and optimality principles to generate new classification units. Here, we explore the generation of discrete classes using the climate space when terrestrial Gross Primary Production (i.e. photosynthetic capacity) is maximum (GPPmax). This allows us to represent an optimal ecosystem process and its relationship with climate conditions at global scale.

60
65
70
2 Methods

2.1 Data

We use FLUXCOM 8-days GPP global product at 0.5 degrees of spatial resolution from 2001 to 2015 that was produced using only remote sensing information, specifically the ensemble of several machine learning techniques (Jung et al., 2019). To filter out pixels where water could be mixed with land, We use the land cover vegetation map generated from MODIS (IGBP, product: MCD12Q1v006 Friedl, Mark and Sulla-Menashe, Damien, 2019) for the year 2015, We resample the product at 0.5 degrees and the percentage of each land cover type is estimated. Each grid cell is assigned with the highest fraction; finally, we filter out pixels with the class "water bodies".

80 We consider precipitation and air temperature as climatic drivers of GPP among others because it allows us to represent if the ecosystems are water or energy limited. Specifically, we use daily ERA-5 global products (Hersbach et al., 2020) at the same spatial resolution, and for the same time period. We compute the mean seasonal cycle per pixel (MSC) for each variable, and perform the following analysis on the calculated MSC.

2.2 Climate space when GPP is maximum

85 We extract the day of the year (DOY) when GPPmax occurs for each grid cell. Then using DOY, we estimate the average
air temperature, and the cumulative precipitation 30 days before GPPmax occurs. We compare these values with the average
air temperature of the MSC, and the average precipitation of the monthly cumulative precipitation per pixel. We define four
Functional classes (FCs) based on if the air temperature, and precipitation (when GPP is maximum) are higher or lower than
the MSC average. **Tair (+) Precip (+)** when both climate values are above the average indicating that GPPmax is limited by
90 both variables, **Tair (-) Precip (+)**, **Tair (+) Precip (-)** when at least one of the values is higher than the average indicating that
ecosystems are limited by at least one the variables. And **Tair (-) Precip (-)** when both values are below the average, indicating
that plant-ecosystems are non-limited by any variables. Finally, to estimate the percentage of agreement of DOY GPPmax and
detecting potential anomalies, we estimate the circular standard deviation (Mardia, 1972) using the original time series from
2001 to 2015.

95 2.3 Comparing Functional classes types with previous classification systems

We compare the four FCs with the current Köppen-Geiger climate classes and the terrestrial biomes proposed by Olson et al.
(2001). We estimate the percentage of agreement between each one of the classification systems through the corrected Rand
index (Hubert and Arabie, 1985) and the Normalized Mutual Information index (Danon et al., 2005). We also count the total
area for each FCs and the co-occurrence with Köppen-Geiger, and ecoregion classes.

100 3 Results

In this section, we first present the global distribution of the timing (phenology) of GPPmax (Figure 1). Then we present
the optimum climate (Air temperature, and Precipitation) space 30 days before GPP is maximum (Figure 2, a,b), and the
geographical distribution of the optimum climate space compared to the average (delta climate space, Figure 2, c,d). In the
last section we present the four FCs based on the sign of the delta climate space (Figure 3, FCs), and finally, we show the
105 relationship between the FCs, with the Köppen-Geiger climate classification and terrestrial biomes proposed by Olson et al.
(2001) (Figure 4).

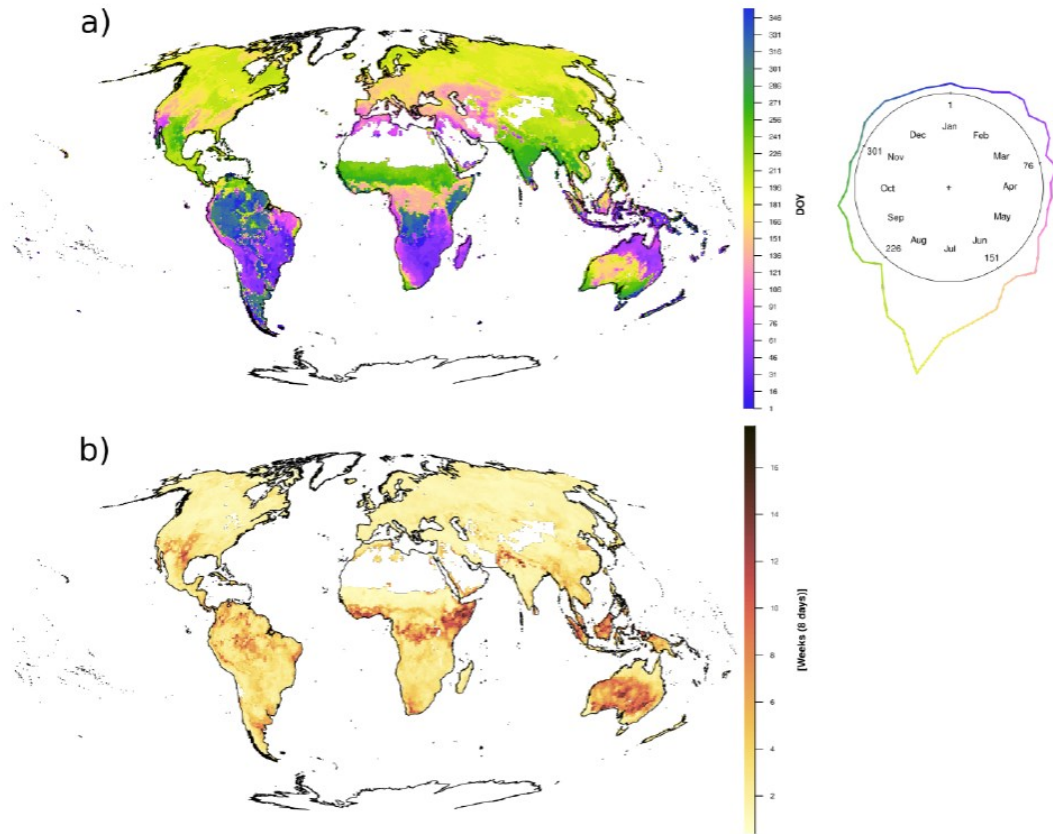


Figure 1. Timing of GPPmax at global scale using the mean seasonal cycle of FLUXCOM GPP (8 daily 2001 - 2015). DOY: Day of the year. The weighted histogram is shown on the right section, where there is a clear prevalence of GPPmax at mid-year because of the highest land extension in the Northern Hemisphere.

In figure 1 a, we present the geographical distribution of the phenology of GPPmax. In the right section, we present the weighted histogram of the timing, where five GPPmax peaks can be recognized. (1) The first peak occurs between February and April and correspond to the Maximum GPP in the Northern section of the Australian territory, South Africa, and Madagascar, South America, excluding the tropical rain forest, and the patagonia, and the Andean mountain range. (2) The second peak occurs between May and June and represent the areas of the Mediterranean region, The northern coast of Eupore, Central Africa, and East, and west coast in the North America. (3) The main peak occurs between June and August and correspond to the northern temperate zone including most of Europe, Russia, Japan, North America, the northern part of the Andes and the transition zone between the Colombian and Venezuelan savannas and tropical rain forest, and central Australia. (4) The four

115 peak occurs between August and November and included the territories of Central America, North of Africa, India, part of
 China, and the South part of Australia. (5) The last peak of the year occurs between November and January and correspond to
 regions in the Amazon and African tropical rain forest, Southern temperate region of South America, and New Zealand. The
 percentage of agreement of the phenology of GPPmax (Figure 1, b) shows that certain geographic patters are generated by
 the high variability of the DOY GPPmax, for instance regions with high standard deviation on North America, partially match
 120 with the change of DOY GPPmax from May to September. The same occurs in other regions of the world, for example in the
 Himalayas, the northern andean region, central australia and Indonesia where the high standard deviation partially match the
 transition zones on DOY.

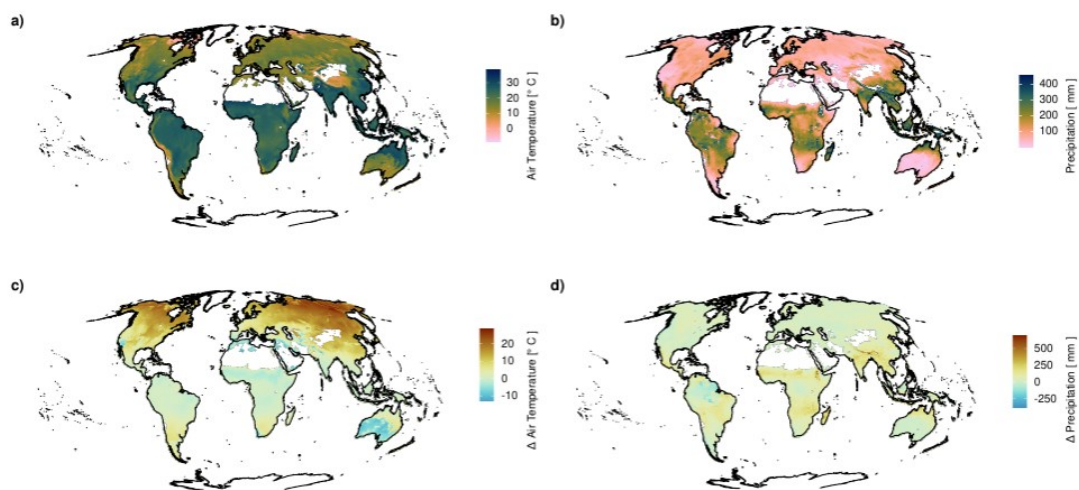


Figure 2. Optimum climate space (30 days before) when GPP is maximum. a) Average air temperature, b) cumulative precipitation (for visualization purpose, values higher than the quantile 0.99 are encoded as NAs). The differences between the optimum value and the average climate condition are represented in the bottom section (c, d). For precipitation, the cumulative precipitation is estimated monthly, and then the mean climatological year is estimated.

In figure 2 we present the optimum climate space 30 days before GPPmax is reached. The air temperature (figure 2, a) varied between -9 and 39 °C where most of the values are between 0 and 30 °C (Figure A1). The mean seasonal cumulative
 125 precipitation (figure 2, b) varied between 0 (mm) and 1678 (mm) where most of the values vary between 0 mm and 500 mm
 A2. When the values 30 days before are compared with average temperature (figure 2, c) we find that for ~78.8% (~90.9
 millions km²) of the total global surface with terrestrial vegetation cover GPPmax occurs when Δ air temperature is positive,
 while for ~21% (~24.5 millions km²) occurs when Δ air temperature is negative, figure 2, c). The negative Δ values are
 located in southern part of Australia, part of India, the tropical rainforest in Africa, part of Brazil, northern Andean region,
 130 Mediterranean region in Europe, and part of the west and east coast in the United States.

When the cumulative precipitation 30 days before GPPmax occurs is compared with the average monthly cumulative precipitation (figure 2, d), we find that for $\sim 85.4\%$ (~ 98.5 millions km^2) of the total global vegetated area, GPPmax occurs when Δ cumulative precipitation is positive. Conversely, for $\sim 14.6\%$ (~ 16.9 millions km^2) of vegetated surface occurs when Δ cumulative precipitation is negative (figure 2, d). The negative Δ values are located in part of the Patagonia (South America), a spot in the tropical rainforest in South America, part of the west and east coast in North America. The northern coast of Europe from Spain to Norway. Northern part of Japan, Part of the Korean peninsula, Iceland, Part of Malaysia and Indonesia, and central Australia.

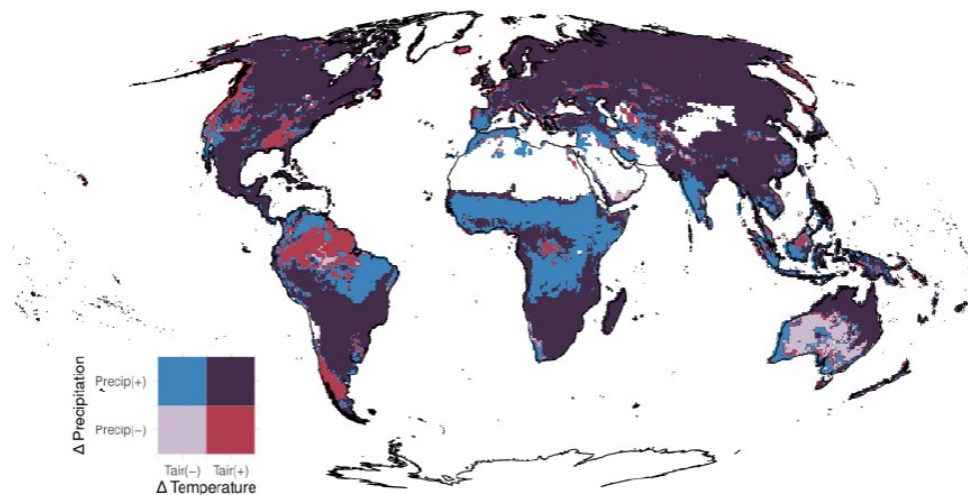


Figure 3. Graphical representation of the relationship between Δ air temperature and Δ precipitation when Gross Primary Production (GPP) is maximum. Four functional classes (FCs) are shown based on sign of the Δ value.

In figure 3 we present four functional classes (FCs) based on the relationship of air temperature and precipitation when GPPmax occurs with the climatological average (figure 2, c, d). In general, we can interpret the sign of Δ as a limiting factor in terms of CO_2 uptake. In this sense, if the sign of the variables is negative, we can say that the variable is not limiting GPPmax. In other words, GPPmax is not occurring when air temperature or monthly cumulative precipitation are maximum. On the other hand, if air temperature or precipitation are positive, we can say that the vegetation for a certain region is mainly driven by temperature or precipitation given that when GPPmax occurs, temperatures or cumulative precipitation for the area are above the average climate regime limiting GPPmax.

The main FC by geographical extension is "Tair (+) Precip (+)" with a $\sim 68\%$ (~ 78.4 millions km^2) of the total global surface with terrestrial vegetation cover. It is distributed across the globe and in all the continents, where most of the vegetation is currently limited by these two variables. The second FC is "Tair (+) Precip (-)" $\sim 17.4\%$ (~ 20.1 millions km^2). In these

areas, GPPmax is limited by temperature but not by precipitation. It is distributed through the east and west coast of North America, where the distribution of temperate coniferous forests explains the pattern. North coast of Europe from Spain to Norway, The Korean peninsula. The northern part of Japan, where there is a transition from humid subtropical climate (cfa) to humid continental climate (Dfa, Köppen-Geiger class). The east coast of Malaysia. Part of the tropical rainforest in the Amazon and partially southern Andean mountains.

The third FC is "Tair (-) Precip (+)" ~10.8% (~12.5 millions km²), and correspond to ecosystems that are mainly limited by precipitation. In this FC, we find the Mediterranean region in Europe that corresponds to Csa (Hot-summer Mediterranean climate) Köppen-Geiger class. Both zones share the same location and extension. We also find large portions of desert and grasslands in the Middle East. Sections of the Andes mountain range, the savannas in Venezuela and Colombia, that limit with the tropical rainforest ("Tair (+) Precip (-)"), part of the Cerrado in Brazil, part of Uruguay that correspond to grassland and croplands. Also, the African savannas, and the Mediterranean region in South Africa. Part of deserts in India that correspond to shrublands, and tropical dry broadleaf forest; the southern Mediterranean coast of Australia; most of the Sumatra island in Indonesia that correspond to palm-oil plantations. The less common FC is "Tair (-) Precip (-)" ~3.8% (~4.4 millions km²), and occurs mainly in Australia in regions that correspond to desert and shrublands, followed by some small patches in the Amazonia, northern Andean, United States and Turkey. These regions correspond to a high standard deviation of the DOY GPPmax estimation (figure 1, b) and can be the result of temporal anomalies in the prediction of GPP.

In figure 4 we present the geographical overlapping between our four Functional Classes, the terrestrial biomes proposed by (Olson et al., 2001), and the Köppen-Geiger climate classification. The FC Tair (+) Precip (+) appears in all biomes, being the most common drivers for most of the ecosystems at global scale. The FC Tair (+) Precip (-) is also distributed across all the biomes, 26.3% of the temperate Conifer forest are in this category, as 23.4% of the Tropical and subtropical dry broadleaf forest, 15.2% of the temperate broadleaf mixed forest, 14.4% of the mediterranean forest, 13.2% of the mangroves, and 12.4% of the Tundra. For the rest of biomes, between 9.9% to 0.4% are included in this category. For the third FC Tair (-) Precip (+) We find that 60.7% of the Mediterranean forests are included in this category as well as 49% of tropical and subtropical, grasslands, savannas, and shrublands, 30% of the Flooded grasslands and savannas, 23.4% of the tropical and subtropical moist broadleaf forests, 20.1% of the Desert and Xeric Shrublands, 16.4% of the Tropical and Subtropical Moist broadleaf Forests, 15% of the mangroves, and between 9% and 0.8% for the other biomes excluding the Boreal forest/Taiga, and the Tundra. For the last category, we find that 13% of the desert and xeric shrublands are in this category, 11.4% of the mediterranean forests, woodlands and scrub, and between 7% and 0.1% of the other biomes excluding the tundra, Flooded grasslands and savannas, Boreal forest/taiga, and tropical and subtropical coniferous forest.

Regarding the Köppen-Geiger climate classification, we find that the most common FC is "Tair (+) Precip (+)" present in all the classes with the exception of Tundra. The second class Tair (+) Precip (-) is the total extension of the ice cap climate (EF) those are patches on the arctic polar circle with an extension of 3859 km², 11.2% of the Tundra climate (ET); 46.4% of the Tropical rainforest climate (Af), 26% of the Tropical monsoon climate (Am); 27.4% of the Temperate dry summer climate (Cs), 20.6% of the Temperate no dry season climate (Cf); 20.5% of the Continental dry summer climate (Ds), and between 6 to 1% from the other climate classes. The third category Tair (-) Precip (+) represents 48.1% of the Tropical savanna dry-winter

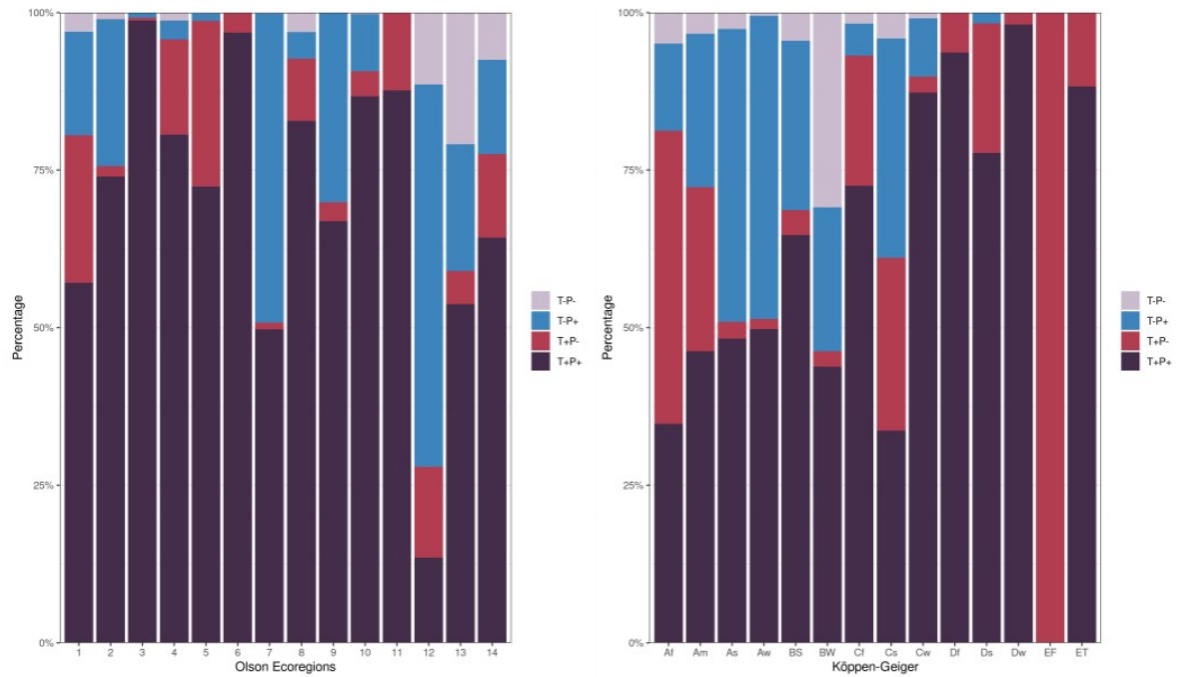


Figure 4. Geographical co-occurrences of the four functional classes with the biomes (left) and the Köppen-Geiger climate classification (right). T+P+ (Tair + Precip +), T+P- (Tair + Precip -), T-P+ (Tair - Precip +), T-P- (Tair - Precip -). Olson biomes: 1: Tropical and Subtropical Moist Broadleaf Forests; 2: Tropical and Subtropical Dry Broadleaf Forests; 3: Tropical and Subtropical Coniferous Forests; 4: Temperate Broadleaf and Mixed Forests; 5: Temperate Conifer Forests; 6: Boreal Forests/Taiga; 7: Tropical and Subtropical Grasslands, Savannas and Shrublands; 8: Temperate Grasslands, Savannas, and Shrublands; 9: Flooded Grasslands and Savannas; 10: Montane Grasslands and Shrublands; 11: Tundra; 12: Mediterranean Forests, Woodlands, and Scrub; 13: Deserts and Xeric Shrublands; 14: Mangroves, Köppen-Geiger climate classification: Af: Tropical rainforest climate; Am: Tropical monsoon climate, BS: Arid steppe climate; BW: Arid desert climate, Cf: Temperate no dry season climate; Cs: Temperate dry summer climate; Cw: Temperate dry winter climate; Df: Continental no dry season climate; Ds: Continental dry summer climate; Dw: Continental dry winter climate; EF: Ice cap climate; ET: Tundra climate.

Table 1. Comparison between the different classification schemes based on climate conditions (Köppen-Geiger), biomes (Olson), and functional response to reach GPPmax (FCs). Two metrics are estimated, to adjust Rand Index and the Normalized Mutual Information, for both metrics vary between 0 and 1 where 0 means the classification system doesn't share any similarity and 1 both classification system are identical.

Comparison	Adjust Rand Index	Normalized Mutual Information
Köppen-Geiger - Olson	0.30	0.42
Köppen-Geiger - FCs	0.11	0.15
Olson - FCs	0.05	0.13

climate (Aw), 46.5% of the Tropical savanna dry-summer climate (As), 13.8% of the Af climate, 34.8% of Temperate dry summer climate, 26.8% of the Arid steppe climate (BS), and 22.7% of Arid desert climate (BW). For the other climate classes, 185 between 9% to 0.05% are represented by this category, excluding ice cap climate and Continental dry winter climate (Dw). The last FC is Tair (-) Precip (-), represents 30% of the BW climate, and between 4.9% to 0.05% of the climate classes excluding Ice cap, and continental climate classes (D).

When we analyze the percentage of agreement between our Functional classes (FCs), the biomes from Olson et al. (2001) and the Köppen-Geiger climate classification 1. We find that, as reference value, there is between a 30% to 42% of agreement 190 between Köppen-Geiger and ecoregion classes, while between our FCs and Köppen-Geiger there is between 11% to 15% percent of agreement, and between 0.5% and 13% when our classes are compared with biomes depending on the metric considered.

4 Discussion

4.1 GPPmax phenology

195 The timing of physiological processes is a relevant metric to understand how ecosystems are changing because of global warming or local transformations (Körner and Basler, 2010; Buitenwerf et al., 2015). Most phenological studies focus on the timing of physiological processes of individual species, but we can also use the timing of ecological processes as gross primary production (GPPmax) to understand the effect of climatic conditions before it reached. Here we present a first approach to the timing of GPPmax at global scale and show how to use it as based for the study of the climatic space that contribute to 200 GPPmax. The geographical distribution of the timing of GPPmax is explained by a combination of the latitudinal gradient of the earth, the local climate regime and the vegetation response. Given that most of the land cover by vegetation is located in the Northern Hemisphere, it is expected that the maximum photosynthetic capacity for these ecosystems will be reached at the middle of the year. Nevertheless, it is well-known that in the Mediterranean region GPP is limited by the water availability during summer, then peaks are expected after the raining seasons (Serrano-Ortiz et al., 2009) as is evident in the Mediterranean 205 region in Europe and the west coast of North America. The combination of the latitudinal and climate effect is also evidence

in North Africa where in the African savannas GPPmax is reached before then the African tropical rainforest, and the gradient continue, being GPPmax reached at the end of the year during the austral summer for the southern part. This gradient can also be observed through the Andes mountain range from North to South. Although in our study the phenology of the GPPmax was only used as a proxy to extract the climate space, a question that remains open is how the climate conditions affect the timing of GPPmax. For example, if temperature increase, the GPPmax is going to occurs earlier or later than expected? Solving this question can give us an idea of how ecosystems are going to respond in terms of maximum productivity to climate change.

4.2 Delta climate space

The estimation of the delta climate space give us a description of how the optimum conditions when GPPmax relates to the mean annual values. In this study, we only consider 2 variables, temperature and precipitation, nevertheless, the inclusion of radiation can give us a better description of the drivers of GPPmax, and can also help us to disentangle areas that in our study are presented as homogeneous at high latitudes.

4.3 Functional classes

Generating new classification systems that consider optimum ecosystem processes can help us to better understand the complex dynamics of the earth system. The low percentage of agreement between our functional classes, the ones from the Olson's biomes and the Köppen-Geiger classes suggest that our classification encoded a functional relationship between ecosystem and climate that was considered on previous classification schemes. It is important to point out that GPP from FLUXCOM is a machine learning product and that the uncertainties from the product are extrapolated to our classification system, in this sense improve our predictions about ecosystem processes will be a necessary step to generate more robust classification systems. The new satellite missions from the Copernicus program can give us a more complete picture of the earth system dynamics, the new red-edge bands available on Sentinel-2 and the radar information from Sentinel-1 can help us to increase the power prediction of the new generation of GPP products.

5 Conclusions

In this study, we explore how synthetic classifications can be done based on optimum ecosystem processes as maximum GPP. Although previous classification systems are based on species distributions as Olson et al. (2001) biomes, and the distribution of different climate regimens as Köppen-Geiger. Considering optimum plant photosynthesis can give us a new perspective of how plants respond to the environment and can contribute to simplify global vegetation models. Nevertheless, the robustness of these classifications systems will be limited by our capacity to up-scale ecosystem processes as net ecosystem exchange or gross primary production. As optimum ecosystem processes occur in a specific time, strengthen our knowledge of the timing of optimum ecosystem processes will open new a window to understand ecosystem dynamics not only in magnitude but also in time.

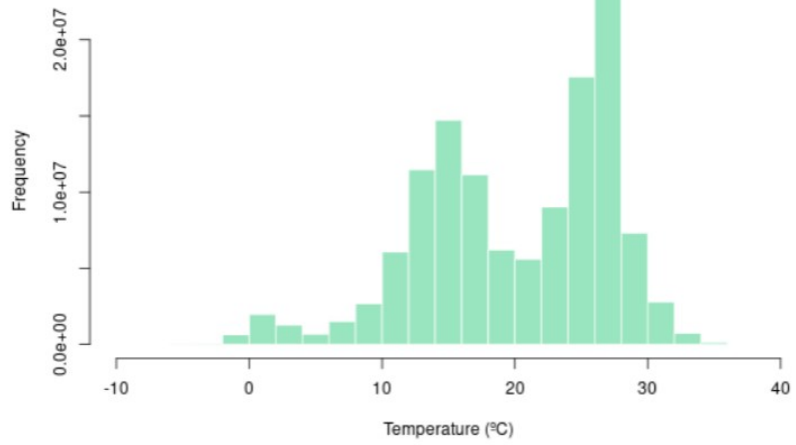


Figure A1. Distribution of the mean seasonal cycle of air temperature 30 days before GPPmax is reached

Appendix A

A1

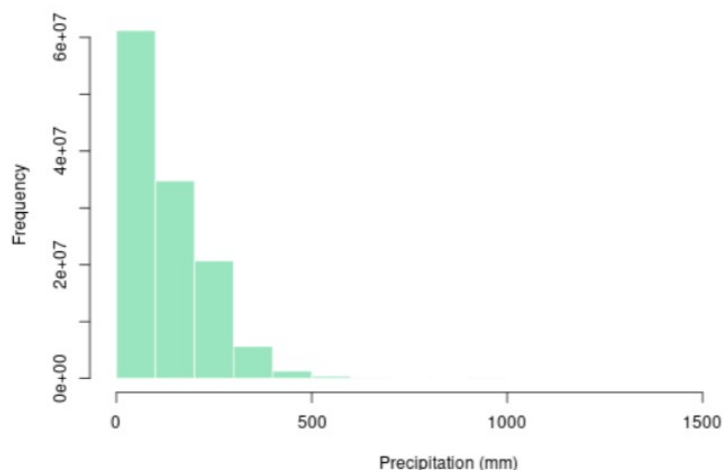


Figure A2. Distribution of the mean seasonal cycle of cumulative precipitation 30 days before GPPmax is reached

References

- Belward, A. S. and Skøien, J. O.: Who Launched What, When and Why; Trends in Global Land-Cover Observation Capacity from Civilian Earth Observation Satellites, *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 115–128, <https://doi.org/10.1016/j.isprsjprs.2014.03.009>, 2015.
- Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., and Reichstein, M.: Upscaled Diurnal Cycles of Land–Atmosphere Fluxes: A New Global Half-Hourly Data Product, *Earth System Science Data*, 10, 1327–1365, <https://doi.org/10.5194/essd-10-1327-2018>, 2018.
- Box, E. E.: *Macroclimate and Plant Forms: An Introduction to Predictive Modeling in Phytogeography*, vol. 1, Springer Science & Business Media, 1981.
- Box, E. O.: Factors Determining Distributions of Tree Species and Plant Functional Types, *Vegetatio*, 121, 101–116, <https://doi.org/10.1007/BF00044676>, 1995.
- Buitenwerf, R., Rose, L., and Higgins, S. I.: Three Decades of Multi-Dimensional Change in Global Leaf Phenology, *Nature Climate Change*, 5, 364–368, <https://doi.org/10.1038/nclimate2533>, 2015.
- Danon, L., Díaz-Guilera, A., Duch, J., and Arenas, A.: Comparing Community Structure Identification, *Journal of Statistical Mechanics: Theory and Experiment*, 2005, P09 008–P09 008, <https://doi.org/10.1088/1742-5468/2005/09/P09008>, 2005.
- Fang, H., Baret, F., Plummer, S., and Schaepman-Strub, G.: An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications, *Reviews of Geophysics*, 57, 739–799, <https://doi.org/10.1029/2018RG000608>, 2019.

- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A.,
 255 Cooper, A., Baccini, A., Gao, F., and Schaaf, C.: Global Land Cover Mapping from MODIS: Algorithms and Early Results, *Remote Sensing of Environment*, 83, 287–302, [https://doi.org/10.1016/S0034-4257\(02\)00078-0](https://doi.org/10.1016/S0034-4257(02)00078-0), 2002.
- Friedl, Mark and Sulla-Menashe, Damien: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006, Tech. rep., NASA EOSDIS Land Processes DAAC, <https://doi.org/10.5067/MODIS/MCD12Q1.006>, 2019.
- Gillison, A.: Towards a Functional Vegetation Classification, *Vegetation classification in Australia*, pp. 30–41, 1981.
- 260 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049,
 265 <https://doi.org/10.1002/qj.3803>, 2020.
- Hubert, L. and Arabie, P.: Comparing Partitions, *Journal of Classification*, 2, 193–218, <https://doi.org/10.1007/BF01908075>, 1985.
- Humboldt, A.: *Ideen Zu Einer Physiognomik Der Gewächse*. Tübingen, Neudruck: Ostwalds Klassiker der exakten Wissenschaften, 247, 1959, 1806.
- Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The
 270 FLUXCOM Ensemble of Global Land-Atmosphere Energy Fluxes, *Scientific Data*, 6, 1–14, <https://doi.org/10.1038/s41597-019-0076-8>, 2019.
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O’Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U.,
 275 and Reichstein, M.: Scaling Carbon Fluxes from Eddy Covariance Sites to Globe: Synthesis and Evaluation of the FLUXCOM Approach, *Biogeosciences*, 17, 1343–1365, <https://doi.org/10.5194/bg-17-1343-2020>, 2020.
- Köppen, W.: Die Wärmezonen Der Erde, Nach Der Dauer Der Heissen, Gemässigten Und Kalten Zeit Und Nach Der Wirkung Der Wärme Auf Die Organische Welt Betrachtet, *Meteorologische Zeitschrift*, 1, 5–226, 1884.
- Körner, C. and Basler, D.: Phenology Under Global Warming, *Science*, 327, 1461–1462, <https://doi.org/10.1126/science.1186473>, 2010.
- 280 Lavorel, S., Díaz, S., Cornelissen, J. H. C., Garnier, E., Harrison, S. P., McIntyre, S., Pausas, J. G., Pérez-Harguindeguy, N., Roumet, C., and Urcelay, C.: Plant Functional Types: Are We Getting Any Closer to the Holy Grail?, in: *Terrestrial Ecosystems in a Changing World*, edited by Canadell, J. G., Pataki, D. E., and Pitelka, L. F., pp. 149–164, Springer Berlin Heidelberg, Berlin, Heidelberg, <https://doi.org/10.1007/978-3-540-32730-113>, 2007.
- Malaterre, C., Dussault, A. C., Mermans, E., Barker, G., Beisner, B. E., Bouchard, F., Desjardins, E., Handa, I. T., Kembel, S. W., Lajoie, G., Maris, V., Munson, A. D., Odenbaugh, J., Poisot, T., Shapiro, B. J., and Suttle, C. A.: Functional Diversity: An Epistemic Roadmap, *BioScience*, 69, 800–811, <https://doi.org/10.1093/biosci/biz089>, 2019.
- Mardia, K. V.: *Statistics of Directional Data*, Probability and Mathematical Statistics, Academic Press, London, New York, 1972.
- Markham, B. L. and Helder, D. L.: Forty-Year Calibrated Record of Earth-Reflected Radiance from Landsat: A Review, *Remote Sensing of Environment*, 122, 30–40, <https://doi.org/10.1016/j.rse.2011.06.026>, 2012.
- 290 Migliavacca, M., Musavi, T., Mahecha, M. D., Nelson, J. A., Knauer, J., Baldocchi, D. D., Perez-Priego, O., Christiansen, R., Peters, J., Anderson, K., Bahn, M., Black, T. A., Blanken, P. D., Bonal, D., Buchmann, N., Caldararu, S., Carrara, A., Carvalhais, N., Cescatti, A.,

- Chen, J., Cleverly, J., Cremonese, E., Desai, A. R., El-Madany, T. S., Farella, M. M., Fernández-Martínez, M., Filippa, G., Forkel, M., Galvagno, M., Gomasasca, U., Gough, C. M., Göckede, M., Ibrom, A., Ikawa, H., Janssens, I. A., Jung, M., Kattge, J., Keenan, T. F., Knohl, A., Kobayashi, H., Kraemer, G., Law, B. E., Liddell, M. J., Ma, X., Mammarella, I., Martini, D., Macfarlane, C., Matteucci, G., Montagnani, L., Pabon-Moreno, D. E., Panigada, C., Papale, D., Pendall, E., Penuelas, J., Phillips, R. P., Reich, P. B., Rossini, M., Rotenberg, E., Scott, R. L., Stahl, C., Weber, U., Wohlfahrt, G., Wolf, S., Wright, I. J., Yakir, D., Zaehle, S., and Reichstein, M.: The Three Major Axes of Terrestrial Ecosystem Function, *Nature*, 598, 468–472, <https://doi.org/10.1038/s41586-021-03939-9>, 2021.
- 295 Musavi, T., Migliavacca, M., van de Weg, M. J., Kattge, J., Wohlfahrt, G., van Bodegom, P. M., Reichstein, M., Bahn, M., Carrara, A., Domingues, T. F., Gavazzi, M., Gianelle, D., Gimeno, C., Granier, A., Gruening, C., Havránková, K., Herbst, M., Hrynkiv, C., Kalhori, A., Kaminski, T., Klumpp, K., Kolari, P., Longdoz, B., Minerbi, S., Montagnani, L., Moors, E., Oechel, W. C., Reich, P. B., Rohatyn, S., Rossi, A., Rotenberg, E., Varlagin, A., Wilkinson, M., Wirth, C., and Mahecha, M. D.: Potential and Limitations of Inferring Ecosystem Photosynthetic Capacity from Leaf Functional Traits, *Ecology and Evolution*, 6, 7352–7366, <https://doi.org/10.1002/ece3.2479>, 2016.
- 300 Netzel, P. and Stepinski, T.: On Using a Clustering Approach for Global Climate Classification, *Journal of Climate*, 29, 3387–3401, <https://doi.org/10.1175/JCLI-D-15-0640.1>, 2016.
- 305 Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D’amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial Ecoregions of the World: A New Map of Life on Earth: A New Global Map of Terrestrial Ecoregions Provides an Innovative Tool for Conserving Biodiversity, *BioScience*, 51, 933–938, [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2), 2001.
- 310 Pavlick, R., Drewry, D. T., Bohn, K., Reu, B., and Kleidon, A.: The Jena Diversity-Dynamic Global Vegetation Model (JeDi-DGVM): A Diverse Approach to Representing Terrestrial Biogeography and Biogeochemistry Based on Plant Functional Trade-Offs, *Biogeosciences*, 10, 4137–4177, <https://doi.org/10.5194/bg-10-4137-2013>, 2013.
- Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D.: Linking Plant and Ecosystem Functional Biogeography, *Proceedings of the National Academy of Sciences*, 111, 13 697–13 702, <https://doi.org/10.1073/pnas.1216065111>, 2014.
- 315 Serrano-Ortiz, P., Domingo, F., Cazorla, A., Were, A., Cuezva, S., Villagarcía, L., Alados-Arboledas, L., and Kowalski, A. S.: Interannual CO₂ Exchange of a Sparse Mediterranean Shrubland on a Carbonaceous Substrate, *Journal of Geophysical Research: Biogeosciences*, 114, <https://doi.org/10.1029/2009JG000983>, 2009.
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms, *Biogeosciences*, 13, 4291–4313, <https://doi.org/10.5194/bg-13-4291-2016>, 2016.
- 320 Wullschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., Kattge, J., Norby, R. J., van Bodegom, P. M., and Xu, X.: Plant Functional Types in Earth System Models: Past Experiences and Future Directions for Application of Dynamic Vegetation Models in High-Latitude Ecosystems, *Annals of Botany*, 114, 1–16, <https://doi.org/10.1093/aob/mcu077>, 2014.
- Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., Ramoino, F., and Arino, O.: ESA WorldCover 10 m 2020 V100, <https://doi.org/10.5281/zenodo.5571936>, 2021.
- 325 Zscheischler, J., Mahecha, M. D., and Harmeling, S.: Climate Classifications: The Value of Unsupervised Clustering, *Procedia Computer Science*, 9, 897–906, <https://doi.org/10.1016/j.procs.2012.04.096>, 2012.

Ecosystem physio-phenology revealed using circular statistics

Manuscript Nr. 2

Title of the Manuscript: Ecosystem physio-phenology revealed using circular statistics

Authors: Pabon-Moreno, Daniel E., Migliavacca, Mirco., Reichstein, Markus., Römermann, Christine., Mahecha Miguel D.

Bibliographic information: Pabon-Moreno, D. E., Musavi, T., Migliavacca, M., Reichstein, M., Römermann, C., & Mahecha, M. D. (2020). Ecosystem physio-phenology revealed using circular statistics. *Biogeosciences*, 17(15), 39914006. <https://doi.org/10.5194/17-3991-2020>

17-3991-2020

The candidate is: First author

Status: Published

Authors' contribution in percentage

Author	Conceptual	Data analysis	Experimental (Numerical)	Writing the manuscript	Provision of material
Daniel E. Pabon-Moreno	37	67	94	60	0
Dr. Talie Musavi	7	7	0	18	0
Dr. Mirco Migliavacca	7	8	3	8	50
Prof. Dr. Markus Reichstein	4	5	0	3	50
Prof. Dr. Christine Römermann	5	3	0	3	0
Prof. Dr. Miguel D. Mahecha	40	10	3	8	0
Total:	100	100	100	100	100

Biogeosciences, 17, 3991–4006, 2020
<https://doi.org/10.5194/bg-17-3991-2020>

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



Ecosystem physio-phenology revealed using circular statistics

Daniel E. Pabon-Moreno¹, Talie Musavi¹, Mirco Migliavacca¹, Markus Reichstein^{1,2}, Christine Römermann^{2,3}, and Miguel D. Mahecha^{1,2,4}

¹Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany

²German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany

³Friedrich Schiller University, Institute of Ecology and Evolution, Philosophenweg 16, 07743 Jena, Germany

⁴Remote Sensing Center For Earth System Research, Leipzig University, 04103 Leipzig, Germany

Correspondence: Daniel E. Pabon-Moreno (dpabon@bgc-jena.mpg.de)

Received: 3 October 2019 – Discussion started: 15 October 2019

Revised: 12 June 2020 – Accepted: 26 June 2020 – Published: 6 August 2020

Abstract. Quantifying how vegetation phenology responds to climate variability is a key prerequisite to predicting how ecosystem dynamics will shift with climate change. So far, many studies have focused on responses of classical phenological events (e.g., budburst or flowering) to climatic variability for individual species. Comparatively little is known on the dynamics of physio-phenological events such as the timing of maximum gross primary production ($\text{DOY}_{\text{GPPmax}}$), i.e., quantities that are relevant for understanding terrestrial carbon cycle responses to climate variability and change. In this study, we aim to understand how $\text{DOY}_{\text{GPPmax}}$ depends on climate drivers across 52 eddy covariance (EC) sites in the FLUXNET network for different regions of the world. Most phenological studies rely on linear methods that cannot be generalized across both hemispheres and therefore do not allow for deriving general rules that can be applied for future predictions. One solution could be a new class of circular–linear (here called circular) regression approaches. Circular regression allows circular variables (in our case phenological events) to be related to linear predictor variables as climate conditions. As a proof of concept, we compare the performance of linear and circular regression to recover original coefficients of a predefined circular model for artificial data. We then quantify the sensitivity of $\text{DOY}_{\text{GPPmax}}$ across FLUXNET sites to air temperature, shortwave incoming radiation, precipitation, and vapor pressure deficit. Finally, we evaluate the predictive power of the circular regression model for different vegetation types. Our results show that the joint effects of radiation, temperature, and vapor pressure deficit are the most relevant controlling factor of $\text{DOY}_{\text{GPPmax}}$ across sites. Woody savannas are an exception, where the most im-

portant factor is precipitation. Although the sensitivity of the $\text{DOY}_{\text{GPPmax}}$ to climate drivers is site-specific, it is possible to generalize the circular regression models across specific vegetation types. From a methodological point of view, our results reveal that circular regression is a robust alternative to conventional phenological analytic frameworks. The analysis of phenological events at the global scale can benefit from the use of circular statistics. Such an approach yields substantially more robust results for analyzing phenological dynamics in regions characterized by two growing seasons per year or when the phenological event under scrutiny occurs between 2 years (i.e., $\text{DOY}_{\text{GPPmax}}$ in the Southern Hemisphere).

1 Introduction

Phenology is the study of the timing of biological events that can be observed at either the organismic level or the ecosystem scale (Lieth, 1974). For the latter, phenology is the study of some integral behavior across phenological states of the integrated canopy reflectance captured by remote sensing (Richardson et al., 2009; Zhang et al., 2003) or vegetation-driven ecosystem–atmosphere CO_2 exchange fluxes (Richardson et al., 2010). Ecosystem-scale physio-phenological processes of this kind are relevant quantities in global biogeochemical cycles and integrate both the seasonal dynamics of biophysical states (e.g., reflected in the canopy development) and the observed photosynthesis at the stand level (i.e., gross primary production). Here we are particularly interested in the timing when ecosystems reach their

maximum CO₂ uptake within a growing season. Ecosystem physio-phenology is influenced by climate conditions but simultaneously contributes to the regulation of different micro- and macrometeorological patterns. Physio-phenological cycles determine the temporal dynamics of land–atmosphere water and energy exchange fluxes. Likewise, the terrestrial carbon cycle is affected by phenological controls on CO₂ uptake and release (Peñuelas et al., 2009).

The eddy covariance (EC) technique allows us to continuously measure the exchange of energy and matter between ecosystems and the atmosphere (Aubinet et al., 2012). The FLUXNET network collects EC data for most ecosystems of the world along with other meteorological variables, i.e., radiation, temperature, precipitation, atmospheric humidity, and often soil moisture (Baldocchi et al., 2001; Baldocchi, 2020). Particularly relevant to phenological studies is the seasonal trajectory of gross primary production (GPP), which allows us to derive phenological transition dates such as start and end of the growing season (e.g., Luo et al., 2018) as well as the timing of the maximum gross primary production, hereafter as referred to as DOY_{GPPmax} (Zhou et al., 2016; Peichl et al., 2018; Wang and Wu, 2019).

In this study we focus on understanding how climate variability affects the time when ecosystems reach their maximum potential for CO₂ absorption. In order to reach this “optimum state”, several preconditions must be met during the preceding part of the growing season. So far several studies have focused on studying the variability of maximum GPP during the growing season (GPPmax). For instance, Zhou et al. (2017) studied how the variability of annual GPP is influenced by GPPmax and the start and the end of the growing season. The authors found that GPPmax is a better explanatory parameter for the interannual variability of annual GPP than the start and end days of the growing season. Bauerle et al. (2012) studied how photoperiod and temperature influence plants’ photosynthetic capacity for 23 tree species in temperate deciduous hardwoods, reporting that the photoperiod explains the variability of photosynthetic capacity better than temperature. So far, to the best of our knowledge, only one study has focused on understanding the temporal variability of GPPmax; Wang and Wu (2019) used a combination of satellite remote-sensing and eddy covariance data to explore how DOY_{GPPmax} is controlled by climatic conditions. The authors reported that higher temperatures advance DOY_{GPPmax}, while the influence of precipitation and radiation were biome-dependent. This study had a geographical focus on China; a global approach considering several ecosystems across the whole latitudinal gradient is still lacking.

The challenge of understanding phenology is generally to characterize a discrete event that repeats periodically. Classically, phenological analyses have been performed using linear regression models (Morente-López et al., 2018; Zhou et al., 2016). Most of these studies analyze ecosystems characterized by one growing season (e.g., temperate or boreal

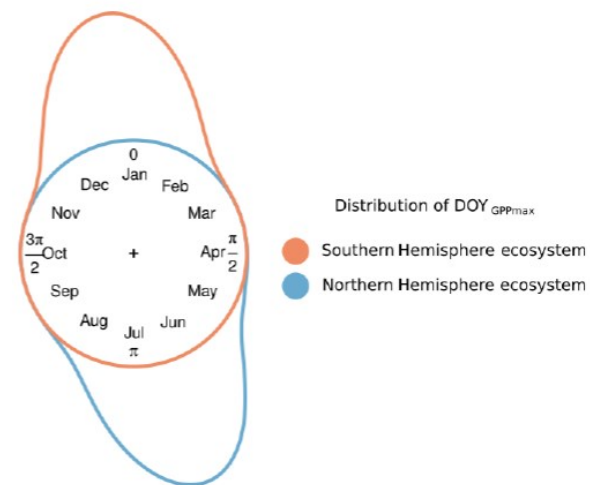


Figure 1. Conceptual distribution of GPPmax timing (DOY_{GPPmax}) for two hypothetical ecosystems: one in the Northern (blue) and one in the Southern Hemisphere (red). The distance between the color line and the circle represents the frequency of the DOY_{GPPmax} observations. The distance between the end and the beginning of the distribution represents the DOY_{GPPmax} interannual variability.

forests) and when the summer is centered around the middle of the calendar year. The existing methods are, however, not sufficiently generic to describe (i) ecosystems in the Southern Hemisphere and (ii) ecosystems with multiple growing seasons per year as is often observed in, for example, semiarid regions.

Figure 1 illustrates the problem of northern vs. southern hemispheric summers from a conceptual point of view, assuming that some discrete event recurs annually, but the timing varies according to some external drivers. We would then need to find a predictive model explaining the interannual variability of phenology, i.e., the probability of this recurrent event in the course of the annual cycle. Figure 1 shows that linear regression models would be inappropriate to predict the day of the year (DOY) of some phenological event in the Southern Hemisphere as the actual target values to predict may alternate between $\gtrsim \frac{3\pi}{2}$ and $\lesssim \frac{\pi}{2}$.

In recent years, circular statistics have gained some attention as they offer a solution to problems of this kind (Morelato et al., 2010; Beyene et al., 2018). Unlike classical statistics, the predicted variables are expressed in terms of angular directions (degrees or radians) across a circumference (Fisher, 1995), allowing us to perform statistical analysis where the data space is not Euclidean. In this framework, point events can be described as a von Mises distribution (Von Mises, 1918), the equivalent to the normal distribution in circular statistics. The von Mises distribution is described by two parameters: the mean angular direction (μ) and the concentration parameter (κ). Circular–linear regressions (in

the following simply named circular regression) allow us to predict circular responses (e.g., the timing of phenological events) from other linear variables (Morellato et al., 2010). Given that any phenological event can be interpreted as an angular direction and should be modeled as such, we assume that these circular regressions are well suited in this context. Despite this evident suitability, circular statistics have not yet been extensively applied in the study of phenology and will therefore be presented here as an alternative to conventional linear techniques.

In this paper, we aim to identify the factors controlling the timing of the maximal seasonal GPP ($\text{DOY}_{\text{GPPmax}}$). The questions we want to answer are as follows: first, can circular statistics describe and predict $\text{DOY}_{\text{GPPmax}}$ per vegetation type? This aspect requires testing the methodological advantages and caveats of circular statistics across hemispheres in comparison with linear methods. Second, can $\text{DOY}_{\text{GPPmax}}$ be explained using cumulative climate conditions? This question needs to consider different possibilities for generating temporally integrating features. And third, how is $\text{DOY}_{\text{GPPmax}}$ affected by the climatic conditions during the growing season? The last question requires a global cross-site analysis. Based on the findings of these three questions, we then discuss the potential of circular regressions beyond this specific application case in related phenological problems and outline future applications.

2 Methods

2.1 Data

We use 52 EC sites (with at least 7 years of data) located throughout the latitudinal gradient of the globe from the FLUXNET2015 database (Table A1; <http://fluxnet.fluxdata.org/>, last access: 11 July 2019 Pastorello et al., 2020). Each FLUXNET site is identified with an abbreviation of the country and the name of the place, e.g., the EC tower AU-How, means that it is located in Howard Springs, Australia. From the dataset we use the GPP data that were derived using the nighttime partitioning method and considering the threshold of the variable u^* to discriminate values of insufficient turbulence (Reichstein et al., 2005). In order to identify maximum daily GPP, we compute the quantile 0.9 for each day based on the half-hourly flux observations. As potential explanatory variables for $\text{DOY}_{\text{GPPmax}}$ we use the daily air temperature (Tair), shortwave incoming radiation (SWin), precipitation (Precip), and vapor pressure deficit (VPD).

Given that the past climate conditions affect the CO_2 exchange between the atmosphere and the ecosystems (ecological memory; Liu et al., 2019; Ryan et al., 2015), we assume that an aggregated form of these climatic variables needs to be considered in the prediction of the phenological responses. We aggregate the original time series of the Tair, SWin, Precip, and VPD for each $\text{DOY}_{\text{GPPmax}}$ using a half-life

decay function (Eq. 1),

$$\langle \mathbf{x}_t \rangle = \frac{\sum_{i=0}^{\tau} x_{t-i} w_i}{\sum_{i=0}^{\tau} w_i}, \quad (1)$$

for estimating an exponentially weighted mean of the observation vector $\mathbf{x}_t = (x_t, x_{t-1}, \dots, x_{t-\tau})^T$ at time step t . The symbol $\langle \dots \rangle$ denotes the weighted average; i indicates the number of days before t , going back to $\tau = 365$ days. The weight decay is represented by

$$w_i = w_0 \exp\left(-i \frac{\ln(2)}{t_{1/2}}\right). \quad (2)$$

The decay function gives the instantaneous value a weight of 1 ($w_0 = 1$), and all preceding values receive an exponentially reduced weight as determined by the half-time parameter $t_{1/2}$. Finally, we make these variables comparable via centering standardization to unit variance. We perform a sensitivity analysis, evaluating the effect of the half-time parameter, and identify the optimum as the value when the variance explained by the circular regression model is at a maximum. The results are presented in Supplement 1.

Due to the high colinearity between the exponential weighted variables of Tair, SWin, and VPD, we perform a principal component analysis (PCA) on the matrix of variables and FLUXNET sites and retain the leading principal component of these variables as well as precipitation as input for the circular statistics model (Hastie et al., 2009). The results of the PCA are presented in Supplement 2.

2.2 Circular statistics

Since units of the circular response variable must be in radians or degrees, we transform the days of the year to radians using Eq. (3). For leap years we remove the last day.

$$\text{rad} = \text{DOY} \frac{360}{365} \frac{\pi}{180}, \quad (3)$$

where DOY means day of the year.

A basic circular regression model was proposed by Fisher and Lee (1992) as follows:

$$y = \mu + 2 \text{atan}(\beta_i x_i), \quad (4)$$

where y is the target variable (i.e., $\text{DOY}_{\text{GPPmax}}$) in radians, μ is the mean angular direction of the target variable, x_i represents the values for the variable i , and β_i is the regression coefficient. The parameters μ and β are fitted via the maximum likelihood method using the reweighted least squares algorithm as proposed by Green (1984).

Relevant interpretations of fitted circular regression models are (1) the sign of the β coefficients, (2) the statistical significance of the coefficients, and (3) the accuracy of the prediction. Regarding the first point, a negative sign of the coefficient would mean that an increasing value of the predictor would lead to an earlier $\text{DOY}_{\text{GPPmax}}$ compared to the

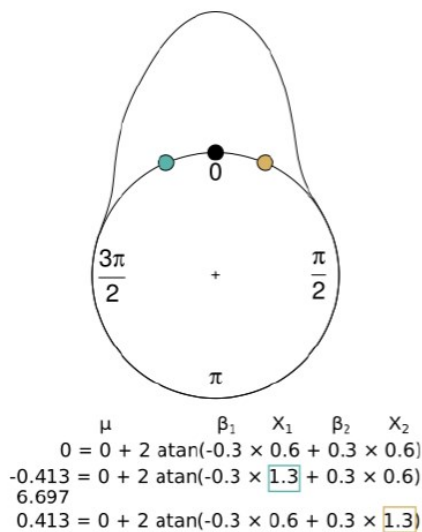


Figure 2. Interpretation of the coefficients in the circular regression considering a reference point (black) generated with a circular-linear model with mean angular direction ($\mu = 0$), two coefficients (β_1, β_2) and two variables (x_1, x_2), where one of the coefficients is negative (β_1), and the other one is positive (β_2). When the coefficient is negative and the value of the parameter increases (blue), the result is an earlier observation compared with the reference point (the equivalent of -0.413 rad is 6.697 rad, as is shown below the equation). On the other hand, when the coefficient is positive and the variable increases (yellow), the observation is later.

mean angular direction. The inverse would happen when the coefficient is positive. Figure 2 conceptually illustrates how the coefficients affect the predictions. Regarding the second aspect, we can state that if a coefficient is not significant, then its contribution would not be relevant to explaining the phenological observation. In our case we define the coefficient to be significant if the median of the distribution of p values is less than 0.05. Finally, we can estimate the accuracy of the prediction using the Jammalamadaka–Sarma (JS) correlation coefficient (Jammalamadaka and Sarma, 1988). As in any other regression framework, this approach helps us to quantify the effect of each climate variable on the interannual variability of $\text{DOY}_{\text{GPPmax}}$.

To estimate the relative sensitivity of $\text{DOY}_{\text{GPPmax}}$ to the leading principal component representing Tair, SWin, and VPD as well as to Precip, we use the implementation of Eq. (4) in the R package “circular” (Agostinelli and Lund, 2017). To increase the robustness of the method we implemented a block bootstrapping per growing season, generating a model parameter average based on 1000 iterations. In each analysis, we estimate the accuracy of the model using the JS correlation coefficient.

2.3 Circular vs. linear regression

To assess the performance of linear versus circular regressions we perform an experiment with simulated data in which we evaluate the accuracy and precision of both approaches to recover original regression coefficients in a circular setting (Eq. 4). We add noise generated with a random von Mises distribution with the parameters $n = 100$ and $\kappa = 30$ to the model to ensure that the result follows a normal distribution. We predefined a range of values for two regression coefficients ($\beta_1 = \langle 0.01, \dots, 3 \rangle$, $\beta_2 = \langle 0.01, \dots, 3 \rangle$). We simulate the variables x_1 and x_2 as normal distributions with $n = 100$; a mean of 10 and 15, respectively; and standard deviations (SDs) of 1 and 2. We evaluate all possible combinations for the regression coefficients 100 times, simulating different x_1 and x_2 . In each iteration we generate y using the setup previously described, and we recover the original regression coefficients using y as a response variable and x_1 and x_2 as predictors. Finally, we analyze two scenarios: (1) when the target timing occurs at the beginning of the year ($\mu = 0$) and (2) when the target timing happens midyear ($\mu = \pi$). The parameters for the entire setup generate realistic data, where the standard deviation of y is not higher than 0.3 rad. An SD of 0.5 rad would be equivalent to having phenological observations across half a year, which would not be realistic.

To quantify the accuracy of each model per coefficient we estimate the mean absolute error per model and coefficient (Eq. 5). To compare the accuracy between models by coefficient we test the mean absolute errors between models (Eq. 6). To generate a single measure that allows us to compare both coefficients and models we estimate the mean difference accuracy (Eq. 7). The results can be understood as follows: if the difference is higher than 0, the circular model has a higher mean accuracy compared to the linear model and vice versa. To quantify which model has higher precision we estimate the difference between the SD of the mean absolute errors per model for each coefficient (Eq. 8). Finally, we estimate the mean differences of precision between the regression coefficients (Eq. 9), where again if the value is higher than 0, the circular model has a higher mean accuracy than the linear model; the inverse is true if the value is lower than 0.

We estimate regression coefficients for the bootstrap sample $i \in \{1, \dots, m\}$ ($m = 100$) for the regression coefficient β_j , $j \in \{1, 2\}$, and the model $M \in \{l, c\}$ (denoted as $\hat{\beta}_{j,i}^M$). The model accuracy can then be estimated as the mean absolute error of the estimated regression parameter $\hat{\beta}_j^M$, $j \in \{1, 2\}$ for the linear model $M = l$ and the circular model $M = c$:

$$a_{M,j} = \frac{1}{m} \sum_{i=1}^m |\hat{\beta}_{j,i}^M - \beta_j|. \quad (5)$$

The difference in accuracy for the coefficient j between the circular and the linear model is shown in

$$\delta_{a,j} = a_{l,j} - a_{c,j}. \quad (6)$$

Finally, the mean difference in accuracy between the linear and the circular model is given by

$$\delta_a = \frac{\delta_{a,1} + \delta_{a,2}}{2}. \quad (7)$$

The difference in precision for the coefficient j between the linear (l) and the circular model (c) is shown in

$$\delta_{p,j} = s_{l,j} - s_{c,j}. \quad (8)$$

The mean difference in precision between the linear and the circular model is given by

$$\delta_p = \frac{\delta_{p,1} + \delta_{p,2}}{2}, \quad (9)$$

where $s_{M,j}$ is the sample SD of the vector $(\hat{\beta}_{j,i}^M)_i$, $M \in \{l, c\}$.

2.4 Analysis setup

The target variable $\text{DOY}_{\text{GPPmax}}$ is the day of the year when GPP reaches its maximum during the growing season. Given that different ecosystems present more than one growing season per year (e.g., semiarid ecosystems), it is necessary to identify the number of growing seasons per year. To identify the number of growing seasons we apply a fast Fourier transformation (FFT; Cooley and Tukey, 1965) to the mean seasonal cycle of the GPP time series. The number of growing seasons is equal to the maximum absolute value of the first four FFT coefficients (excluding the first one). For each FLUXNET site, we reconstruct the GPP time series, taking the real numbers of the inverse FFT. We use these reconstructed time series to calculate the expected mean timing of $\text{DOY}_{\text{GPPmax}}$ and use this value as a template. To recover the real $\text{DOY}_{\text{GPPmax}}$ from the original time series, we define a window around the template of length inversely proportional to the number of cycles (180 d/number of growing seasons). To increase the robustness of the analysis we identify the days with the 10 highest GPP values. These days are used in the block bootstrapping mentioned above. Finally, since most of the sites are located in the Northern Hemisphere we expect that, in most cases, $\text{DOY}_{\text{GPPmax}}$ will be reached by the middle of the calendar year.

To quantify the contribution of each climate variable, we count the number of sites per vegetation type where the regression coefficient is statistically significant. We perform a leave-one-out cross-validation per vegetation type to evaluate the predictive power of the circular regression using climate conditions. We only consider vegetation types with more than five sites. In this case the standardization of the climate variables is not applied. Finally, we use the mean of the optimum half-time parameter per vegetation type to weigh the climate conditions.

3 Results

Here, we first report results from simulated data to describe the performance of the circular regression approach compared to a linear model. Second, we compare the performance of circular and linear regression using empirical data. Third, we analyze the sensitivity of $\text{DOY}_{\text{GPPmax}}$ across vegetation types and climate classes. Finally, we show the results of the predictive power of circular regression per vegetation type.

3.1 Circular vs. linear regression

Figure 3a and c show that for $\mu = 0$ ($\text{DOY}_{\text{GPPmax}}$ at the beginning of the year), circular regression has a higher accuracy and precision compared to the linear regression for the entire space of regression coefficient values, with a maximum difference of the order of 0.1 in terms of accuracy and of the order of 1 for precision. For $\mu = \pi$ ($\text{DOY}_{\text{GPPmax}}$ midyear) the linear model has a higher accuracy in most of the evaluated space, with a maximum difference of the order of 0.001 compared with the circular regression, while circular regression has a higher precision for most of the regression coefficients of the order of 0.001. These results show that circular regression has a higher precision in recovering the original regression coefficients than linear regression no matter the moment of the year. On the other hand, circular regression has a higher accuracy than the linear model at the beginning of the year. While linear is better midyear, the differences are of the order of 0.001.

To illustrate the method in practice, we compare the circular and linear models using data from two sites: US-Ha1 (Northern Hemisphere, deciduous broadleaf forest) and AU-How (Southern Hemisphere, woody savanna). We relate the climate variables with $\text{DOY}_{\text{GPPmax}}$ (see Methods) and reconstruct the $\text{DOY}_{\text{GPPmax}}$ using the linear and circular regression models. We compare observed and predicted $\text{DOY}_{\text{GPPmax}}$ using JS correlation for the circular model and the Pearson product moment for the linear model. For US-Ha1 both methods show similar performance in predicting $\text{DOY}_{\text{GPPmax}}$ (Fig. 4), while for AU-How, the circular model retrieves the original data better than the linear model, explaining 30% more of the variance. In the event that the $\text{DOY}_{\text{GPPmax}}$ is reached at the beginning of the year, linear methods produce a strong bias that predicts the timing across the entire year (Fig. 4b).

3.2 Sensitivity of $\text{DOY}_{\text{GPPmax}}$ to climate variables

From the 52 sites analyzed in this study, only one site (ES-LJu) shows bimodal growing seasons (see Supplement 1.2). As expected, in most cases $\text{DOY}_{\text{GPPmax}}$ occurs in the middle of the calendar year (Fig. S6 in the Supplement), reflecting the uneven site distribution in FLUXNET (Schimel et al., 2015). However, some ecosystems in the Northern Hemi-

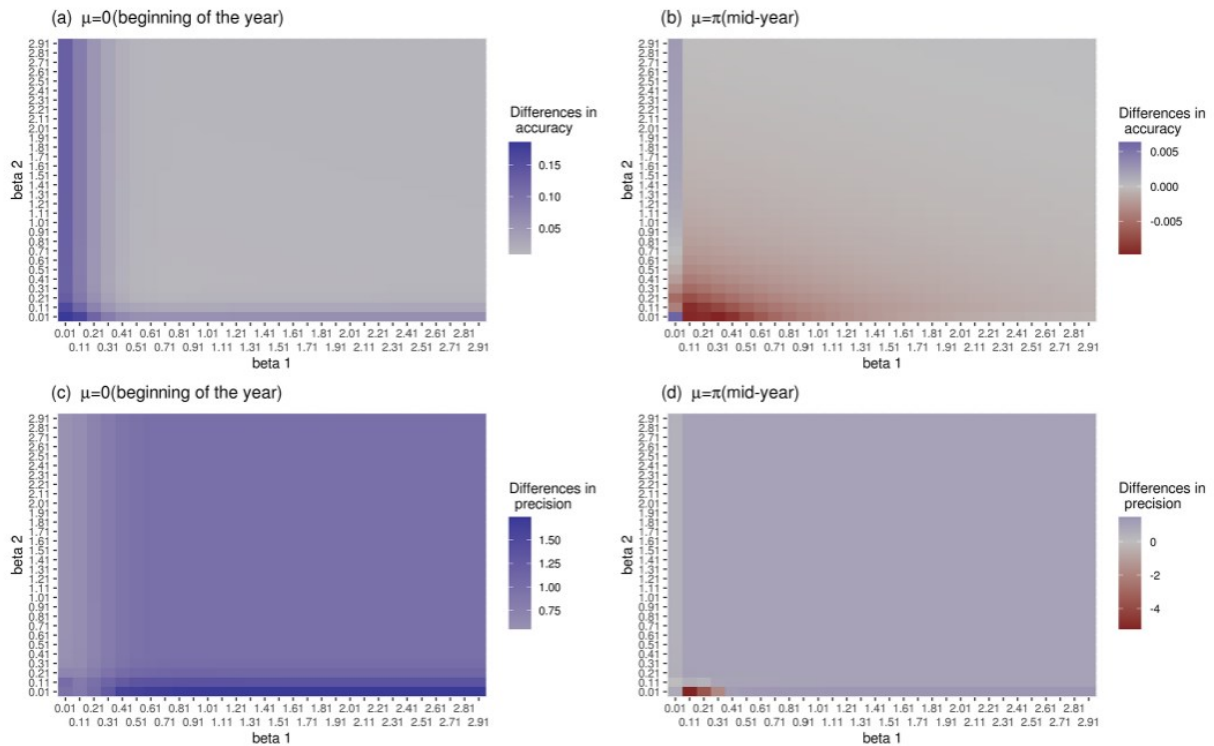


Figure 3. Accuracy and precision of linear and circular regression models by recovering the original regression coefficients of a circular regression. **(a, c)** $\mu = 0$ (maximum at the beginning of the year). **(b, d)** $\mu = \pi$ (maximum midyear). Panels **(a)** and **(b)** correspond to the differences in accuracy between the models. Panels **(c)** and **(d)** correspond to the differences in the precision between the models. Blue means better performance of the circular model compared with the linear model, and red means higher performance of the linear model.

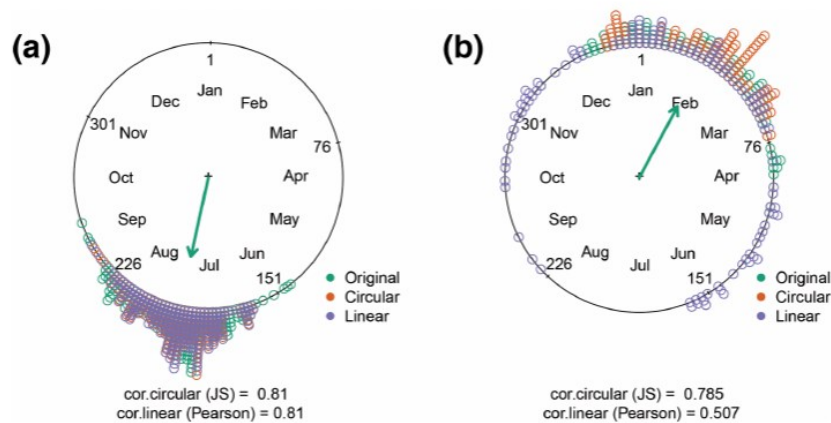


Figure 4. Correlation coefficient between the observed and predicted DOY_{GPPmax} using climatic variables. Two sites are presented: **(a)** US-Ha1 and **(b)** AU-How. The observed DOY_{GPPmax} (green) is compared with the data retrieved using circular (orange) and linear (purple) regressions. Two correlation coefficients are used: Jammalamadaka–Sarma (JS) and the Pearson product moment (Pearson). In the circular plot the months and the day of the year (DOY) are also plotted every 75 d. The green arrow indicates the mean angular direction of the original data distribution.

Table 1. Number of FLUXNET sites where each regression coefficient is statistically significant to explain the physio-phenology of GPP_{max} ($DOY_{GPP_{max}}$). The table is divided by the sign of the coefficient. The first column is the coefficient for the dimensionality reduction between air temperature (Tair), shortwave incoming radiation (SWin), and vapor pressure deficit (VPD); the second column is the coefficient for precipitation (Precip).

Climatic variable			
Sign	Tair, SWin, VPD	Precip	
(+)		8	2
(-)		38	14

sphere do reach $DOY_{GPP_{max}}$ at the beginning of the year; these are Mediterranean sites such as US-Var and ES-LJu. In general terms, most of the sites have an SD between 10 d and 40 d. The maximal SD is 46.9 d for the AU-Tum site. A detailed table with the mean angular direction and SD of $DOY_{GPP_{max}}$ of each site is presented in Sect. S1.2.

For half of the sites, the JS correlation coefficients are between 0.70 and 0.97 (Supplement 1, Fig. S5), showing that the interannual variability of $DOY_{GPP_{max}}$ is mainly explained by the cumulative effect of the climate variables. Nineteen sites have a JS coefficient of less than 0.7 (DK-Sor, FI-Hyy, US-MMS, DK-ZaH, FR-Pue, US-UMB, AU-Tum, US-Ton, FR-LBr, US-Me2, IT-Lav, AT-Neu, DE-Gri, IT-MBo, IT-Ro2, US-Wkg, BR-Sa1, FR-Fon, CZ-wet). For ES-LJu the JS coefficient is 0.77 for the first growing season and 0.78 for the second one (Table S2 in the Supplement).

We find that air temperature, shortwave incoming radiation, and vapor pressure deficit appear as the dominant drivers worldwide at 43 of the total sites (84%; Supplement 3). Precipitation is the main driver for five sites (AU-How US-Ton ZA-Kru US-SRM US-Wkg; Supplement 3). Interestingly, precipitation was the most important factor for all the woody savanna sites (Supplement 3). For three sites (DE-Gri, IT-Ro2, BR-Sa1), any climatic variable is significant. In terms of the sign of the coefficients, all the variables are predominantly negative (Table 1). This means that higher values of radiation, air temperature, VPD, and precipitation lead to an earlier $DOY_{GPP_{max}}$. Individual sensitivities per site are shown in Supplement 3.

The PCA between shortwave incoming radiation, air temperature, and vapor pressure deficit has the highest frequency of significant correlation coefficients by number of sites for all the vegetation types with the exception of woody savannas (WSAs), where precipitation is shown to be more important for most sites than the dimensionality reduction between Tair, SWin, and VPD (Fig. 5). For closed shrublands (CSHs) and savannas (SAVs), both drivers have the same number of sites where the coefficients are statistically significant.

A special case for understanding the sensitivity of $DOY_{GPP_{max}}$ to climate variables is the site “Llano de los

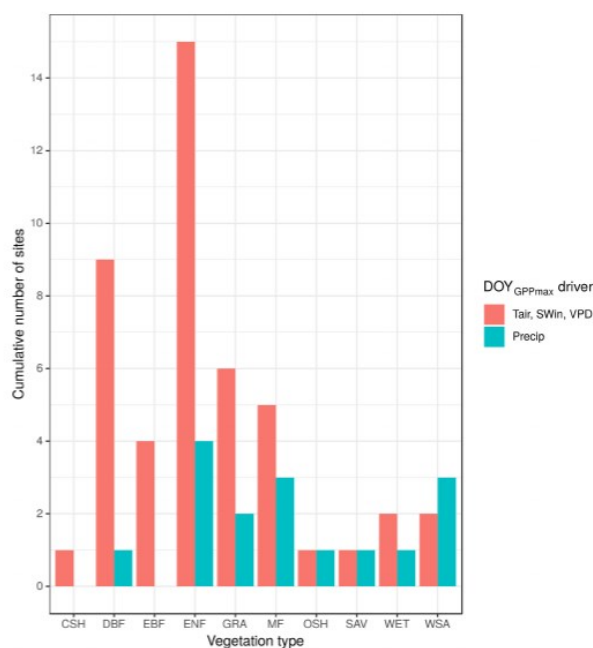


Figure 5. Contribution of each climate variable to explain the inter-annual variation in $DOY_{GPP_{max}}$ per vegetation type. CSHs: closed shrublands ($n = 1$); DBF: deciduous broadleaf forest ($n = 10$); EBF: evergreen broadleaf forest ($n = 5$); ENF: evergreen needleleaf forest ($n = 15$); GRA: grassland ($n = 8$); MF: mixed forest ($n = 5$); OSHs: open shrublands ($n = 1$); SAV: savannas ($n = 1$); WET: permanent wetlands ($n = 2$); WSAs: woody savannas ($n = 3$). Each bar shows the cumulative number of sites where each climate variable is statistically significant.

Juanes” (ES-LJu), an open shrubland ecosystem in Spain. It is the only clearly bimodal ecosystem in our study (Fig. 6). In this case precipitation is not statistically significant, while the combination of Tair, SWin, and VPD is significant for both seasons. Furthermore, in both growing seasons Tair, SWin, and VPD have a negative coefficient.

The leave-one-out cross-validation for several vegetation types shows that the predictive power of the model for grassland (GRA) and evergreen broadleaf forest (EBF) is -0.3 and -0.31 , respectively. For deciduous broadleaf forest (DBF) it is 0.46, and for evergreen needleleaf forest (ENF) it is 0.4, while for mixed forest (MF) the predictive power of the model is 0.88 (Fig. 7).

4 Discussion

4.1 Circular vs. linear regression

We explored whether circular regression is a suitable tool for analyzing phenological events. Our results suggest that circular regressions can recover predefined coefficients in a set

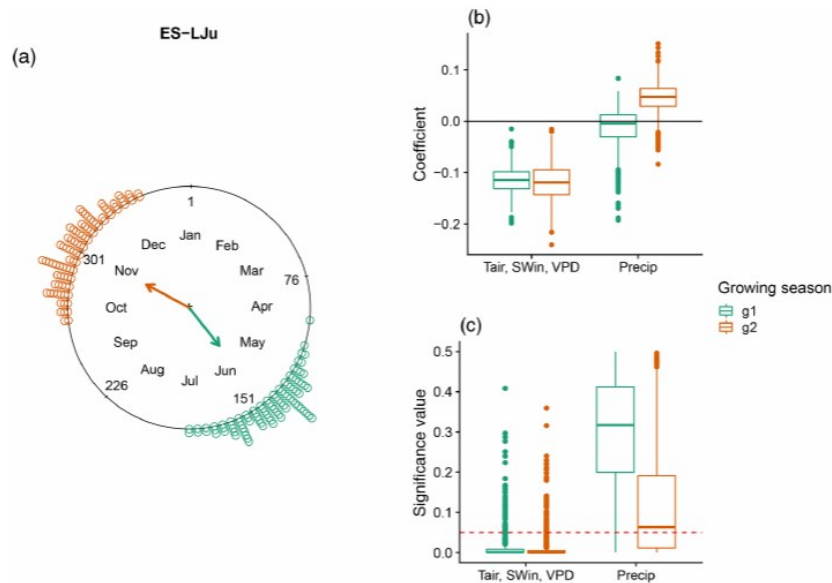


Figure 6. $\text{DOY}_{\text{GPPmax}}$ sensitivity to different climate drivers in a Mediterranean ecosystem: Llano de los Juanes (ES-LJu), Spain, with two growing seasons (green and orange). **(a)** $\text{DOY}_{\text{GPPmax}}$ distribution across the year. The arrows indicate the mean angular direction of the growing season. **(b)** Regression coefficients for each growing season and **(c)** the significance values for each variable. The red line in panel **(c)** represents a p value of 0.05.

of simulations with higher accuracy and precision than linear regressions. Hence, we would generally suggest that circular regressions may be advantageous when the aim is analyzing the effect of climatic variables on phenological events. We also found cases where the classical linear regression may be either more robust or equally suitable, e.g., when phenological events are reached close to midyear. In the overall view, however, we consider that circular regressions are to be preferred over linear regression for their conceptual capacity to analyze the physio-phenology of ecosystems regardless of the day of the year when an event of interest occurs. This allows us to analyze phenological studies at the global scale regardless of geographic location or the distribution of the observations during the year.

Different phenological models have been developed, ranging from empirical approaches (Richardson et al., 2013) to process models (Asse et al., 2020) over the last decades. As we demonstrate here, circular statistics open new opportunities to increase the robustness of phenological models, allowing us to analyze ecosystems across hemispheres within the same consistent framework. In fact, the results of the phenological sensitivity of $\text{DOY}_{\text{GPPmax}}$ indicate the complexity of ecosystem responses to climate variability. Our approach provides motivation to integrate circular regressions into more complex statistical techniques like regression trees, Gaussian processes, or artificial neural networks, targeting a circular response variable.

4.2 Sensitivity of $\text{DOY}_{\text{GPPmax}}$ to climate variables

The geographical location of the FLUXNET2015 sites represents an advantage when capturing the $\text{DOY}_{\text{GPPmax}}$ variability at the global scale (Supplement 1, Fig. S6). Most of the analyzed sites (47) are located in the Northern Hemisphere. Two sites (GF-Guy and BR-Sa1) are located in the tropical region, and three sites (ZA-Kru, AU-How, AU-Tum) are in the Southern Hemisphere. However, because of the low number of sites reported in the tropical and southern region with more than 7 years of data, our understanding of the $\text{DOY}_{\text{GPPmax}}$ variability in these regions is still limited. Increasing the number of tropical and Southern Hemisphere sites should be considered a high priority in the near future to complement our knowledge about the physio-phenological ecosystem state.

The high values of the JS correlation coefficients for most of the sites demonstrate that the interannual variability of $\text{DOY}_{\text{GPPmax}}$ can be explained as the cumulative effect of the climate variables during the growing season. Sites where it was not possible to explain the variations in $\text{DOY}_{\text{GPPmax}}$ with enough confidence (JS correlation < 0.7) might require the incorporation of biotic variables (e.g., species composition; Peichl et al., 2018) or soil property information that can improve the predictive power of the model.

Our results suggest that there is no pattern between the $\text{DOY}_{\text{GPPmax}}$ sensitivity across vegetation types and climate classes (Sect. Fig. S1.7). In other words, the $\text{DOY}_{\text{GPPmax}}$

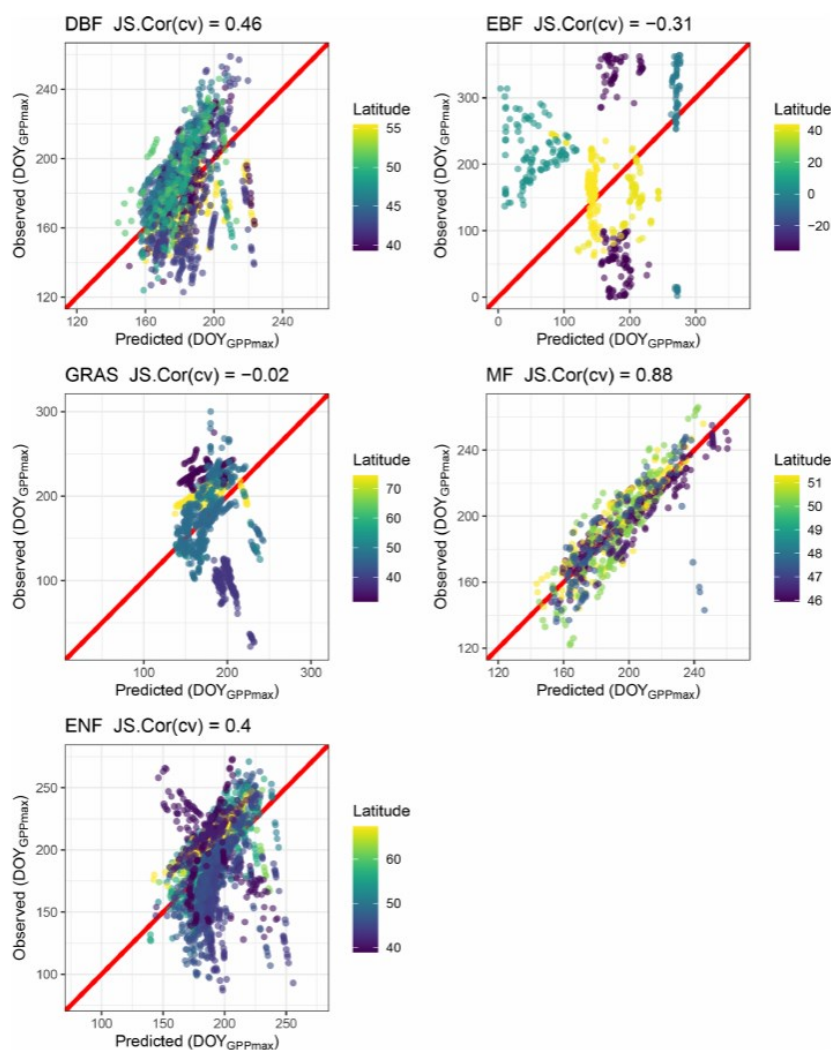


Figure 7. Cross-validation of the circular regression model to predict $\text{DOY}_{\text{GPPmax}}$ for different vegetation types using air temperature, shortwave incoming radiation, precipitation and vapor pressure deficit (see Methods). Deciduous broadleaf forest (DBF), evergreen broadleaf forest (EBF), grassland (GRA), mixed forest (MF), and evergreen needleleaf forest (ENF). For each vegetation type the Jammalamadaka–Sarma (JS) correlation coefficient is shown in the title of each plot. The red line represents the perfect fit.

sensitivity is site-specific, probably produced by the unique combination of biotic (e.g., species composition, species phenology, species interaction, and phenotypic plasticity) factors that are not evaluated in our study. Several studies that focused on ecosystem phenology suggest that species composition plays a fundamental role in ecosystem physio-phenology of the CO_2 uptake (Gonsamo et al., 2017; Peichl et al., 2018).

While there is no clear relationship between the $\text{DOY}_{\text{GPPmax}}$ sensitivity and the vegetation type, we find a predominant role of the combined effects of shortwave incoming radiation (SWin), air temperature (Tair), and vapor

pressure deficit (VPD) at the global scale on the $\text{DOY}_{\text{GPPmax}}$ interannual variability, where for most of the sites these variables have a negative regression coefficient. This means that if the SWin, Tair, and VPD increase during the growing season, the $\text{DOY}_{\text{GPPmax}}$ will be reached earlier. This effect can be a consequence of $\text{DOY}_{\text{GPPmax}}$ being reached when SWin and Tair are at a maximum.

On a global scale, our analysis shows that the combination of air temperature, shortwave incoming radiation, and vapor pressure deficit as well as precipitation has a negative sign. This means that if these variables increase during the growing season, the GPPmax will be reached earlier. Our re-

sults are similar to those obtained by Wang and Wu (2019), who were the authors to conclude that an increase in the temperature produces an earlier $\text{DOY}_{\text{GPPmax}}$. This phenomenon is likely explained by the leaf-out advancing during spring. Nevertheless, there is still no consensus on whether the increase in temperature will produce an earlier end of the growing season. Several studies have demonstrated for different vegetation types that when temperature increases, spring onset is earlier, and autumn senescence is later (Stocker et al., 2013; Linkosalo et al., 2009; Migliavacca et al., 2012; Morin et al., 2010; Post and Forchhammer, 2008), increasing the length of the growing season and the amount of CO_2 that is taken up by ecosystems (Richardson et al., 2013).

Ecosystems with two growing seasons per year represent a very interesting case of the effect of climate drivers on $\text{DOY}_{\text{GPPmax}}$ across different growing seasons. In Llano de los Juanes, Spain (ES-LJu; Fig. 6), $\text{DOY}_{\text{GPPmax}}$ is reached in the first growing season, when the rainy season is finishing, while in the second growing season $\text{DOY}_{\text{GPPmax}}$ is reached in the middle of the rainy season (data not shown). The effect of shortwave incoming radiation, temperature, and vapor pressure deficit for both growing seasons is negative, suggesting that if we increase these variables during the period before, the $\text{DOY}_{\text{GPPmax}}$ will happen earlier.

Phenology in Mediterranean ecosystems is mainly controlled by water availability (Kramer et al., 2000; Luo et al., 2018; Peñuelas et al., 2009). However, our results suggest that $\text{DOY}_{\text{GPPmax}}$ is mainly sensitive to SWin , T_{air} , and VPD . These results agree with the analysis performed by Gordo and Sanz (2005), who were the authors to evaluate the phenological sensitivity of Mediterranean ecosystems to temperature and precipitation. They concluded that temperature was the most important driver. Although water is a limiting factor in Mediterranean ecosystems, its influence on plant physiology and plant phenology can be completely different. In terms of physiology, the GPPmax value can decrease, but in terms of phenology, $\text{DOY}_{\text{GPPmax}}$ can still be the same.

Complex interactions between climate variables and phenological response and the interspecificity of the sensitivity at the site level explain in part the poor predictive power of the model for grasslands, evergreen broadleaf forest, evergreen needleleaf forest, and deciduous broadleaf forests in the cross-validation analysis (Fig. 7). However, the predictive power for mixed forest is high, even when the distribution of the latitudinal gradient is not the same for all the sites. These results reflect the fact that the circular regression model can be extrapolated from different sites to predict the $\text{DOY}_{\text{GPPmax}}$ interannual variability. This advantage could be a way to solve the common criticism that phenological models cannot be extrapolated by only generating ad hoc hypotheses (Richardson et al., 2013).

5 Conclusions

In this study we explored the potential of “circular regressions” to explain the physio-phenology of maximal CO_2 uptake rates. We conclude that (1) shortwave incoming radiation, temperature, and vapor pressure deficit are the main drivers of the timing of maximal CO_2 uptake at the global scale (precipitation only plays a secondary role, with the exception of woody savannas, where the most important variable is precipitation), and (2) although the sensitivity of the $\text{DOY}_{\text{GPPmax}}$ to the climate drivers is site-specific, it is possible to extrapolate the circular regression model for different sites with the same vegetation type and similar latitudes. Finally, we used simulated and empirical data to demonstrate that circular regression produces more accurate results than linear regression, in particular in cases when data need to be explored across hemispheres.

Appendix A: FLUXNET sites

Table A1. FLUXNET sites used in our study. We report the name of the sites, the time period used for the analysis, and the climate class of each site following Köppen–Geiger classification: tropical monsoon climate (Am), tropical savanna climate (Aw), cold semiarid climate (BSk), humid subtropical climate (Cfa), oceanic climate (Cfb), hot summer Mediterranean climate (Csa), warm summer Mediterranean climate (Csb), humid subtropical climate (Cwa), humid continental climate (Dfb), subarctic climate (Dfc, Dsc), and tundra climate (ET). We also report the vegetation type of the sites: closed shrubland (CSH), deciduous broadleaf forest (DBF), evergreen broadleaf forest (EBF), evergreen needleleaf forest (ENF), grassland (GRA), mixed forest (MF), open shrubland (OSH), savanna (SAV), permanent wetland (WET), and woody savanna (WSA).

Site name	Köppen–Geiger class	Vegetation type	Period	No. years analyzed	Citation	Data DOI
AT-Neu	Dfc	GRA	2002:2012	11	Wohlfahrt et al. (2008)	https://doi.org/10.18140/FLX/1440121
AU-How	Aw	WSA	2002:2014	13	Beringer et al. (2007)	https://doi.org/10.18140/FLX/1440125
AU-Tum	Cfb	EBF	2001:2014	14	Leuning et al. (2005)	https://doi.org/10.18140/FLX/1440126
BE-Bra	Cfb	MF	1999:2002, 2004:2014	15	Carrara et al. (2004)	https://doi.org/10.18140/FLX/1440128
BE-Vie	Cfb	MF	1997:2014	18	Aubinet et al. (2001)	https://doi.org/10.18140/FLX/1440130
BR-Sa1	Am	EBF	2002:2005, 2009:2011	7	Saleska et al. (2003)	https://doi.org/10.18140/FLX/1440032
CA-Man	Dfc	ENF	1994:1996, 1998:2003	12	Brooks et al. (1997)	https://doi.org/10.18140/FLX/1440035
CH-Cha	Cfb	GRA	2005:2014	10	Merbold et al. (2014)	https://doi.org/10.18140/FLX/1440131
CH-Dav	ET	ENF	1997:2014	18	Zielis et al. (2014)	https://doi.org/10.18140/FLX/1440178
CH-Fru	Cfb	GRA	2005:2014	10	Imer et al. (2013)	https://doi.org/10.18140/FLX/1440133
CH-Lae	Cfb	MF	2004:2014	11	Etzold et al. (2011)	https://doi.org/10.18140/FLX/1440134
CZ-wet	Cfb	WET	2006:2014	9	Dušek et al. (2012)	https://doi.org/10.18140/FLX/1440145
DE-Gri	Cfb	GRA	2004:2014	11	Prescher et al. (2010)	https://doi.org/10.18140/FLX/1440147
DE-Hai	Cfb	DBF	2000:2012	13	Knohl et al. (2003)	https://doi.org/10.18140/FLX/1440148
DE-Tha	Cfb	ENF	1996:2014	19	Grünwald and Bernhofer (2007)	https://doi.org/10.18140/FLX/1440152
DK-Sor	Cfb	DBF	1996:2014	19	Pilegaard et al. (2011)	https://doi.org/10.18140/FLX/1440155
DK-ZaH	ET	GRA	2000:2010, 2012:2014	14	Lund et al. (2012)	https://doi.org/10.18140/FLX/1440224
ES-LJu	Csa	OSH	2005:2013	9	Serrano-Ortiz et al. (2009)	https://doi.org/10.18140/FLX/1440226
FI-Hyy	Dfc	ENF	1996:2014	19	Suni et al. (2003)	https://doi.org/10.18140/FLX/1440158
FI-Sod	Dfc	ENF	2001:2014	14	Thum et al. (2007)	https://doi.org/10.18140/FLX/1440160
FR-Fon	Cfb	DBF	2005:2014	10	Delpierre et al. (2016)	https://doi.org/10.18140/FLX/1440161
FR-LBr	Cfb	ENF	1996:2008	13	Berbigier et al. (2001)	https://doi.org/10.18140/FLX/1440163
FR-Pue	Csa	EBF	2000:2015	15	Rambal et al. (2004)	https://doi.org/10.18140/FLX/1440164
GF-Guy	Am	EBF	2004:2014	11	Bonal et al. (2008)	https://doi.org/10.18140/FLX/1440165
IT-Col	Csa	DBF	1996:2014	19	Valentini et al. (1996)	https://doi.org/10.18140/FLX/1440167
IT-Cpz	Csa	EBF	2000:2008	9	Garbulsky et al. (2008)	https://doi.org/10.18140/FLX/1440168
IT-Lav	Cfb	ENF	2003:2014	12	Marcolla et al. (2003)	https://doi.org/10.18140/FLX/1440169
IT-MBo	Dfb	GRA	2003:2013	11	Marcolla et al. (2011)	https://doi.org/10.18140/FLX/1440170
IT-Noe	Csa	CSH	2004:2014	11	Marras et al. (2011)	https://doi.org/10.18140/FLX/1440171
IT-Ren	Dfc	ENF	1999, 2002:2003, 2005:2013	12	Montagnani et al. (2009)	https://doi.org/10.18140/FLX/1440173
IT-Ro1	Csa	DBF	2001:2008	8	Rey et al. (2002)	https://doi.org/10.18140/FLX/1440174
IT-Ro2	Csa	DBF	2002:2008, 2010:2012	10	Tedeschi et al. (2006)	https://doi.org/10.18140/FLX/1440175
IT-SRo	Csa	ENF	1999:2012	14	Chiesi et al. (2005)	https://doi.org/10.18140/FLX/1440176
NL-Loo	Cfb	ENF	1996:2014	18	Moors (2012)	https://doi.org/10.18140/FLX/1440178
RU-Cok	Dsc	OSH	2003:2013	11	van der Molen et al. (2007)	https://doi.org/10.18140/FLX/1440182
RU-Fyo	Dfb	ENF	1998:2014	17	Kurbatova et al. (2008)	https://doi.org/10.18140/FLX/1440183
US-Blo	Csa	ENF	1997:2007	11	Baker et al. (1999)	https://doi.org/10.18140/FLX/1440068
US-GLE	Dfc	ENF	2005:2014	10	McDowell et al. (2000)	https://doi.org/10.18140/FLX/1440069
US-Ha1	Dfb	DBF	1992:2012	21	Urbanski et al. (2007)	https://doi.org/10.18140/FLX/1440071
US-Los	Dfb	WET	2001:2008, 2010, 2014	10	Davis et al. (2003)	https://doi.org/10.18140/FLX/1440076
US-Me2	Csb	ENF	2002:2014	13	Treuhaf et al. (2004)	https://doi.org/10.18140/FLX/1440079
US-MMS	Cfa	DBF	1999:2014	16	Schmid et al. (2000)	https://doi.org/10.18140/FLX/1440083
US-NR1	Dfc	ENF	1999:2014	16	Monson et al. (2002)	https://doi.org/10.18140/FLX/1440087
US-PFa	Dfb	MF	1996:2014	19	Berger et al. (2001)	https://doi.org/10.18140/FLX/1440089
US-SRM	BSk	WSA	2004:2014	11	Scott et al. (2008)	https://doi.org/10.18140/FLX/1440090
US-Syv	Dfb	MF	2001:2007, 2012:2014	10	Desai et al. (2005)	https://doi.org/10.18140/FLX/1440091
US-Ton	Csa	WSA	2001:2014	14	Xu and Baldocchi (2003)	https://doi.org/10.18140/FLX/1440092
US-UMB	Dfb	DBF	2000:2014	15	Curtis et al. (2002)	https://doi.org/10.18140/FLX/1440093
US-Var	Csa	GRA	2001:2014	14	Xu and Baldocchi (2004)	https://doi.org/10.18140/FLX/1440094
US-WCr	Dfb	DBF	1999:2006, 2011:2014	12	Curtis et al. (2002)	https://doi.org/10.18140/FLX/1440095
US-Wkg	BSk	GRA	2004:2014	11	Emmerich (2003)	https://doi.org/10.18140/FLX/1440096
ZA-Kru	Cwa	SAV	2000:2005, 2007:2013	13	Archibald et al. (2009)	https://doi.org/10.18140/FLX/1440188

Code availability. Code is available under GPL-3 license at <https://github.com/dpabon/ecosystem-physio-phenology-repo>, last access: 29 July 2020 (<https://doi.org/10.5281/zenodo.3921892> Pabon-Moreno et al., 2020).

Data availability. FLUXNET database is available online at <https://fluxnet.fluxdata.org/>, last access: 11 July 2019 (<https://doi.org/10.1038/s41597-020-0534-3> Pastorello et al., 2020).

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/bg-17-3991-2020-supplement>.

Author contributions. DEPM, TM, MM, and MDM designed the study in collaboration with MR and CR. DEPM conducted the analysis and wrote the manuscript, with substantial contributions from all coauthors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank the reviewers for their helpful suggestions and Guido Kraemer for his help with the mathematical notation. This project has received funding from the European Union's Horizon 2020 research and innovation program via the TRuStEE project under the Marie Skłodowska-Curie grant agreement no. 721995. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization were carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and the Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, as well as the OzFlux, ChinaFlux, and AsiaFlux offices.

Financial support. This research has been supported by the H2020 Marie Skłodowska-Curie Actions (TRuStEE grant no. 721995).

The article processing charges for this open-access publication were covered by the Max Planck Society.

Review statement. This paper was edited by David Bowling and reviewed by two anonymous referees.

References

- Agostinelli, C. and Lund, U.: R package circular: Circular Statistics (version 0.4-93), CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy; UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA, available at: <https://r-forge.r-project.org/projects/circular/>, last access: 29 June 2017.
- Archibald, S. A., Kirton, A., van der Merwe, M. R., Scholes, R. J., Williams, C. A., and Hanan, N.: Drivers of inter-annual variability in Net Ecosystem Exchange in a semi-arid savanna ecosystem, South Africa, *Biogeosciences*, 6, 251–266, <https://doi.org/10.5194/bg-6-251-2009>, 2009.
- Asse, D., Randin, C. F., Bonhomme, M., Delestrade, A., and Chuine, I.: Process-based models outcompete correlative models in projecting spring phenology of trees in a future warmer climate, *Agr. Forest Meteorol.*, 285–286, 107931, <https://doi.org/10.1016/j.agrformet.2020.107931>, 2020.
- Aubinet, M., Chermanne, B., Vandenhaute, M., Longdoz, B., Yernaux, M., and Laitat, E.: Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes, *Agr. Forest Meteorol.*, 108, 293–315, [https://doi.org/10.1016/S0168-1923\(01\)00244-1](https://doi.org/10.1016/S0168-1923(01)00244-1), 2001.
- Aubinet, M., Vesala, T., and Papale, D. (Eds.): *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, Springer Atmospheric Sciences, Springer Netherlands, 2012.
- Baker, B., Guenther, A., Greenberg, J., Goldstein, A., and Fall, R.: Canopy fluxes of 2-methyl-3-buten-2-ol over a ponderosa pine forest by relaxed eddy accumulation: Field data and model comparison, *J. Geophys. Res.-Atmos.*, 104, 26107–26114, <https://doi.org/10.1029/1999JD900749>, 1999.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, *B. Am. Meteorol. Soc.*, 82, 2415–2434, [https://doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2), 2001.
- Baldocchi, D. D.: How eddy covariance flux measurements have contributed to our understanding of Global Change Biology, *Glob. Change Biol.*, 26, 242–260, <https://doi.org/10.1111/gcb.14807>, 2020.
- Bauerle, W. L., Oren, R., Way, D. A., Qian, S. S., Stoy, P. C., Thornton, P. E., Bowden, J. D., Hoffman, F. M., and Reynolds, R. F.: Photoperiodic regulation of the seasonal pattern of photosynthetic capacity and the implications for carbon cycling, *P. Natl. Acad. Sci. USA*, 109, 8612–8617, <https://doi.org/10.1073/pnas.1119131109>, 2012.
- Berbigier, P., Bonnefond, J.-M., and Mellmann, P.: CO₂ and water vapour fluxes for 2 years above Euroflux forest site, *Agr. Forest Meteorol.*, 108, 183–197, [https://doi.org/10.1016/S0168-1923\(01\)00240-4](https://doi.org/10.1016/S0168-1923(01)00240-4), 2001.
- Berger, B. W., Davis, K. J., Yi, C., Bakwin, P. S., and Zhao, C. L.: Long-Term Carbon Dioxide Fluxes from a Very Tall Tower in a Northern Forest: Flux Measurement Methodology, *J. Atmos. Ocean. Technol.*, 18, 529–542, [https://doi.org/10.1175/1520-0426\(2001\)018<0529:LTCDFF>2.0.CO;2](https://doi.org/10.1175/1520-0426(2001)018<0529:LTCDFF>2.0.CO;2), 2001.

- Beringer, J., Hutley, L. B., Tapper, N. J., and Cernusak, L. A.: Savanna fires and their impact on net ecosystem productivity in North Australia, *Glob. Change Biol.*, 13, 990–1004, <https://doi.org/10.1111/j.1365-2486.2007.01334.x>, 2007.
- Beyene, M. T., Jain, S., and Gupta, R. C.: Linear-Circular Statistical Modeling of Lake Ice-Out Dates, *Water Resour. Res.*, 54, 7841–7858, <https://doi.org/10.1029/2017WR021731>, 2018.
- Bonal, D., Bosc, A., Ponton, S., Goret, J.-Y., Burban, B., Gross, P., Bonnefond, J.-M., Elbers, J., Longdoz, B., Epron, D., Guehl, J.-M., and Granier, A.: Impact of severe dry season on net ecosystem exchange in the Neotropical rainforest of French Guiana, *Glob. Change Biol.*, 14, 1917–1933, <https://doi.org/10.1111/j.1365-2486.2008.01610.x>, 2008.
- Brooks, J. R., Flanagan, L. B., Varney, G. T., and Ehleringer, J. R.: Vertical gradients in photosynthetic gas exchange characteristics and refixation of respired CO₂ within boreal forest canopies, *Tree Physiol.*, 17, 1–12, <https://doi.org/10.1093/treephys/17.1.1>, 1997.
- Carrara, A., Janssens, I. A., Curiel Yuste, J., and Ceulemans, R.: Seasonal changes in photosynthesis, respiration and NEE of a mixed temperate forest, *Agr. Forest Meteorol.*, 126, 15–31, <https://doi.org/10.1016/j.agrformet.2004.05.002>, 2004.
- Chiesi, M., Maselli, F., Bindi, M., Fibbi, L., Cherubini, P., Arlotta, E., Tirone, G., Matteucci, G., and Seufert, G.: Modelling carbon budget of Mediterranean forests using ground and remote sensing measurements, *Agr. Forest Meteorol.*, 135, 22–34, <https://doi.org/10.1016/j.agrformet.2005.09.011>, 2005.
- Cooley, J. W. and Tukey, J. W.: An Algorithm for the Machine Calculation of Complex Fourier Series, *Math. Comput.*, 19, 297–301, <https://doi.org/10.2307/2003354>, 1965.
- Curtis, P. S., Hanson, P. J., Bolstad, P., Barford, C., Randolph, J. C., Schmid, H. P., and Wilson, K. B.: Biometric and eddy-covariance based estimates of annual carbon storage in five eastern North American deciduous forests, *Agr. Forest Meteorol.*, 113, 3–19, [https://doi.org/10.1016/S0168-1923\(02\)00099-0](https://doi.org/10.1016/S0168-1923(02)00099-0), 2002.
- Davis, K. J., Bakwin, P. S., Yi, C., Berger, B. W., Zhao, C., Teclaw, R. M., and Isebrands, J. G.: The annual cycles of CO₂ and H₂O exchange over a northern mixed forest as observed from a very tall tower, *Glob. Change Biol.*, 9, 1278–1293, <https://doi.org/10.1046/j.1365-2486.2003.00672.x>, 2003.
- Delpierre, N., Berveiller, D., Granda, E., and Duffrêne, E.: Wood phenology, not carbon input, controls the interannual variability of wood growth in a temperate oak forest, *New Phytol.*, 210, 459–470, <https://doi.org/10.1111/nph.13771>, 2016.
- Desai, A. R., Bolstad, P. V., Cook, B. D., Davis, K. J., and Carey, E. V.: Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA, *Agr. Forest Meteorol.*, 128, 33–55, <https://doi.org/10.1016/j.agrformet.2004.09.005>, 2005.
- Dušek, J., Čížková, H., Stellner, S., Czerný, R., and Květ, J.: Fluctuating water table affects gross ecosystem production and gross radiation use efficiency in a sedge-grass marsh, *Hydrobiologia*, 692, 57–66, <https://doi.org/10.1007/s10750-012-0998-z>, 2012.
- Emmerich, W. E.: Carbon dioxide fluxes in a semiarid environment with high carbonate soils, *Agr. Forest Meteorol.*, 116, 91–102, [https://doi.org/10.1016/S0168-1923\(02\)00231-9](https://doi.org/10.1016/S0168-1923(02)00231-9), 2003.
- Etzold, S., RUEHR, N. K., Zweifel, R., Dobbertin, M., Zingg, A., Pluess, P., Häslner, R., Eugster, W., and Buchmann, N.: The Carbon Balance of Two Contrasting Mountain Forest Ecosystems in Switzerland: Similar Annual Trends, but Seasonal Differences, *Ecosystems*, 14, 1289–1309, <https://doi.org/10.1007/s10021-011-9481-3>, 2011.
- Fisher, N. I.: *Statistical Analysis of Circular Data*, Cambridge University Press, Cambridge, 1995.
- Fisher, N. I. and Lee, A. J.: Regression Models for an Angular Response, *Biometrics*, 48, 665–677, <https://doi.org/10.2307/2532334>, 1992.
- Garbulsky, M. F., Peñuelas, J., Papale, D., and Filella, I.: Remote estimation of carbon dioxide uptake by a Mediterranean forest, *Glob. Change Biol.*, 14, 2860–2867, <https://doi.org/10.1111/j.1365-2486.2008.01684.x>, 2008.
- Gonsamo, A., D’Odorico, P., Chen, J. M., Wu, C., and Buchmann, N.: Changes in vegetation phenology are not reflected in atmospheric CO₂ and 13C/12C seasonality, *Glob. Change Biol.*, 23, 4029–4044, <https://doi.org/10.1111/gcb.13646>, 2017.
- Gordo, O. and Sanz, J. J.: Phenology and climate change: a long-term study in a Mediterranean locality, *Oecologia*, 146, 484–495, <https://doi.org/10.1007/s00442-005-0240-z>, 2005.
- Green, P. J.: Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives, *J. Roy. Stat. Soc. B*, 46, 149–192, 1984.
- Grünwald, T. and Bernhofer, C.: A decade of carbon, water and energy flux measurements of an old spruce forest at the Anchor Station Tharandt, *Tellus B*, 59, 387–396, <https://doi.org/10.1111/j.1600-0889.2007.00259.x>, 2007.
- Hastie, T., Tibshirani, R., and Friedman, J. H.: *The elements of statistical learning: data mining, inference, and prediction*, Springer Series in Statistics, 2nd Edn., Springer, New York, NY, 2009.
- Imer, D., Merbold, L., Eugster, W., and Buchmann, N.: Temporal and spatial variations of soil CO₂, CH₄ and N₂O fluxes at three differently managed grasslands, *Biogeosciences*, 10, 5931–5945, <https://doi.org/10.5194/bg-10-5931-2013>, 2013.
- Jammalamadaka, S. R. and Sarma, Y.: A correlation coefficient for angular variables, *Statistical theory and data analysis II*, North-Holland, Amsterdam, 349–364, 1988.
- Knohl, A., Schulze, E.-D., Kolle, O., and Buchmann, N.: Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany, *Agr. Forest Meteorol.*, 118, 151–167, [https://doi.org/10.1016/S0168-1923\(03\)00115-1](https://doi.org/10.1016/S0168-1923(03)00115-1), 2003.
- Kramer, K., Leinonen, I., and Loustau, D.: The importance of phenology for the evaluation of impact of climate change on growth of boreal, temperate and Mediterranean forests ecosystems: an overview, *Int. J. Biometeorol.*, 44, 67–75, <https://doi.org/10.1007/s004840000066>, 2000.
- Kurbatova, J., Li, C., Varlagin, A., Xiao, X., and Vygodskaya, N.: Modeling carbon dynamics in two adjacent spruce forests with different soil conditions in Russia, *Biogeosciences*, 5, 969–980, <https://doi.org/10.5194/bg-5-969-2008>, 2008.
- Leuning, R., Cleugh, H. A., Zegelin, S. J., and Hughes, D.: Carbon and water fluxes over a temperate Eucalypt forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates, *Agr. Forest Meteorol.*, 129, 151–173, <https://doi.org/10.1016/j.agrformet.2004.12.004>, 2005.
- Lieth, H. (Ed.): *Phenology and Seasonality Modeling*, Ecological Studies, Springer-Verlag, Berlin Heidelberg, 1974.
- Linkosalo, T., Häkkinen, R., Terhivuo, J., Tuomenvirta, H., and Hari, P.: The time series of flowering and leaf bud burst of bo-

- real trees (1846–2005) support the direct temperature observations of climatic warming, *Agr. Forest Meteorol.*, 149, 453–461, <https://doi.org/10.1016/j.agrformet.2008.09.006>, 2009.
- Liu, Y., Schwalm, C. R., Samuels-Crow, K. E., and Ogle, K.: Ecological memory of daily carbon exchange across the globe and its importance in drylands, *Ecol. Lett.*, 22, 1806–1816, <https://doi.org/10.1111/ele.13363>, 2019.
- Lund, M., Falk, J. M., Friborg, T., Mbufong, H. N., Sigsgaard, C., Soegaard, H., and Tamstorf, M. P.: Trends in CO₂ exchange in a high Arctic tundra heath, 2000–2010, *J. Geophys. Res.-Biogeo.*, 117, G02001, <https://doi.org/10.1029/2011JG001901>, 2012.
- Luo, Y., El-Madany, T. S., Filippa, G., Ma, X., Ahrens, B., Carrara, A., Gonzalez-Cascon, R., Cremonese, E., Galvagno, M., Hammer, T. W., Pacheco-Labrador, J., Martín, M. P., Moreno, G., Perez-Priego, O., Reichstein, M., Richardson, A. D., Römermann, C., and Migliavacca, M.: Using Near-Infrared-Enabled Digital Repeat Photography to Track Structural and Physiological Phenology in Mediterranean Tree–Grass Ecosystems, *Remote Sens.*, 10, 1293, <https://doi.org/10.3390/rs10081293>, 2018.
- Marcolla, B., Pitacco, A., and Cescatti, A.: Canopy Architecture and Turbulence Structure in a Coniferous Forest, *Bound.-Layer Meteorol.*, 108, 39–59, <https://doi.org/10.1023/A:1023027709805>, 2003.
- Marcolla, B., Cescatti, A., Manca, G., Zorer, R., Cavagna, M., Fiora, A., Gianelle, D., Rodeghiero, M., Sottocornola, M., and Zampedri, R.: Climatic controls and ecosystem responses drive the inter-annual variability of the net ecosystem exchange of an alpine meadow, *Agr. Forest Meteorol.*, 151, 1233–1243, <https://doi.org/10.1016/j.agrformet.2011.04.015>, 2011.
- Marras, S., Pyles, R. D., Sirca, C., Paw U, K. T., Snyder, R. L., Duce, P., and Spano, D.: Evaluation of the Advanced Canopy–Atmosphere–Soil Algorithm (ACASA) model performance over Mediterranean maquis ecosystem, *Agr. Forest Meteorol.*, 151, 730–745, <https://doi.org/10.1016/j.agrformet.2011.02.004>, 2011.
- McDowell, N. G., Marshall, J. D., Hooker, T. D., and Musselman, R.: Estimating CO₂ flux from snowpacks at three sites in the Rocky Mountains, *Tree Physiol.*, 20, 745–753, <https://doi.org/10.1093/treephys/20.11.745>, 2000.
- Merbold, L., Eugster, W., Stieger, J., Zahniser, M., Nelson, D., and Buchmann, N.: Greenhouse gas budget (CO₂, CH₄ and N₂O) of intensively managed grassland following restoration, *Glob. Change Biol.*, 20, 1913–1928, <https://doi.org/10.1111/gcb.12518>, 2014.
- Migliavacca, M., Sonntag, O., Keenan, T. F., Cescatti, A., O’Keefe, J., and Richardson, A. D.: On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model, *Biogeosciences*, 9, 2063–2083, <https://doi.org/10.5194/bg-9-2063-2012>, 2012.
- Monson, R. K., Turnipseed, A. A., Sparks, J. P., Harley, P. C., Scott-Denton, L. E., Sparks, K., and Huxman, T. E.: Carbon sequestration in a high-elevation, subalpine forest, *Glob. Change Biol.*, 8, 459–478, <https://doi.org/10.1046/j.1365-2486.2002.00480.x>, 2002.
- Montagnani, L., Manca, G., Canepa, E., Georgieva, E., Acosta, M., Feigenwinter, C., Janous, D., Kerschbaumer, G., Lindroth, A., Minach, L., Minerbi, S., Mölder, M., Pavelka, M., Seufert, G., Zerri, M., and Ziegler, W.: A new mass conservation approach to the study of CO₂ advection in an alpine forest, *J. Geophys. Res.-Atmos.*, 114, D07306, <https://doi.org/10.1029/2008JD010650>, 2009.
- Moors, E.: Water use of forests in the Netherlands, Tech. rep., Vrije Universiteit, Amsterdam, available at: <http://edepot.wur.nl/213926> (last access: 26 April 2019), 2012.
- Morellato, L. P. C., Alberti, L. F., and Hudson, I. L.: Applications of Circular Statistics in Plant Phenology: a Case Studies Approach, in: *Phenological Research*, Springer, Dordrecht, https://doi.org/10.1007/978-90-481-3335-2_16, 339–359, 2010.
- Morente-López, J., Lara-Romero, C., Ormosa, C., and Iriondo, J. M.: Phenology drives species interactions and modularity in a plant–flower visitor network, *Sci. Rep.*, 8, 9386, <https://doi.org/10.1038/s41598-018-27725-2>, 2018.
- Morin, X., Roy, J., Sonié, L., and Chuine, I.: Changes in leaf phenology of three European oak species in response to experimental climate change, *New Phytol.*, 186, 900–910, <https://doi.org/10.1111/j.1469-8137.2010.03252.x>, 2010.
- Pabon-Moreno, D. E., Musavi, T., Migliavacca, M., Reichstein, M., Römermann, C., and Mahecha, M. D.: Code-repository for the research paper: “11Ecosystem Physio-phenology revealed using circular statistics”, Zenodo, <https://doi.org/10.5281/zenodo.3921892>, 2020.
- Pastorello, G., Trotta, C., Canfora, E., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, *Scientific Data*, 7, 225 pp., <https://doi.org/10.1038/s41597-020-0534-3>, 2020.
- Peichl, M., Gažovič, M., Vermeij, I., Goede, E. d., Sonntag, O., Limpens, J., and Nilsson, M. B.: Peatland vegetation composition and phenology drive the seasonal trajectory of maximum gross primary production, *Sci. Rep.*, 8, 8012, <https://doi.org/10.1038/s41598-018-26147-4>, 2018.
- Peñuelas, J., Rutishauser, T., and Filella, I.: Phenology Feedbacks on Climate Change, *Science*, 324, 887–888, <https://doi.org/10.1126/science.1173004>, 2009.
- Pilegaard, K., Ibrom, A., Courtney, M. S., Hummelshøj, P., and Jensen, N. O.: Increasing net CO₂ uptake by a Danish beech forest during the period from 1996 to 2009, *Agr. Forest Meteorol.*, 151, 934–946, <https://doi.org/10.1016/j.agrformet.2011.02.013>, 2011.
- Post, E. and Forchhammer, M. C.: Climate change reduces reproductive success of an Arctic herbivore through trophic mismatch, *Philos. T. Roy. Soc. B-Biol.*, 363, 2367–2373, <https://doi.org/10.1098/rstb.2007.2207>, 2008.
- Prescher, A.-K., Grünwald, T., and Bernhofer, C.: Land use regulates carbon budgets in eastern Germany: From NEE to NBP, *Agr. Forest Meteorol.*, 150, 1016–1025, <https://doi.org/10.1016/j.agrformet.2010.03.008>, 2010.
- Rambal, S., Joffre, R., Ourcival, J. M., Cavender-Bares, J., and Rocheteau, A.: The growth respiration component in eddy CO₂ flux from a *Quercus ilex* mediterranean forest, *Glob. Change Biol.*, 10, 1460–1469, <https://doi.org/10.1111/j.1365-2486.2004.00819.x>, 2004.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havráňková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On

- the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Glob. Change Biol.*, 11, 1424–1439, <https://doi.org/10.1111/j.1365-2486.2005.001002.x>, 2005.
- Rey, A., Pegoraro, E., Tedeschi, V., Parri, I. D., Jarvis, P. G., and Valentini, R.: Annual variation in soil respiration and its components in a coppice oak forest in Central Italy, *Glob. Change Biol.*, 8, 851–866, <https://doi.org/10.1046/j.1365-2486.2002.00521.x>, 2002.
- Richardson, A. D., Braswell, B. H., Hollinger, D. Y., Jenkins, J. P., and Ollinger, S. V.: Near-surface remote sensing of spatial and temporal variation in canopy phenology, *Ecol. Appl.*, 19, 1417–1428, <https://doi.org/10.1890/08-2022.1>, 2009.
- Richardson, A. D., Andy Black, T., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., Hollinger, D. Y., Kutsch, W. L., Longdoz, B., Luysaert, S., Migliavacca, M., Montagnani, L., William Munger, J., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri, E., Vargas, R., and Varlagin, A.: Influence of spring and autumn phenological transitions on forest ecosystem productivity, *Philos. T. Roy. Soc. B-Biol.*, 365, 3227–3246, <https://doi.org/10.1098/rstb.2010.0102>, 2010.
- Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., and Toomey, M.: Climate change, phenology, and phenological control of vegetation feedbacks to the climate system, *Agr. Forest Meteorol.*, 169, 156–173, <https://doi.org/10.1016/j.agrformet.2012.09.012>, 2013.
- Ryan, E. M., Ogle, K., Zelikova, T. J., LeCain, D. R., Williams, D. G., Morgan, J. A., and Pendall, E.: Antecedent moisture and temperature conditions modulate the response of ecosystem respiration to elevated CO₂ and warming, *Glob. Change Biol.*, 21, 2588–2602, <https://doi.org/10.1111/gcb.12910>, 2015.
- Saleska, S. R., Miller, S. D., Matross, D. M., Goulden, M. L., Wofsy, S. C., da Rocha, H. R., de Camargo, P. B., Crill, P., Daube, B. C., de Freitas, H. C., Hutyla, L., Keller, M., Kirchhoff, V., Menton, M., Munger, J. W., Pyle, E. H., Rice, A. H., and Silva, H.: Carbon in Amazon Forests: Unexpected Seasonal Fluxes and Disturbance-Induced Losses, *Science*, 302, 1554–1557, <https://doi.org/10.1126/science.1091165>, 2003.
- Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K., and Cox, P.: Observing terrestrial ecosystems and the carbon cycle from space, *Glob. Change Biol.*, 21, 1762–1776, <https://doi.org/10.1111/gcb.12822>, 2015.
- Schmid, H. P., Grimmer, C. S. B., Copley, F., Offerle, B., and Su, H.-B.: Measurements of CO₂ and energy fluxes over a mixed hardwood forest in the mid-western United States, *Agr. Forest Meteorol.*, 103, 357–374, [https://doi.org/10.1016/S0168-1923\(00\)00140-4](https://doi.org/10.1016/S0168-1923(00)00140-4), 2000.
- Scott, R. L., Cable, W. L., and Hultine, K. R.: The hydrologic significance of hydraulic redistribution in a semiarid savanna, *Water Resour. Res.*, 44, W02440, <https://doi.org/10.1029/2007WR006149>, 2008.
- Serrano-Ortiz, P., Domingo, F., Cazorla, A., Were, A., Cuezva, S., Villagarcía, L., Alados-Arboledas, L., and Kowalski, A. S.: Interannual CO₂ exchange of a sparse Mediterranean shrubland on a carbonaceous substrate, *J. Geophys. Res.-Biogeosci.*, 114, G04015, <https://doi.org/10.1029/2009JG000983>, 2009.
- Stocker, T. F., Qin, D., Plattner, G.-K. et al.: Climate change 2013: The physical science basis, Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, 1535, Cambridge university press Cambridge, United Kingdom and New York, NY, USA, 2013.
- Suni, T., Rinne, J., Reissell, A., Altimir, N., Keronen, P., Rannik, U., Dal Maso, M., Kulmala, M., and Vesala, T.: Long-term measurements of surface fluxes above a Scots pine forest in Hyytiälä, southern Finland, 1996–2001, *Boreal Environ. Res.*, 8, 287–301, 2003.
- Tedeschi, V., Rey, A., Manca, G., Valentini, R., Jarvis, P. G., and Borghetti, M.: Soil respiration in a Mediterranean oak forest at different developmental stages after coppicing, *Glob. Change Biol.*, 12, 110–121, <https://doi.org/10.1111/j.1365-2486.2005.01081.x>, 2006.
- Thum, T., Aalto, T., Laurila, T., Aurela, M., Kolari, P., and Hari, P.: Parametrization of two photosynthesis models at the canopy scale in a northern boreal Scots pine forest, *Tellus B*, 59, 874–890, <https://doi.org/10.1111/j.1600-0889.2007.00305.x>, 2007.
- Treuhaft, R. N., Law, B. E., and Asner, G. P.: Forest Attributes from Radar Interferometric Structure and Its Fusion with Optical Remote Sensing, *BioScience*, 54, 561–571, 2004.
- Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czirkowsky, M., and Munger, J. W.: Factors controlling CO₂ exchange on timescales from hourly to decadal at Harvard Forest, *J. Geophys. Res.-Biogeosci.*, 112, G02020, <https://doi.org/10.1029/2006JG000293>, 2007.
- Valentini, R., Angelis, P. D., Matteucci, G., Monaco, R., Dore, S., and Mucnozza, G. E. S.: Seasonal net carbon dioxide exchange of a beech forest with the atmosphere, *Glob. Change Biol.*, 2, 199–207, <https://doi.org/10.1111/j.1365-2486.1996.tb00072.x>, 1996.
- van der Molen, M. K., van Huissteden, J., Parmentier, F. J. W., Petrescu, A. M. R., Dolman, A. J., Maximov, T. C., Kononov, A. V., Karsanaev, S. V., and Suzdalov, D. A.: The growing season greenhouse gas balance of a continental tundra site in the Indigirka lowlands, NE Siberia, *Biogeosciences*, 4, 985–1003, <https://doi.org/10.5194/bg-4-985-2007>, 2007.
- Von Mises, R.: Über die “Ganzzahligkeit” der Atomgewichte und verwandte Fragen, *Phys. Z.*, 19, 490–500, 1918.
- Wang, X. and Wu, C.: Estimating the peak of growing season (POS) of China’s terrestrial ecosystems, *Agr. Forest Meteorol.*, 278, 107639, <https://doi.org/10.1016/j.agrformet.2019.107639>, 2019.
- Wohlfahrt, G., Hammerle, A., Haslwanter, A., Bahn, M., Tappeiner, U., and Cernusca, A.: Seasonal and inter-annual variability of the net ecosystem CO₂ exchange of a temperate mountain grassland: Effects of weather and management, *J. Geophys. Res.-Atmos.*, 113, D08110, <https://doi.org/10.1029/2007JD009286>, 2008.
- Xu, L. and Baldocchi, D. D.: Seasonal trends in photosynthetic parameters and stomatal conductance of blue oak (*Quercus douglasii*) under prolonged summer drought and high temperature, *Tree Physiol.*, 23, 865–877, <https://doi.org/10.1093/treephys/23.13.865>, 2003.
- Xu, L. and Baldocchi, D. D.: Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California, *Agr. Forest Meteorol.*, 123, 79–96, <https://doi.org/10.1016/j.agrformet.2003.10.004>, 2004.
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., and Huete, A.: Monitoring vegetation phenology using MODIS, *Remote Sensing Environ.*, 84, 471–475, [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9), 2003.

Zhou, S., Zhang, Y., Caylor, K. K., Luo, Y., Xiao, X., Ciais, P., Huang, Y., and Wang, G.: Explaining inter-annual variability of gross primary productivity from plant phenology and physiology, *Agr. Forest Meteorol.*, 226–227, 246–256, <https://doi.org/10.1016/j.agrformet.2016.06.010>, 2016.

Zhou, S., Zhang, Y., Ciais, P., Xiao, X., Luo, Y., Caylor, K. K., Huang, Y., and Wang, G.: Dominant role of plant physiology in trend and variability of gross primary productivity in North America, *Sci. Rep.*, 7, 41366, <https://doi.org/10.1038/srep41366>, 2017.

Zielis, S., Etzold, S., Zweifel, R., Eugster, W., Haeni, M., and Buchmann, N.: NEP of a Swiss subalpine forest is significantly driven not only by current but also by previous year's weather, *Biogeosciences*, 11, 1627–1635, <https://doi.org/10.5194/bg-11-1627-2014>, 2014.

CHAPTER 4

On the Potential of Sentinel-2 for Estimating Gross Primary Production

Manuscript Nr. 3

Title of the Manuscript: On the Potential of Sentinel-2 for Estimating Gross Primary Production

Authors: Pabon-Moreno, Daniel E., Migliavacca, Mirco., Reichstein, Markus., Mahecha Miguel D.

Bibliographic information: Pabon-Moreno, D. E., Migliavacca, M., Reichstein, M., & Mahecha, M. D. (2022). On the Potential of Sentinel-2 for Estimating Gross Primary Production. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 112. <https://doi.org/10.1109/TGRS.2022.3152272>

The candidate is: First author

Status: Published

Authors' contribution in percentage

Author	Conceptual	Data analysis	Experimental (Numerical)	Writing the manuscript	Provision of material
Daniel E. Pabon-Moreno	40	80	92	70	90
Dr. Mirco Migliavacca	20	6	3	12	5
Prof. Dr. Markus Reichstein	3	4	2	8	5
Prof. Dr. Miguel D. Mahecha	37	10	3	10	0
Total:	100	100	100	100	100

On the Potential of Sentinel-2 for Estimating Gross Primary Production

Daniel E. Pabon-Moreno¹, Mirco Migliavacca, Markus Reichstein, and Miguel D. Mahecha²

Abstract—Estimating gross primary production (GPP), the gross uptake of CO₂ by vegetation, is a fundamental prerequisite for understanding and quantifying the terrestrial carbon cycle. Over the last decade, multiple approaches have been developed to derive spatiotemporal dynamics of GPP combining *in situ* observations and remote sensing data using machine learning techniques or semiempirical models. However, no high spatial resolution GPP product exists so far that is derived entirely from satellite-based remote sensing data. Sentinel-2 satellites are expected to open new opportunities to analyze ecosystem processes with spectral bands chosen to study vegetation between 10- and 20-m spatial resolutions with five-day revisit frequency. Of particular relevance is the availability of red-edge bands that are suitable for deriving estimates of canopy chlorophyll content that are expected to be much better than any previous global mission. Here, we analyzed whether red-edge-based and near-infrared-based vegetation indices (VIs) or machine learning techniques that consider VIs, all spectral bands, and their nonlinear interactions could predict daily GPP derived from 58 eddy covariance sites. Using linear regressions based on classic VIs, including near-infrared reflectance of vegetation (NIRv), we achieved prediction powers of $R^2 = 0.51$ and an $\text{RMSE}_{10\text{-fold}} = 2.95 \text{ } [\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}]^{10\text{-fold}}$ in a 10-fold cross validation. Chlorophyll index red (CIR) and the novel kernel NDVI (kNVDI) achieved significantly higher prediction powers of around $R^2 \approx 0.61$ and $\text{RMSE}_{10\text{-fold}} \approx 2.57 \text{ } [\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}]^{10\text{-fold}}$. Using all spectral bands and VIs jointly in a machine learning prediction framework allowed us to predict GPP with $R^2 = 0.71$ and $\text{RMSE}_{10\text{-fold}} = 2.68 \text{ } [\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}]^{10\text{-fold}}$. Despite the high-power prediction when machine learning techniques are used, under water-stress scenarios or heat waves, optical information alone is not enough to predict GPP properly. In general, our analyses show the potential of nonlinear combinations of spectral bands and VIs for monitoring GPP across ecosystems at a level of accuracy comparable to previous works, which, however, required additional meteorological drivers.

Index Terms—Gross primary production, red edge, Sentinel-2.

Manuscript received October 13, 2021; revised January 5, 2022; accepted January 17, 2022. Date of publication February 16, 2022; date of current version March 31, 2022. This work was supported by the European Union's Horizon 2020 Research and Innovation Program via the TRuStEE Project through the Marie Skłodowska-Curie Grant under Agreement 721995. (Corresponding author: Daniel E. Pabon-Moreno.)

Daniel E. Pabon-Moreno is with the Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany (e-mail: dpabon@bgc-jena.mpg.de).

Mirco Migliavacca is with the European Commission, Joint Research Centre (JRC), 21027 Ispra, Italy, and also with the Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany (e-mail: mirco.migliavacca@ec.europa.eu).

Markus Reichstein is with the Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany, and also with the Michael-Stifel-Center Jena for Data-Driven and Simulation Science, Friedrich-Schiller-University Jena, 07743 Jena, Germany (e-mail: mreichstein@bgc-jena.mpg.de).

Miguel D. Mahecha is with the Remote Sensing Center for Earth System Research, Leipzig University, 04103 Leipzig, Germany, and also with the Helmholtz Centre for Environmental Research, 04318 Leipzig, Germany (e-mail: miguel.mahecha@uni-leipzig.de).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TGRS.2022.3152272>, provided by the authors. Digital Object Identifier 10.1109/TGRS.2022.3152272

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

I. INTRODUCTION

GROSS primary production (GPP), the amount of carbon absorbed by the ecosystem via plant photosynthesis, is the largest single flux in the global carbon cycle [1]. GPP varies in response to several abiotic (e.g., radiation, temperature, and precipitation; 2 and 3) and biotic factors (e.g., metabolic pathway, vegetation type, leaf chemical traits, and species composition; 4). However, GPP cannot be directly observed and needs to be derived from *in situ* measurements of net CO₂ exchanges using the eddy covariance (EC) technique over canopies [5], [6]. Using different flux partitioning methods, it is possible to estimate the amount of carbon that is taken up by the ecosystem (GPP) or released through ecosystem respiration (RECO) [7]–[11]. Nevertheless, EC can only measure the exchange of energy and matter between the ecosystem and the atmosphere at the scale of the climatology footprint, which can vary between a few hundred meters to a few kilometers (e.g., 12). Today, EC data are available globally in multiple regional networks (Integrated Carbon Observation System: ICOS, The National Ecological Observatory Network: NEON, AmeriFlux, AsiaFlux) and the meta-network Fluxnet [13], [14]. The flux database networks enable studies into local processes understanding [6], [15]–[17], evaluating biotic and abiotic relationships on multiple time scales (e.g., 18 and 19), and evaluating terrestrial biosphere models [20]–[23].

In the last decades, many process-based, semiempirical, and data-driven models have been developed to upscale GPP using remote sensing data, and climate information, in order to understand the spatiotemporal dynamics of the global carbon cycle [3], [24]–[26]. For instance, the MODIS MOD17 product is based on a semiempirical model that estimates GPP as the product between the light-use efficiency and the absorbed photosynthetically active radiation (APAR) [27]. The maximum light-use efficiency is a plant functional-type-dependent parameter, and it is downregulated by stress factors that depend on temperature and vapor pressure deficit that need to be parameterized. The Breathing Earth System Simulator (BESS) [28] is a process-based approach, which relies on a radiative-transfer model coupled with several remote sensing products to predict GPP and evapotranspiration (ET) at a global scale with a temporal resolution of eight days. Jung *et al.* [29] showed that machine learning methods can likewise efficiently upscale fluxes from *in situ* data to the globe. Building on this work, Tramontana *et al.* [30] used the FLUXNET dataset and MODIS remote sensing information to train multiple machine learning techniques to predict monthly GPP at a global scale. Later, Bodesheim *et al.* [31] produced GPP global products at half-hour temporal resolution using different settings, but of low spatial resolution (0.5°). The state-of-the-art machine learning-based upscaling of GPP is described in [26].

A more direct approach to predicting GPP is to identify vegetation indices that are highly correlated with GPP dynamics. Badgley *et al.* [32], for instance, found that the near-infrared reflectance of vegetation (NIRv) index strongly correlates with monthly estimates of sun-induced chlorophyll fluorescence (SIF), rendering it a potential predictor for GPP at the global scale. Later on, Badgley *et al.* [33] showed that NIRv can explain 68% of the monthly GPP variability at the FLUXNET sites. Recently, Camps-Valls *et al.* [34] presented a nonlinear version of the normalized difference vegetation index (NDVI) based on kernel methods (kNDVI) that correlates better with GPP and SIF products than NIRv and NDVI. The advantage of such approaches is that they rely purely on remote sensing data and circumvent the parameterization of light-use efficiency models. However, relying on reflectance values alone means that the detection of physiological regulation of photosynthesis via meteorological conditions is not detectable unless they last long enough to affect vegetation pigments and structure.

Today, new satellite missions have increased the information available to characterize vegetation properties and ecosystem processes [35], [36]. Specifically, the satellite missions from the Copernicus program have opened new ways to monitor ecosystem processes with unprecedented spatial, temporal, and spectral resolution [37], [38]. For instance, it has been shown that Copernicus missions allow deriving plant traits such as chlorophyll and nitrogen content along with other biophysical parameters [39]–[42]. To the best of our knowledge, only three studies have evaluated the prediction capacities of GPP using Sentinel-2: Wolanin *et al.* [43] used the SCOPE model and machine learning techniques to predict GPP of C3 crops. Lin *et al.* [44] evaluated the potential prediction of GPP as a function of the vegetation index multiplied by the incident photosynthetic active radiation (PARin). They analyzed the performance of five red-edge vegetation indices and three nonred-edge vegetation indices. They found that chlorophyll index red (CIR) showed the highest correlation with GPP from the EC tower for two grassland sites. Finally, Cai *et al.* [45] compared GPP predictions using Sentinel-2 and MODIS for several EC-sites in Northern Europe. The authors did not find any improvement for the prediction of GPP when using Sentinel-2 compared to MODIS using the enhanced vegetation index (EVI2). Despite these advances, there is a lack of systematic comparison between novel red-edge vegetation indices and vegetation indices based on the classic red and NIR bands (i.e., NDVI, kNDVI, and NIRv) in terms of their predictive power regarding GPP. Likewise, the question of whether a machine learning approach considering all Sentinel-2 bands could improve the satellite-based predictions of GPP remains unresolved.

In this study, we aim at understanding the potential of Sentinel-2 mission for monitoring GPP across European and North American major biomes at high spatial resolution. First, we want to understand, which vegetation indices or spectral bands available from Sentinel-2 are the most relevant for predicting GPP. Second, we investigate what is the difference in prediction performance between different approaches based on state-of-the-art vegetation indices (e.g., NIRv, kNDVI, red-edge based, and nonred-edge indices) and machine learning using all spectral bands.

II. METHODS

A. Eddy Covariance Sites

We used 58 EC sites compiled by the ICOS Drought 2018 Team (49 sites) and the Ameriflux/ONEFLUX (9 sites) initiatives from 2015 to 2018 (Appendix A). We used half-hourly GPP data (GPP_NT_VUT_USTAR50) estimated using the FLUXNET2015 workflow [14]. GPP is calculated in FLUXNET with the night-time partitioning method [8] using a variable u^* threshold for each year. The annual u^* threshold is derived from the 50th percentile of u^* threshold distribution obtained by bootstrapping the original night-time net ecosystem exchange data [14]. Daily GPP values are estimated as the mean of the half-hourly values where net ecosystem exchange is observed or gap-filled with good quality (e.g., NEE_VUT_USTAR50_QC = 0 and 1). In our analysis, days with less than 70% of good quality half-hourly data were set to “NA.” Finally, we smoothed the time series using a moving window mean with a window size of seven days.

The EC sites span across Europe and United States from a latitude of 34.3°N to 67.8°N and include a variety of vegetation types: croplands (9 sites), deciduous broadleaf forests (9 sites), evergreen needleleaf forests (18 sites), grasslands (7 sites), mixed forest (4 sites), open shrublands (2 sites), savannas (4 sites), and wetlands (5 sites). The sites’ locations represent a variety of climatic regimes, including Mediterranean, humid subtropical, temperate oceanic, humid continental, subarctic, and tundra (Appendix A).

B. Sentinel-2 Imagery

We downloaded Sentinel-2 L1C products for the EC sites from 2015 to 2018 using the SciHub Copernicus portal (<https://scihub.copernicus.eu/>, last accessed October 2020). We performed atmospheric corrections for all products using Sen2Cor 2.5.5 [46]. All bands were resampled to 20-m resolution using the nearest neighbor approach for upsampling and median for downsampling. Finally, we computed several vegetation indices (see Supplementary Material 9) such as NDVI, kNDVI, NIRv, and multiple red-edge vegetation indices as the inverted red-edge chlorophyll index (IRECI) and CIR. Among these indices, kNDVI requires a specific parameterization of the kernel width σ , which was here set to the median distance between the near-infrared band (NIR) and the red band per spatial pixel; for Sentinel-2, $\sigma = \text{median}(0.5 \times (B8 + B4))$. Postprocessing of the images was performed using SNAP v7.0 [47] and automatized using the graph processing framework and the graph processing tool. The scripts for the postprocessing of the products are available at a Zenodo repository (see code availability).

We defined a buffer area of 100 m radius around the EC towers to ensure that the flux footprint climatology lies within this area (Supplementary Material 1). We used the scene classification generated by Sen2Cor to filter out images with: “no data,” “saturated or defective pixels,” “dark areas,” “cloud shadows,” “water,” “cloud,” “thin cirrus,” and “snow.” To reduce the effect of shadows or saturated pixels that are not correctly classified by Sen2Cor, we implemented an outlier detection approach that consists of three steps. First, we computed z -scores (data centered and scaled to unit variance) per image and removed pixels of the buffer area with an absolute residual value higher than quantile

0.99 [48]. Second, to detect potential images with clouds, we used the time series of the spectral bands per site. We then estimated the average of the buffer area for each image/band and decomposed the time series of each band into a seasonal and a trend component using locally estimated scatterplot smoothing [LOESS 49]. Next, we applied an inner quantile range technique over the residual of the time series decomposition [50]. Residuals with values higher or lower value than three times the quantiles 0.25 and 0.75, respectively, were also classified as outliers. This analysis was performed using the “anomalize” R package [50]. Third, we defined a bigger buffer area of 900 m, where we estimated the percentage of clouds. We removed observations where the percentage of clouds was above 70%. We also identified 16 additional images with clouds by visual inspection (Supplementary Material 3). We present the complete description of the time series decomposition and the outlier detection in Supplementary Material 2 (the R scripts are available in the Zenodo repository, see code availability). The minimum number of images per site detected as an outlier is 1, the maximum is 20, and the mean across sites is 6 images. Finally, we selected the daily GPP values for the days when we also have valid images from Sentinel-2.

C. Dataset Balancing

The imbalanced representation of different categories in a dataset can influence the weighting of the observations during the training process and consequently in the quality of the prediction [51]. In the last decades, several methods have been developed to solve this issue, mainly for classifications problems, but recently also for regression analysis [52], [53]. To address this problem for the prediction of GPP through different vegetation types that are not all equally represented (Fig. 1), we implemented three methods to balance the dataset given the differences in the number of observations per vegetation type.

- 1) *Undersampling Balancing*: All observations are grouped by vegetation type and are resampled without replacement, to the number of observations of the vegetation type with the least observations.
- 2) *Oversampling Balancing*: All observations are again grouped by vegetation type. Each category is completed until reaching the number of observations of the maximum category (sampling with replacement). The replacement technique is only applied when the total observations of the category are less than the difference between the number of observations of the category with the maximum number of observations and the total number of observations of the current vegetation type.
- 3) *Synthetic Minority Oversampling TEchnique for Regression (SMOTER) Balancing*: It is a balancing technique proposed by Torgo *et al.* [52], where the idea behind the method is to undersample observations with high frequency. In contrast, values with a low frequency (rare observations) are oversampled. In this form, rare observations will have a higher weight during the training. All the following analyzes were applied considering all three balancing techniques as well as to the imbalanced case.

D. Linear Regression-Based GPP Prediction

We evaluated the performance of red-edge vegetation indices to predict GPP using linear regression using the

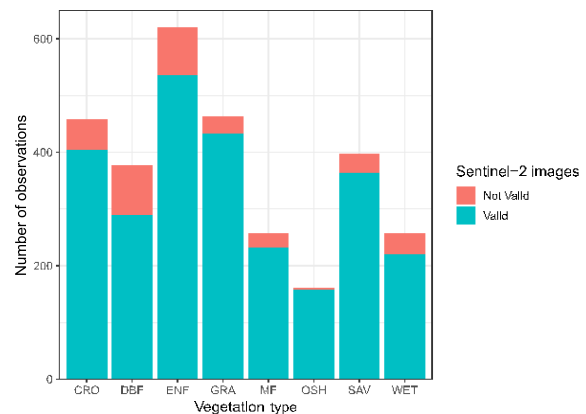


Fig. 1. Number of Sentinel-2 images used for the prediction of GPP (2015–2018) per vegetation type. Each observation corresponds to a Sentinel-2 image at a 100 m radius around the EC tower. Red color indicates the images with no data, saturated or defective pixels, dark areas, cloud shadows, water, clouds, thin cirrus, or snow effects that were removed. Blue color represents the number of valid images. Evergreen needleleaf forests (ENF), croplands (CRO), deciduous broadleaf forests (DBF), grasslands (GRA), wetlands (WET), mixed forest (MF), savannas (SAV), and open shrublands (OSH).

balanced and imbalanced datasets. We compared the performance of NDVI, NIRv, and kNDVI [34], as well as red-edge vegetation indices such as IRECI and CIR (for an overview, see Supplementary Material 9). All evaluations were based on leave-location-and-time-out tenfold cross validation as proposed by Meyer *et al.* [54] and implemented in the “CAST” R package [55]. To increase the robustness of the analysis, the generation of tenfolds was repeated 50 times. In this approach, the partitions for the cross validation are semirandomly generated to minimize spatial and temporal autocorrelation. We evaluated the performances of the different models using the coefficient of determination (R^2) of the linear regression between observed and predicted GPP, the root-mean-square error (RMSE), and the mean absolute error (MAE). Finally, we compared the distributions of the model evaluation metrics between the vegetation indices using the Wilcoxon test [56].

E. Machine Learning-Based GPP Prediction

We used random forests [57] as prediction approach for GPP for each balanced and imbalanced dataset. A detailed description of how to use RF for upscaling land surface fluxes can be found in [31]. We explored what variables are the most relevant for predicting GPP. For this, we evaluated the radiometric indices presented in Supplementary Material 9, additionally to the spectral bands B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, and B12 (Supplementary Material 8) resulting in a total of 35 predictor variables. kNDVI was not included here since it is a nonlinear transformation of the NDVI using kernel methods, and its inclusion would have added no information when applying machine learning techniques. We performed a forward feature selection as suggested by Meyer *et al.* [58], where the models are generated based on the pairs’ combination of predictors, allowing us here to compare nonlinear combinations of spectral bands and vegetation indices, as we may expect that they could reduce model complexity. The power prediction of each model was estimated using a tenfold leave-location-and-time-out cross validation [54], where the tenfolds were generated 50 times to increase the robustness

of the analysis. The idea is that the model with highest R^2 is selected first, and then, new variables are iteratively added to this initial model. The process finishes when none of the remaining variables increases model performance.

III. RESULTS AND DISCUSSION

In the following, we first report the results of the GPP prediction using different vegetation indices in linear regressions, where we specifically discuss the performance of GPP estimates based on red-edge vegetation indices compared with the ones based on NIRv, NDVI, and kNDVI. We also discuss the effect of the balancing techniques on the performance of the prediction. Second, we present the results of the GPP prediction using Sentinel-2 spectral bands and vegetation indices using random forests, where we present examples of the prediction for different EC sites and an entire Sentinel-2 tile. Finally, we discuss the possibilities and limitations of predicting GPP using remote sensing information only and how such prediction can be improved in the future and provide globally continuous flux estimates.

A. GPP Prediction Using Linear Regressions

In Fig. 2, we compare the performance of linear GPP predictions using red-edge-based vegetation indices (CIR and IRECI, see Supplementary Material 9), NIRv, NDVI, and kNDVI. Red-edge vegetation indices perform better than NDVI and NIRv in all considered metrics (Fig. 2), while kNDVI performs as well as IRECI. According to the Wilcoxon test, the differences in the performance distribution of each index are statistically significant. In general, CIR explains on average 3% more of the GPP variance than kNDVI, 4% more than IRECI, 10% more than NIRv, and 11% more variance than NDVI. kNDVI explains an average 1% more than IRECI, 7% more than NIRv, and 8% more than NDVI. The prediction of GPP using IRECI shows that 6% more variance in GPP is explained compared to NIRv and 7% more than NDVI. NIRv only explains 1% more of the GPP variance than NDVI. The RMSE shows smaller errors in GPP estimated with CIR, kNDVI, and IRECI compared to the estimates based on NIRv and NDVI (Fig. 2). As expected, when balanced datasets are used, the explained variance increases 2% for CIR, from 2% to 4% for IRECI, from 2% to 5% for NIRv, from 1% to 3% for kNDVI, and from 2% to 3% for NDVI (Table I and Supplementary Material 4).

Badgley *et al.* [32] introduced the NIRv as an alternative to SIF for the estimation of monthly GPP. Compared to machine learning products or radiative-transfer models, the advantage of this approach is that it could be used to estimate global GPP easily using global and long-term time series products such as MODIS. However, our results suggest that the red-edge vegetation index CIR yields significantly higher prediction powers of GPP compared to NIRv. This finding could be interpreted as an important argument for relying on the novel Sentinel-2 data for GPP prediction.

Red edge is the region around 710 nm, which marks the sharp transition between the red region (700 nm), where the absorption of chlorophyll occurs, and the near-infrared region (730 nm), where the reflectance is produced by the internal structures of the leaf [59, p. 180]. This region is highly sensitive to the leaf chlorophyll content [60], [61]. At the same time, chlorophyll content is a controlling factor of the fraction of photosynthetically active radiation absorbed by

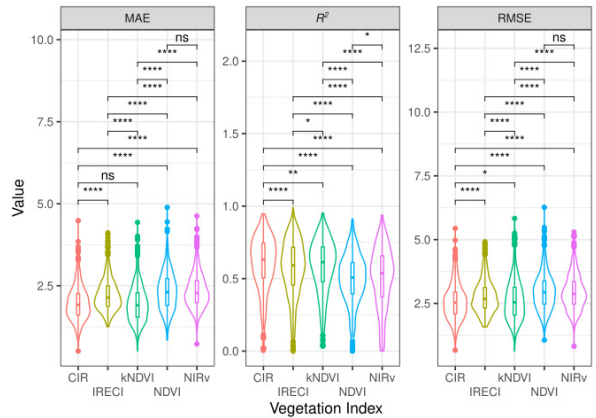


Fig. 2. Prediction of GPP using linear regression and different vegetation indices (CIR: chlorophyll index red, IRECI: inverted red-edge chlorophyll index, NIRv: near-infrared vegetation, and NDVI: normalized difference vegetation index) as predictors. Results are shown for the imbalanced (original) dataset only. The vertical lines correspond to the results of the Wilcoxon test in pairs, where ns is the nonsignificant differences, *: $p \leq 0.05$, **: $p \leq 0.01$, and ***: $p \leq 0.001$.

plants (APAR). This is one possible explanation why CIR is strongly correlated with GPP [62], even if it cannot reflect the fast variations of the photosynthesis itself. For these reasons, VIs based on red-edge bands might generally have advantages for estimating GPP over VIs that do not rely on the red edge. Lin *et al.* [44] found that CIR multiplied by PAR can explain slightly more variability of GPP than NIRv multiplied by PAR for two grasslands sites. However, we would argue that the PAR effect could be dominant in their study, while our aim here was to focus on the spectral information only.

We also tested the predictive performance of kNDVI [34], which was reported to predict monthly GPP better than NIRv. The idea behind kNDVI is to solve the saturation problem of NDVI at high values by exploring the nonlinear relations of the two bands of the NDVI. Even though no red-edge information is used, we found that kNDVI performed at the level of IRECI in our study. One interpretation of this finding is that most of the information contained in the red-edge bands can be captured by an appropriate transformation of the distance between near-infrared and red bands. However, there is no direct mechanistic argument, and it is unclear to what extent this observation is general and further research will be necessary. However, our results may imply that kernel versions of classical vegetation indices could derive relevant information from satellite missions that do not have red-edge indices.

B. GPP Prediction Using Random Forest

Another question of this study was whether machine learning could outperform even the new generation of vegetation indices. In Table II, we present the results of the variable selection analysis where a different number of predictors are selected depending on the balancing technique. From 35 predictors that included Sentinel-2 spectral bands (Supplementary Material 8) and derived vegetation indices (Supplementary Material 9), CIR, S2REP, and B1 are selected for all datasets, while GNDVI, PSSRA B3, and B4 are selected at least in three cases. ARVI, MTCI, MCARI, B2, and B5 are selected at least in two datasets. IRECI, NDI45, RVI, TNDVI, TSAVI,

Chapter 4. On the Potential of Sentinel-2 for Estimating Gross Primary Production

TABLE I
AVERAGE PERFORMANCE OF THE GPP PREDICTION USING LINEAR REGRESSION (TENFOLD TEMPORAL–SPATIAL CROSS VALIDATION) WITH RED-EDGE AND NONRED-EDGE VEGETATION INDICES. THE COLUMN DATASET REFERS TO THE BALANCING TECHNIQUE USED TO BALANCE THE REPRESENTATION OF DIFFERENT VEGETATION TYPES

Dataset	Red-edge Vegetation indices						Non Red-edge vegetation indices								
	IRECI			CIR			NIRv			NDVI			kNDVI		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Imbalanced	0.57	2.77	2.23	0.61	2.57	1.98	0.51	2.95	2.36	0.50	2.98	2.34	0.58	2.65	1.97
Undersampling	0.61	2.51	2.05	0.62	2.43	1.91	0.56	2.70	2.15	0.53	2.84	2.31	0.61	2.48	1.91
Oversampling	0.59	2.63	2.09	0.61	2.51	1.93	0.53	2.80	2.20	0.50	2.95	2.36	0.59	2.59	1.94
SMOTER	0.60	3.29	2.75	0.62	3.19	2.64	0.55	3.45	2.84	0.52	3.74	3.07	0.54	3.53	2.83

TABLE II
VARIABLES SELECTED FOR THE PREDICTION OF GPP USING THE FORWARD FEATURE SELECTION PRESENTED BY MEYER *et al.* [58]. THE DATASET COLUMN REPRESENTS THE BALANCING TECHNIQUE USED TO BALANCE THE DIFFERENT VEGETATION TYPES IN THE ORIGINAL (IMBALANCED) DATASET. A TENFOLD CROSS VALIDATION IS PERFORMED TO ESTIMATE: R^2 , RMSE, AND MAE. THE OPTIMUM NUMBER OF VARIABLES RANDOMLY SAMPLED AS CANDIDATES FOR EACH SPLIT (MTRY) IS ALSO SHOWN. THE INCREASE OF R^2 AND THE DECREASE IN THE STANDARD ERROR WHEN EACH VARIABLE IS ADDED TO THE INITIAL MODEL ARE SHOWN, WHERE THE FIRST VALUE USING THE FIRST TWO PREDICTORS IN THE COLUMN VARIABLES SELECTED AND THE LAST VALUE CORRESPONDS TO THE VALUE OF THE FINAL MODEL. RED-EDGE CHLOROPHYLL INDEX POSITION INDEX (S2REP), ATMOSPHERICALLY RESISTANT VEGETATION CHLOROPHYLL INDEX (MTCI), GREEN NORMALIZED DIFFERENCE NORMALIZED DIFFERENCE VEGETATION INDEX (NDI45), INFRARED PERCENTAGE VEGETATION (PSSRA), TRANSFORMED SOIL ADJUSTED ABSORPTION RATIO INDEX (CIR), SENTINEL-2 RED-EDGE INDEX (ARVI), MERIS TERRESTRIAL VEGETATION INDEX (GNDVI), TRANSFORMED (TNDVI), NORMALIZED DIFFERENCE INDEX 45 INDEX (IPVI), PIGMENT SPECIFIC SIMPLE RATIO VEGETATION INDEX (TSAVI), MODIFIED CHLOROPHYLL (MCARI), AND GREEN CHLOROPHYLL INDEX (CIG)

Dataset	Number of observations	Number of variables selected	$R^2_{10-fold}$ final model	$RMSE_{10-fold}$ final model	$MAE_{10-fold}$ final model	mtry	Variables selected	$R^2_{10-fold}$	Standard Error
Imbalanced	2636	9	0.66	2.34	1.76	2	CIR, B1	0.593	0.008
							B3	0.632	0.008
							B4	0.649	0.008
							B2	0.653	0.008
							B5	0.655	0.008
							PSSRA	0.655	0.008
							S2REP	0.657	0.008
GNDVI	0.659	0.007							
Undersampling	1264	12	0.68	2.20	1.68	2	CIR, B1	0.620	0.010
							B5	0.652	0.009
							TNDVI	0.664	0.009
							PSSRA	0.669	0.009
							NDVI45	0.671	0.009
							GNDVI	0.671	0.009
							IRECI	0.671	0.009
							MTCI	0.673	0.009
							RVI	0.674	0.009
							S2REP	0.674	0.009
ARVI	0.675	0.009							
Oversampling	4288	9	0.67	2.28	1.70	2	CIR, B1	0.582	0.008
							B3	0.632	0.008
							B4	0.656	0.007
							GNDVI	0.661	0.007
							PSSRA	0.665	0.007
							S2REP	0.668	0.007
							MCARI	0.669	0.007
							ARVI	0.670	0.007
SMOTER	2635	11	0.71	2.68	2.10	2	CIR, B4,	0.635	0.009
							B3	0.664	0.009
							B2	0.685	0.008
							MTCI	0.690	0.008
							S2REP	0.697	0.008
							B12	0.700	0.008
							B1	0.702	0.008
							TSAVI	0.703	0.008
							MCARI	0.705	0.008
							CIG	0.706	0.008

CIG, and B12 are selected at least once (Table II). The variable selection analysis shows that even when nonlinear combinations of spectral bands are possible, vegetation indices are still selected as they probably would simplify the machine learning model. Yet, not all information required for predicting GPP seems to be encoded in vegetation indices alone. Bands B1, B2, B3, B4, B5, and B12 also appear to provide

information that is useful for the predictions. A surprising result is the selection of band B1. This band is typically used for aerosol detection and correction purposes. We speculate that B1 is a proxy for radiation dynamics (e.g., direct and diffuse radiation) that are important for GPP. However, we note that Penuelas *et al.* [63] had considered this spectral region earlier in their structure insensitive pigment index (SIPI)

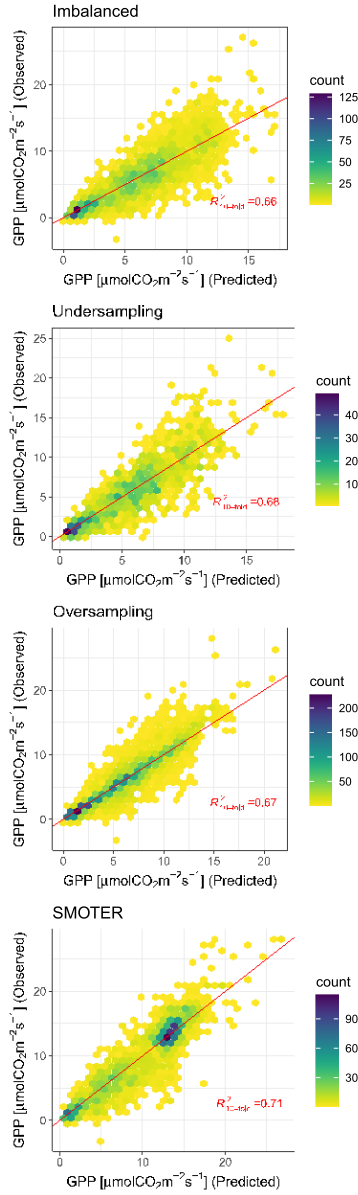


Fig. 3. Prediction of GPP using different data balancing techniques. In each case, the observed values are shown on the y -axis, and the predicted values are shown on the x -axis. The red line represents the 1:1 line. Imbalanced makes reference to the original dataset. Undersampling, oversampling, and SMOTER make reference to each technique used to balance the dataset (see Section II for further details).

that has, however, not been developed further for vegetation monitoring. The additional selection of bands B2 (blue), B3 (green), B4 (red), and B5 (vegetation red-edge) suggests that there is space for the development of new vegetation indices that can capture the GPP variability beyond the existing indices.

In Fig. 3, we present the prediction of GPP using random forest regression, where GPP can be predicted with

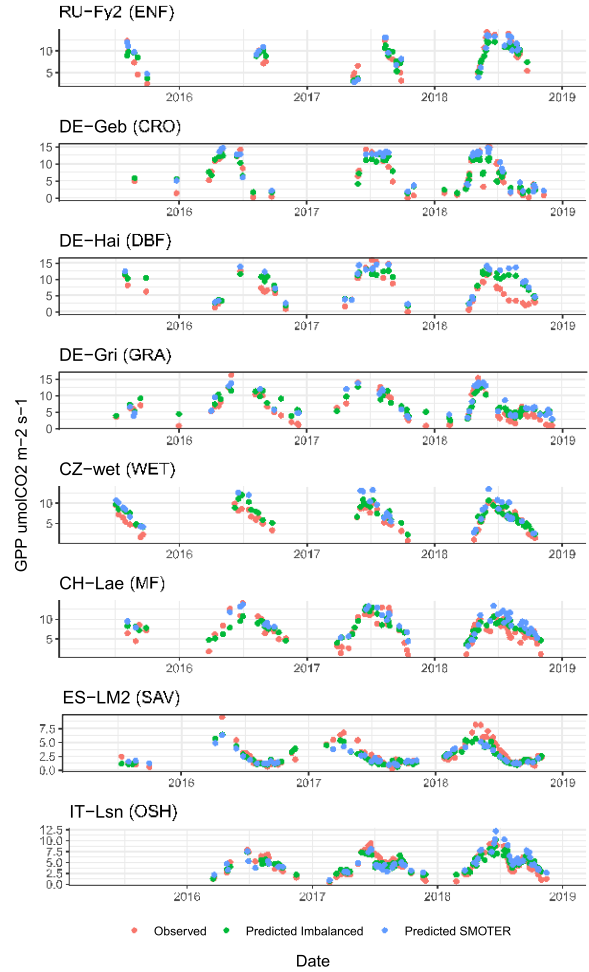


Fig. 4. Observed and predicted GPP values at selected EC sites representing different vegetation types: evergreen needleleaf forests (ENF), croplands (CRO), deciduous broadleaf forests (DBF), grasslands (GRA), wetlands (WET), mixed forest (MF), savannas (SAV), and open shrublands (OSH). Models were trained using a leave-one-site-out cross-validation strategy.

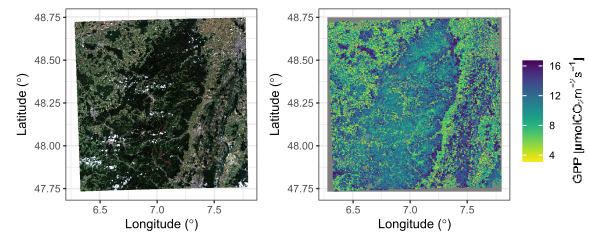


Fig. 5. GPP product for a Sentinel-2 tile, over the Ballons des Vosges Regional Nature Park (France, June 23, 2020). The land cover classification generated by Sen2cor was applied before the prediction, where pixels considered as nonvegetation are encoded as NAs.

$R_{10\text{-fold}}^2 = 0.66$ and $\text{RMSE}_{10\text{-fold}} = 2.34$ [$\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$] for the imbalanced dataset. There are improvements in the variance explained using the balanced dataset. $R_{10\text{-fold}}^2 = 0.68$ and $\text{RMSE}_{10\text{-fold}} = 2.20$ [$\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$]

using undersampling, $R_{10\text{-fold}}^2 = 0.67$ and $\text{RMSE}_{10\text{-fold}} = 2.28$ [$\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$] using oversampling technique, and $R_{10\text{-fold}}^2 = 0.71$ and $\text{RMSE}_{10\text{-fold}} = 2.68$ [$\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$] using the SMOTER technique (Table II). The comparison between the distribution of the metrics shows that there are significant differences between the imbalanced and balanced datasets (Supplementary Material 10). The results of the cross validation for each fold and balancing technique are presented in Supplementary Material 7. Tramontana *et al.* [30] reported that spectral information with machine learning techniques can explain around 78% of the GPP variability across sites. One of the advantages of our approach is that it does not require a previous vegetation-type classification [64]. In comparison with the estimation of GPP using biophysical parameters as, e.g., in [44], we show that it GPP can be estimated more directly with high accuracy.

In Fig. 4, we present the examples of predicted and observed GPP representing different vegetation types. The prediction for each site is presented in Supplementary Material 5. Despite the overall high variances explained by random forests, there are indeed cases when GPP cannot be predicted correctly. For instance, the maximum GPP is underestimated in savannas and evergreen needleleaf forest ecosystems. Our study period covers the 2018 heat wave, an extreme event where northwestern Europe vegetation was highly affected [65]–[67]. We find, however, that the reduction in CO_2 uptake during this event was not well captured for mixed forest and deciduous broadleaf forest (Fig. 4). This can also be seen when comparing the time series of 2016 and 2017 to 2018 (see Fig. 4). This means that the prediction of ecosystem fluxes during extreme events remains an open issue that needs to be addressed with high priority as discussed in [68]. However, our finding that GPP dynamics during drought events cannot be well represented is in-line with earlier findings. For instance, Bodesheim *et al.* [31] showed that GPP was not properly predicted during dry summers for several EC sites and attributed this to the poor representation of water availability in their dataset. Different from our study, their study also used climate information, which, in theory, increases the model performance for water-stress scenarios. One general problem could be the time lag between the change of photosynthesis rates and the decline in the concentration of the pigments, including chlorophyll content, in the leaves. However, given that the data generated here are based on vegetation reflectance properties only, it is expected that they can only pick up changes in GPP that are primarily driven by changes in APAR and pigment concentrations but are not apt to capture the fast response of photosynthesis mediated, e.g., by stomatal closure. This limitation is inherent to all reflectance-based methods and the reason why, in some sites, we are not able to reproduce GPP dynamics under stress.

Nevertheless, the overall seasonal dynamics are represented very well in our GPP estimates across sites and vegetation types. Future studies should investigate whether the inclusion of thermal information from Sentinel-3 or radar information from Sentinel-1 can help to indirectly address the water deficit in the ecosystems during drought periods [69] and lead to the next generation of operational GPP products based on

remote sensing data only. In addition, the unique combination of red-edge vegetation indices in Sentinel-2, radar information from Sentinel-1, or multispectral and thermal information from the bands available in Sentinel-3 may open unprecedented possibilities for vegetation monitoring in the near future [35].

Previous studies used plant functional classes as a spatial feature to upscale GPP [30], [70]. To use vegetation types as a predictor of GPP, a necessary step will be to improve the land cover maps to match the resolution of Sentinel-2. The ESA WorldCover consortium gave the first steps, producing the first global land cover map at 10-m resolution for 2020 using radar information from Sentinel-1 and optical information from Sentinel-2 [71]. Future research will have to test the added value of these upcoming products for predicting carbon fluxes at high spatial resolution.

To give a taste of what the mapping of carbon fluxes might look like in the future, in Fig. 5, we present an example of the upscaling of GPP for a Sentinel-2 tile over the Ballons des Vosges Regional Nature Park (France, June 23, 2020; Supplementary Material 6). The area contains different types of deciduous broadleaf forest, wetlands, grasslands, and croplands. Even though our model does not use vegetation type as a predictor, it does clearly differentiate GPP dynamics of crops, wetlands, and forests. The high spatial resolution of Sentinel-2 could be a nice avenue to monitor forests with a high degree of fragmentation [72] or even green areas in cities [73]. A tutorial of how to use the final models produced in our study to upscale GPP using any Sentinel-2 L2A product provided by Copernicus-ESA is presented in the code repository.

IV. CONCLUSION

In this study, we explore how remote sensing information provided by Sentinel-2 can be used to predict GPP across a variety of vegetation types. We find that the CIR explains an average 10% more of the variability of GPP at daily scale than NIRv and 11% more than NDVI using linear regressions. The high correspondence between kNDVI and IRECI is unanticipated and requires further physical exploration. The prediction power of vegetation indices can be slightly outperformed using machine learning: using random forests, the spectral information provided by Sentinel-2 alone can predict an average 68% of GPP variability (cross-validated). However, under extreme climate conditions such as the 2018 drought/heat wave, meteorological data or thermal information might be necessary to improve the prediction of short-term reduction of GPP that is not associated with changes in APAR or the decline of chlorophyll content. From a methodological point of view, we also explored whether balancing techniques can help to represent vegetation types and rare observations. Furthermore, we found that improvements in the prediction accuracy of GPP are associated with the use of balanced datasets for training. Overall, our study presents a first attempt to assess the capability of Sentinel-2 data alone to predict GPP. Despite the discussed limitations, Sentinel-2 generally offers a highly relevant perspective to map fluxes at high spatial resolution, opening new ways to understand ecosystem processes and responses from local to global scale.

APPENDIX A: EDDY COVARIANCE SITES

SITES FROM THE ICOS DROUGHT 2018 TEAM AND ONEFLUX INITIATIVES USED IN THIS STUDY. THE NUMBER OF OBSERVATIONS CORRESPONDS TO THE NUMBER OF VALID SENTINEL-2 IMAGES RECOVERED FOR THE SITE DURING THE TIME PERIOD. THE VEGETATION TYPE FOR EACH SITE IS PRESENTED: MF = MIXED FORESTS, CRO = CROPLANDS, GRA = GRASSLANDS, ENF = EVERGREEN NEEDLELEAF FORESTS, DBF = DECIDUOUS BROADLEAF FORESTS, WET = WETLANDS, SAV = SAVANNAS, AND OSH = OPEN SHRUBLANDS

Site name	Vegetation type	Number of observations	Years	DOI	Reference
BE-Bra	MF	56	2015-2018	https://doi.org/10.18160/F738-634R	[74]
BE-Lon	CRO	21	2015-2018	https://doi.org/10.18160/6SM0-NFES	[75]
BE-Vie	MF	12	2016-2018	https://doi.org/10.18160/MK3Q-BBEK	[76]
CH-Aws	GRA	31	2016-2018	https://doi.org/10.18160/3YQE-7BR8	NA
CH-Cha	GRA	98	2015-2018	https://doi.org/10.18160/GMMW-5E2D	[77]
CH-Dav	ENF	13	2015-2018	https://doi.org/10.18160/R86M-H3HX	[78]
CH-Fru	GRA	79	2015-2018	https://doi.org/10.18160/J938-0MKS	[79]
CH-Lac	MF	80	2015-2018	https://doi.org/10.18160/FABD-SVJJ	[80]
CH-Oc2	CRO	44	2015-2018	https://doi.org/10.18160/N01Y-R7DF	[81]
CZ-BK1	ENF	22	2015-2018	https://doi.org/10.18160/7QXR-AYEE	[82]
CZ-Lnz	MF	84	2015-2018	https://doi.org/10.18160/84SN-YBSD	NA
CZ-RAJ	ENF	50	2015-2018	https://doi.org/10.18160/HFS9-JBTG	NA
CZ-Stn	DBF	22	2015-2018	https://doi.org/10.18160/V2JN-DQPJ	NA
CZ-wct	WET	68	2015-2018	https://doi.org/10.18160/W4YS-463W	[83]
DE-Akm	WET	39	2015-2018	https://doi.org/10.18160/24B5-J44F	NA
DE-Geb	CRO	60	2015-2018	https://doi.org/10.18160/ZK18-3YW3	[84]
DE-Gri	GRA	73	2015-2018	https://doi.org/10.18160/EN60-T3FG	[85]
DE-Hai	DBF	48	2015-2018	https://doi.org/10.18160/D4ET-BFPS	[86]
DE-HoH	DBF	37	2015-2018	https://doi.org/10.18160/J1YB-YEHC	NA
DE-Hte	WET	44	2015-2018	https://doi.org/10.18160/J1YB-YEHC	NA
DE-Hzd	DBF	28	2015-2018	https://doi.org/10.18160/PJEC-43XB	NA
DE-Kli	CRO	56	2015-2018	https://doi.org/10.18160/STT9-TBJZ	[85]
DE-Obe	ENF	6	2015-2018	https://doi.org/10.18160/FSM3-RC5F	NA
DE-RuR	GRA	39	2015-2018	https://doi.org/10.18160/HPV9-K8R1	[87]
DE-RuS	CRO	28	2015-2018	https://doi.org/10.18160/A2TK-QD5U	[88]
DE-RuW	ENF	20	2015-2018	https://doi.org/10.18160/H7Y6-2R1H	NA
DE-Tha	ENF	45	2015-2018	https://doi.org/10.18160/BSE6-EMVJ	[89]
DK-Sor	DBF	53	2015-2018	https://doi.org/10.18160/BFDT-7HYE	[90]
ES-Abr	SAV	171	2015-2018	https://doi.org/10.18160/11TP-MX4F	[91]
ES-LM1	SAV	80	2015-2018	https://doi.org/10.18160/FDSD-GVRS	[92]
ES-LM2	SAV	92	2015-2018	https://doi.org/10.18160/3SVJ-XSB7	[92]
FI-Hyy	ENF	22	2015-2018	https://doi.org/10.18160/0JHQ-BZMU	
FI-Let	ENF	17	2017-2018	https://doi.org/10.18160/0JHQ-BZMU	[93]
FI-Sii	WET	23	2016-2018	https://doi.org/10.18160/0RE3-DTWD	NA
FI-Var	ENF	38	2016-2018	https://doi.org/10.18160/NYH7-5JEB	NA
FR-EM2	CRO	59	2017-2018	https://doi.org/10.18160/HC1V-8VKJ	NA

Chapter 4. On the Potential of Sentinel-2 for Estimating Gross Primary Production

(Continued.) SITES FROM THE ICOS DROUGHT 2018 TEAM AND ONEFLUX INITIATIVES USED IN THIS STUDY. THE NUMBER OF OBSERVATIONS CORRESPONDS TO THE NUMBER OF VALID SENTINEL-2 IMAGES RECOVERED FOR THE SITE DURING THE TIME PERIOD. THE VEGETATION TYPE FOR EACH SITE IS PRESENTED: MF = MIXED FORESTS, CRO = CROPLANDS, GRA = GRASSLANDS, ENF = EVERGREEN NEEDLELEAF FORESTS, DBF = DECIDUOUS BROADLEAF FORESTS, WET = WETLANDS, SAV = SAVANNAS, AND OSH = OPEN SHRUBLANDS

FR-Hes	DBF	76	2015-2018	https://doi.org/10.18160/WTYC-JVQV	NA
IT-BCi	CRO	38	2015-2018	https://doi.org/10.18160/T25N-PDIH	[94]
IT-Lsn	OSH	113	2016-2018	https://doi.org/10.18160/RTKZ-VTDJ	NA
IT-Tor	GRA	68	2015-2018	https://doi.org/10.18160/ERMH-PSVW	[95]
NL-Loo	ENF	43	2015-2018	https://doi.org/10.18160/MV3K-WM09	[96]
RU-Fy2	ENF	44	2015-2018	https://doi.org/10.18160/WEV2-WQXY	[97]
RU-Fyo	ENF	46	2015-2018	https://doi.org/10.18160/4J2N-DY7S	[98]
SE-Deg	WET	46	2015-2018	https://doi.org/10.18160/0T47-MEEU	NA
SE-Htm	ENF	35	2015-2018	https://doi.org/10.18160/17FF-96RT	NA
SE-Lnn	CRO	45	2015-2018	https://doi.org/10.18160/5GZQ-S6Z0	NA
SE-Nor	ENF	37	2015-2018	https://doi.org/10.18160/K57M-TVGE	NA
SE-Ros	ENF	58	2015-2018	https://doi.org/10.18160/ZF2F-82Q7	NA
SE-Svb	ENF	38	2015-2018	https://doi.org/10.18160/X57W-HWTE	NA
US-ARM	CRO	53	2016-2018	https://doi.org/10.17190/AMF/1246027	[99]
US-Bar	DBF	4	2016-2018	https://doi.org/10.17190/AMF/1246030	[100]
US-Ho1	ENF	14	2015-2018	https://doi.org/10.17190/AMF/1246061	[101]
US-MMS	DBF	8	2015-2018	https://doi.org/10.17190/AMF/1246080	[102]
US-Seg	GRA	45	2015-2018	https://doi.org/10.17190/AMF/1246124	[103]
US-Ses	OSH	45	2016-2017	https://doi.org/10.17190/AMF/1246125	[104]
US-UMB	DBF	14	2015-2017	https://doi.org/10.17190/AMF/1246107	[105]
US-Vcm	ENF	10	2016-2017	https://doi.org/10.17190/AMF/1246121	[106]
US-Wjs	SAV	20	2015-2017	https://doi.org/10.17190/AMF/1246120	[107]

CODE AVAILABILITY

Code is available under GPL-3 license at: https://github.com/dpabon/Sentinel-2_GPP.

DATA AVAILABILITY

ICOS data are available on the web-site: <https://www.icos-cp.eu/data-products/YVR0-4898>. Ameriflux data are available on the website: <https://ameriflux.lbl.gov/data/download-data-oneflux-beta/>.

ACKNOWLEDGMENT

Markus Reichstein and Miguel D. Mahecha thank the European Space Agency for funding the project DeepExtremes-AI for Science, Multi-Hazards, Compounds and Cascade events. This work used eddy covariance data acquired and shared by Ameriflux, and the Integrated Carbon Observation System (ICOS). In addition, funding for AmeriFlux data resources was provided by the U.S. Department of Energy's Office of Science. The authors thank Prof. Dr. Gustau Camps-Valls for his feedback on kNDVI.

REFERENCES

- [1] F. S. Chapin *et al.*, "Reconciling carbon-cycle concepts, terminology, and methods," *Ecosystems*, vol. 9, no. 7, pp. 1041–1050, Nov. 2006.
- [2] A. D. Richardson, D. Y. Hollinger, J. D. Aber, S. V. Ollinger, and B. H. Braswell, "Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange," *Global Change Biol.*, vol. 13, no. 4, pp. 788–803, Jan. 2007.
- [3] C. Beer *et al.*, "Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate," *Science*, vol. 329, pp. 834–838, Aug. 2010.
- [4] T. Musavi *et al.*, "Stand age and species richness dampen interannual variation of ecosystem-level photosynthetic capacity," *Nature Ecol. Evol.*, vol. 1, no. 2, pp. 1–7, Jan. 2017.
- [5] D. D. Baldocchi, "Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future," *Global Change Biol.*, vol. 9, no. 4, pp. 479–492, Apr. 2003.
- [6] D. D. Baldocchi *et al.*, "Outgoing near-infrared radiation from vegetation scales with canopy photosynthesis across a spectrum of function, structure, physiological capacity, and weather," *J. Geophys. Res., Biogeosci.*, vol. 125, no. 7, Jul. 2020, Art. no. e2019JG005534.
- [7] E. Falge *et al.*, "Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements," *Agricult. Forest Meteorol.*, vol. 113, nos. 1–4, pp. 53–74, Dec. 2002.
- [8] M. Reichstein *et al.*, "On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm," *Global Change Biol.*, vol. 11, no. 9, pp. 1424–1439, Sep. 2005.

- [9] G. Lasslop *et al.*, "Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation," *Global Change Biol.*, vol. 16, no. 1, pp. 187–208, Jan. 2010.
- [10] T. F. Keenan *et al.*, "Widespread inhibition of daytime ecosystem respiration," *Nature Ecol. Evol.*, vol. 3, no. 3, pp. 407–415, Mar. 2019.
- [11] G. Tramontana *et al.*, "Partitioning net carbon dioxide fluxes into photosynthesis and respiration using neural networks," *Global Change Biol.*, vol. 26, no. 9, pp. 5235–5253, 2020.
- [12] M. Aubinet, T. Vesala, and D. Papale, Eds., *Eddy Covariance: A Practical Guide to Measurement and Data Analysis* (Springer Atmospheric Sciences). Amsterdam, The Netherlands: Springer, 2012.
- [13] D. Baldocchi *et al.*, "FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities," *Bull. Amer. Meteorol. Soc.*, vol. 82, no. 82, pp. 2415–2434, 2001.
- [14] G. Pastorello *et al.*, "The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data," *Sci. Data*, vol. 7, no. 1, pp. 1–27, 2020.
- [15] D. Baldocchi, "'Breathing' of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems," *Austral. J. Botany*, vol. 56, no. 1, pp. 1–26, Feb. 2008.
- [16] D. E. Pabon-Moreno, T. Musavi, M. Migliavacca, M. Reichstein, C. Römermann, and M. D. Mahecha, "Ecosystem physio-phenology revealed using circular statistics," *Biogeosciences*, vol. 17, no. 15, pp. 3991–4006, Aug. 2020.
- [17] M. Migliavacca *et al.*, "The three major axes of terrestrial ecosystem function," *Nature*, vol. 598, no. 7881, pp. 468–472, Oct. 2021.
- [18] M. D. Mahecha *et al.*, "Characterizing ecosystem-atmosphere interactions from short to interannual time scales," *Biogeosciences*, vol. 4, no. 5, pp. 743–758, Sep. 2007.
- [19] C. Krich *et al.*, "Estimating causal networks in biosphere-atmosphere interaction with the PCMC1 approach," *Biogeosciences*, vol. 17, no. 4, pp. 1033–1061, Feb. 2020.
- [20] A. D. Friend *et al.*, "FLUXNET and modelling the global carbon cycle," *Global Change Biol.*, vol. 13, no. 3, pp. 610–633, Mar. 2007.
- [21] M. D. Mahecha *et al.*, "Comparing observations and process-based simulations of biosphere-atmosphere exchanges on multiple timescales," *J. Geophys. Res., Biogeosci.*, vol. 115, no. G2, Jun. 2010, Art. no. G02003.
- [22] G. B. Bonan *et al.*, "Improving canopy processes in the community land model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data," *J. Geophys. Res.*, vol. 116, no. G2, 2011, Art. no. G02014.
- [23] M. Sjöström *et al.*, "Evaluation of MODIS gross primary productivity for Africa using eddy covariance data," *Remote Sens. Environ.*, vol. 131, pp. 275–286, Apr. 2013.
- [24] F. A. Heinsch *et al.*, *Users Guide: GPP and NPP (MOD17A2/A3) Products NASA MODIS Land Algorithm*. Missoula, MT, USA: Univ. Montana, 2003.
- [25] Y. Ryu, J. A. Berry, and D. D. Baldocchi, "What is global photosynthesis? History, uncertainties and opportunities," *Remote Sens. Environ.*, vol. 223, pp. 95–114, Mar. 2019.
- [26] M. Jung *et al.*, "Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach," *Biogeosciences*, vol. 17, no. 5, pp. 1343–1365, Mar. 2020.
- [27] S. W. Running, R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto, "A continuous satellite-derived measure of global terrestrial primary production," *BioScience*, vol. 54, no. 6, pp. 547–560, Jun. 2004.
- [28] C. Jiang and Y. Ryu, "Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from breathing Earth system simulator (BESS)," *Remote Sens. Environ.*, vol. 186, pp. 528–547, Dec. 2016.
- [29] M. Jung *et al.*, "Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations," *J. Geophys. Res.*, vol. 116, pp. 1–16, Sep. 2011.
- [30] G. Tramontana *et al.*, "Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms," *Biogeosciences*, vol. 13, no. 14, pp. 4291–4313, Jul. 2016.
- [31] P. Bodesheim, M. Jung, F. Gans, M. D. Mahecha, and M. Reichstein, "Upscaled diurnal cycles of land-atmosphere fluxes: A new global half-hourly data product," *Earth Syst. Sci. Data*, vol. 10, no. 3, pp. 1327–1365, Jul. 2018.
- [32] G. Badgley, C. B. Field, and J. A. Berry, "Canopy near-infrared reflectance and terrestrial photosynthesis," *Sci. Adv.*, vol. 3, no. 3, Mar. 2017, Art. no. e1602244.
- [33] G. Badgley, L. D. L. Anderegg, J. A. Berry, and C. B. Field, "Terrestrial gross primary production: Using NIRV to scale from site to globe," *Global Change Biol.*, vol. 25, no. 11, pp. 3731–3740, 2019.
- [34] G. Camps-Valls *et al.*, "A unified vegetation index for quantifying the terrestrial biosphere," *Sci. Adv.*, vol. 7, no. 9, Feb. 2021, Art. no. eabc7447.
- [35] X. Ma *et al.*, "Monitoring plant functional diversity using the reflectance and echo from space," *Remote Sens.*, vol. 12, no. 8, p. 1248, Apr. 2020.
- [36] Y. Zhang, M. Migliavacca, J. Penuelas, and W. Ju, "Advances in hyperspectral remote sensing of vegetation traits and functions," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112121.
- [37] R. Fernandez-Beltran, F. Pla, J. Kang, J. Moreno, and A. Plaza, "Sentinel-3/FLEX biophysical product confidence using Sentinel-2 land-cover spatial distributions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3447–3461, 2021.
- [38] X. Ma *et al.*, "Inferring plant functional diversity from space: The potential of Sentinel-2," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111368.
- [39] W. J. Frampton, J. Dash, G. Watmough, and E. J. Milton, "Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation," *ISPRS J. Photogramm. Remote Sens.*, vol. 82, pp. 83–92, Aug. 2013.
- [40] R. Darvishzadeh *et al.*, "Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 79, pp. 58–70, Jul. 2019.
- [41] T. W. Gara, R. Darvishzadeh, A. K. Skidmore, T. Wang, and M. Heurich, "Accurate modelling of canopy traits from seasonal Sentinel-2 imagery based on the vertical distribution of leaf traits," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 108–123, Nov. 2019.
- [42] A. M. Ali *et al.*, "Comparing methods for mapping canopy chlorophyll content in a mixed mountain forest using Sentinel-2 data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 87, May 2020, Art. no. 102037.
- [43] A. Wolanin *et al.*, "Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations," *Remote Sens. Environ.*, vol. 225, pp. 441–457, May 2019.
- [44] S. Lin, J. Li, Q. Liu, L. Li, J. Zhao, and W. Yu, "Evaluating the effectiveness of using vegetation indices based on red-edge reflectance from Sentinel-2 to estimate gross primary productivity," *Remote Sens.*, vol. 11, no. 11, p. 1303, May 2019.
- [45] Z. Cai *et al.*, "Modelling daily gross primary productivity with Sentinel-2 data in the Nordic region—comparison with data from MODIS," *Remote Sens.*, vol. 13, no. 3, p. 469, Jan. 2021.
- [46] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, "Sen2Cor for Sentinel-2," *Proc. SPIE*, vol. 10427, Oct. 2017, Art. no. 1042704.
- [47] *SNAP-ESA Sentinel Application Platform*, Brockmann Consult, Sky-Watch, and C-S, Hamburg, Germany, 2020.
- [48] R. E. Shiffler, "Maximum Z scores and outliers," *Amer. Statist.*, vol. 42, no. 1, pp. 79–80, Feb. 1988, doi: 10.1080/00031305.1988.10475530.
- [49] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition," *J. Off. Statist.*, vol. 6, no. 1, pp. 3–73, 1990.
- [50] M. Dancho and D. Vaughan, *Anomalize: Tidy Anomaly Detection*, R Package, 2020.
- [51] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
- [52] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, "SMOTE for regression," in *Progress in Artificial Intelligence*, L. Correia, L. P. Reis, and J. Cascalho, Eds. Berlin, Germany: Springer, 2013, pp. 378–389.
- [53] P. Branco, L. Torgo, and R. P. Ribeiro, "SMOGR: A pre-processing approach for imbalanced regression," in *Proc. 1st Int. Workshop Learn. With Imbalanced Domains, Theory Appl.*, 2017, pp. 36–50.
- [54] H. Meyer, C. Reudenbach, T. Hengl, M. Katurji, and T. Nauss, "Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation," *Environ. Model. Softw.*, vol. 101, pp. 1–9, Mar. 2018.
- [55] H. Meyer, *CAST: 'Caret' Applications for Spatial-Temporal Models*, R Package, 2020.

- [56] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.
- [57] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [58] H. Meyer, C. Reudenbach, S. Wöllauer, and T. Nauss, "Importance of spatial predictor variable selection in machine learning applications—Moving from data reproduction to spatial prediction," *Ecological Model.*, vol. 411, Nov. 2019, Art. no. 108815.
- [59] H. G. Jones and R. A. Vaughan, *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. Oxford, U.K.: OUP Oxford, Jul. 2010.
- [60] J. R. Miller, E. W. Hare, and J. Wu, "Quantitative characterization of the vegetation red edge reflectance 1. An inverted-Gaussian reflectance model," *Int. J. Remote Sens.*, vol. 11, no. 10, pp. 1755–1773, 1990.
- [61] A. Gitelson and M. N. Merzlyak, "Quantitative estimation of chlorophyll—A using reflectance spectra: Experiments with autumn chestnut and Maple leaves," *J. Photochem. Photobiol. B, Biol.*, vol. 22, no. 3, pp. 247–252, Mar. 1994.
- [62] A. A. Gitelson, Y. Gritz, and M. N. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," *J. Plant Physiol.*, vol. 160, no. 3, pp. 271–282, 2003.
- [63] J. Penuelas, F. Baret, and I. Filella, "Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance," *Photosynthetica*, vol. 31, no. 2, pp. 221–230, 1995.
- [64] S. Lin *et al.*, "Improved global estimations of gross primary productivity of natural vegetation types by incorporating plant functional type," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 100, Aug. 2021, Art. no. 102328.
- [65] C. Albergel *et al.*, "Monitoring and forecasting the impact of the 2018 summer heatwave on vegetation," *Remote Sens.*, vol. 11, no. 5, p. 520, Mar. 2019.
- [66] W. Peters, A. Bastos, P. Ciaï, and A. Vermeulen, "A historical, geographical and ecological perspective on the 2018 European summer drought," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 375, no. 1810, Oct. 2020, Art. no. 20190505.
- [67] M. Flach, A. Brenning, F. Gans, M. Reichstein, S. Sippel, and M. D. Mahecha, "Vegetation modulates the impact of climate extremes on gross primary production," *Biogeosciences*, vol. 18, no. 1, pp. 39–53, Jan. 2021.
- [68] S. Sippel *et al.*, "Drought, heat, and the carbon cycle: A review," *Current Climate Change Rep.*, vol. 4, no. 3, pp. 266–286, Sep. 2018.
- [69] R. Guzinski and H. Nieto, "Evaluating the feasibility of using Sentinel-2 and Sentinel-3 satellites for high-resolution evapotranspiration estimations," *Remote Sens. Environ.*, vol. 221, pp. 157–172, Feb. 2019.
- [70] M. Jung *et al.*, "The FLUXCOM ensemble of global land-atmosphere energy fluxes," *Sci. Data*, vol. 6, no. 1, pp. 1–14, May 2019.
- [71] D. Zanaga *et al.*, "ESA WorldCover 10 m 2020 V100," Zenodo, Oct. 2021, doi: [10.5281/zenodo.5571936](https://doi.org/10.5281/zenodo.5571936).
- [72] S. Bolívar-Santamaría and B. Reu, "Detection and characterization of agroforestry systems in the Colombian Andes using Sentinel-2 imagery," *Agroforestry Syst.*, vol. 95, no. 3, pp. 499–514, Mar. 2021.
- [73] C. Ludwig, R. Hecht, S. Lautenbach, M. Schorcht, and A. Zipf, "Mapping public urban green spaces based on OpenStreetMap and Sentinel-2 imagery using belief functions," *ISPRS Int. J. Geo-Geoinf.*, vol. 10, no. 4, p. 251, Apr. 2021.
- [74] A. Carrara, I. A. Janssens, J. Curiel Yuste, and R. Ceulemans, "Seasonal changes in photosynthesis, respiration and NEE of a mixed temperate forest," *Agricult. Forest Meteorol.*, vol. 126, nos. 1–2, pp. 15–31, Nov. 2004.
- [75] C. Moureaux, A. Debacq, B. Bodson, B. Heinesch, and M. Aubinet, "Annual net ecosystem carbon exchange by a sugar beet crop," *Agricult. Forest Meteorol.*, vol. 139, nos. 1–2, pp. 25–39, Sep. 2006.
- [76] M. Aubinet, B. Chermanne, M. Vandenhaute, B. Longdoz, M. Yernaux, and E. Laitat, "Long term carbon dioxide exchange above a mixed forest in the Belgian ardennes," *Agricult. Forest Meteorol.*, vol. 108, no. 4, pp. 293–315, Jul. 2001.
- [77] L. Merbold, W. Eugster, J. Stieger, M. Zahniser, D. Nelson, and N. Buchmann, "Greenhouse gas budget (CO₂, CH₄ and N₂O) of intensively managed grassland following restoration," *Global Change Biol.*, vol. 20, no. 6, pp. 1913–1928, Jun. 2014.
- [78] S. Zielis, S. Etzold, R. Zweifel, W. Eugster, M. Haeni, and N. Buchmann, "NEP of a Swiss subalpine forest is significantly driven not only by current but also by previous year's weather," *Biogeosciences*, vol. 11, no. 6, pp. 1627–1635, Mar. 2014.
- [79] D. Imer, L. Merbold, W. Eugster, and N. Buchmann, "Temporal and spatial variations of soil CO₂, CH₄ and N₂O fluxes at three differently managed grasslands," *Biogeosciences*, vol. 10, no. 9, pp. 5931–5945, Sep. 2013.
- [80] S. Etzold *et al.*, "The carbon balance of two contrasting mountain forest ecosystems in Switzerland: Similar annual trends, but seasonal differences," *Ecosystems*, vol. 14, no. 8, pp. 1289–1309, Dec. 2011.
- [81] D. Dietiker, N. Buchmann, and W. Eugster, "Testing the ability of the DNDC model to predict CO₂ and water vapour fluxes of a Swiss cropland site," *Agricult. Ecosyst. Environ.*, vol. 139, no. 3, pp. 396–401, Nov. 2010.
- [82] M. Acosta *et al.*, "Soil surface CO₂ efflux measurements in Norway spruce forests: Comparison between four different sites across Europe—From boreal to Alpine forest," *Geoderma*, vol. 192, pp. 295–303, Jan. 2013.
- [83] J. Dušek, H. Cížková, S. Stellner, R. Czerný, and J. Květ, "Fluctuating water table affects gross ecosystem production and gross radiation use efficiency in a sedge-grass Marsh," *Hydrobiologia*, vol. 692, no. 1, pp. 57–66, Aug. 2012.
- [84] P. M. Anthoni *et al.*, "Forest and agricultural land-use-dependent CO₂ exchange in Thuringia, Germany," *Global Change Biol.*, vol. 10, no. 12, pp. 2005–2019, Dec. 2004.
- [85] A.-K. Prescher, T. Grünwald, and C. Bernhofer, "Land use regulates carbon budgets in eastern Germany: From NEE to NBP," *Agricult. Forest Meteorol.*, vol. 150, nos. 7–8, pp. 1016–1025, Jul. 2010.
- [86] A. Knohl, E.-D. Schulze, O. Kolle, and N. Buchmann, "Large carbon uptake by an unmanaged 250-year-old deciduous forest in central Germany," *Agricult. Forest Meteorol.*, vol. 118, nos. 3–4, pp. 151–167, Sep. 2003.
- [87] H. Post, H. J. Hendricks Franssen, A. Graf, M. Schmidt, and H. Vereecken, "Uncertainty analysis of eddy covariance CO₂ flux measurements for different EC tower distances using an extended two-tower approach," *Biogeosciences*, vol. 12, no. 4, pp. 1205–1221, Feb. 2015.
- [88] M. Mauder *et al.*, "A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements," *Agricult. Forest Meteorol.*, vol. 169, pp. 122–135, Feb. 2013.
- [89] T. Grünwald and C. Bernhofer, "A decade of carbon, water and energy flux measurements of an old spruce forest at the anchor station tharandt," *Tellus B, Chem. Phys. Meteorol.*, vol. 59, no. 3, pp. 387–396, Jan. 2007.
- [90] K. Pilegaard, A. Ibrom, M. S. Courtney, P. Hummelshøj, and N. O. Jensen, "Increasing net CO₂ uptake by a Danish beech forest during the period from 1996 to 2009," *Agricult. Forest Meteorol.*, vol. 151, no. 7, pp. 934–946, Jul. 2011.
- [91] T. S. El-Madany *et al.*, "Drought and heatwave impacts on semi-arid ecosystems' carbon fluxes along a precipitation gradient," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 375, no. 1810, Oct. 2020, Art. no. 20190519.
- [92] Y. Luo *et al.*, "Using near-infrared-enabled digital repeat photography to track structural and physiological phenology in Mediterranean Tree-Grass ecosystems," *Remote Sens.*, vol. 10, no. 8, p. 1293, Aug. 2018.
- [93] M. Koskinen, K. Minkinen, P. Ojanen, M. Kämäräinen, T. Laurila, and A. Lohila, "Measurements of CO₂ exchange with an automated chamber system throughout the year: Challenges in measuring nighttime respiration on porous peat soil," *Biogeosciences*, vol. 11, no. 2, pp. 347–363, Jan. 2014.
- [94] L. Vitale, P. Di Tommasi, G. D'Urso, and V. Magliulo, "The response of ecosystem carbon fluxes to LAI and environmental drivers in a maize crop grown in two contrasting seasons," *Int. J. Biometeorol.*, vol. 60, no. 3, pp. 411–420, Mar. 2016.
- [95] M. Galvagno *et al.*, "Phenology and carbon dioxide source/sink strength of a subalpine grassland in response to an exceptionally short snow season," *Environ. Res. Lett.*, vol. 8, no. 2, Apr. 2013, Art. no. 025008.
- [96] E. J. Moors, "Water use of forests in The Netherlands," Vrije Universiteit, Amsterdam, The Netherlands, Tech. Rep. 41, 2012.
- [97] I. M. Milyukova, O. Kolle, A. V. Varlagin, N. N. Vygodskaya, E.-D. Schulze, and J. Lloyd, "Carbon balance of a southern Taiga spruce stand in European Russia," *Tellus B: Chem. Phys. Meteorol.*, vol. 54, no. 5, pp. 429–442, Nov. 2002.
- [98] J. Kurbatova, C. Li, A. Varlagin, X. Xiao, and N. Vygodskaya, "Modeling carbon dynamics in two adjacent spruce forests with different soil conditions in Russia," *Biogeosciences*, vol. 5, no. 4, pp. 969–980, Jul. 2008.

- [99] M. L. Fischer, D. P. Billesbach, J. A. Berry, W. J. Riley, and M. S. Torn, "Spatiotemporal variations in growing season exchanges of CO₂, H₂O, and sensible heat in agricultural fields of the southern great plains," *Earth Interact.*, vol. 11, no. 17, pp. 1–21, Oct. 2007.
- [100] A. P. Ouimette *et al.*, "Carbon fluxes and interannual drivers in a temperate forest ecosystem assessed through comparison of top-down and bottom-up approaches," *Agricult. Forest Meteorol.*, vols. 256–257, pp. 420–430, Jun. 2018, doi: 10.1016/j.agrformet.2018.03.017.
- [101] H. Chu *et al.*, "Representativeness of eddy-covariance flux footprints for areas surrounding AmeriFlux sites," *Agricult. Forest Meteorol.*, vols. 301–302, p. 108350, May 2021, doi: 10.1016/j.agrformet.2021.108350.
- [102] D. T. Roman, K. A. Novick, E. R. Brzostek, D. Dragoni, F. Rahman, and R. P. Phillips, "The role of isohydric and anisohydric species in determining ecosystem-scale response to severe drought," *Oecologia*, vol. 179, no. 3, pp. 641–654, Nov. 2015, doi: 10.1007/s00442-015-3380-9.
- [103] M. Litvak, "AmeriFlux U.S.-Seg sevillea grassland from 2007-present," Tech. Rep., 2016.
- [104] M. Litvak, "AmeriFlux U.S.-Ses sevillea shrubland from 2007-present," Tech. Rep., 2016.
- [105] C. M. Gough, *et al.*, "Sustained carbon uptake and storage following moderate disturbance in a great lakes forest," *Ecol. Appl.*, vol. 23, no. 5, pp. 1202–1215, 2013.
- [106] M. Litvak, "AmeriFlux U.S.-Vcm valles caldera mixed conifer from 2007-Present," Tech. Rep., 2016.
- [107] M. Litvak, "AmeriFlux U.S.-Wjs willard juniper savannah from 2007-present," Tech. Rep., 2016.
- [108] Y. J. Kaufman and D. Tanre, "Atmospherically resistant vegetation index (ARVI) for EOS-MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 261–270, Mar. 1992.
- [109] C. J. Tucker, "Remote sensing of leaf water content in the near infrared," *Remote Sens. Environ.*, vol. 10, no. 1, pp. 23–32, Aug. 1980.
- [110] B. Pinty and M. M. Verstraete, "GEMI: A non-linear index to monitor global vegetation from satellites," *Vegetatio*, vol. 101, no. 1, pp. 15–20, Jul. 1992.
- [111] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 289–298, 1996.
- [112] R. E. Crippen, "Calculating the vegetation index faster," *Remote Sens. Environ.*, vol. 34, no. 1, pp. 71–73, 1990.
- [113] C. S. T. Daughtry, C. L. Walthall, M. S. Kim, E. B. de Colstoun, and J. E. McMurtrey, III, "Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance," *Remote Sens. Environ.*, vol. 74, no. 2, pp. 229–239, Nov. 2000.
- [114] J. Qi, A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian, "A modified soil adjusted vegetation index," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 119–126, 1994.
- [115] J. Dash and P. J. Curran, "The MERIS terrestrial chlorophyll index," *Int. J. Remote Sens.*, vol. 25, no. 23, pp. 5403–5413, Dec. 2004.
- [116] J. Delegado, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors*, vol. 11, no. 7, pp. 7063–7081, 2011.
- [117] G. A. Blackburn, "Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches," *Remote Sens. Environ.*, vol. 66, no. 3, pp. 273–285, Dec. 1998.
- [118] A. J. Richardson and C. L. Wiegand, "Distinguishing vegetation from soil background information," *Photogramm. Eng. Remote Sens.*, vol. 43, no. 12, pp. 1541–1552, Dec. 1977.
- [119] D. J. Major, F. Baret, and G. Guyot, "A ratio vegetation index adjusted for soil brightness," *Int. J. Remote Sens.*, vol. 11, no. 5, pp. 727–740, May 1990.
- [120] G. Guyot and F. Baret, "Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux," in *Spectral Signatures of Objects in Remote Sensing*, vol. 287, Apr. 1988, p. 279.
- [121] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, 1988.
- [122] G. M. Senseman, C. F. Bagley, and S. A. Tweddale, "Correlation of rangeland cover measures to satellite-imagery-derived vegetation indices," *Geocarto Int.*, vol. 11, no. 3, pp. 29–38, Sep. 1996.
- [123] F. Baret, G. Guyot, and D. J. Major, "TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation," in *Proc. 12th Can. Symp. Remote Sens. Geosci. Remote Sens. Symp.*, Jul. 1989, pp. 1355–1358.
- [124] J. G. P. W. Clevers, "The derivation of a simplified reflectance model for the estimation of leaf area index," *Remote Sens. Environ.*, vol. 25, no. 1, pp. 53–69, Jun. 1988.
- [125] A. A. Gitelson *et al.*, "Relationship between gross primary production and chlorophyll content in crops: Implications for the synoptic monitoring of vegetation productivity," *J. Geophys. Res.*, vol. 111, no. D8, 2006, Art. no. D08S11.
- [126] A. A. Gitelson, A. Viña, T. J. Arkebauer, D. C. Rundquist, G. Keydan, and B. Leavitt, "Remote estimation of leaf area index and green leaf biomass in maize canopies," *Geophys. Res. Lett.*, vol. 30, no. 5, p. 1248, Mar. 2003.
- [127] J. Rouse Jr., R. Haas, J. Schell, and D. Deering, "Paper a 20," in *Proc. 3rd Earth Resour. Technol. Satellite Symp. Symposium Held Goddard Space Flight Center Washington, Prepared Goddard Space Flight Center*, vol. 351. Washington, DC, USA: Scientific and Technical Information Office, National Aeronautics and Space, 1974, p. 309.



Daniel E. Pabon-Moreno received the B.Sc. degree in biology from the Industrial University of Santander, Bucaramanga, Colombia, in 2017. He is currently pursuing the Ph.D. degree with Friedrich Schiller University Jena, Jena, Germany.

He is currently with the Department Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena. His research interests include vegetation response to climate change, ecosystem monitoring using remote sensing information, and ecosystem physiophenology modeling using circular statistics.



Mirco Migliavacca received the Ph.D. degree from the University of Milano-Bicocca, Milan, Italy, in 2009.

He worked as a postdoc with the European Commission and then as a Group Leader with the Max Planck Institute for Biogeochemistry, Jena, Germany. He is currently an Environmental Scientist with the University of Milano-Bicocca. Currently, he is a Scientific Officer with the European Commission, Joint Research Centre. His research interests involve broad questions in the fields of global change ecology, biometeorology, and biogeochemistry related to the carbon and water cycles in terrestrial ecosystems and climate-biosphere interactions.



Markus Reichstein received the Ph.D. degree from the Department of Plant Ecology, University of Bayreuth, Bayreuth, Germany, in 2001.

Since 2013, he has been a Professor of global geocology with Friedrich Schiller University, Jena, Germany, and the Founding Director of the Michael Stifel Center Jena for Data-Driven and Simulation Science. He is currently the Director of the Biogeochemical Integration Department, Max Planck Institute for Biogeochemistry, Jena. His research interests include around the response and feedback of ecosystems (vegetation and soils) to climatic variability with an Earth system perspective, considering coupled carbon, water, and nutrient cycles, as well as specific interest in the interplay of climate extremes with ecosystem and societal resilience.

Dr. Reichstein has been serving as a Lead Author for the IPCC Special Report on Climate Extremes (SREX) and a member for the German Committee Future Earth on Sustainability Research and Thuringian Panel on Climate.



Miguel D. Mahecha received the Ph.D. degree from ETH Zürich, Zürich, Switzerland, in 2009.

He was a Group Leader with the Max Planck Institute for Biogeochemistry, Jena, Germany. Since 2020, he has been a Full Professor with the Remote Sensing Centre for Earth System Research, Leipzig University, and the Helmholtz Centre for Environmental Research, UFZ, Germany. His research interests revolve around global ecosystem responses to climate variability with an emphasis on the role of climate extremes. He also works on quantifying biodiversity patterns, and he has a genuine interest in Earth system data science methods.

CHAPTER 5

Concluding Discussion

The current scenario of climate change and global warming has been better understood thanks to estimating energy and matter fluxes between the atmosphere and the biosphere (Baldocchi, 2020). In this direction, the development of concepts that try to reconcile ecosystem processes at different spatial and temporal scales, such as the functional properties of ecosystems, has also been fundamental (Reichstein et al., 2014; Migliavacca et al., 2021).

In this dissertation, I frame our understanding of terrestrial vegetation activity, represented by gross primary production, using three axes: Magnitude, Time, and Space (Figure 5.1). In the first axis, Magnitude, the maximum gross primary production represents the optimal photosynthesis rate at the ecosystem level (GPP_{max}). It is expected that understanding the limiting factors of GPP_{max} can help to understand what are the effects of climatic variability on the entire ecosystem. I find that for ~68% (~78.4 millions km^2) of the total global surface with terrestrial vegetation cover, air temperature and precipitation are equally limiting GPP_{max} . For ~17.4% (~20.1 millions km^2) of the land surface with terrestrial vegetation cover GPP_{max} is mainly limited by temperature. I also find that for many regions of the world the

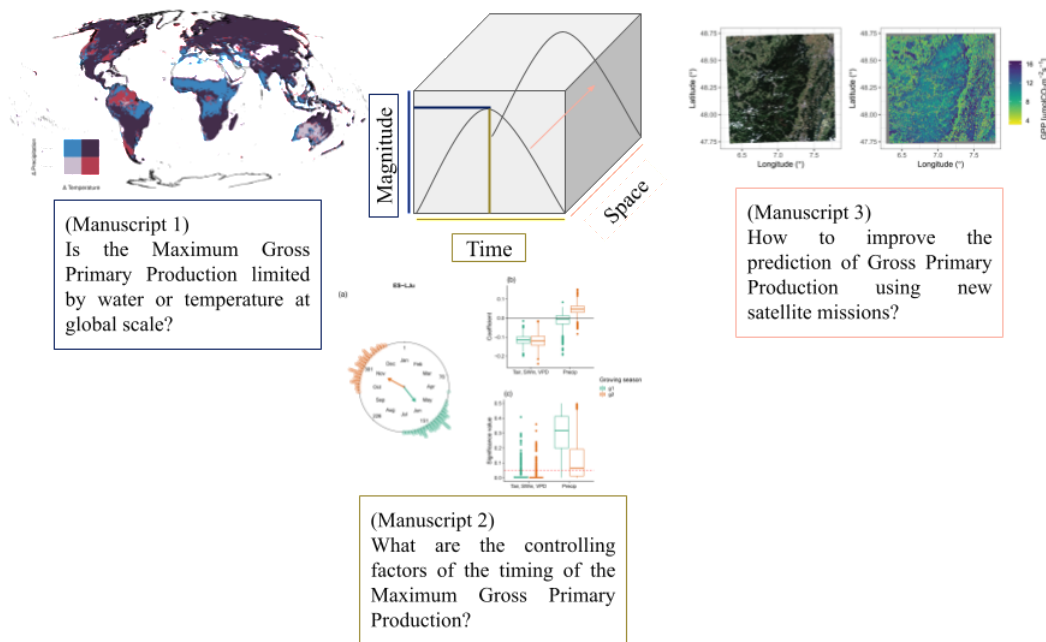


Figure 5.1: Conceptual representation of the three axes of study on this dissertation. The gray line represents the gross primary production (GPP) variation during the growing season. The first axis **Magnitude** refers to the value of GPP and the maximum GPP (GPP_{max}) *per-se*. The second axis **Time** is the timing when GPP_{max} is reached. The third axis, **Space**, represents the prediction/upscaling of GPP using new remote sensing information. The arrow represents the increase in spatial resolution. From local Eddy Covariance Towers (EC) to regional and global predictions of GPP

classification system built based on the GPP_{max} controlling factors matches previous ecological classifications systems such as Koppen-Geiger (Although, this is now considered a climate classification system). Evaluating the relationship between an optimum ecosystem state as GPP_{max} and climate variables allows us to generate a functional classification system that represents vegetation's response regarding ecosystem fluxes to climate variations. Nevertheless, a complete comparison year by year for the last decades evaluating this relationship can give us more information on how ecosystems respond each year to multiple pressure factors (e.g., climate change, land-use change, heatwaves).

In the second axis, Time, I evaluate the effect of climate variables on the timing of GPP_{max} ($DOYGPP_{max}$). I find that for most ecosystems, an increase in short-wave incoming radiation, temperature, and vapor pressure deficit will produce that $DOYGPP_{max}$ will be reached earlier (when compared with the mean $DOYGPP_{max}$)

(Pabon-Moreno et al., 2020). I also find that although the sensitivity of DOYGPP_{\max} is site-specific, it is possible to formulate models for the same vegetation type at similar latitudes. In scenarios where temperature and precipitation increase for the same region, we expect to find no changes in the timing of GPP_{\max} . Otherwise, if one of the variables increases, we expect a change following the coefficient sign in the circular-linear regression model. From a methodological perspective, I show that circular statistics increase the robustness of the timing analysis compared to linear statistics. Even more circular statistics show to be more versatile in analyzing data sets across the globe regardless of whether the observations are from the Northern or Southern Hemisphere (Pabon-Moreno et al., 2020).

On the third axis, Space, I show how the new information from Sentinel-2 improves the prediction of gross primary production. When vegetation indices based on the red-edge spectral region (700 nm - 730 nm) as chlorophyll index red (CIR) and the inverted red-edge chlorophyll index (IRECI) are used, the accuracy on the prediction of GPP increases by ~10% more of variance explained. This improvement represents a step forward when compared with previous vegetation indices as Normalized vegetation index (NDVI) or novel ones as the kernelized version of NDVI (kNDVI Camps-Valls et al., 2021). From a methodological perspective, I assess the bias produced by the imbalanced representation of observations by vegetation type and seasons on the prediction of GPP. I use different statistical techniques to balance the number of observations per vegetation type, the frequency of observations per season, and even the presence of rare observations (observations with a low frequency) that are difficult to predict (Torgo et al., 2013). Here I demonstrate that when these techniques are applied, the predictions' robustness improves, allowing the formulation of more generalized models (Pabon-Moreno et al., 2022).

The exploration of the axes proposed in this dissertation allows us to have a more holistic view of the effects of climate change and global warming on ecosystem functions and the ecosystem functional properties. The analysis presented ranges from local observation (a few hundred meters around the EC towers) and daily

frequency to regional analysis thanks to new satellite missions such as Sentinel-2, and global analysis at the 15-day temporal resolution using previous GPP products.

In the first axis, Magnitude, understanding the limiting factors of GPP_{max} give us a better idea of how climate change may translate into changes in optimum ecosystem processes. Furthermore, my approach allows to generate a functional map that consider optimum ecosystem states (Reichstein et al., 2007; Mahecha et al., 2007). Current plant functional types were generated by combining plant traits, species distributions maps, and vegetation activity from satellite imagery (Box, 1995, 1996; Diaz and Cabido, 1997). Using an optimal ecosystem process as GPP_{max} and its limiting factors, my approach allow us to generate new functional classes that relate optimum ecosystem processes with climate information. These new classes can be used as input for future global-scale modeling studies. Specifically, ecosystem functional properties can describe the interaction of the environment with the vegetation in a smaller number of parameters.

Equally relevant to understand the effect of climate change on ecosystems is the timing of optimum ecosystem processes, such as $DOYGPP_{max}$. Most of the studies evaluated the timing of a biological process as periodic oscillations that can be decomposed into different signals. In this dissertation, I show that time can also be interpreted as the relationship between abiotic components and the day of the year when optimum ecosystem processes are reached. It is well known that climate change has affected plant phenology on a global scale (Richardson et al., 2013); however, there is still no clear consensus regarding how climate change affects the beginning, end, and peak of productivity in ecosystems on a global scale. My study (Pabon-Moreno et al., 2020) is a first step to understand the effect of climate variables in the timing of GPP_{max} . Circular statistics may represent the first step in analyzing these phenomena. Even if I only focused on the peak of the growing season (i.e., $DOYGPP_{max}$), the same analysis can be performed for the beginning and the end of the growing season. Another important analysis is to assess the correlation between the times of the beginning, the end, and the peak of the growing season

using circular-circular regressions. These new analyses may provide more clarity and could generate consensus on the response of vegetation to climate change in terms of timing.

In the last axis, Space, improving the GPP predictions is necessary to improve our knowledge of the ecosystem processes and the changes produced by global warming and climate change. The Normalized Difference Vegetation Index (NDVI) was developed in the 70s as part of the first satellite missions focused on tracking changes on the Earth's surface (Rouse et al., 1974). Since then, several vegetation indices have been developed to correlate the photosynthetic activity of plants and their spectral characteristics (Bannari et al., 1995). Many recent approaches rely on the classic NDVI proposed in the 70s (e.g., Badgley et al., 2017, 2019; Camps-Valls et al., 2021). The basic premise of vegetation indices, is that the amount of plant biomass is proportional to the ratio between absorption in the red region of the electromagnetic spectrum, and reflection in the infrared region of the spectrum. The absorption in the red region is caused by chlorophyll, and reflection in the infrared region is caused by the cell's structure of plants (Myneni et al., 1995). The Sentinel-2 satellite mission included two red-edge bands that were not included in previous satellite missions (Martimort et al., 2007). Previous studies found that vegetation indices based on the red-edge spectral region would produce more accurate estimations of plant photosynthesis (Delegido et al., 2011). Nevertheless, a robust-empirical comparison between the different vegetation indices and estimations of GPP from EC towers was missing before my study (Pabon-Moreno et al., 2022). While my study has shown to improve the prediction of GPP, some questions remain open: for example, what is the performance of red-edge-based vegetation indices when kernel methods are applied? What is the performance of GPP prediction based on Sentinel-2 and new red-edge vegetation indices compared to previous satellite missions such as MODIS and Landsat? Answering these questions will be relevant before performing global-scale prediction exercises using Sentinel-2 imagery.

In recent decades, understanding the functional biogeography of plants has be-

come increasingly relevant to the role of plants in climate change and global warming (Friend, 2010; Mahmood et al., 2014). Mainly, the research has focused on plant trait variability (Violle et al., 2014), and most of it on the fluxes of energy and matter between the biosphere and the atmosphere (Musavi et al., 2015, 2016; Migliavacca et al., 2021). While defining the function of an organism as a serie of processes that can be imputed to them not only ends up being ambiguous, but also useful for framing research. Nevertheless, what is the limit in the definition of functions? For example, I could define the shadow cast by trees as a function of trees. In other words, could someone refute that the function of trees is not to give shade to humans? In this sense, the defined function can be used and abused depending on my interests. Whether those interests are relevant or not is a more sociological question than a purely natural science one. Defining functions in nature inevitably goes hand in hand with trying to represent our interpretation of economy, necessity, and relationship with nature (Lockwood, 1999). When we move from physical and biochemical phenomena to biological and ecological ones, the mechanism and causality schemes presented in physics and chemistry tend not to work completely on biological/ecological problems (Ross, 2021). The explanation for this is the evolutionary mechanism inherent to any living being. As a biologist, a key element not considered in functional biogeography and the proposed three axes is the role of evolution.

Outlook

Although, evolution and its mechanism are beyond this dissertation, evolution can help to solve important questions for functional biogeography. For example, Do ecosystems always tend to have a degree of homeostasis in terms of matter and energy fluxes? Do species change represent changes in energy and matter flows, or are these flows used as an arm for competition? Do the leaf economy of plants and the ecosystem functional properties concepts go in the same direction that evolution fitness concepts? While evolution acts at different scales, assessing and

understanding its role in functional biogeography is an ongoing task. In the last decades, several studies proposed models and evidence about some of the previous questions (Ernest and Brown, 2001; Zakharov and Trofimov, 2014; Kikuchi et al., 2018; Dyke and Weaver, 2013). For example, Ernest and Brown (2001) present evidence that despite environmental changes, species diversity tend to homeostatically regulate ecosystem energy and matter fluctuations. Furthermore, the authors state that total population estimates, biomass, and vegetation cover do not provide precise measures of energy and matter resources usage by the ecosystems. In this direction, bio-meteorological techniques such as eddy covariance towers and remote sensing information can help to reconcile the evolution mechanism of species and ecosystem dynamics. Currently, most of the eddy covariance sites on FLUXNET initiative are located on managed ecosystems in Europe and North-America. During the development of western civilization and industrialization, most of the forests in Europe were intervened by humans. Therefore, the effects of natural selection in the long term and evolution itself are more difficult to evidence on Europe. Nevertheless, forest plots that are currently monitored in Europe and North America are a good tool to evaluate and try to predict the outcome of evolution processes in the following decades and centuries. To bring eddy covariance towers and remote sensing technology to current monitored forest plots is a top-priority task to develop. Some first steps in this direction are the studies of Ma et al. (2019) and Pacheco-Labrador et al. (2022), where the authors evaluated the relation between spectral diversity and functional diversity. Another important step will be to generate and apply a unified protocol for systematic forest inventories, including the collection of diversity metrics for the current eddy covariance towers that are part of FLUXNET initiative. A further step will be to establish the systematic and periodic genetic sampling of the species in the eddy covariance sites that can help to link genetic expression and metabolic regulation pathways with ecosystem fluxes and functional diversity. Combining the knowledge and data from biometeorology, phylogenetic, ecology, and systematic biology will help us to develop new models and theories to reconcile

evolution mechanism and explain changes on ecosystem at all organization-levels from genes to ecosystem fluxes at different temporal scales. A central key in the puzzle of reconciling functional biology and classic biogeography will be to increase the investment of human resources and technological transaction to deploy new eddy covariance towers, as well genome, and gene expression technologies in the tropics. Such approaches will also help to understand the effect of climate change and global warming in ecosystems. It is not a coincidence that naturalists from last centuries developed their ideas after visiting and collecting information in the tropics (Schiebinger and Swan, 2007; Wilke, 2010; Baber, 2016). A way to compensate the colonialism and bio-piracy (for which European and North American countries still do not pay royalties, Schiebinger and Swan, 2007; Baber, 2016) is to help research groups located in the tropics to develop their own ideas and deploy their research infrastructure (i.e. transferring patents and industrial knowledge necessary to develop the research equipment). Reconciling evolution theories and functional biogeography including the ideas and data from tropical research groups will not only contribute to the progress on both disciplines, but also to advance our understanding of the effect of climate change and global warming on the terrestrial ecosystems around the globe.

Bibliography

- Andres, R. J., Boden, T. A., Bréon, F.-M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P., and Treanton, K. (2012). A synthesis of carbon dioxide emissions from fossil-fuel combustion. *Biogeosciences*, 9(5):1845–1871.
- Aubinet, M., Vesala, T., and Papale, D., editors (2012). *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer Atmospheric Sciences. Springer Netherlands.
- Baber, Z. (2016). The Plants of Empire: Botanic Gardens, Colonial Power and Botanical Knowledge. *Journal of Contemporary Asia*, 46(4):659–679.
- Badgley, G., Anderegg, L. D. L., Berry, J. A., and Field, C. B. (2019). Terrestrial gross primary production: Using NIRV to scale from site to globe. *Global Change Biology*, 25(11):3731–3740.
- Badgley, G., Field, C. B., and Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances*, 3(3):e1602244.
- Baldocchi, D. (2008). ‘Breathing’ of the terrestrial biosphere: Lessons learned from

- a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*, 56(1):1–26.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, 82(11):2415–2434.
- Baldocchi, D. D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future. *Global Change Biology*, 9(4):479–492.
- Baldocchi, D. D. (2020). How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Global Change Biology*, 26(1):242–260.
- Bannari, A., Morin, D., Bonn, F., and Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, 13(1-2):95–120.
- Belward, A. S. and Skøien, J. O. (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:115–128.
- Box, E. O. (1995). Factors determining distributions of tree species and plant functional types. *Vegetatio*, 121(1):101–116.
- Box, E. O. (1996). Plant functional types and climate at the global scale. *Journal of Vegetation Science*, 7(3):309–320.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

- Buitenwerf, R., Rose, L., and Higgins, S. I. (2015). Three decades of multi-dimensional change in global leaf phenology. *Nature Climate Change*, 5(4):364–368.
- Campbell, J. E., Berry, J. A., Seibt, U., Smith, S. J., Montzka, S. A., Launois, T., Belviso, S., Bopp, L., and Laine, M. (2017). Large historical growth in global terrestrial gross primary production. *Nature*, 544(7648):84–87.
- Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Mari, J., García-Haro, F. J., Guanter, L., Jung, M., Gamon, J. A., Reichstein, M., and Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9):eabc7447.
- Canadell, J. G., Monteiro, P. M., Costa, M. H., Cunha, L. C. D., Cox, P. M., Eliseev, A. V., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P. K., Piao, S., Syampungani, S., Zaehle, S., Zickfeld, K., Alexandrov, G. A., Bala, G., Bopp, L., Boysen, L., Cao, L., Chandra, N., Ciais, P., Denisov, S. N., Dentener, F. J., Douville, H., Fay, A., Forster, P., Fox-Kemper, B., Friedlingstein, P., Fu, W., Fuss, S., Garçon, V., Gier, B., Gillett, N. P., Gregor, L., Haustein, K., Haverd, V., He, J., Hewitt, H. T., Hoffman, F. M., Ilyina, T., Jackson, R., Jones, C., Keller, D. P., Kwiatkowski, L., Lamboll, R. D., Lan, X., Laufkötter, C., Quéré, C. L., Lenton, A., Lewis, J., Liddicoat, S., Lorenzoni, L., Lovenduski, N., Macdougall, A. H., Mathesius, S., Matthews, D. H., Meinshausen, M., Mokhov, I. I., Naik, V., Nicholls, Z. R. J., Nurhati, I. S., O’sullivan, M., Peters, G., Pongratz, J., Poulter, B., Sallée, J.-B., Saunois, M., Schuur, E. A., Seneviratne, S., Stavert, A., Suntharalingam, P., Tachiiri, K., Terhaar, J., Thompson, R., Tian, H., Turnbull, J., Vicente-Serrano, S. M., Wang, X., Wanninkhof, R. H., Williamson, P., Brovkin, V., Feely, R. A., and Lebehot, A. D. (2021). Global Carbon and other Biogeochemical Cycles and Feedbacks. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report*

- of the Intergovernmental Panel on Climate Change*, pages 673–816. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Chapin, F. S., Matson, P. A., and Vitousek, P. M. (2011a). Carbon Inputs to Ecosystems. In Chapin, F. S., Matson, P. A., and Vitousek, P. M., editors, *Principles of Terrestrial Ecosystem Ecology*, pages 123–156. Springer, New York, NY.
- Chapin, F. S., Matson, P. A., and Vitousek, P. M. (2011b). Trophic Dynamics. In Chapin, F. S., Matson, P. A., and Vitousek, P. M., editors, *Principles of Terrestrial Ecosystem Ecology*, pages 297–320. Springer, New York, NY.
- Chase, J. M., Leibold, M. A., Downing, A. L., and Shurin, J. B. (2000). The Effects of Productivity, Herbivory, and Plant Species Turnover in Grassland Food Webs. *Ecology*, 81(9):2485–2497.
- Chen, M. and Blankenship, R. E. (2021). Photosynthesis | Photosynthesis. In Jez, J., editor, *Encyclopedia of Biological Chemistry III (Third Edition)*, pages 150–156. Elsevier, Oxford.
- Delegido, J., Verrelst, J., Alonso, L., and Moreno, J. (2011). Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors*, 11(7):7063–7081.
- Diaz, S. and Cabido, M. (1997). Plant functional types and ecosystem function in relation to global change. *Journal of Vegetation Science*, 8(4):463–474.
- Dokulil, M. T. (2019). Gross and Net Production in Different Environments. In Fath, B., editor, *Encyclopedia of Ecology (Second Edition)*, pages 334–345. Elsevier, Oxford.
- Duckworth, J. C., Kent, M., and Ramsay, P. M. (2000). Plant functional types: An alternative to taxonomic plant community description in biogeography? *Progress in Physical Geography: Earth and Environment*, 24(4):515–542.

- Dyke, J. G. and Weaver, I. S. (2013). The Emergence of Environmental Homeostasis in Complex Ecosystems. *PLOS Computational Biology*, 9(5):e1003050.
- Ernest, S. K. M. and Brown, J. H. (2001). Homeostasis and Compensation: The Role of Species and Resources in Ecosystem Stability. *Ecology*, 82(8):2118–2132.
- Eyring, V., Gillet, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S.-K., Morgenstern, O., and Sun, Y. (2021). Human Influence on the Climate System. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 423–552. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Friedlingstein, P., Jones, M. W., O’Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S. (2019). Global Carbon Budget 2019. *Earth System Science Data*, 11(4):1783–1838.
- Friend, A. D. (2010). Terrestrial plant production and climate change. *Journal of Experimental Botany*, 61(5):1293–1309.

- Goward, S. N., Markham, B., Dye, D. G., Dulaney, W., and Yang, J. (1991). Normalized difference vegetation index measurements from the advanced very high resolution radiometer. *Remote Sensing of Environment*, 35(2):257–277.
- Goward, S. N., Tucker, C. J., and Dye, D. G. (1985). North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. *Vegetatio*, 64(1):3–14.
- Gulev, S., Thorne, P., Ahn, J., Dentener, F., Domingues, C., Gerland, S., and Vose, R. (2021). Changing state of the climate system. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 287–422. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, 6(1):1–14.
- Jung, M., Reichstein, M., and Bondeau, A. (2009). Towards global empirical up-scaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, 6(10):2001–2013.
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O’Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, 17(5):1343–1365.
- Kikuchi, J., Ito, K., and Date, Y. (2018). Environmental metabolomics with data

- science for investigating ecosystem homeostasis. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 104:56–88.
- Lockwood, M. (1999). Humans Valuing Nature: Synthesising Insights from Philosophy, Psychology and Economics. *Environmental Values*, 8(3):381–401.
- Ma, X., Mahecha, M. D., Migliavacca, M., van der Plas, F., Benavides, R., Ratcliffe, S., Kattge, J., Richter, R., Musavi, T., Baeten, L., Barnoaiea, I., Bohn, F. J., Bouriaud, O., Bussotti, F., Coppi, A., Domisch, T., Huth, A., Jaroszewicz, B., Joswig, J., Pabon-Moreno, D. E., Papale, D., Selvi, F., Laurin, G. V., Valladares, F., Reichstein, M., and Wirth, C. (2019). Inferring plant functional diversity from space: The potential of Sentinel-2. *Remote Sensing of Environment*, 233:111368.
- Mahecha, M. D., Reichstein, M., Lange, H., Carvalhais, N., Bernhofer, C., Grünwald, T., Papale, D., and Seufert, G. (2007). Characterizing ecosystem-atmosphere interactions from short to interannual time scales. *Biogeosciences*, 4(5):743–758.
- Mahmood, R., Pielke Sr., R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., McAlpine, C., Carleton, A. M., Hale, R., Gameda, S., Beltrán-Przekurat, A., Baker, B., McNider, R., Legates, D. R., Shepherd, M., Du, J., Blanken, P. D., Frauenfeld, O. W., Nair, U., and Fall, S. (2014). Land cover changes and their biogeophysical effects on climate. *International Journal of Climatology*, 34(4):929–953.
- Malaterre, C., Dussault, A. C., Mermans, E., Barker, G., Beisner, B. E., Bouchard, F., Desjardins, E., Handa, I. T., Kembel, S. W., Lajoie, G., Maris, V., Munson, A. D., Odenbaugh, J., Poisot, T., Shapiro, B. J., and Suttle, C. A. (2019). Functional Diversity: An Epistemic Roadmap. *BioScience*, 69(10):800–811.
- Martimort, P., Arino, O., Berger, M., Biasutti, R., Carnicero, B., Bello, U. D., Fernandez, V., Gascon, F., Greco, B., Silvestrin, P., Spoto, F., and Sy, O. (2007). Sentinel-2 optical high resolution mission for GMES operational services. In *2007*

- IEEE International Geoscience and Remote Sensing Symposium*, pages 2677–2680.
- Migliavacca, M., Meroni, M., Busetto, L., Colombo, R., Zenone, T., Matteucci, G., Manca, G., and Seufert, G. (2009). Modeling Gross Primary Production of Agro-Forestry Ecosystems by Assimilation of Satellite-Derived Information in a Process-Based Model. *Sensors*, 9(2):922–942.
- Migliavacca, M., Musavi, T., Mahecha, M. D., Nelson, J. A., Knauer, J., Baldocchi, D. D., Perez-Priego, O., Christiansen, R., Peters, J., Anderson, K., Bahn, M., Black, T. A., Blanken, P. D., Bonal, D., Buchmann, N., Caldararu, S., Carrara, A., Carvalhais, N., Cescatti, A., Chen, J., Cleverly, J., Cremonese, E., Desai, A. R., El-Madany, T. S., Farella, M. M., Fernández-Martínez, M., Filippa, G., Forkel, M., Galvagno, M., Gomasasca, U., Gough, C. M., Göckede, M., Ibrom, A., Ikawa, H., Janssens, I. A., Jung, M., Kattge, J., Keenan, T. F., Knohl, A., Kobayashi, H., Kraemer, G., Law, B. E., Liddell, M. J., Ma, X., Mammarella, I., Martini, D., Macfarlane, C., Matteucci, G., Montagnani, L., Pabon-Moreno, D. E., Panigada, C., Papale, D., Pendall, E., Penuelas, J., Phillips, R. P., Reich, P. B., Rossini, M., Rotenberg, E., Scott, R. L., Stahl, C., Weber, U., Wohlfahrt, G., Wolf, S., Wright, I. J., Yakir, D., Zaehle, S., and Reichstein, M. (2021). The three major axes of terrestrial ecosystem function. *Nature*, 598(7881):468–472.
- Montgomery, R. B. (1948). VERTICAL EDDY FLUX OF HEAT IN THE ATMOSPHERE. *Journal of the Atmospheric Sciences*, 5(6):265–274.
- Musavi, T., Mahecha, M. D., Migliavacca, M., Reichstein, M., van de Weg, M. J., van Bodegom, P. M., Bahn, M., Wirth, C., Reich, P. B., Schrod, F., and Kattge, J. (2015). The imprint of plants on ecosystem functioning: A data-driven approach. *International Journal of Applied Earth Observation and Geoinformation*, 43:119–131.
- Musavi, T., Migliavacca, M., van de Weg, M. J., Kattge, J., Wohlfahrt, G., van Bodegom, P. M., Reichstein, M., Bahn, M., Carrara, A., Domingues, T. F., Gavazzi,

- M., Gianelle, D., Gimeno, C., Granier, A., Gruening, C., Havránková, K., Herbst, M., Hrynkiw, C., Kalhori, A., Kaminski, T., Klumpp, K., Kolari, P., Longdoz, B., Minerbi, S., Montagnani, L., Moors, E., Oechel, W. C., Reich, P. B., Rohatyn, S., Rossi, A., Rotenberg, E., Varlagin, A., Wilkinson, M., Wirth, C., and Mahecha, M. D. (2016). Potential and limitations of inferring ecosystem photosynthetic capacity from leaf functional traits. *Ecology and Evolution*, 6(20):7352–7366.
- Myneni, R. B., Hall, F. G., Sellers, P. J., and Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):481–486.
- Ohtaki, E. (1984). Application of an infrared carbon dioxide and humidity instrument to studies of turbulent transport. *Boundary-Layer Meteorology*, 29(1):85–107.
- Ohtaki, E. and Matsui, T. (1982). Infrared device for simultaneous measurement of fluctuations of atmospheric carbon dioxide and water vapor. *Boundary-Layer Meteorology*, 24(1):109–119.
- Pabon-Moreno, D. E., Migliavacca, M., Reichstein, M., and Mahecha, M. D. (2022). On the Potential of Sentinel-2 for Estimating Gross Primary Production. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12.
- Pabon-Moreno, D. E., Musavi, T., Migliavacca, M., Reichstein, M., Römermann, C., and Mahecha, M. D. (2020). Ecosystem physio-phenology revealed using circular statistics. *Biogeosciences*, 17(15):3991–4006.
- Pacheco-Labrador, J., Migliavacca, M., Ma, X., Mahecha, M. D., Carvalhais, N., Weber, U., Benavides, R., Bouriaud, O., Barnoaiea, I., Coomes, D. A., Bohn, F. J., Kraemer, G., Heiden, U., Huth, A., and Wirth, C. (2022). Challenging the link between functional and spectral diversity with radiative transfer modeling and data. *Remote Sensing of Environment*, 280:113170.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Ribeca, A., van Ingen, C., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., De Cinti, B., de Grandcourt, A., De Ligne, A., De Oliveira, R. C., Delpeire, N., Desai, A. R., Di Bella, C. M., di Tommasi, P., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gittelsohn, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J. H.,

- Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., Moreaux, V., Moureaux, C., Munger, J. W., Nakai, T., Neiryneck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Noretto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Ourcival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Reed, D., de Dios, V. R., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S. R., Sánchez-Cañete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tiedemann, F., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., and Papale, D. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, 7(1):225.
- Pörtner, H. O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R. A., Betts, R., Kerr, R. B., Biesbroek, R., et al. (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability*. Cambridge University Press. In Press.
- Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D. (2014). Linking plant and ecosystem functional biogeography. *Proceedings of the National Academy of Sciences*, 111(38):13697–13702.

- Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S., Viovy, N., Cramer, W., Granier, A., Ogée, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., Grünwald, T., Heimann, M., Heinesch, B., Knohl, A., Kutsch, W., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Pilegaard, K., Pumpanen, J., Rambal, S., Schaphoff, S., Seufert, G., Soussana, J.-F., Sanz, M.-J., Vesala, T., and Zhao, M. (2007). Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: A joint flux tower, remote sensing and modelling analysis. *Global Change Biology*, 13(3):634–651.
- Richardson, A. D., Andy Black, T., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., Hollinger, D. Y., Kutsch, W. L., Longdoz, B., Luysaert, S., Migliavacca, M., Montagnani, L., William Munger, J., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri, E., Vargas, R., and Varlagin, A. (2010). Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1555):3227–3246.
- Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., and Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, 169:156–173.
- Ross, L. N. (2021). Causal Concepts in Biology: How Pathways Differ from Mechanisms and Why It Matters. *The British Journal for the Philosophy of Science*, 72(1):131–158.
- Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H. (2004). A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *BioScience*, 54(6):547–560.

- Ryu, Y., Berry, J. A., and Baldocchi, D. D. (2019). What is global photosynthesis? History, uncertainties and opportunities. *Remote Sensing of Environment*, 223:95–114.
- Schiebinger, L. and Swan, C. (2007). *Colonial Botany: Science, Commerce, and Politics in the Early Modern World*. University of Pennsylvania Press.
- Schmid, H. P. (1994). Source areas for scalars and scalar fluxes. *Boundary-Layer Meteorology*, 67(3):293–318.
- Schmid, H. P. and Lloyd, C. R. (1999). Spatial representativeness and the location bias of flux footprints over inhomogeneous areas. *Agricultural and Forest Meteorology*, 93(3):195–209.
- Swinbank, W. C. (1951). THE MEASUREMENT OF VERTICAL TRANSFER OF HEAT AND WATER VAPOR BY EDDIES IN THE LOWER ATMOSPHERE. *Journal of the Atmospheric Sciences*, 8(3):135–145.
- Thépaut, J.-N., Dee, D., Engelen, R., and Pinty, B. (2018). The Copernicus Programme and its Climate Change Service. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1591–1593.
- Torgo, L., Ribeiro, R. P., Pfahringer, B., and Branco, P. (2013). SMOTE for regression. In Correia, L., Reis, L. P., and Cascalho, J., editors, *Progress in Artificial Intelligence*, pages 378–389, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, 13(14):4291–4313.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150.

- Verma, S. B., Baldocchi, D. D., Anderson, D. E., Matt, D. R., and Clement, R. J. (1986). Eddy fluxes of CO₂, water vapor, and sensible heat over a deciduous forest. *Boundary-Layer Meteorology*, 36(1):71–91.
- Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., and Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences*, 111(38):13690–13696.
- Warne, R. W., Pershall, A. D., and Wolf, B. O. (2010). Linking precipitation and C₃–C₄ plant production to resource dynamics in higher-trophic-level consumers. *Ecology*, 91(6):1628–1638.
- Wilke, S. (2010). *Performing Tropics Alexander von Humbolt's Ansichten Der Natur and the Colonial Roots of Nature Writing*. Univ Press Virginia, Charlottesville.
- Wolkovich, E. M., Cook, B. I., and Davies, T. J. (2014). Progress towards an interdisciplinary science of plant phenology: Building predictions across space, time and species diversity. *New Phytologist*, 201(4):1156–1162.
- Yu, R., Ruddell, B. L., Kang, M., Kim, J., and Childers, D. (2019). Anticipating global terrestrial ecosystem state change using FLUXNET. *Global Change Biology*, 25(7):2352–2367.
- Zakharov, V. M. and Trofimov, I. E. (2014). Homeostatic mechanisms of biological systems: Development homeostasis. *Russian Journal of Developmental Biology*, 45(3):105–116.

APPENDIX A

Complementary results: Empirical comparison between GPP_{\max} and GPP_{sat}

Theoretically, ecosystem functional properties are conceptualized to link plant traits with ecosystem biogeochemical processes (Reichstein et al., 2014). Nevertheless, I consider that its definition on the framework of functional biogeography can also include the empirical optimum ecosystem processes and not only the theoretical ones; as the functionality concept is to delimit what organisms do instead of what organisms are composed (Violle et al., 2014). In this section I made a empirical comparison between GPP_{sat} and GPP_{\max} . GPP_{sat} is an ecosystem functional property, where GPP estimation is derived assuming optimal conditions of photosynthetic active radiation (PAR) and absorbed photosynthetic radiation (APAR) (Musavi et al., 2016). GPP_{\max} is the maximum observed GPP during the growing season of an ecosystem. I hypothesize that the empirical values of GPP_{\max} will be highly similar to estimations of GPP_{sat} and then can be used as an ecosystem functional property to describe an ecosystem optimum state.

To compare GPP_{sat} and GPP_{\max} I used the GPP_{sat} values reported by Migliavacca et al. (2021) for 46 eddy covariance sites from FLUXNET 2015 dataset

(Pastorello et al., 2020). I estimated the GPP_{max} values following the methodology proposed by Pabon-Moreno et al. (2020) for the same sites, and the same time period of Migliavacca et al. (2021) study. It is important to clarify that Migliavacca et al. (2021) report one GPP_{sat} value per site (90th percentile), while Pabon-Moreno et al. (2020) track 10 maximum days/GPP per growing season. Then it is expected a higher variability from Pabon-Moreno et al. (2020), for this reason, the mean of the GPP_{max} values were estimated per site. A detailed comparison between the values per side EC site is presented below. Finally, I compared the mean GPP_{max} values (across years) with the GPP_{sat} values using linear regression and estimating the Person-correlation coefficient.

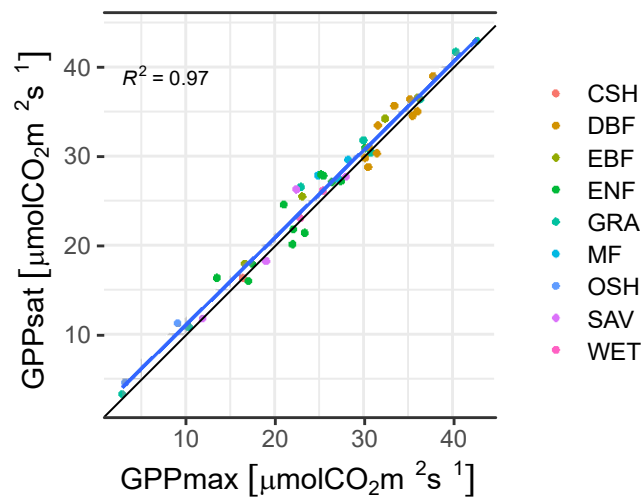


Figure A.1: Comparison of GPP_{max} (Pabon-Moreno et al., 2020), and GPP_{sat} (Migliavacca et al., 2021) Colors represent the vegetation type of each Eddy Covariance site. CSH: Closed Shrublands; DBF: Deciduous Broadleaf Forests; EBF: Evergreen Broadleaf Forests; ENF: Evergreen Needleleaf Forests; GRA: Grasslands; MF: Mixed Forests; OSH: Open Shrublands; SAV: Savannas; WET: Permanent Wetlands. The black line represents a perfect fit. The blue line represents the fit of the observations.

When the mean GPP_{max} is compared with GPP_{sat} values, I found a high correlation between both metrics as evidenced in Figure A.1 and a Person correlation result of 0.987. When the comparison is performed per vegetation types, I found

that in Mixed Forest, Open Shrublands, and Wetlands, the GPP_{sat} distribution is slightly higher than the GPP_{max} distribution (Figure A.3). While, for Deciduous Broadleaf Forests, Evergreen Broadleaf Forests, Evergreen Needleleaf Forests, Grasslands, and Savanas, both GPP_{max} and GPP_{sat} have similar distributions. For Closed Shrublands, where only one site is reported, the differences are in the $0.01 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ order. When the values are aggregated by climatic classes, I found that for tropical monsoon, humid subtropical, and subarctic climate GPP_{max} is lower than GPP_{sat} . In contrast, for tropical savanna, humid subtropical, and Mediterranean climate GPP_{max} is higher than GPP_{sat} . For Mediterranean, humid continental, subarctic, and tundra climate the distributions of GPP_{max} and GPP_{sat} share similar quantile ranges.

Despite the expected higher variability of GPP_{max} produced by the influence of climatic conditions. The high correlation between GPP_{max} and GPP_{sat} and the similarities of its distributions validate the hypothesis that GPP_{max} values can be used to represent an optimum ecosystem state, even if these are not derived from optimal conditions (i.e., optimal light conditions in the case of GPP_{sat}). Nevertheless, it is important to clarify that each is designed to answer and evaluate different hypotheses. Based on estimations of GPP_{sat} , it is possible to reduce the effect of climate regimens in estimating the value. Then, GPP_{sat} allows us to explore the effect of vegetation structure on the magnitude of GPP. While GPP_{max} estimates allow us to explore the effect of climatic variations on an optimum ecosystem process in time and space.

Appendix A. Complementary results: Empirical comparison between GPP_{max} and GPP_{sat}

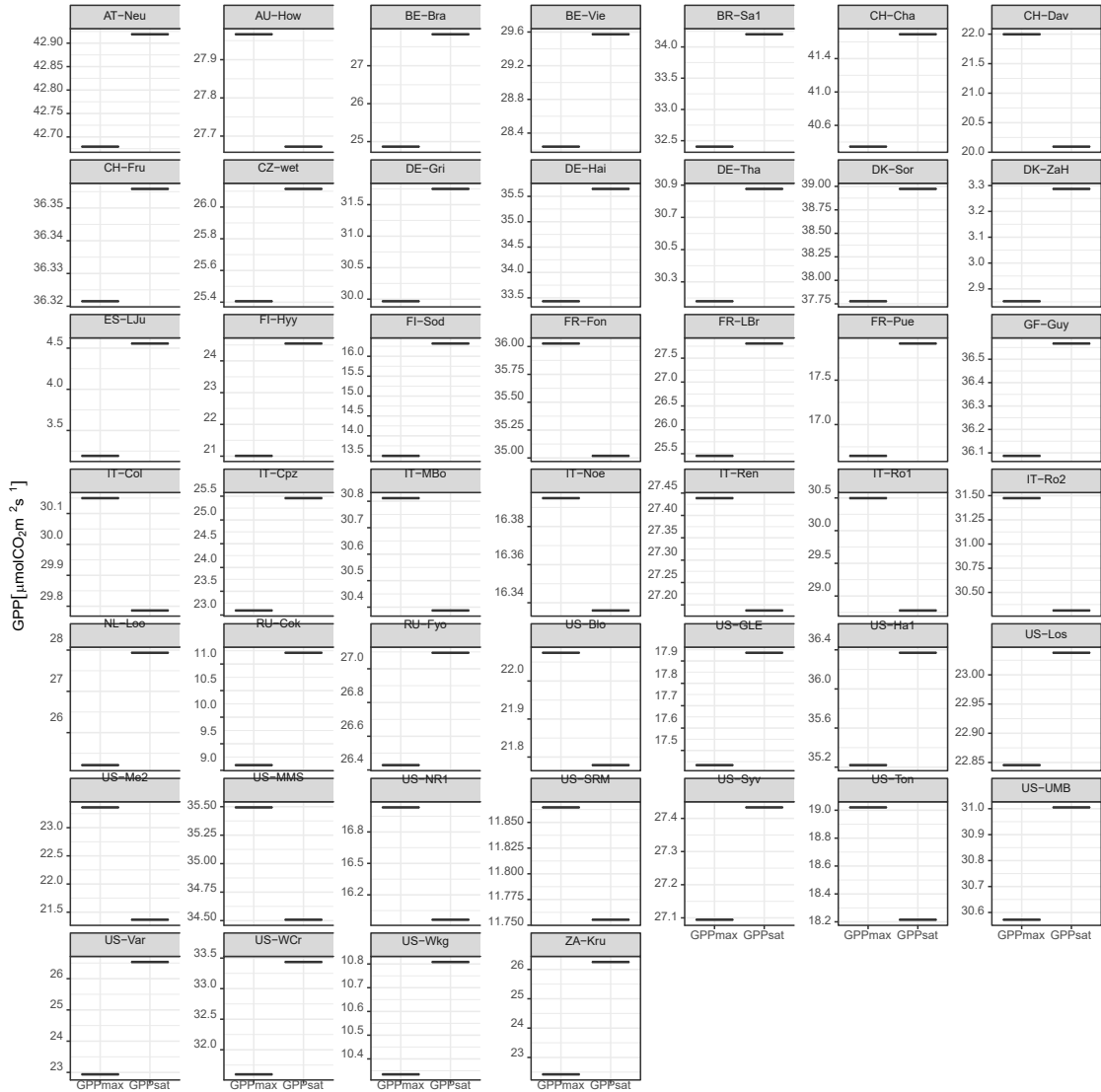


Figure A.2: Comparison of GPP_{max} (Pabon-Moreno et al., 2020), and GPP_{sat} (Migliavacca et al., 2009) per eddy covariance site. For GPP_{max} (Pabon-Moreno et al., 2020) 10 values are reported by growing season while for GPP_{sat} (Migliavacca et al., 2021) a single value is reported by site.

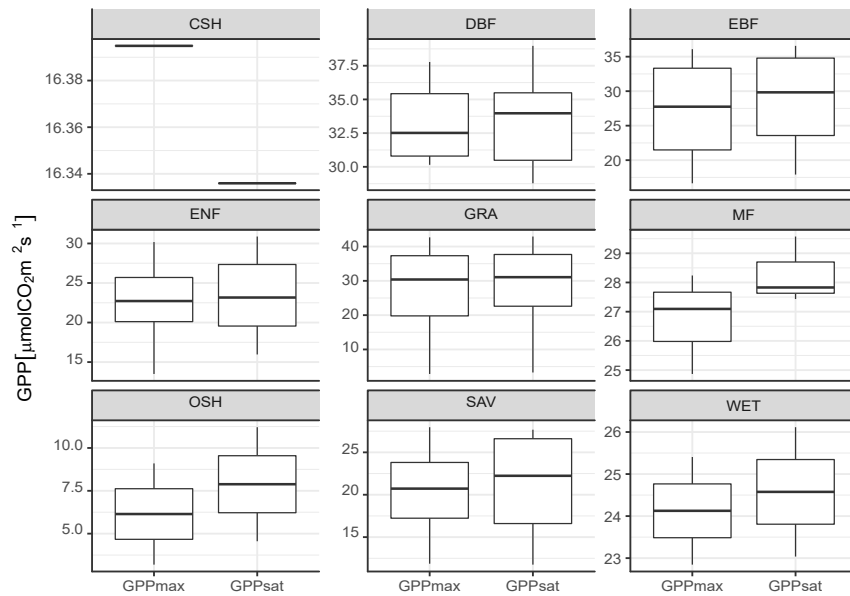


Figure A.3: Comparison of GPP_{max} (Pabon-Moreno et al., 2020), and GPP_{sat} (Migliavacca et al., 2009) per vegetation type. CSH: Closed Shrublands; DBF: Deciduous Broadleaf Forests; EBF: Evergreen Broadleaf Forests; ENF: Evergreen Needleleaf Forests; GRA: Grasslands; MF: Mixed Forests; OSH: Open Shrublands; SAV: Savannas; WET: Permanent Wetlands.

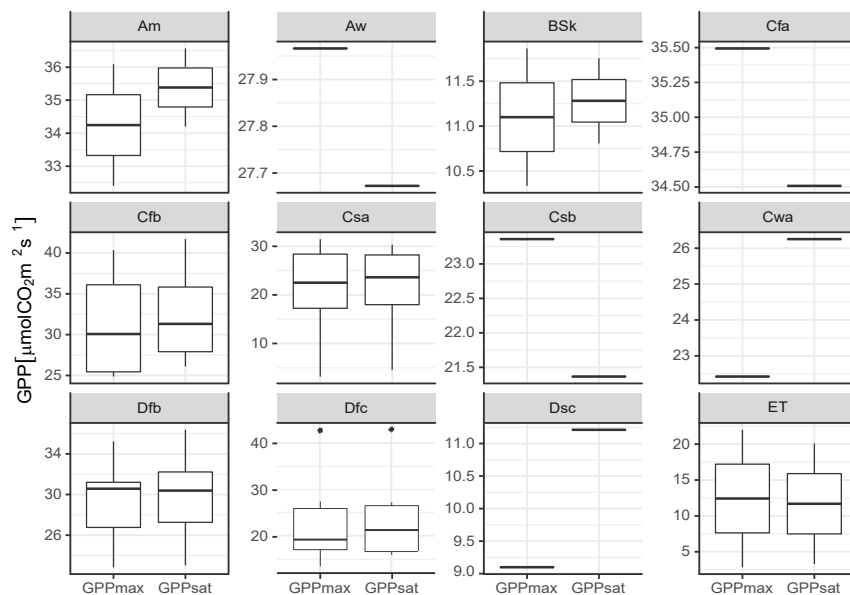


Figure A.4: Comparison of GPP_{max} (Pabon-Moreno et al., 2020), and GPP_{sat} (Migliavacca et al., 2009) per climate class. Tropical monsoon climate (Am), tropical savanna climate (Aw), cold semiarid climate (BSk), humid subtropical climate (Cfa), oceanic climate (Cfb), hot summer Mediterranean climate (Csa), warm summer Mediterranean climate (Csb), humid subtropical climate (Cwa), humid continental climate (Dfb), subarctic climate (Dfc, Dsc), and tundra climate (ET).

APPENDIX B

Supplementary information: Ecosystem physio-phenology
revealed using circular statistics

Ecosystem physio-phenology revealed using circular statistics

Supplementary information (I)

Daniel E. Pabon-Moreno¹, Talie Musavi¹, Mirco Migliavacca¹, Markus Reichstein^{1,2},
Christine Römermann^{2,3}, and Miguel D. Mahecha^{1,2}

¹Max Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, 07745 Jena, Germany.

²German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany.

³Friedrich Schiller University, Institute of Ecology and Evolution, Philosophenweg 16, D-07743 Jena, Germany.

1 Half-time sensitivity analysis (System memory to explain DOY_{GPPmax})

The optimum halftime parameter is estimated for each site showing that for most of the unimodal sites JS correlation is maximum when the halftime parameter takes values between 2 and 100 days. Per vegetation type Deciduous Broadleaf Forest (DBF), Evergreen Needleleaf Forest (ENF) and Grassland (GRA) have similar values for the optimum half-time parameter
5 (Figure S3). Per climate classes, the oceanic climate (Cfb) has the highest variation. There are only 2 unimodal sites with an optimum half-time greater than 180 days: IT-MBo and FR-Fon. On the other hand, for the bimodal site (ES-LJu) the maximum JS coefficient is obtained with a half-time of 13 for the first growing season and 21 days for the second one (Table S2). Estimating the halftime of the drivers per site is a prerequisite for optimizing the predictions with the circular regressions in the next step. For most of the sites, the JS correlation coefficient is maximum between 0.98 and 0.7 (Figure S5).

10 Our results for the optimum halftime parameter between 2 and 100 days for all sites are similar to the time window length of 15 to 120 days required to explain the variations on the leaf unfolding for different tree species in Europe (Fu et al., 2015). Or, the time window length of 45 to 95 days to explain the flowering day of different plant species in Switzerland (Güsewell et al., 2017). No matter what phenological event is being analyzed all studies agree that phenological events are influenced by past climatic conditions in a cumulative form.

15 In our case, the use of a half-life decay function changes the interpretation of the optimum halftime day parameter indicating that half of the contribution is given before the halftime day in an exponential form and that the contribution of the rest of the days is low, but not equal to 0. Finally, we find that the optimum half-time is not necessarily related to the vegetation type or the predominant climate class in each site. We suggest that it could be more related to the species dominance and the spatial arrangement and the intraspecific climatic variability of each site.

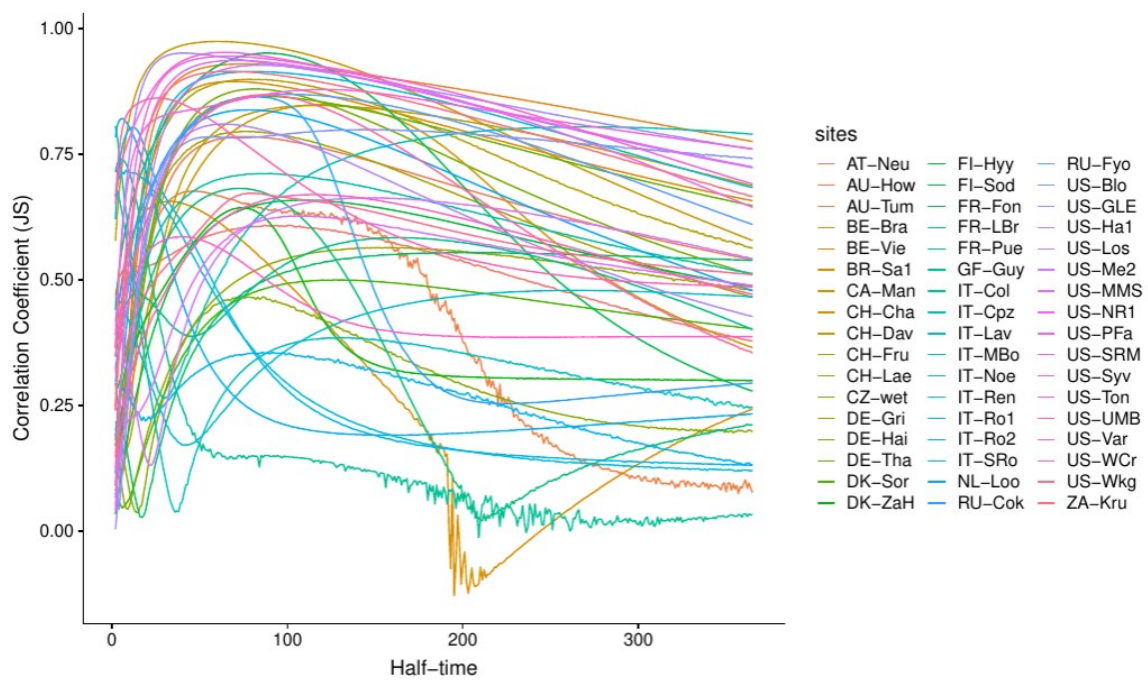


Figure S1. Half-time sensitivity analysis. The correlation coefficient (JS) between the observed and predicted DOY_{GPPmax} using different half-time values. Each FLUXNET site is represented with a color.

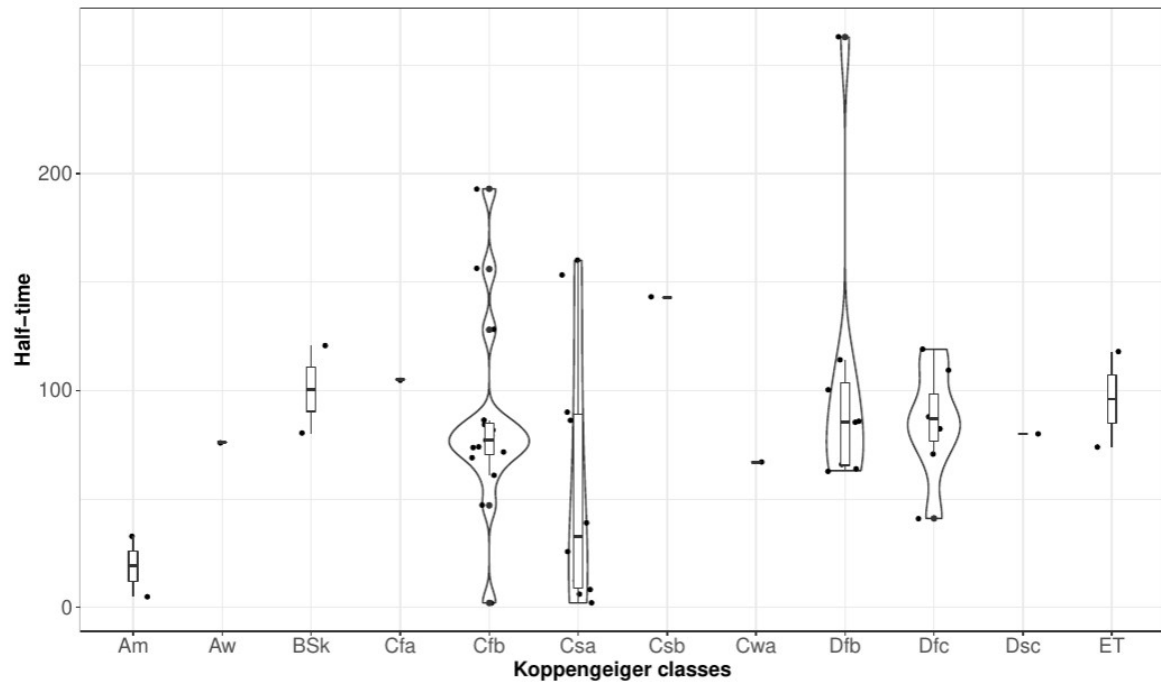


Figure S2. Half-time sensitivity analysis. Distribution of the half-time when the Jammalamanaka-Sarna (JS) coefficient is maximum for each FLUXNET site using as classification criterium the Koppen climate classes: Tropical monsoon climate (Am), Tropical savanna climate (Aw), Cold semi-arid climates (BSk), Humid subtropical climate (Cfa), Oceanic climate (Cfb), Hot-summer mediterranean climate (Csa), Warm-summer mediterranean climate (Csb), Humid subtropical climate (Cwa), humid continental climate (Dfb), Subarctic climate (Dfc, Dsc), and Tundra climate (ET)

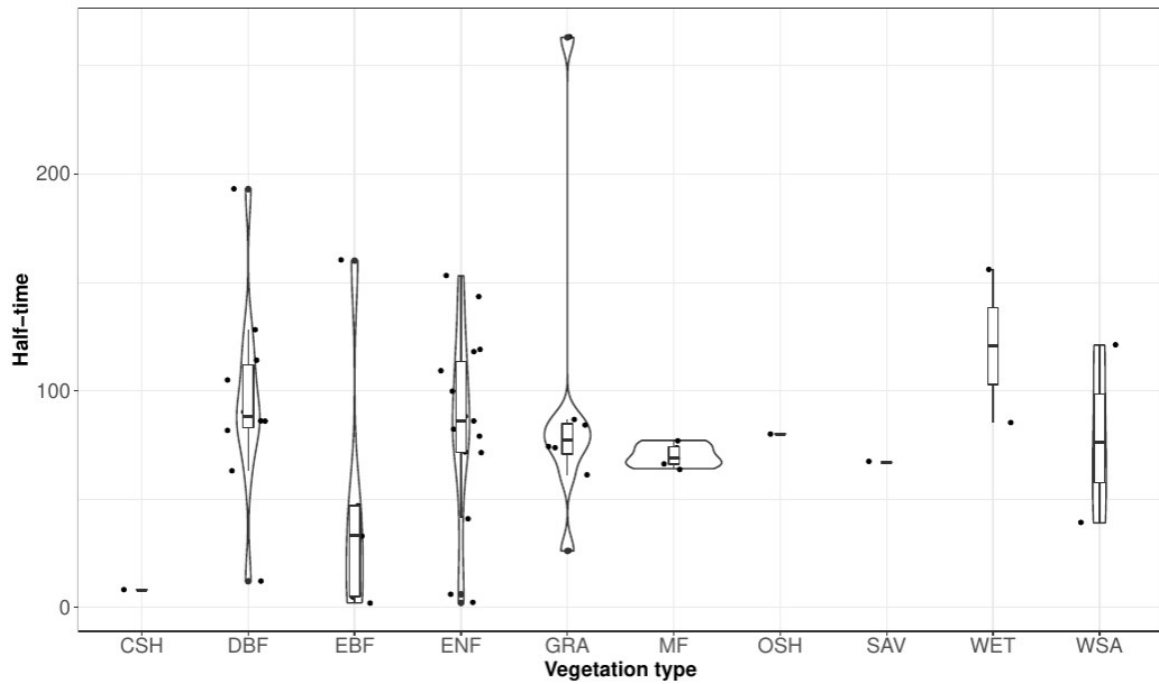


Figure S3. Half-time sensitivity analysis. Distribution of the optimum half-time parameter when the Jammalamadaka-Sarna (JS) coefficient is maximum per vegetation type. Closed Shrublands (CSH), Deciduous Broadleaf Forests (DBF), Evergreen Broadleaf Forests (EBF), Evergreen Needleleaf Forests (ENF), Grasslands (GRA), Mixed Forests (MF), Open Shrublands (SAV) Savannas, Permanent Wetlands (WET), Woody Savannas (WSA)

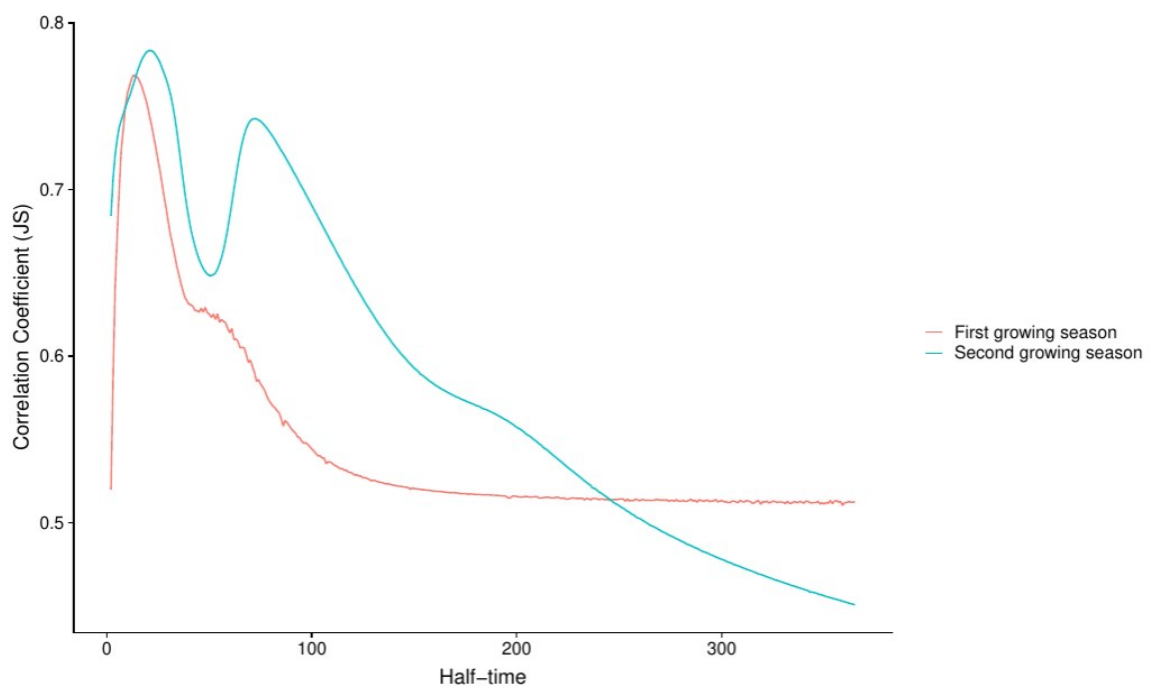


Figure S4. Half-time sensitivity analysis for ES-LJu the unique FLUXNET site analyzed with two growing seasons.

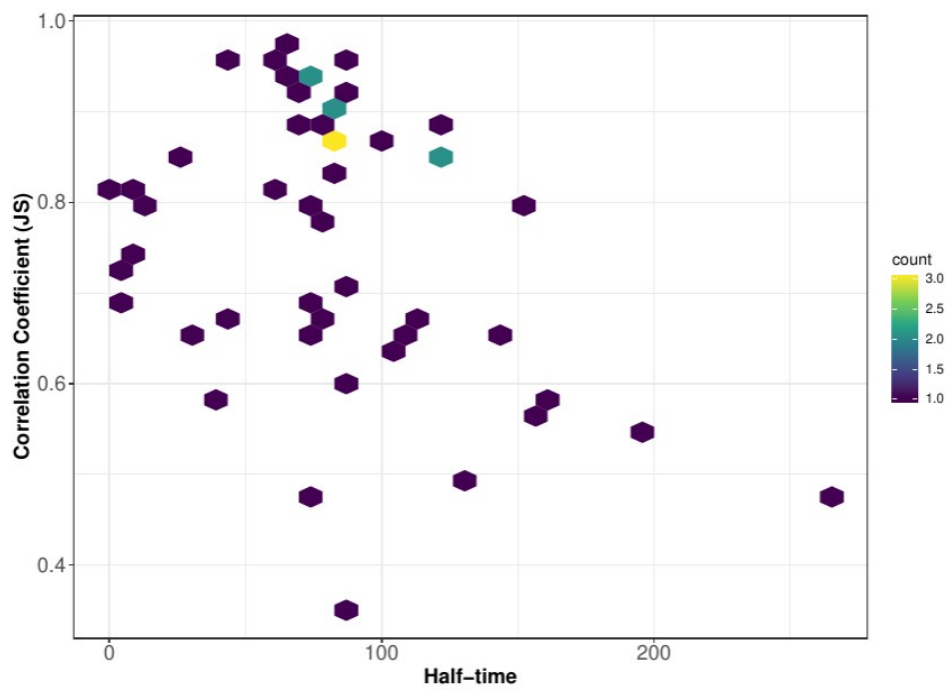


Figure S5. Distribution of the maximum correlation coefficient values when the optimum haltime has been used. Most of the sites have the maximum correlation coefficient when half-time is between 5 and 100 days.

Appendix B. Supplementary information: Ecosystem physio-phenology revealed using circular statistics

Table S1: Optimum half-time coefficient and correlation coefficient per FLUXNET site. We report the name of sites, the climate class of the site following the Köppen-Geiger classification: Tropical monsoon climate (Am), Tropical savanna climate (Aw), Cold semi-arid climates (BSk), Humid subtropical climate (Cfa), Oceanic climate (Cfb), Hot-summer mediterranean climate (Csa), Warm-summer mediterranean climate (Csb), Humid subtropical climate (Cwa), humid continental climate (Dfb), Subarctic climate (Dfc, Dsc), and Tundra climate (ET). We also report the vegetation type where: We also report the Vegetation type: Closed Shrublands (CSH), Deciduous Broadleaf Forests (DBF), Evergreen Broadleaf Forest (EBF), Evergreen Needle-leaf Forests (ENF), Grasslands (GRA), Mixed Forests (MF), Open Shrublands (OSH), Savannas (SAV), Permanent Wetlands (WET), Woody Savannas (WSA).

Site name	Köppen-Geiger class	Vegetation type	Half-time	Correlation coefficient (JS)
US-Ha1	Dfb	DBF	63	0.81
US-PFa	Dfb	MF	66	0.94
BE-Bra	Cfb	MF	74	0.93
BE-Vie	Cfb	MF	69	0.89
DE-Tha	Cfb	ENF	86	0.86
DK-Sor	Cfb	DBF	128	0.5
FI-Hyy	Dfc	ENF	109	0.66
IT-Col	Csa	DBF	90	0.71
NL-Loo	Cfb	ENF	79	0.84
CH-Dav	ET	ENF	118	0.85
RU-Fyo	Dfb	ENF	100	0.87
US-NR1	Dfc	ENF	71	0.94
IT-Ren	Dfc	ENF	82	0.91
US-MMS	Cfa	DBF	105	0.63
US-WCr	Dfb	DBF	86	0.87
CA-Man	Dfc	ENF	119	0.85
DK-ZaH	ET	GRA	74	0.64
FR-Pue	Csa	EBF	160	0.58
US-Los	Dfb	WET	85	0.93
US-UMB	Dfb	DBF	114	0.67
US-Var	Csa	GRA	26	0.86
AU-How	Aw	WSA	76	0.78
AU-Tum	Cfb	EBF	47	0.68
FI-Sod	Dfc	ENF	88	0.95

IT-SRo	Csa	ENF	6	0.82
US-Syv	Dfb	MF	64	0.95
US-Ton	Csa	WSA	39	0.59
ZA-Kru	Cwa	SAV	67	0.91
DE-Hai	Cfb	DBF	82	0.88
FR-LBr	Cfb	ENF	72	0.68
IT-Cpz	Csa	EBF	2	0.81
US-Me2	Csb	ENF	143	0.66
IT-Lav	Cfb	ENF	2	0.68
RU-Cok	Dsc	OSH	80	0.86
AT-Neu	Dfc	GRA	87	0.61
CH-Lae	Cfb	MF	77	0.8
DE-Gri	Cfb	GRA	74	0.47
GF-Guy	Am	EBF	5	0.74
IT-MBo	Dfb	GRA	263	0.48
IT-Noe	Csa	CSH	8	0.72
IT-Ro2	Csa	DBF	86	0.36
US-Blo	Csa	ENF	153	0.8
US-GLE	Dfc	ENF	41	0.95
US-SRM	Bsk	WSA	121	0.88
US-Wkg	Bsk	GRA	80	0.67
BR-Sa1	Am	EBF	33	0.66
CH-Cha	Cfb	GRA	61	0.97
CH-Fru	Cfb	GRA	84	0.9
FR-Fon	Cfb	DBF	193	0.55
CZ-wet	Cfb	WET	156	0.56
IT-Ro1	Csa	DBF	12	0.8

Table S2. Results of the optimum half-time and the maximum correlation coefficient for " Llano de los Juanes", Spain with Köppen-Geiger class: Hot-summer Mediterranean climate (Csa) and vegetation type: Open Shrublands (OSH)

Site name	Köppen-Geiger class	Vegetation type	DOY _{GPPmax} (GS 1)		DOY _{GPPmax} (GS 2)	
			Optimum Halftime	Correlation coefficient (JS)	Optimum Halftime	Correlation coefficient (JS)
ES-Lju	Csa	OSH	13	0.77	21	0.78

20 2 Phenological summary of the FLUXNET sites (Recovering DOY_{GPPmax})

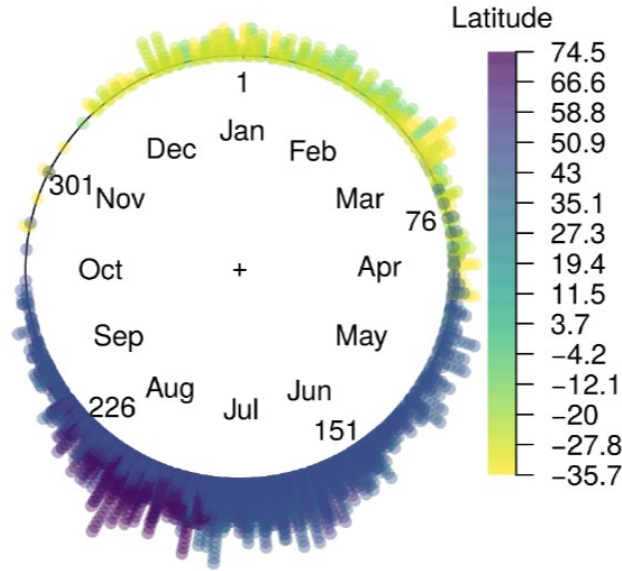


Figure S6. DOY_{GPPmax} distribution across the latitudinal gradient. Most of the DOY_{GPPmax} is reached at the middle of the year. This pattern is generated by the geographical trend of the location of the FLUXNET network in the Northern hemisphere.

Table S3: Mean angular direction and the standard deviation of DOY_{GPPmax} for ecosystems with one growing season per year.

Site name	Mean DOY_{GPPmax} (days)	SD DOY_{GPPmax} (days)
US-Ha1	195.2	19.5
US-PFa	196.6	21.6
BE-Bra	196.71	25.35
BE-Vie	192.28	29.9
DE-Tha	182.91	27.4
DK-Sor	169.96	13.96
FI-Hyy	199.93	17.74
IT-Col	187.75	23.84
NL-Loo	210.27	25.82
CH-Dav	180.99	38.69
RU-Fyo	192.14	23.17

Appendix B. Supplementary information: Ecosystem physio-phenology revealed using circular statistics

US-NR1	201.98	28.79
IT-Ren	193.55	32.34
US-MMS	183.59	22.71
US-WCr	198.26	20.02
CA-Man	215.87	19.96
DK-ZaH	204.39	10.3
FR-Pue	159.97	34.07
US-Los	194.81	13.46
US-UMB	189.13	20.5
US-Var	95.95	21.82
AU-How	28.77	30.87
AU-Tum	24.29	46.9
FI-Sod	214.35	15.66
IT-SRo	142.6	28.3
US-Syv	194.9	25.65
US-Ton	114.63	20
ZA-Kru	15.35	37.09
DE-Hai	190.26	22.84
FR-LBr	177.35	23.83
IT-Cpz	154.5	46.04
US-Me2	182.5	26.15
IT-Lav	167.19	37.11
RU-Cok	209.61	11.48
AT-Neu	161.46	33.73
CH-Lae	185.32	30.71
DE-Gri	178.61	33.51
GF-Guy	214.86	41.17
IT-MBo	169.51	12.22
IT-Noe	120.71	28.96
IT-Ro2	155.48	18.69
US-Blo	199.04	33.38
US-GLE	209.55	17.46
US-SRM	227.65	17.66

US-Wkg	228.54	11.26
BR-Sa1	325.24	38.49
CH-Cha	201.24	38.2
CH-Fru	163.9	39.97
FR-Fon	179.49	24.11
CZ-wet	169.88	17.68
IT-Ro1	148.23	11.33

Table S4. Mean angular direction and standard deviation of DOY_{GPPmax} for ecosystems with two growing seasons

Site name	Koppen	Vegetation type	DOY_{GPPmax} (GS 1)		DOY_{GPPmax} (GS 2)	
			Mean (DOY)	SD (days)	Mean (DOY)	SD (days)
ES-Lju	Csa	OSH	143.62	17.84	302.63	19.18

3 Similarity of regression coefficients per vegetation type and climate classes

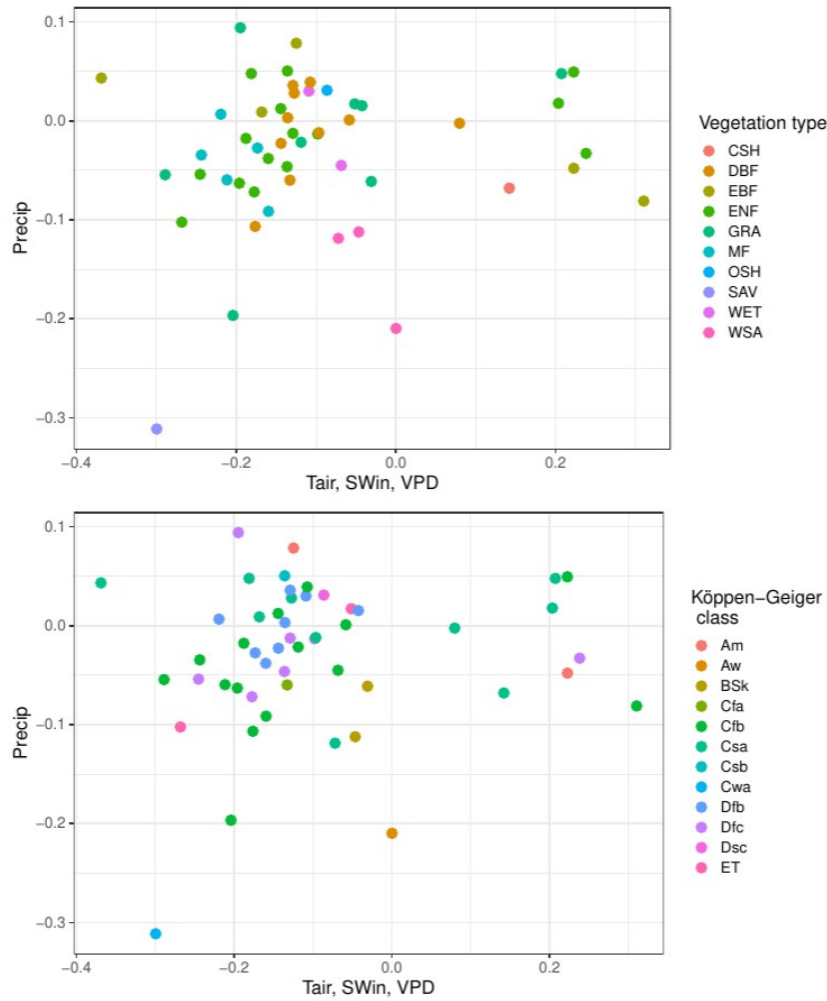


Figure S7. Regression coefficient values of the first component of the PCA of air temperature, shortwave incoming radiation, and vapor pressure deficit, and precipitation. a). Colors per the vegetation: Closed Shrublands (CSH), Deciduous Broadleaf Forests (DBF), Evergreen Broadleaf Forest (EBF), Evergreen Needleleaf Forests (ENF), Grasslands (GRA), Mixed Forests (MF), Open Shrublands (OSH), Savannas (SAV), Permanent Wetlands (WET), Woody Savannas (WSA). c) Colors per Köppen-Geiger climate classes: Tropical monsoon climate (Am), Tropical savanna climate (Aw), Cold semi-arid climates (BSk), Humid subtropical climate (Cfa), Oceanic climate (Cfb), Hot-summer mediterranean climate (Csa), Warm-summer mediterranean climate (Csb), Humid subtropical climate (Cwa), humid continental climate (Dfb), Subarctic climate (Dfc, Dsc), and Tundra climate (ET)

References

- Fu, Y. H., Zhao, H., Piao, S., Peaucelle, M., Peng, S., Zhou, G., Ciais, P., Huang, M., Menzel, A., Peñuelas, J., Song, Y., Vitasse, Y., Zeng, Z., and Janssens, I. A.: Declining global warming effects on the phenology of spring leaf unfolding, *Nature*, 526, 104–107, 25 <https://doi.org/10.1038/nature15402>, <https://www.nature.com/articles/nature15402>, 2015.
- Güsewell, S., Furrer, R., Gehrig, R., and Pietragalla, B.: Changes in temperature sensitivity of spring phenology with recent climate warming in Switzerland are related to shifts of the preseason, *Global Change Biology*, 23, 5189–5202, <https://doi.org/10.1111/gcb.13781>, <http://onlinelibrary.wiley.com/doi/10.1111/gcb.13781/abstracthttp://onlinelibrary.wiley.com/store/10.1111/gcb.13781/asset/gcb13781.pdf?v=1&t=jdmyu5yf&s=8101659af8b310121d3227eb531cf4166cbfc8a8>, 2017.

Ecosystem physio-phenology revealed using circular statistics Supplementary information (II)

Daniel E. Pabon-Moreno¹, Talie Musavi¹, Mirco Migliavacca¹, Markus Reichstein^{1,2},
Christine Römermann^{2,3}, and Miguel D. Mahecha^{1,2}

¹Max Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, 07745 Jena, Germany.

²German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany.

³Friedrich Schiller University, Institute of Ecology and Evolution, Philosophenweg 16, D-07743 Jena, Germany.

1 Principal Component Analysis for Air temperature, Short-wave Incoming Radiation, and Vapor pressure deficit

The principal component analysis shows that for most of the sites the first component explain more than 70 % of the variation for air temperature, short-wave incoming radiation and vapor pressure deficit (Fig. S1, Tab. S1). For the first component the contribution of each variable is similar between the sites (Fig. S2, Tab. S2). The results show that to some extent all variables

5 are represented in the first component.

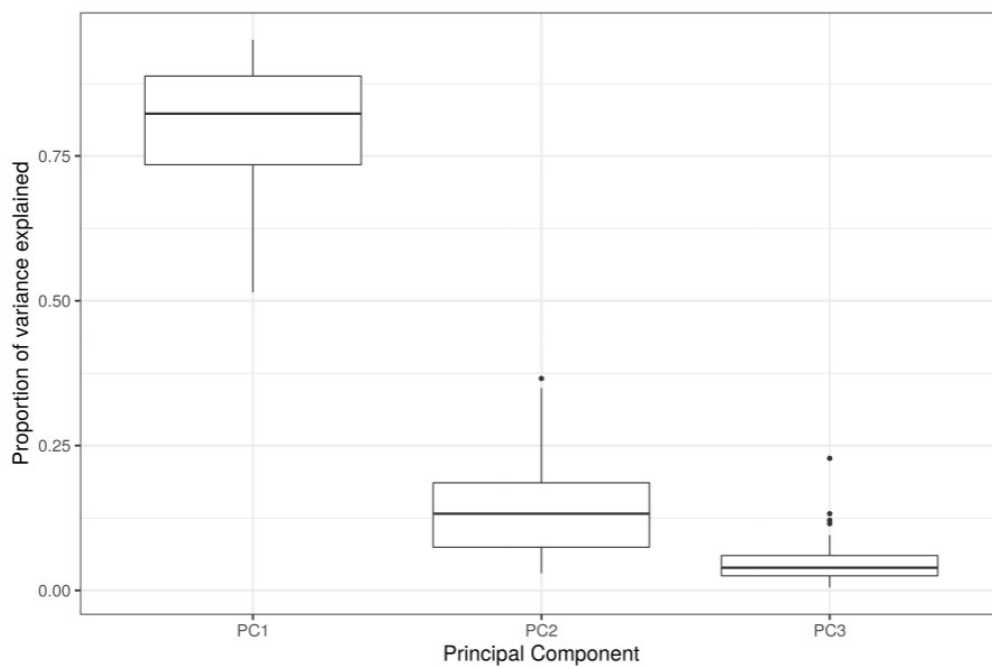


Figure S1. Contribution of each principal component to explain the variance of air temperature, short-wave incoming radiation and vapor pressure deficit. Each category represent the distribution for all the FLUXNET sites analyzed in this study.

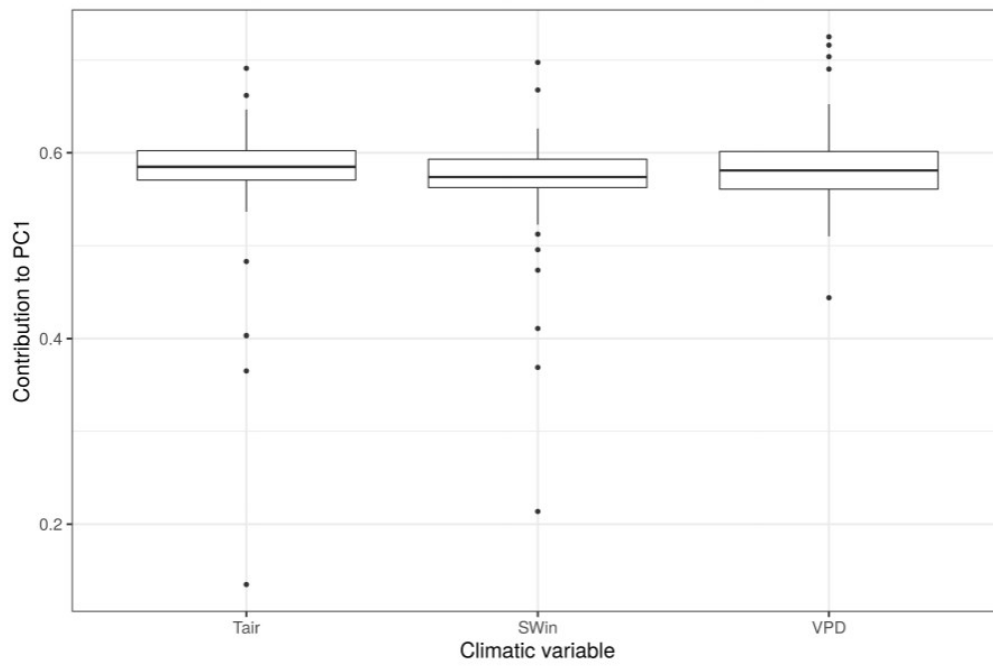


Figure S2. Contribution of the climate variables to PC1. Tair = Air temperature, SWin = Short-wave Incoming radiation, VPD = Vapor pressure deficit. For visualization purposes if the sign of the contribution was negative for all variables of the site, the sign was changed to positive

Table S1: Contribution of the principal components to explain the variance of air temperature, short-wave incoming radiation, and vapor pressure deficit.

Site name	PC1	PC2	PC3
US-Ha1	0.81	0.14	0.05
US-PFa	0.79	0.17	0.04
BE-Bra	0.89	0.07	0.04
BE-Vie	0.87	0.11	0.02
DE-Tha	0.87	0.09	0.04
DK-Sor	0.83	0.12	0.04
FI-Hyy	0.73	0.17	0.1
IT-Col	0.87	0.07	0.06
NL-Loo	0.78	0.15	0.08
CH-Dav	0.82	0.14	0.04
RU-Fyo	0.82	0.13	0.05
US-NR1	0.94	0.04	0.03
IT-Ren	0.82	0.14	0.03
US-MMS	0.82	0.13	0.05
US-WCr	0.75	0.18	0.07
CA-Man	0.64	0.35	0.01
DK-ZaH	0.7	0.24	0.06
FR-Pue	0.73	0.21	0.06
US-Los	0.71	0.24	0.04
US-UMB	0.85	0.13	0.03
US-Var	0.95	0.03	0.02
AU-How	0.62	0.29	0.09
AU-Tum	0.9	0.07	0.03
FI-Sod	0.77	0.21	0.03
IT-SRo	0.9	0.08	0.02
US-Syv	0.91	0.05	0.04
US-Ton	0.92	0.07	0.01
ZA-Kru	0.76	0.19	0.05
DE-Hai	0.86	0.1	0.04
FR-LBr	0.73	0.15	0.12

IT-Cpz	0.78	0.15	0.08
US-Me2	0.86	0.13	0.01
IT-Lav	0.77	0.17	0.06
RU-Cok	0.63	0.37	0.01
AT-Neu	0.88	0.09	0.03
CH-Lae	0.92	0.06	0.03
DE-Gri	0.91	0.05	0.04
GF-Guy	0.84	0.11	0.05
IT-MBo	0.54	0.33	0.13
IT-Noe	0.92	0.05	0.03
IT-Ro2	0.9	0.07	0.03
US-Blo	0.88	0.12	0.01
US-GLE	0.92	0.06	0.02
US-SRM	0.51	0.26	0.23
US-Wkg	0.58	0.29	0.12
BR-Sa1	0.94	0.05	0.01
CH-Cha	0.92	0.06	0.02
CH-Fru	0.73	0.2	0.07
ES-LJu 1	0.82	0.15	0.03
ES-LJu 2	0.73	0.2	0.07
FR-Fon	0.82	0.14	0.04
CZ-wet	0.88	0.1	0.02
IT-Ro1	0.66	0.32	0.02

Table S2: Contribution of the climate variables to the first component (PC1). Tair = Air Temperature, SWin = Short-wave Incoming radiation, VPD = Vapor Pressure Deficit

Site name	Tair	SWin	VPD
US-Ha1	0.57	0.61	0.55
US-PFa	0.62	0.58	0.53
BE-Bra	0.57	0.59	0.57
BE-Vie	0.58	0.6	0.55
DE-Tha	0.58	0.56	0.59
DK-Sor	0.57	0.56	0.61

FI-Hyy	0.54	0.59	0.6
IT-Col	0.58	0.57	0.58
NL-Loo	0.54	0.6	0.59
CH-Dav	0.61	0.54	0.58
RU-Fyo	0.57	0.61	0.56
US-NR1	-0.58	-0.58	-0.58
IT-Ren	0.61	0.58	0.53
US-MMS	0.59	0.59	0.54
US-WCr	0.61	0.6	0.52
CA-Man	0.4	0.57	0.72
DK-ZaH	0.6	0.47	0.65
FR-Pue	0.6	0.5	0.62
US-Los	0.65	0.62	0.44
US-UMB	0.59	0.53	0.6
US-Var	0.58	0.57	0.58
AU-How	0.37	0.67	0.65
AU-Tum	-0.59	-0.57	-0.57
FI-Sod	0.48	0.6	0.63
IT-SRo	0.59	0.57	0.56
US-Syv	0.58	0.57	0.58
US-Ton	0.6	0.56	0.57
ZA-Kru	0.6	0.62	0.51
DE-Hai	0.59	0.58	0.56
FR-LBr	0.57	0.57	0.59
IT-Cpz	-0.61	-0.56	-0.56
US-Me2	0.6	0.53	0.6
IT-Lav	0.62	0.58	0.54
RU-Cok	0.55	0.41	0.73
AT-Neu	0.58	0.59	0.56
CH-Lae	0.57	0.57	0.59
DE-Gri	0.58	0.57	0.58
GF-Guy	0.57	0.57	0.6
IT-MBo	0.13	0.7	0.7

Appendix B. Supplementary information: Ecosystem physio-phenology revealed using circular statistics

IT-Noe	0.59	0.58	0.57
IT-Ro2	0.59	0.57	0.58
US-Blo	0.6	0.53	0.59
US-GLE	0.58	0.57	0.58
US-SRM	0.61	0.56	0.56
US-Wkg	0.66	0.37	0.65
BR-Sa1	0.58	0.57	0.58
CH-Cha	0.57	0.58	0.58
CH-Fru	0.59	0.63	0.51
ES-LJu 1	-0.61	-0.52	-0.6
ES-LJu 2	0.63	0.51	0.59
FR-Fon	0.54	0.6	0.59
CZ-wet	0.56	0.57	0.6
IT-Ro1	0.69	0.21	0.69

Ecosystem physio-phenology revealed using circular statistics

Supplementary information (III)

Daniel E. Pabon-Moreno¹, Talie Musavi¹, Mirco Migliavacca¹, Markus Reichstein^{1,2},
Christine Römermann^{2,3}, and Miguel D. Mahecha^{1,2}

¹Max Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, 07745 Jena, Germany.

²German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany.

³Friedrich Schiller University, Institute of Ecology and Evolution, Philosophenweg 16, D-07743 Jena, Germany.

1 DOY_{GPPmax} sensitivity to different climate drivers

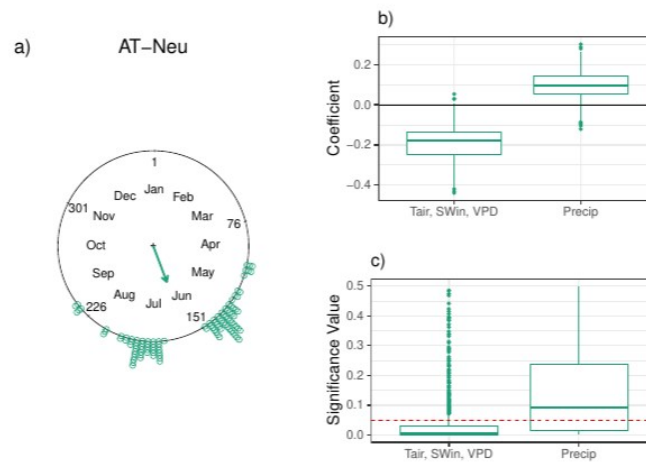


Figure S1. DOY_{GPPmax} sensitivity to different climate drivers for the Neustift (AT-Neu) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

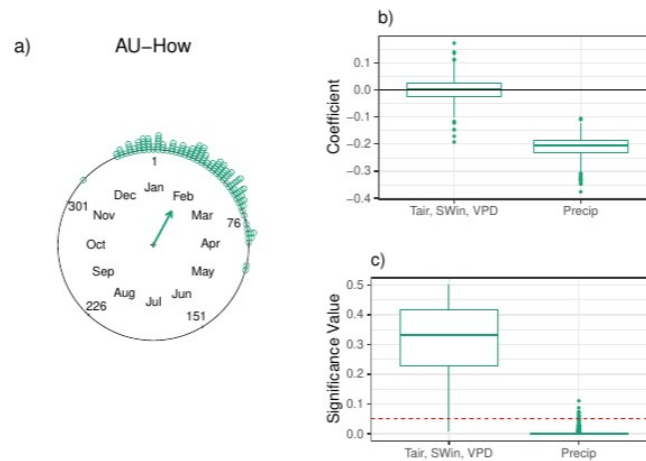


Figure S2. DOY_{GPPmax} sensitivity to different climate drivers for the Howard Springs (AU-How) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

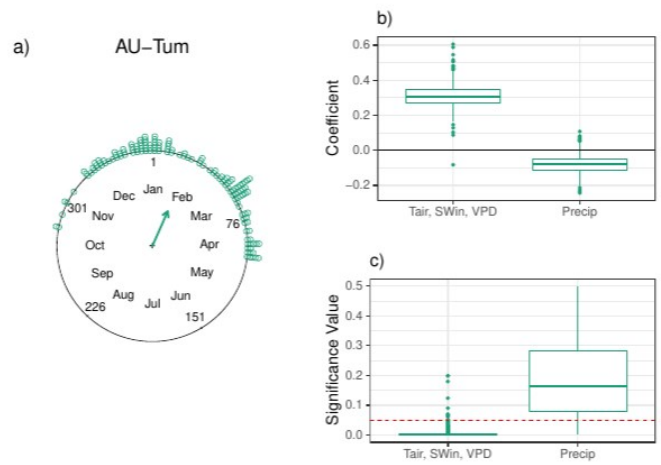


Figure S3. DOY_{GPPmax} sensitivity to different climate drivers for the Tumarumba (AU-Tum) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

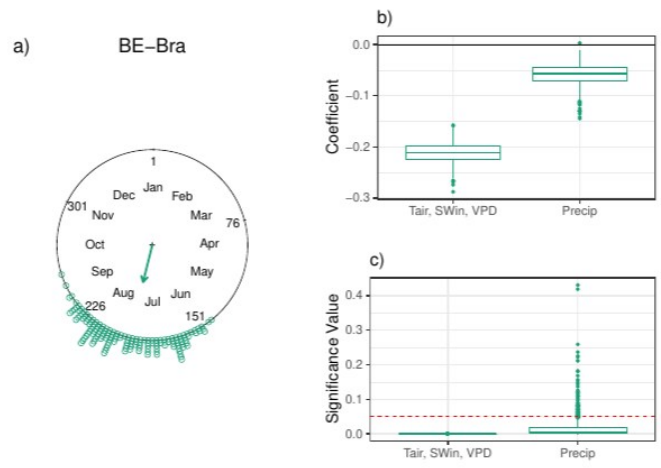


Figure S4. DOY_{GPPmax} sensitivity to different climate drivers for the Brasschaat (BE-Bra) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

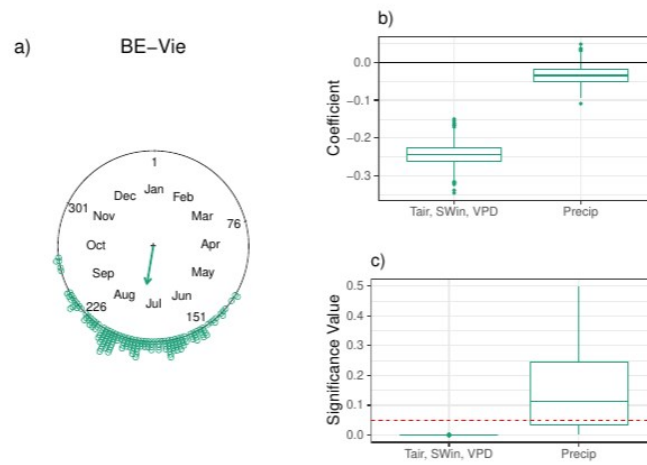


Figure S5. DOY_{GPPmax} sensitivity to different climate drivers for the Vielsalm (BE-Vie) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

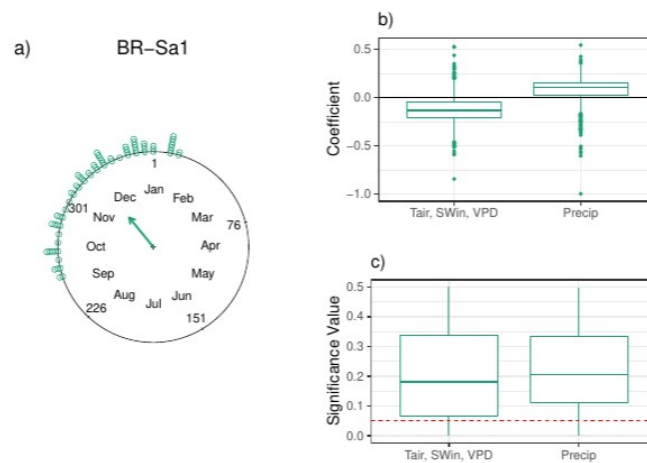


Figure S6. DOY_{GPPmax} sensitivity to different climate drivers for the Santarem-Km67-Primary Forest (BR-Sa1) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

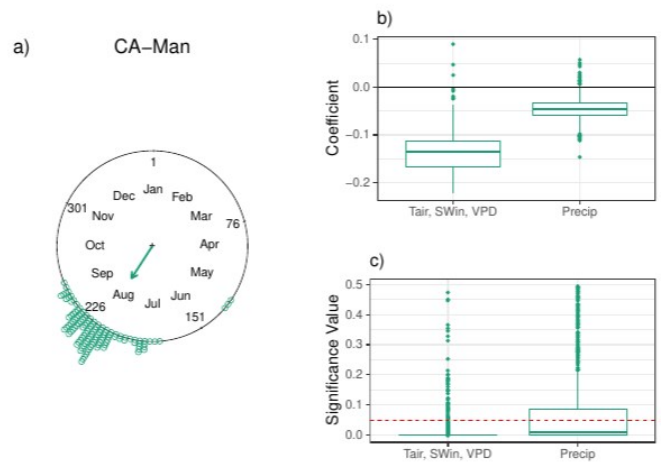


Figure S7. DOY_{GPPmax} sensitivity to different climate drivers for the Manitoba - Northern Old Black Spruce (CA-Man) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

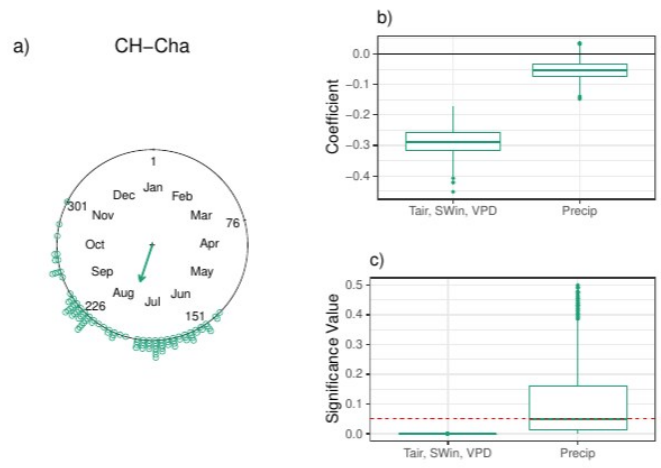


Figure S8. DOY_{GPPmax} sensitivity to different climate drivers for the Chamau (CH-Cha) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

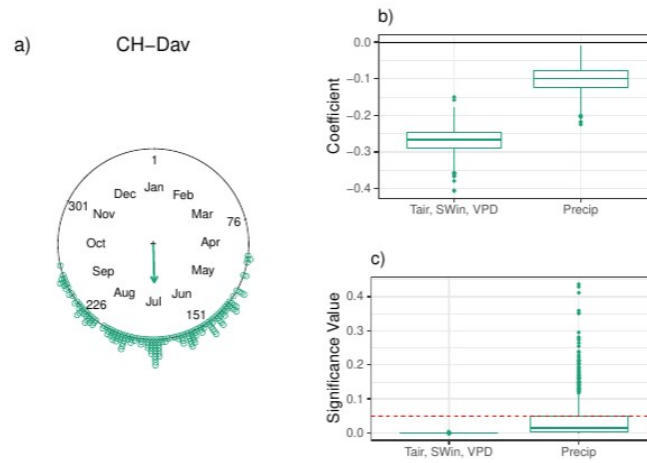


Figure S9. DOY_{GPPmax} sensitivity to different climate drivers for the Davos (CH-Dav) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

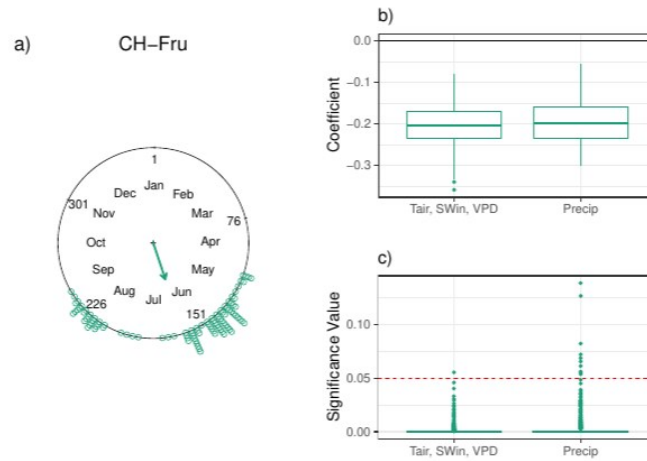


Figure S10. DOY_{GPPmax} sensitivity to different climate drivers for the Frübüel (CH-Fru) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

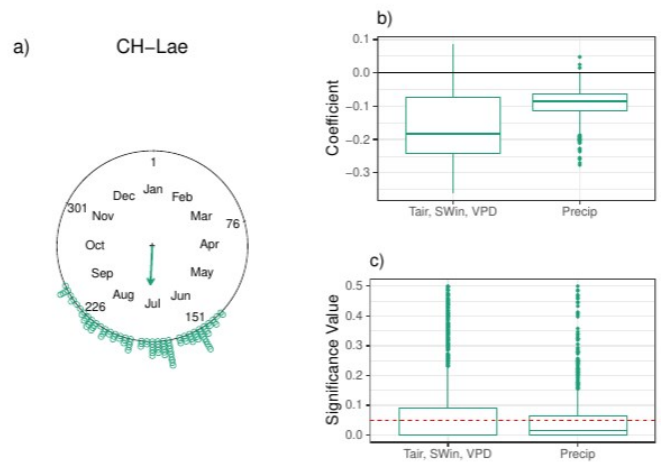


Figure S11. DOY_{GPPmax} sensitivity to different climate drivers for the Laegern (CH-Lae) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

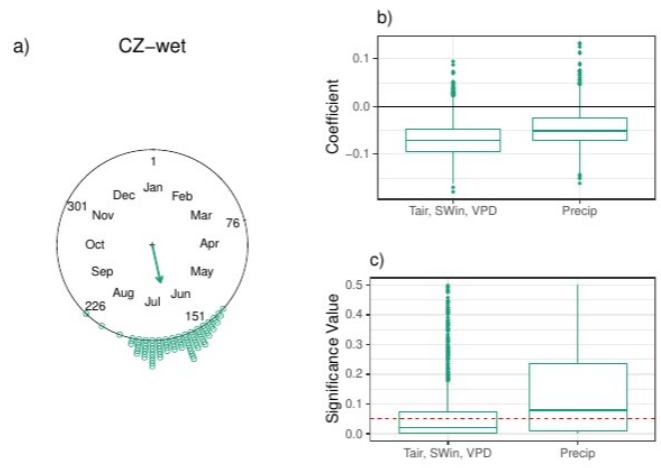


Figure S12. DOY_{GPPmax} sensitivity to different climate drivers for the Trebon (CZECHWET) (CZ-wet) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

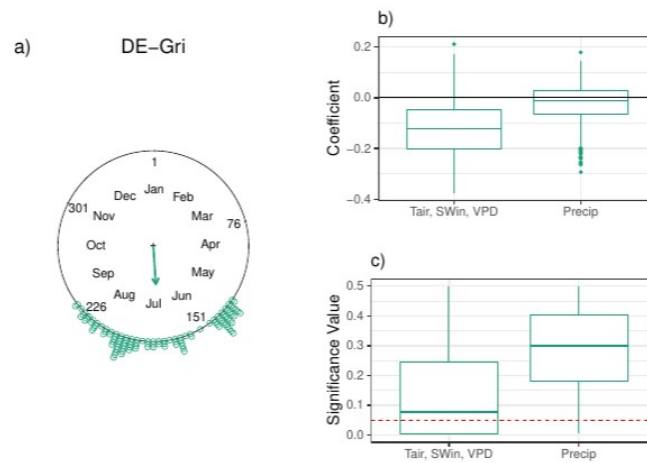


Figure S13. DOY_{GPPmax} sensitivity to different climate drivers for the Grillenburg (DE-Gri) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

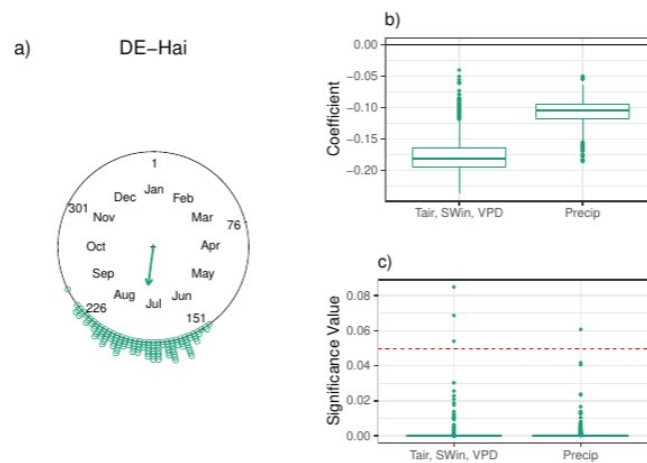


Figure S14. DOY_{GPPmax} sensitivity to different climate drivers for the Hainich (DE-Hai) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

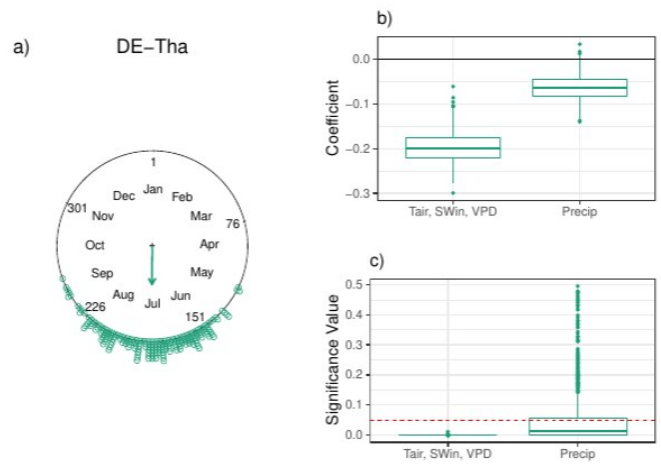


Figure S15. DOY_{GPPmax} sensitivity to different climate drivers for the Tharandt (DE-Tha) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

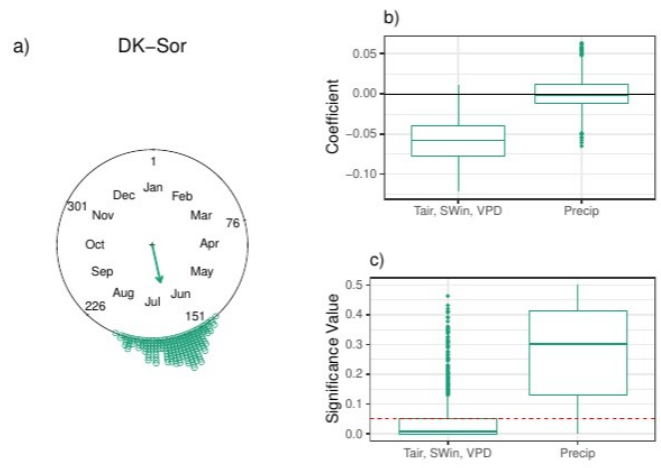


Figure S16. DOY_{GPPmax} sensitivity to different climate drivers for the Soroe (DK-Sor) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

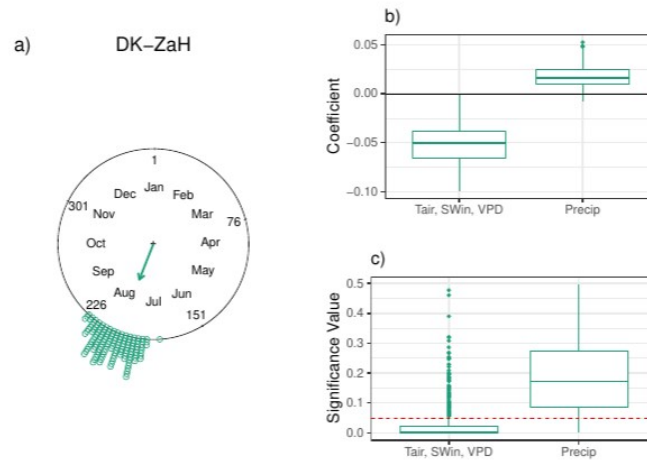


Figure S17. DOY_{GPPmax} sensitivity to different climate drivers for the Zackenberg Heath (DK-ZaH) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

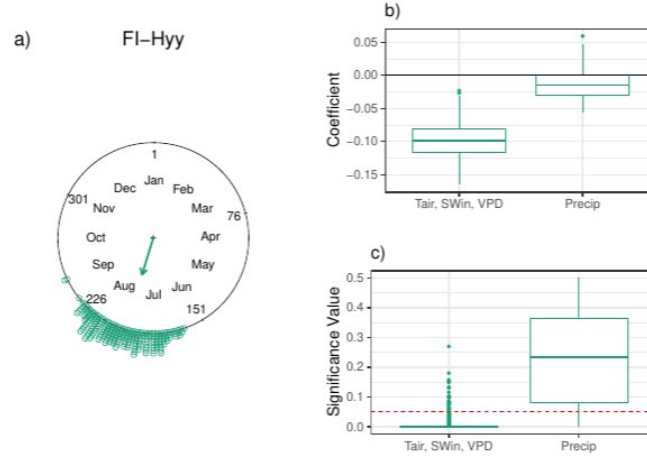


Figure S18. DOY_{GPPmax} sensitivity to different climate drivers for the Hyytiala (FI-Hyy) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

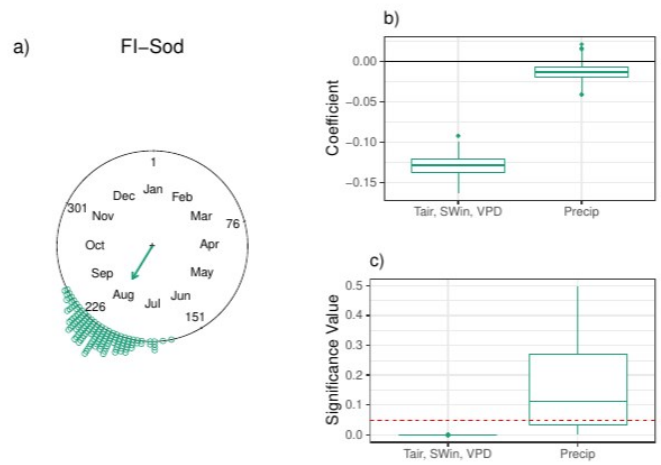


Figure S19. DOY_{GPPmax} sensitivity to different climate drivers for the Sodankyla (FI-Sod) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

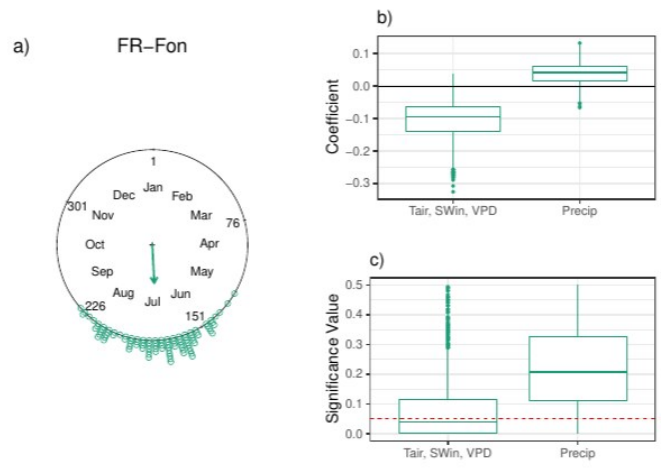


Figure S20. DOY_{GPPmax} sensitivity to different climate drivers for the Fontainebleau-Barbeau (FR-Fon) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

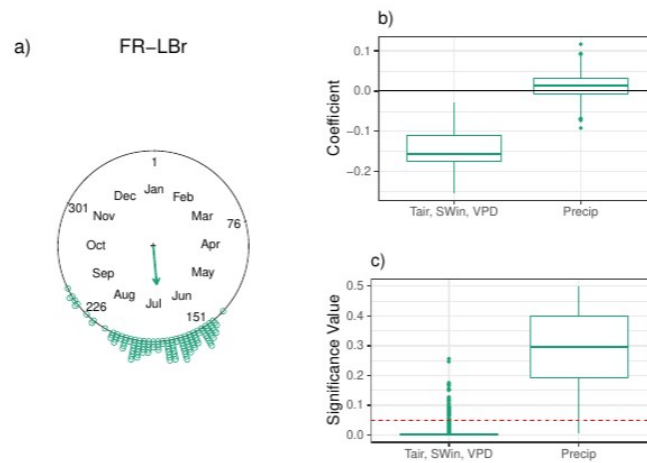


Figure S21. DOY_{GPPmax} sensitivity to different climate drivers for the Le Bray (FR-LBr) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

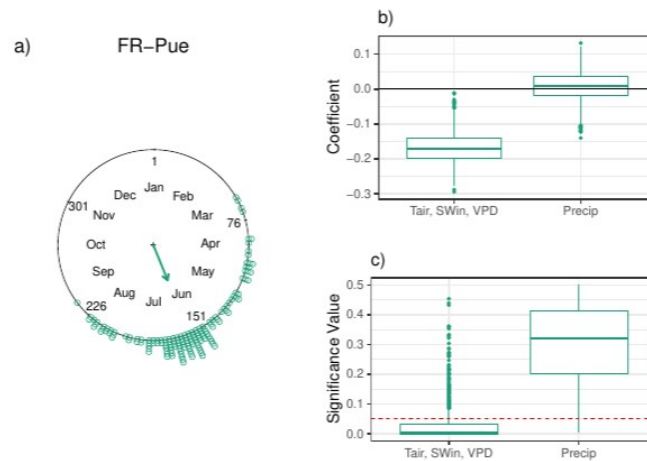


Figure S22. DOY_{GPPmax} sensitivity to different climate drivers for the Puechabon (FR-Pue) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

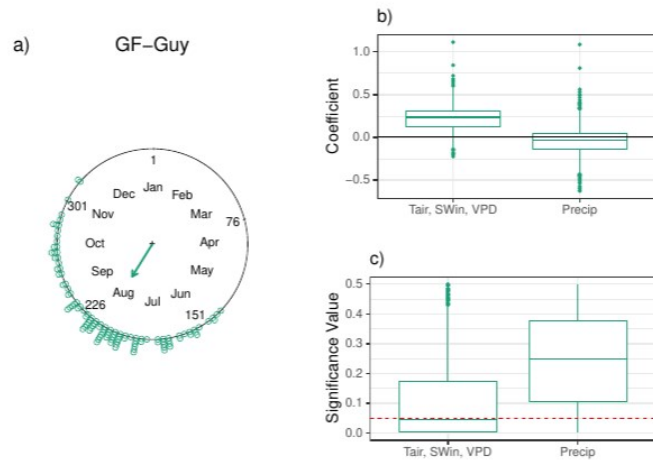


Figure S23. DOY_{GPPmax} sensitivity to different climate drivers for the Guyaflux (GF-Guy) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

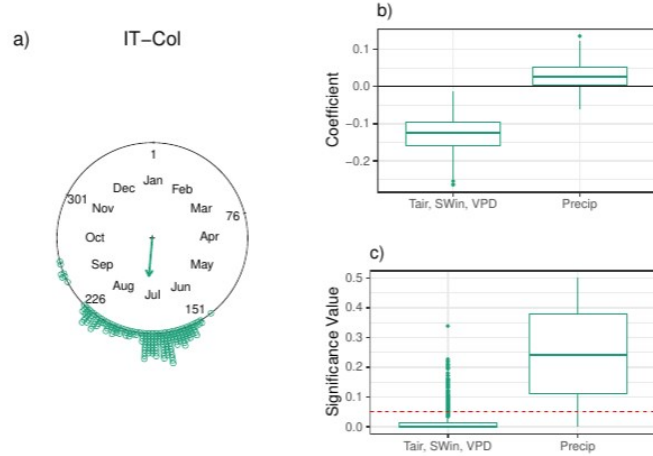


Figure S24. DOY_{GPPmax} sensitivity to different climate drivers for the Collelongo (IT-Col) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

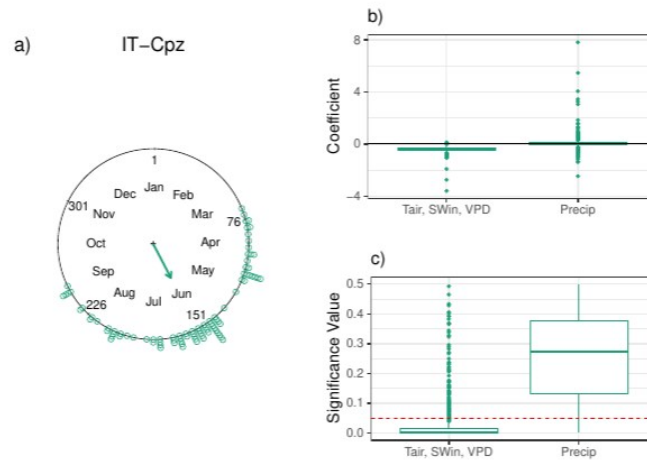


Figure S25. DOY_{GPPmax} sensitivity to different climate drivers for the Castelporziano (IT-Cpz) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

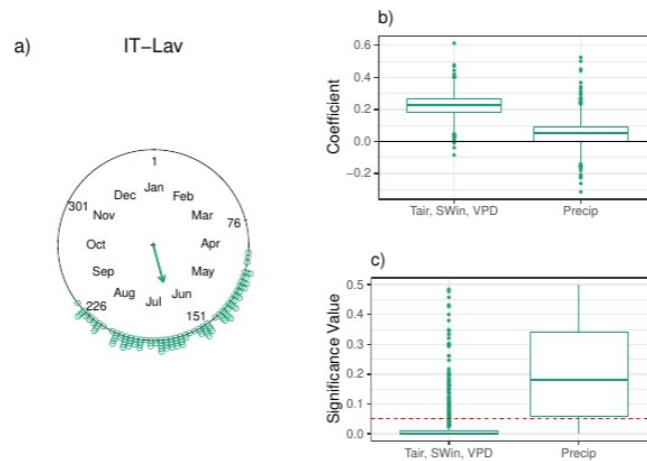


Figure S26. DOY_{GPPmax} sensitivity to different climate drivers for the Lavarone (IT-Lav) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

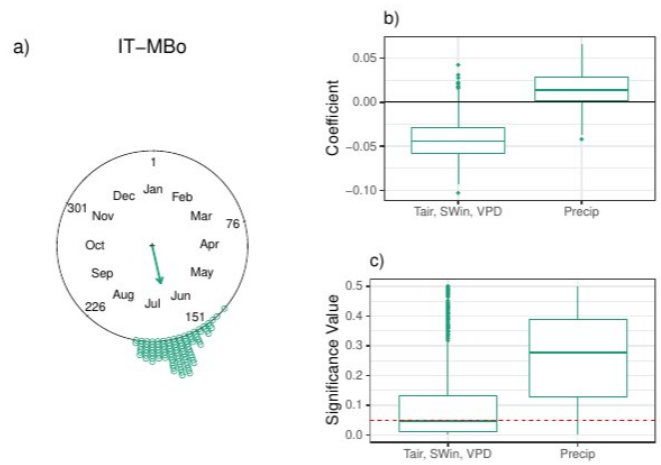


Figure S27. DOY_{GPPmax} sensitivity to different climate drivers for the Monte Bondone (IT-MBo) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

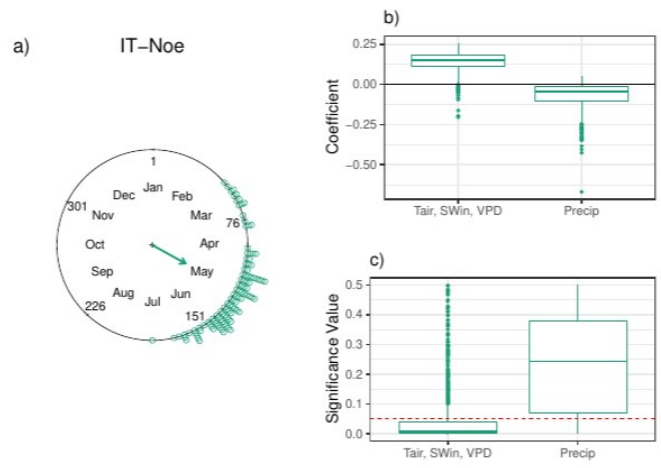


Figure S28. DOY_{GPPmax} sensitivity to different climate drivers for the Arca di Noe - Le Prigionette (IT-Noe) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

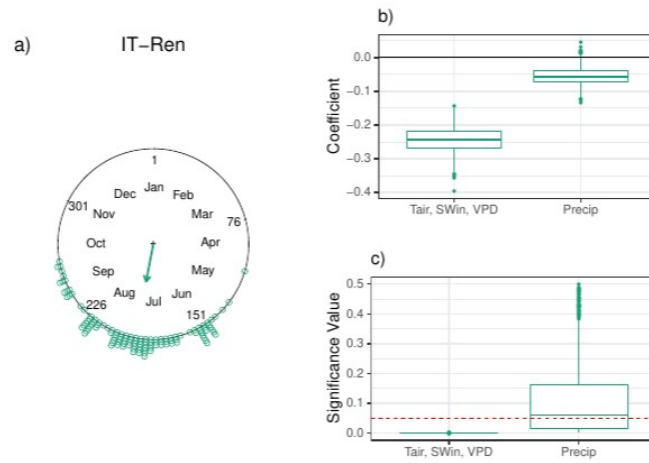


Figure S29. DOY_{GPPmax} sensitivity to different climate drivers for the Renon (IT-Ren) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

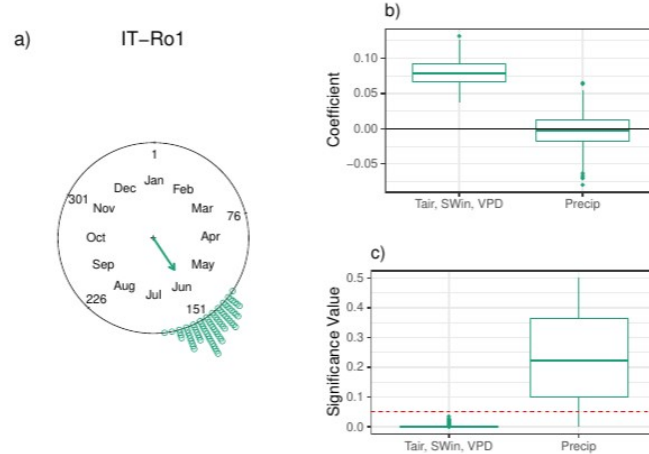


Figure S30. DOY_{GPPmax} sensitivity to different climate drivers for the Roccaespanpani 1 (IT-Ro1) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

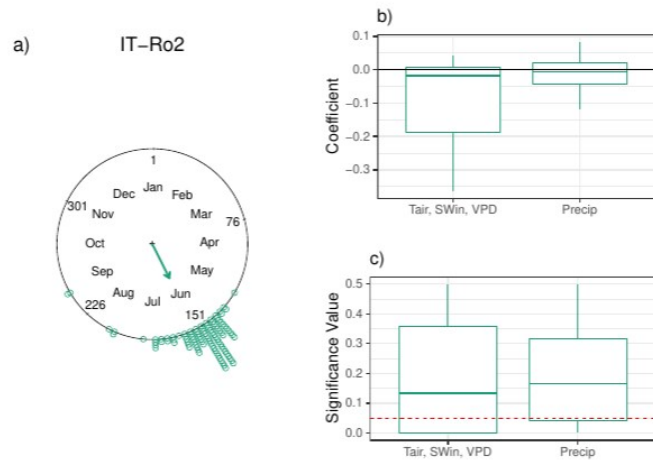


Figure S31. DOY_{GPPmax} sensitivity to different climate drivers for the Roccarespampani 2 (IT-Ro2) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

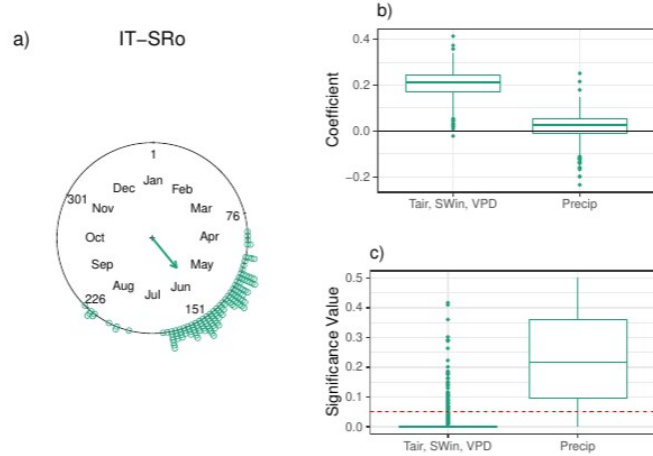


Figure S32. DOY_{GPPmax} sensitivity to different climate drivers for the San Rossore (IT-SRo) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

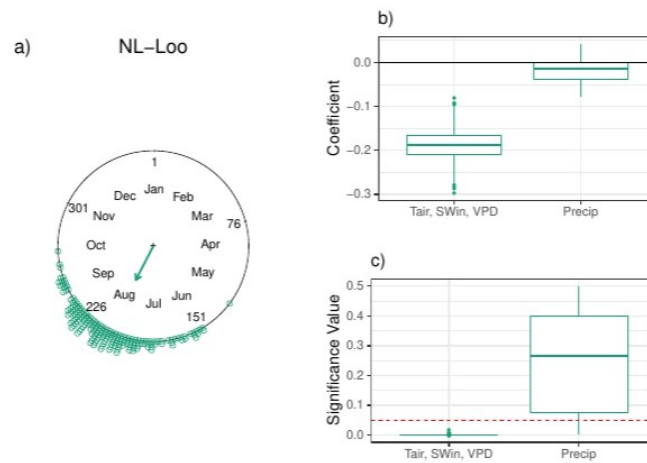


Figure S33. DOY_{GPPmax} sensitivity to different climate drivers for the Loobos (NL-Loo) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

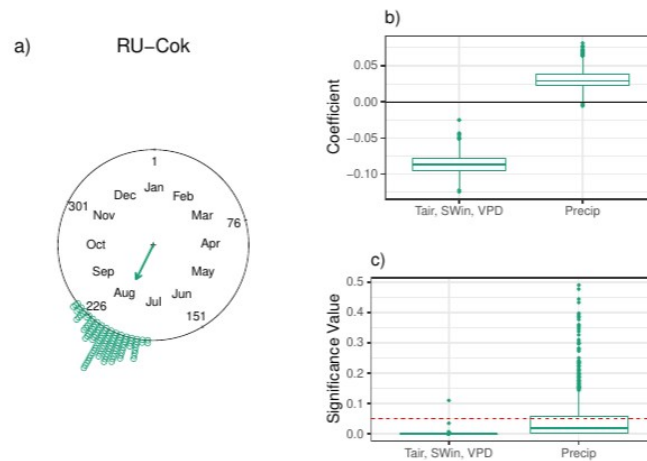


Figure S34. DOY_{GPPmax} sensitivity to different climate drivers for the Chokurdakh (RU-Cok) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

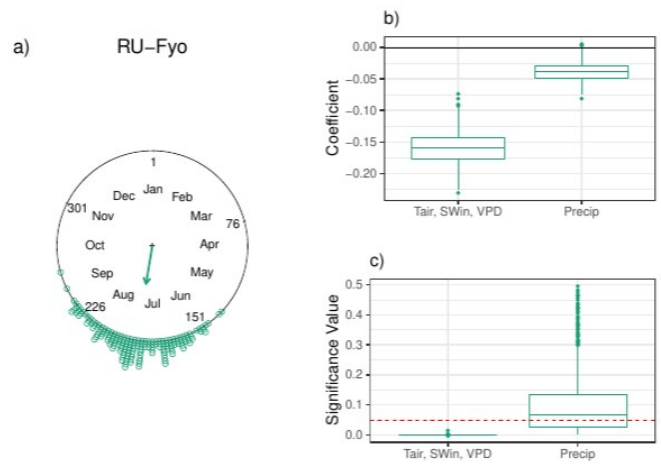


Figure S35. DOY_{GPPmax} sensitivity to different climate drivers for the Fyodorovskoye dry spruce (RU-Fyo) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

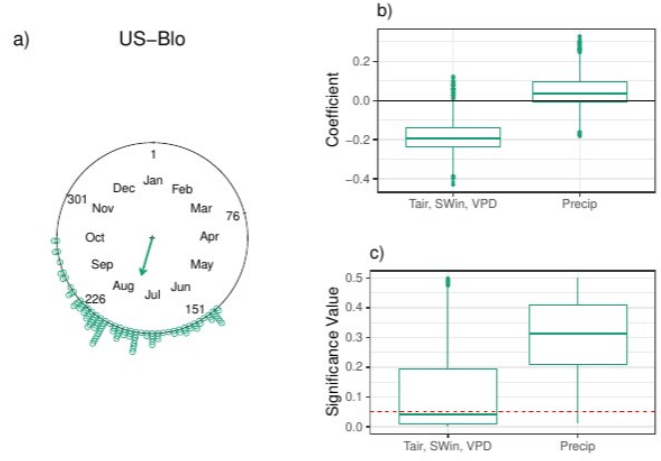


Figure S36. DOY_{GPPmax} sensitivity to different climate drivers for the Blodgett Forest (US-Blo) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

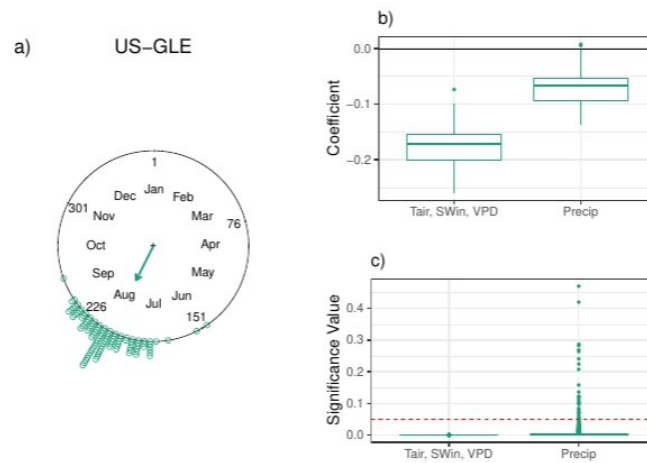


Figure S37. DOY_{GPPmax} sensitivity to different climate drivers for the GLEES (US-GLE) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

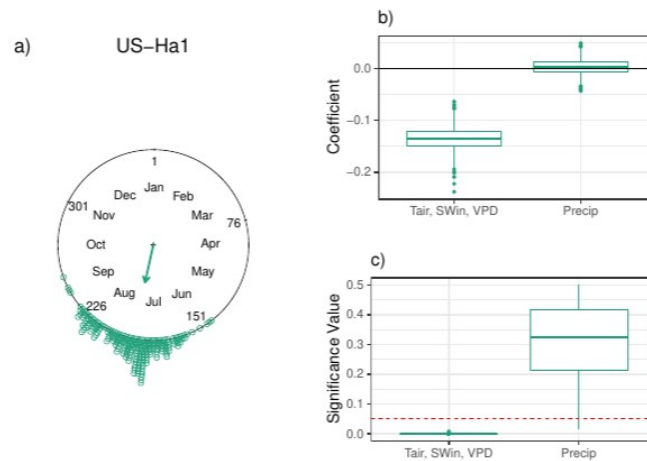


Figure S38. DOY_{GPPmax} sensitivity to different climate drivers for the Harvard Forest EMS Tower (HFR1) (US-Ha1) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

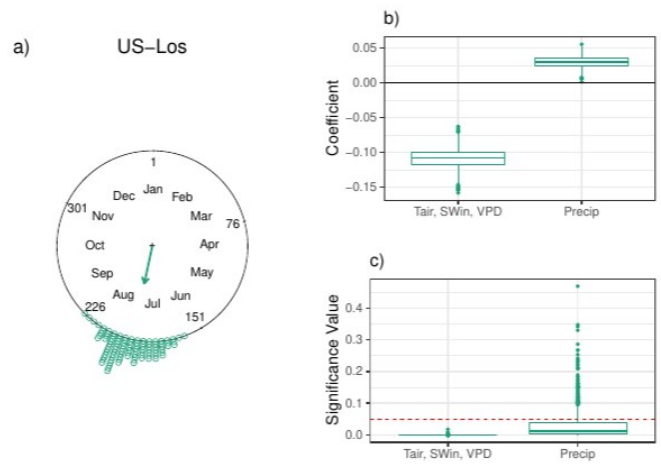


Figure S39. DOY_{GPPmax} sensitivity to different climate drivers for the Lost Creek (US-Los) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

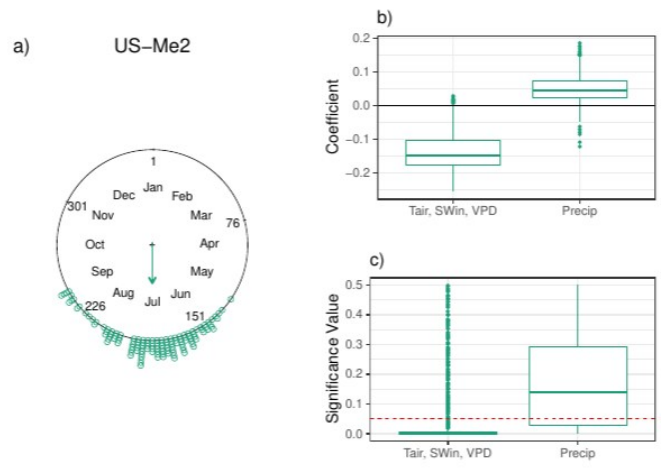


Figure S40. DOY_{GPPmax} sensitivity to different climate drivers for the Metolius mature ponderosa pine (US-Me2) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

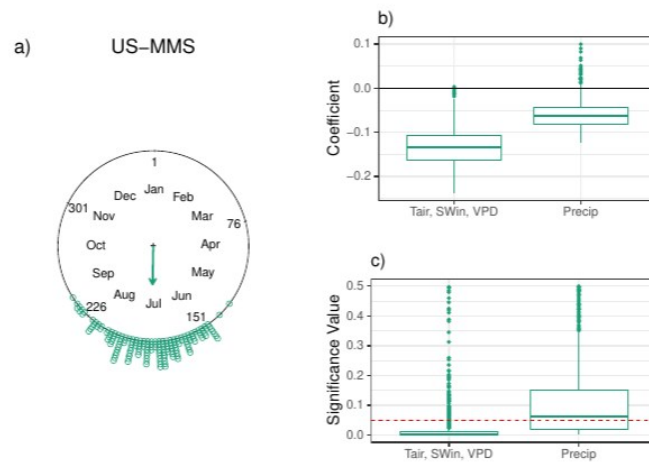


Figure S41. DOY_{GPPmax} sensitivity to different climate drivers for the Morgan Monroe State Forest (US-MMS) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

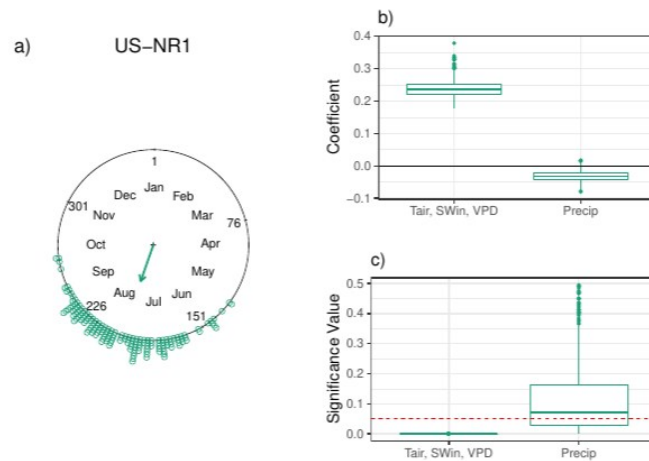


Figure S42. DOY_{GPPmax} sensitivity to different climate drivers for the Niwot Ridge Forest (LTER NWT1) (US-NR1) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

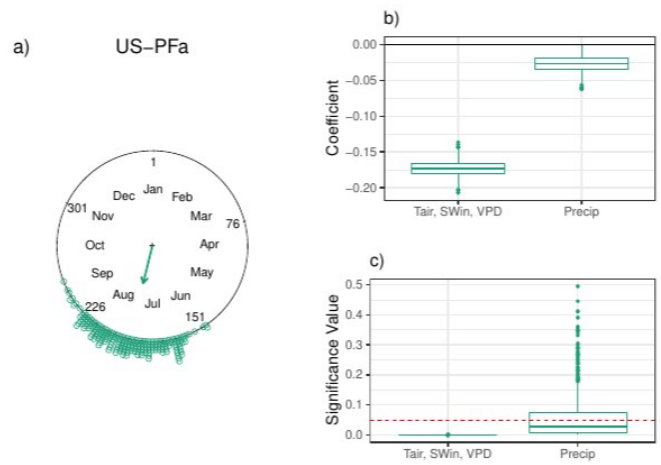


Figure S43. DOY_{GPPmax} sensitivity to different climate drivers for the Park Falls/WLEF (US-PFa) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

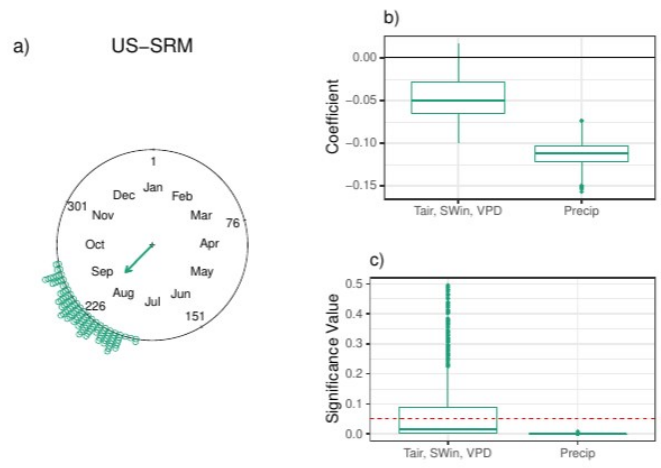


Figure S44. DOY_{GPPmax} sensitivity to different climate drivers for the Santa Rita Mesquite (US-SRM) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

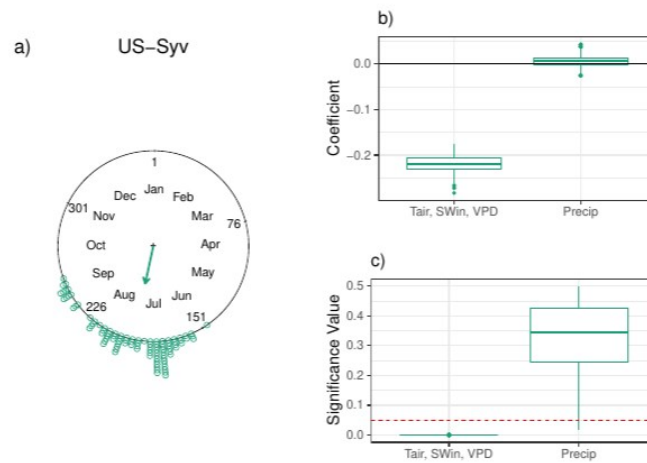


Figure S45. DOY_{GPPmax} sensitivity to different climate drivers for the Sylvania Wilderness Area (US-Syv) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

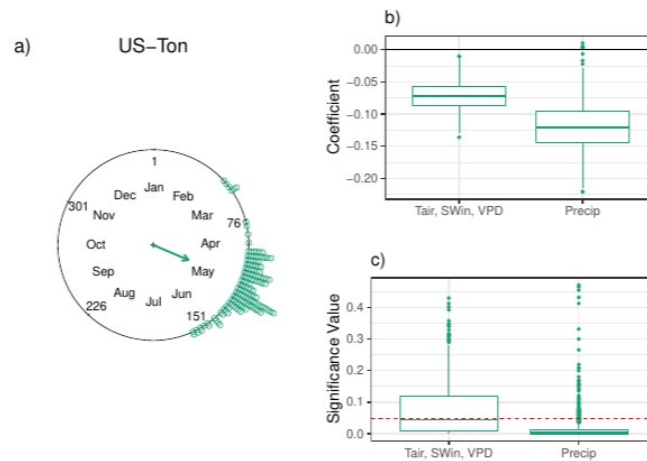


Figure S46. DOY_{GPPmax} sensitivity to different climate drivers for the Tonzi Ranch (US-Ton) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

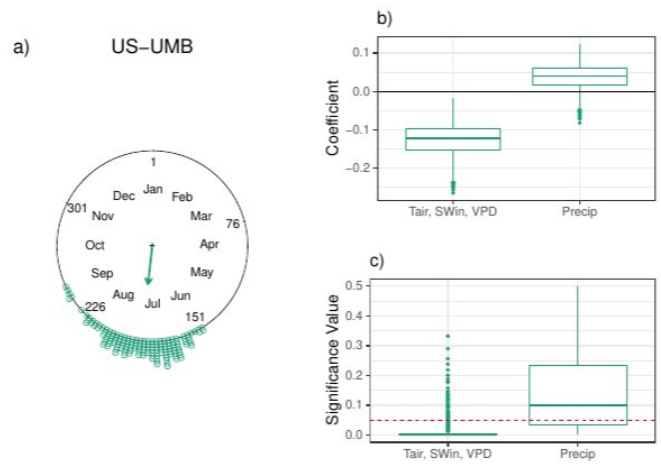


Figure S47. DOY_{GPPmax} sensitivity to different climate drivers for the Univ. of Mich. Biological Station (US-UMB) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

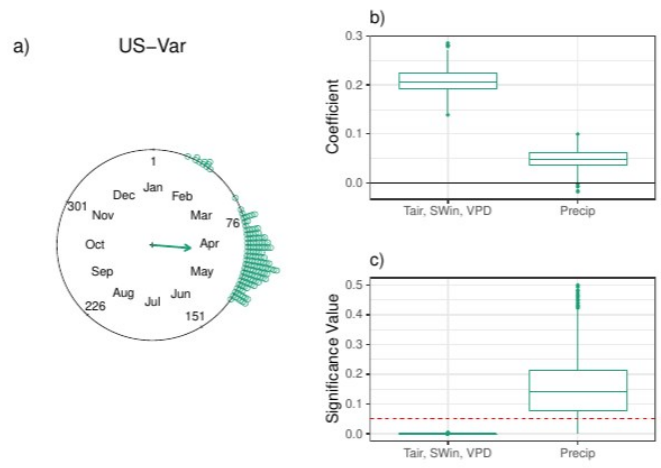


Figure S48. DOY_{GPPmax} sensitivity to different climate drivers for the Vaira Ranch- Ione (US-Var) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

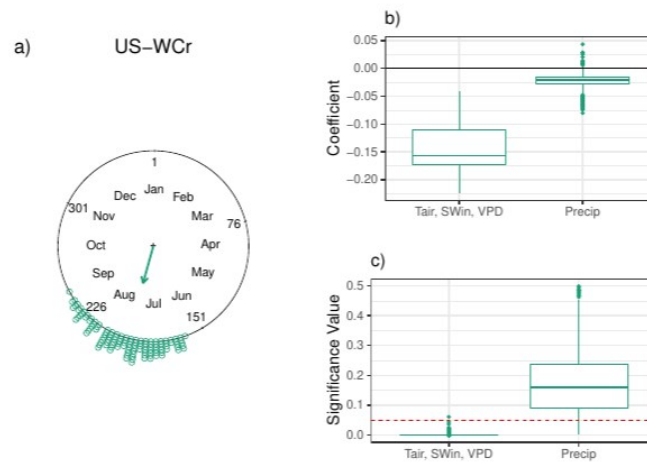


Figure S49. DOY_{GPPmax} sensitivity to different climate drivers for the Willow Creek (US-WCr) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

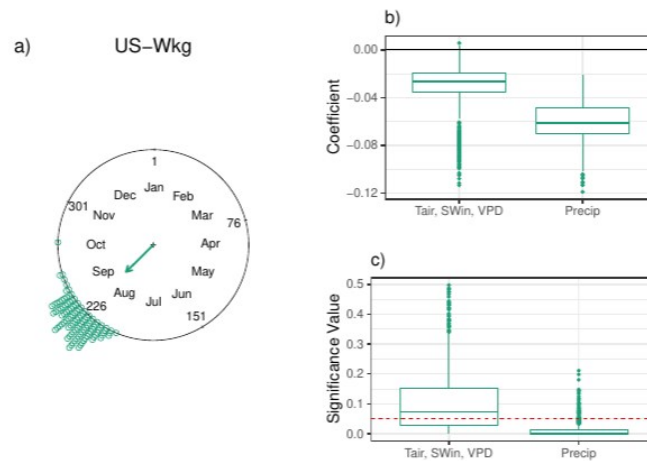


Figure S50. DOY_{GPPmax} sensitivity to different climate drivers for the Santa Rita Walnut Gulch Kendall Grasslands (US-Wkg) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

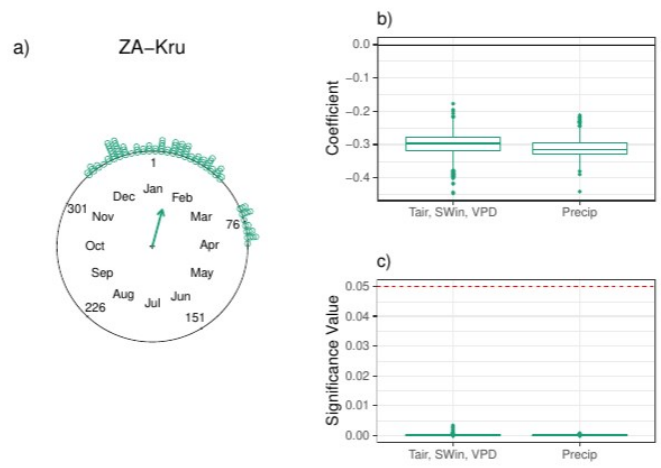


Figure S51. DOY_{GPPmax} sensitivity to different climate drivers for the Skukuza (ZA-Kru) FLUXNET site. a) DOY_{GPPmax} distribution across the year. The arrow indicates the mean angular direction of the growing season. b) Regression coefficients for each variable. c) Significance values for each variable. The red line represents a p-value of 0.05.

APPENDIX C

Supplementary information: On the Potential of Sentinel-2 for
Estimating Gross Primary Production

Description: The supplementary files are provided in seven parts. Supplementary material 1 provides the buffer areas for the eddy covariance sites. Supplementary material 2 provides outliers detection using the time series. Supplementary material 3 provides images filtered manually. Supplementary material 4 provides the prediction of GPP using linear regressions and balanced datasets. Supplementary material 5 provides GPP prediction using random forest and 10-folds spatio-temporal cross validation. Supplement material 6 provides an example of the upscaling of GPP for an entire Sentinel-2 tile. Supplementary material 7 includes raw results of the variable importance analysis for the imbalanced and balanced datasets.

Size: 97.4 MB

Packing List: tgrs-3152272-mm.zip

Player Information: any PDF reader; any .csv reader; any .png reader

Contact Information:

Mr. Daniel E. Pabon-Moreno

Max Planck Institute for Biogeochemistry, Biogeochemical Integration Department

Jena, Germany

Email: dpabon@bgc-jena.mpg.de

Supplement Material 2

Outlier detection using time series

1 Methods

The following is a description of the algorithm implemented in by the “anomalize” R package to detect outliers in irregular time series.

1.1 Time series decomposition by LOESS

1. Given the irregular time series, a new time series object is generated per site where the frequency of the new time series is estimated as follows:
 - (a) The median time difference between the observations is estimated.
 - (b) The scale of median time difference is estimated based on magnitude of the value following the first column of the table 1 (e.g. median time difference of seven days is considered “week”, a median time difference of 23 days is considered “month”)
 - (c) The new periodicity is estimate as the corresponding category in the frequency column in the table 1, for the scale of the median time difference estimated in the step before (e.g. if the time scale is month, the frequency selected is 1 year).
 - (d) The dataset is aggregate again base on the period selected. And the frequency is estimated as the median frequency of the observations per group.
 - (e) If the total number of observations is less than 3 times the frequency, a new aggregation period is selected as the one before to the correspondent frequency column in table 1. (e.g. instance of 1 year, 1 quarter is selected as aggregation metric).
 - (f) The aggregation is performed again with the new period and again the frequency is estimated as the median frequency of the observations per group.
 - (g) If still the number of observations is less than 3 times the frequency estimated the frequency is defined as 1.
2. The span of the LOESS window for trend extraction (trend) is estimated as follows:

- (a) The median time difference between the observations is estimated.
 - (b) The scale of median time difference is estimated based on magnitude of the value following the first column of the table 1 (e.g. median time difference of seven days is considered “week”, a median time difference of 23 days is considered “month”)
 - (c) The new periodicity is estimate as the corresponding category in the trend column in the table 1, for the scale of the median time difference estimated in the step before (e.g. if the timescale is month, the trend selected is 1 year).
 - (d) The dataset is aggregate again base on the period selected. And the trend is estimated as the median frequency of the observations per group.
 - (e) If the total number of observations divided by the trend is less than 2, a new aggregation period is selected as the one before to the correspondent trend column in table 1. (e.g. instance of 5 years, 1 year is selected as aggregation metric).
 - (f) The aggregation is performed again with the new period and again the trend is estimated as the ceiling median frequency of the observations per group.
 - (g) If still the number of observations divided by the trend is less than 2 the trend is defined as the number of observations.
3. A *ts* object is created where the frequency of the object is equal to the one defined in the step one.
 4. The time series decomposition by LOESS is applied to the new dataset where the seasonal adjustment is the estimated frequency in the step 1, and the trend window is estimated in step 2. The rest of parameters is the default parameters for the *stl* function.

1.2 Innerquartile range for outlier detection

1. The residuals of the time series decomposition estimated in the step 4, are used as input.
2. The threshold is defined as 3 times the quantile 0.25 and 0.75 of the distribution of the residuals.
3. The observations with a lower and higher value than 3 times are classified as outliers, where not more than 20% of the observations can be classified as outliers.

Table 1: Equivalences to define the frequency and the trend of a time series given the median time difference of the observations.

timescale	frequency	trend
second	1 hour	12 hours
minute	1 day	14 days
hour	1 day	1 month
day	1 week	3 months
week	1 quarter	1 year
month	1 year	5 years
quarter	1 year	10 years
year	5 years	30 years

2 Results

2.1 Percent of images filtered out per site

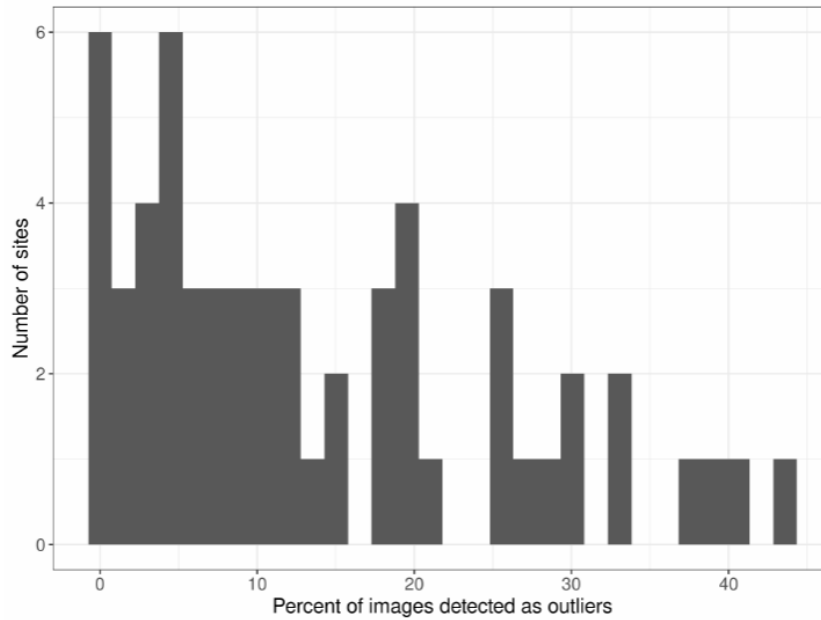


Figure 1: Percent of images detected as outliers per site

2.2 Outlier detection per site using time series

For three sites the outlier detection using time series was not possible given the low number of observations: US-Bar, US-MMS, US-Vcm. For the rest of the sites the plots are presented below.

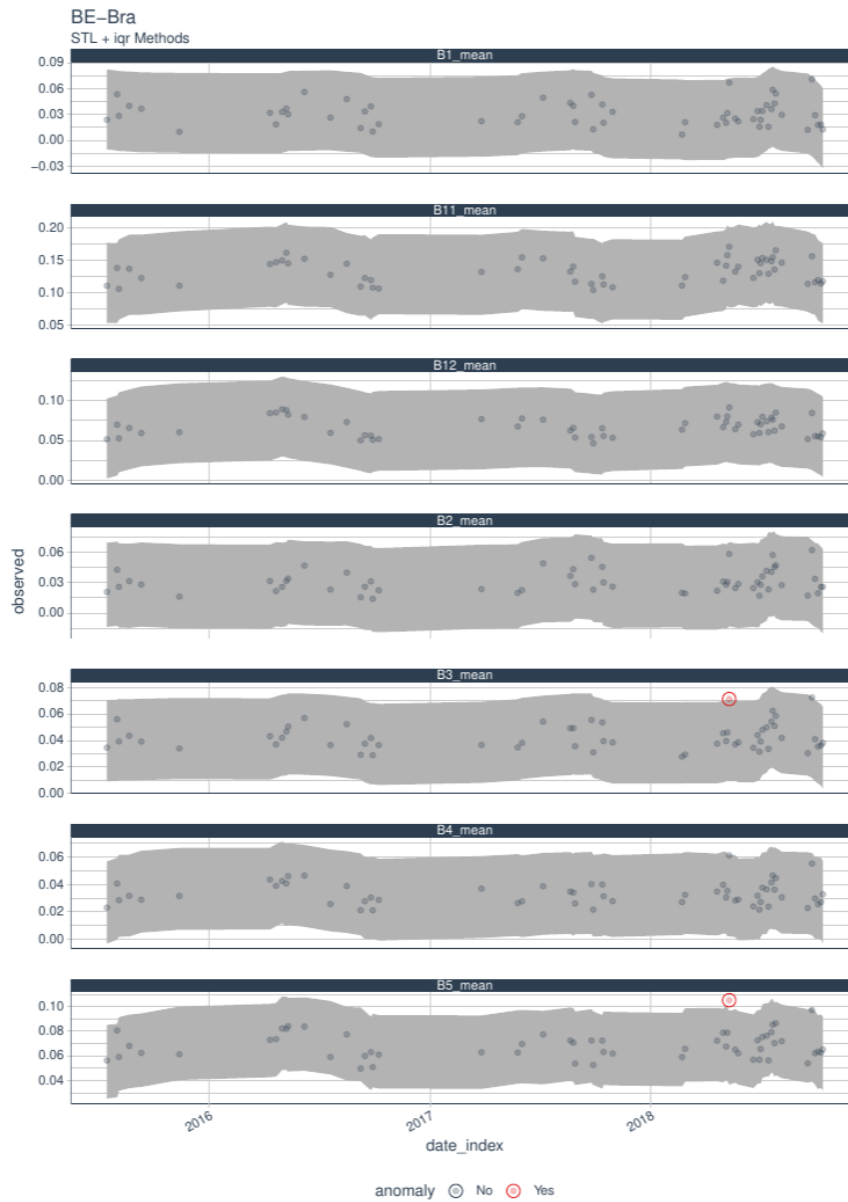


Figure 2: Outliers detection in Sentinel-2 images for the Brasschaat (BE-Bra) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

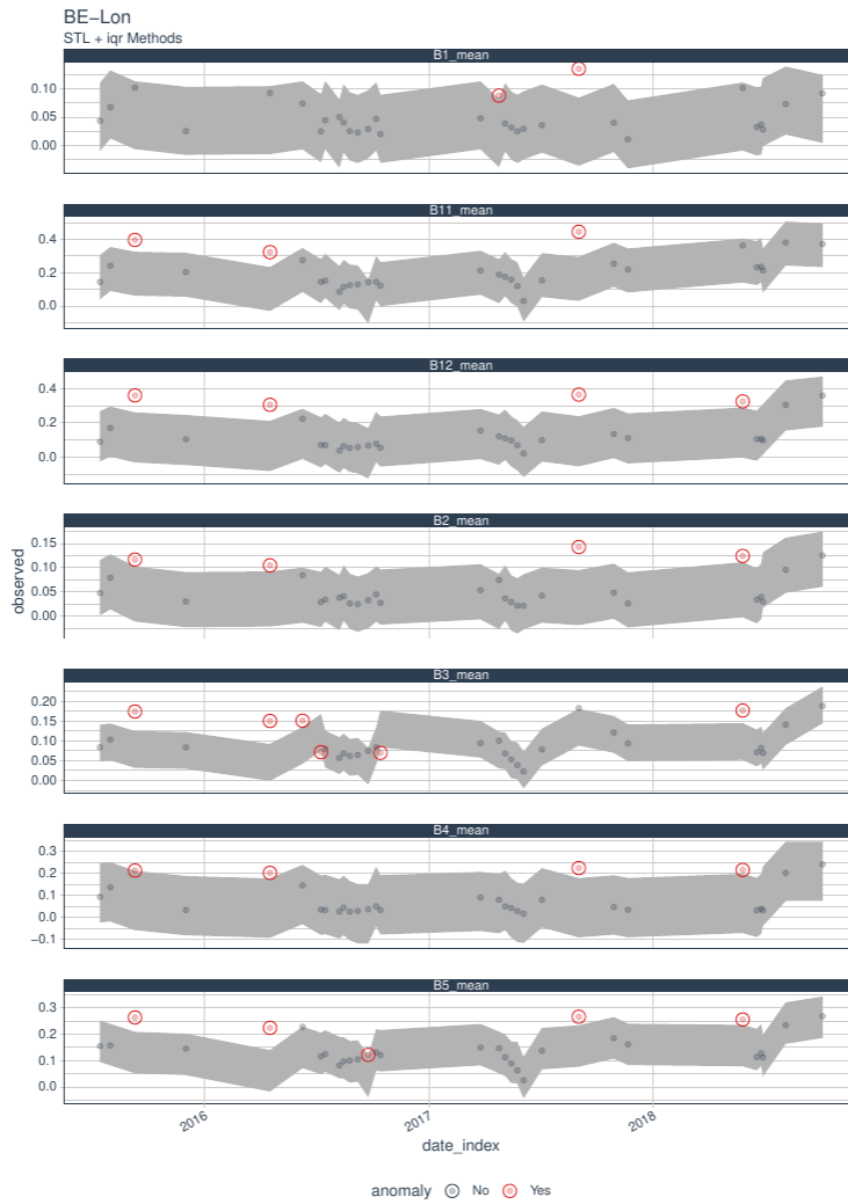


Figure 3: Outliers detection in Sentinel-2 images for the Lonzee (BE-Lon) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

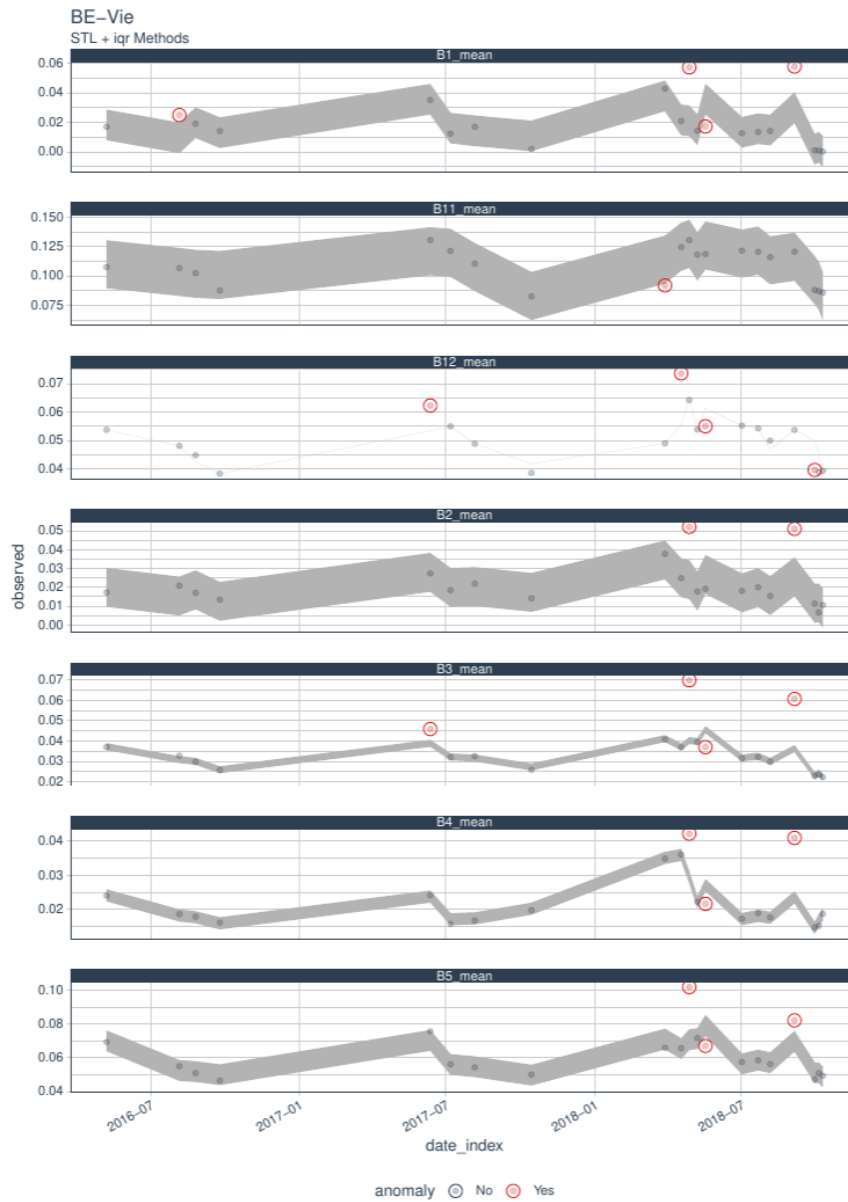


Figure 4: Outliers detection in Sentinel-2 images for the Vielsalm (BE-Vie) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

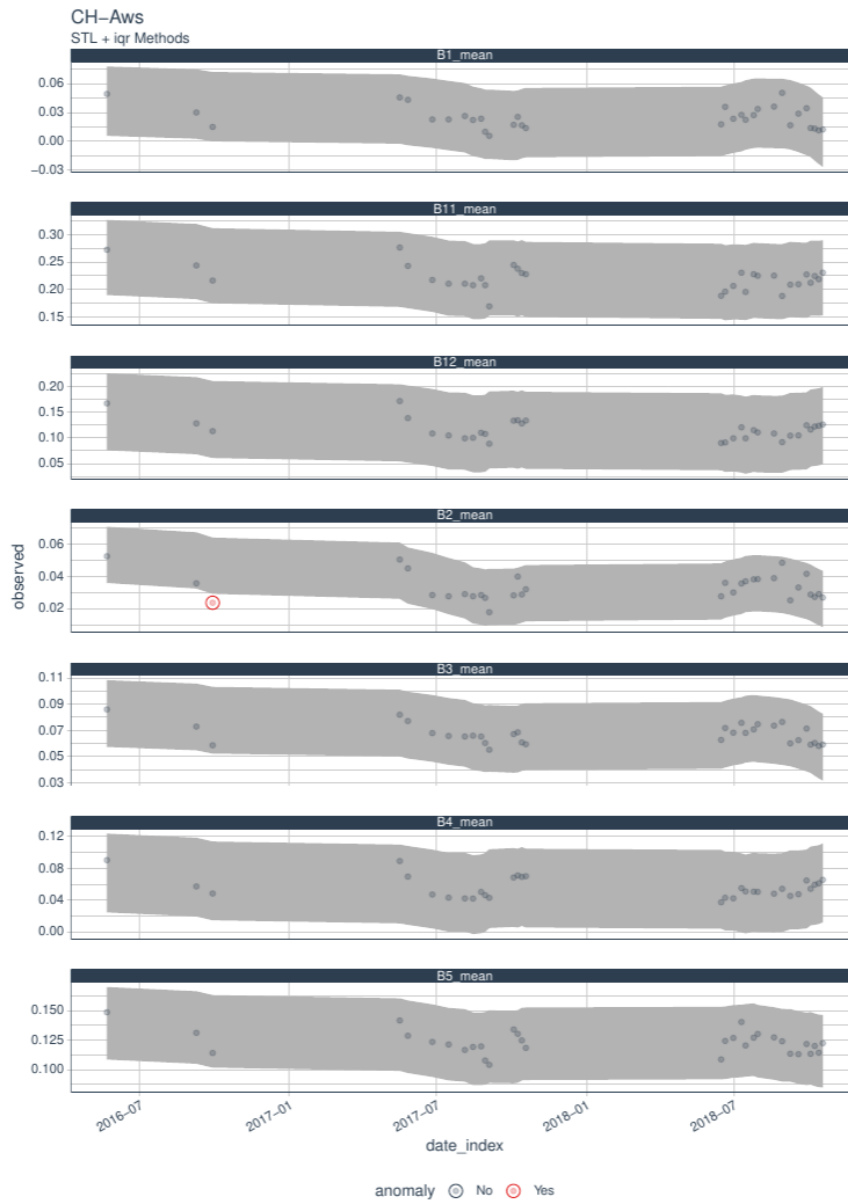


Figure 5: Outliers detection in Sentinel-2 images for the Alp Weissenstein (CH-Aws) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

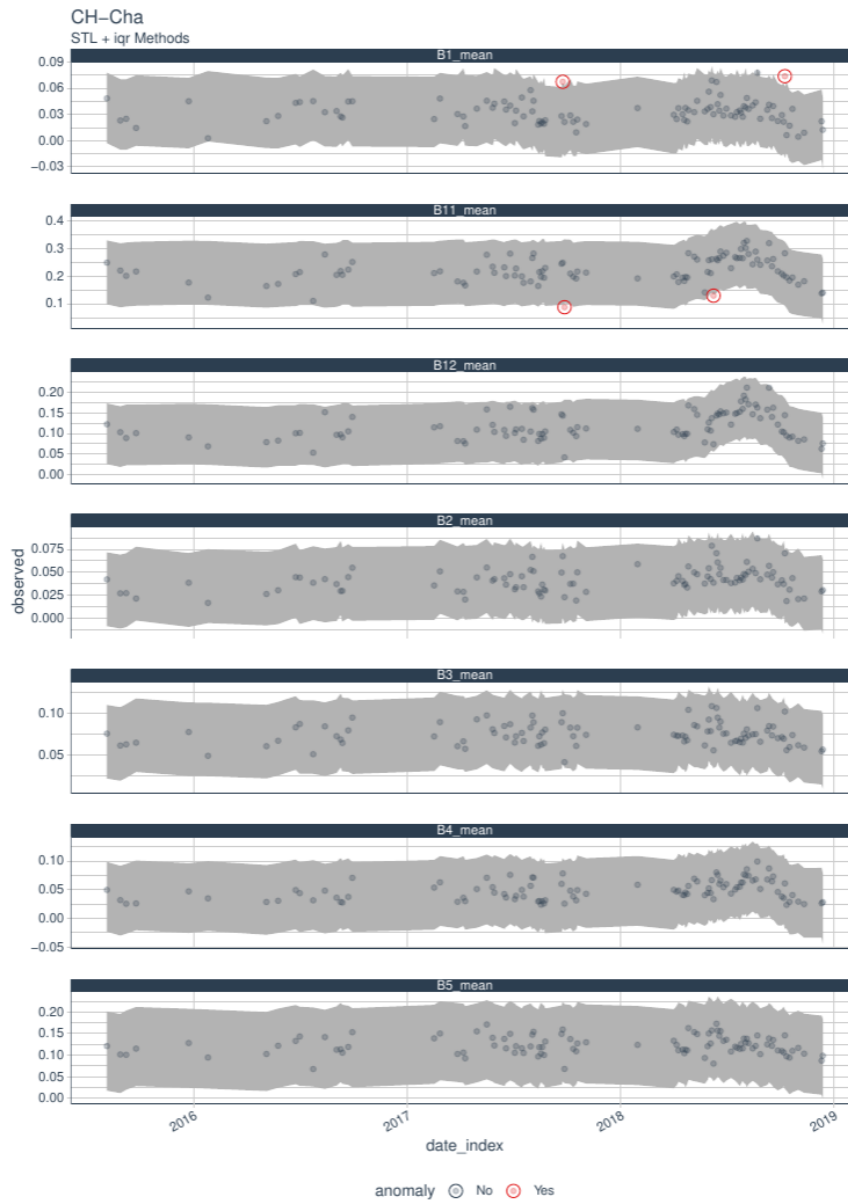


Figure 6: Outliers detection in Sentinel-2 images for the Chamau (CH-Cha) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

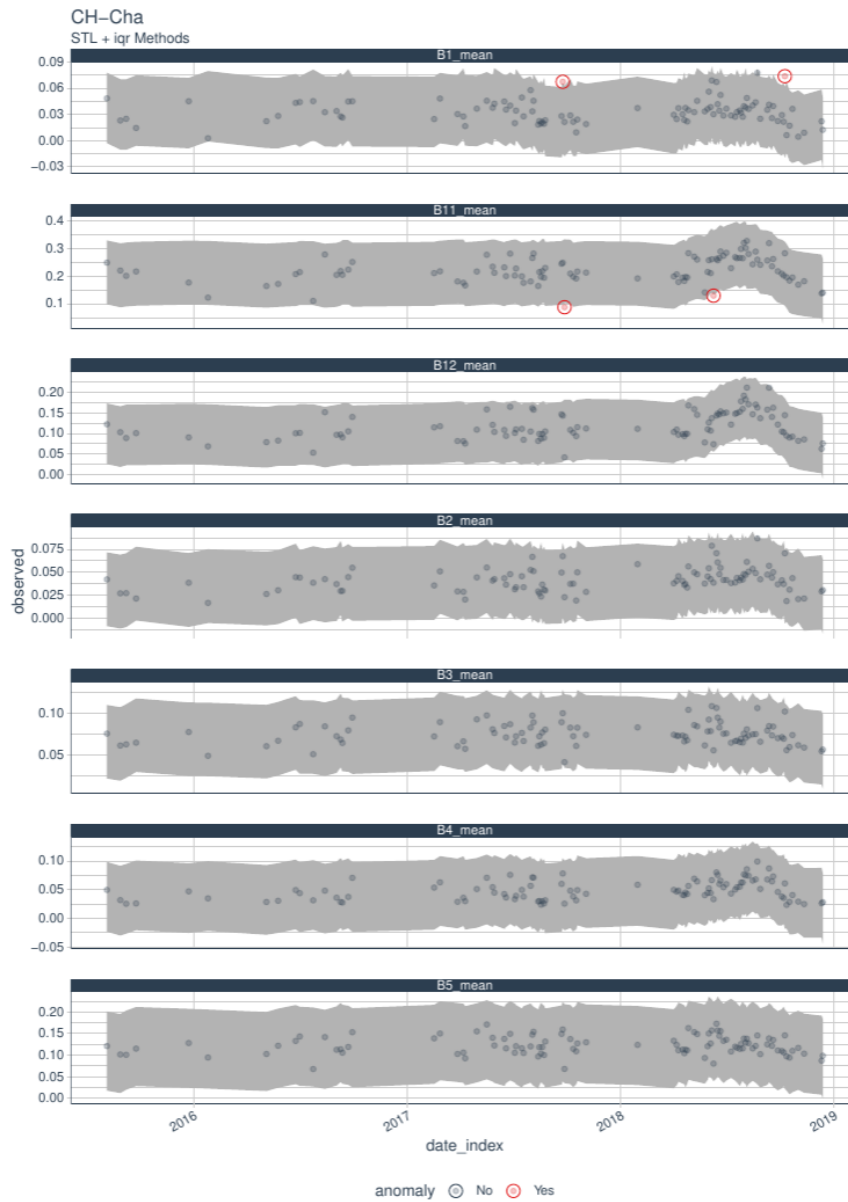


Figure 7: Outliers detection in Sentinel-2 images for the Davos (CH-Dav) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

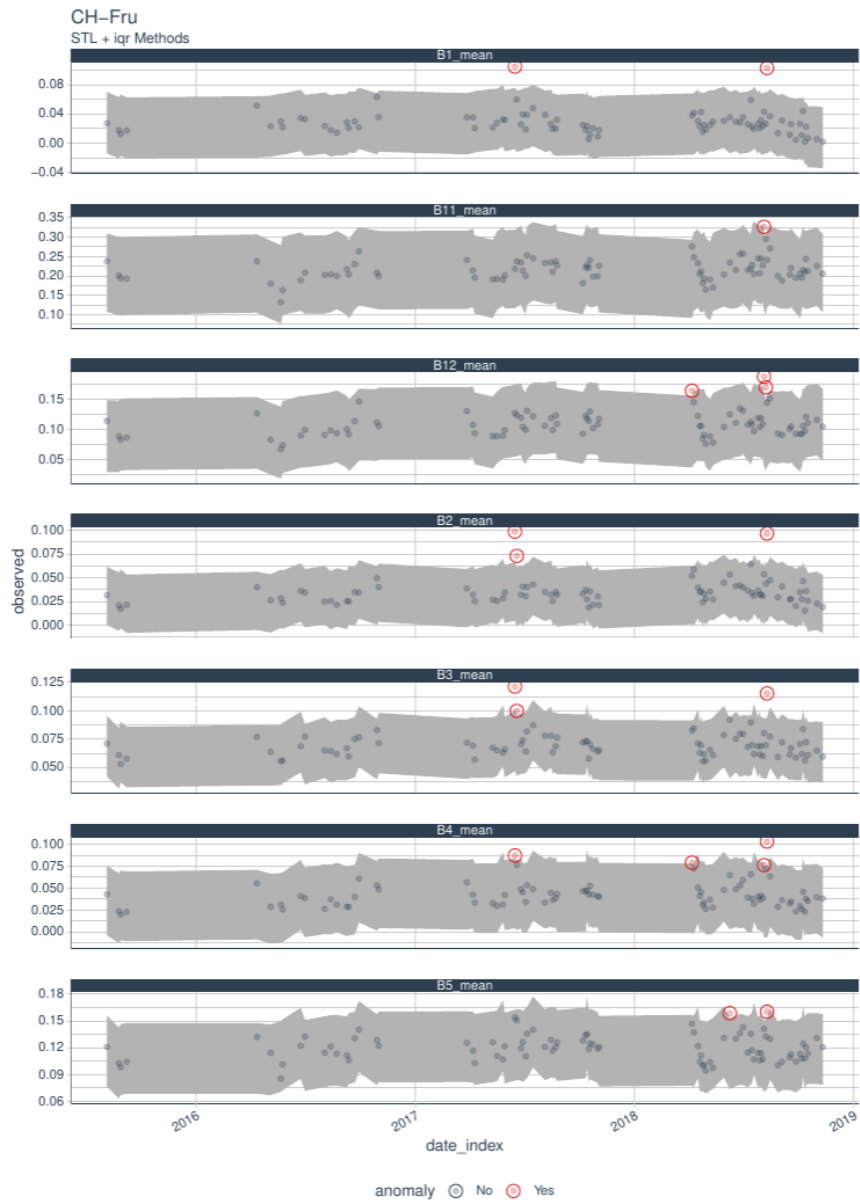


Figure 8: Outliers detection in Sentinel-2 images for the Fruebüel (CH-Fru) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

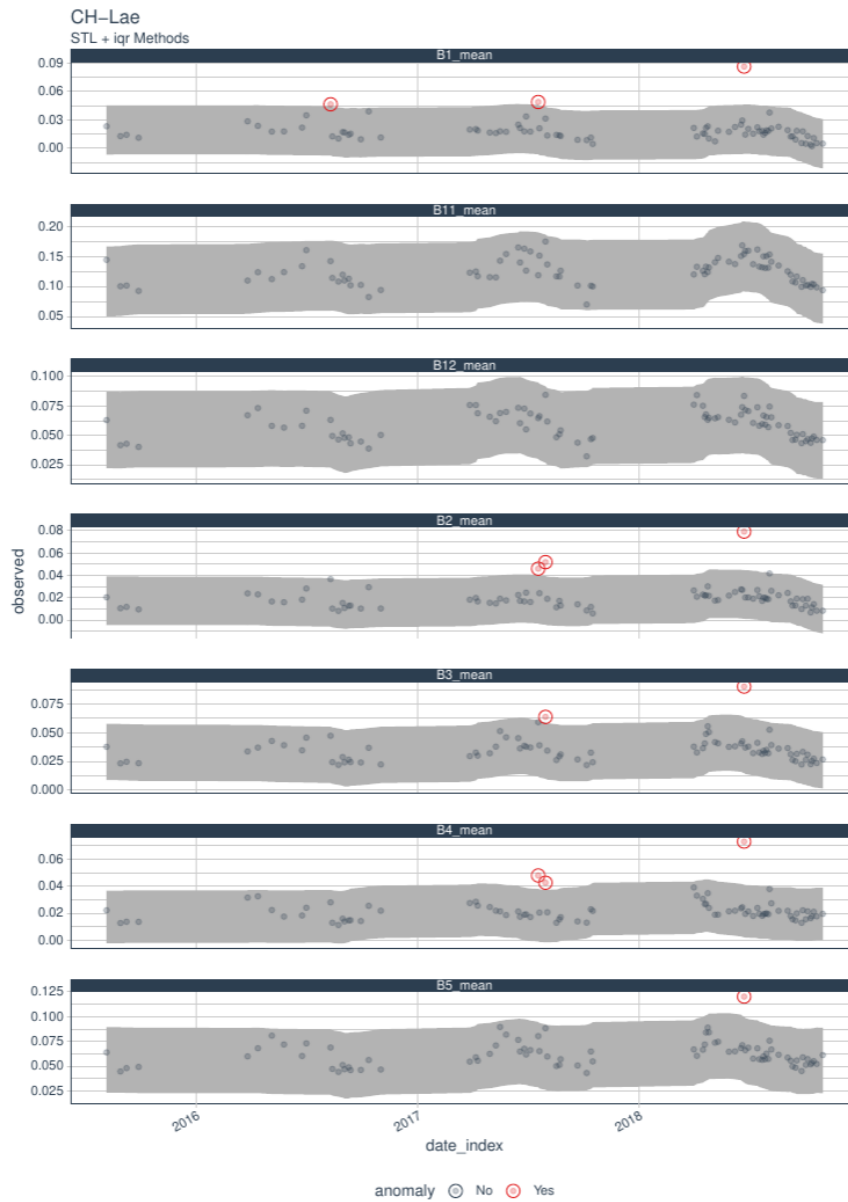


Figure 9: Outliers detection in Sentinel-2 images for the Laegern (CH-Lae) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

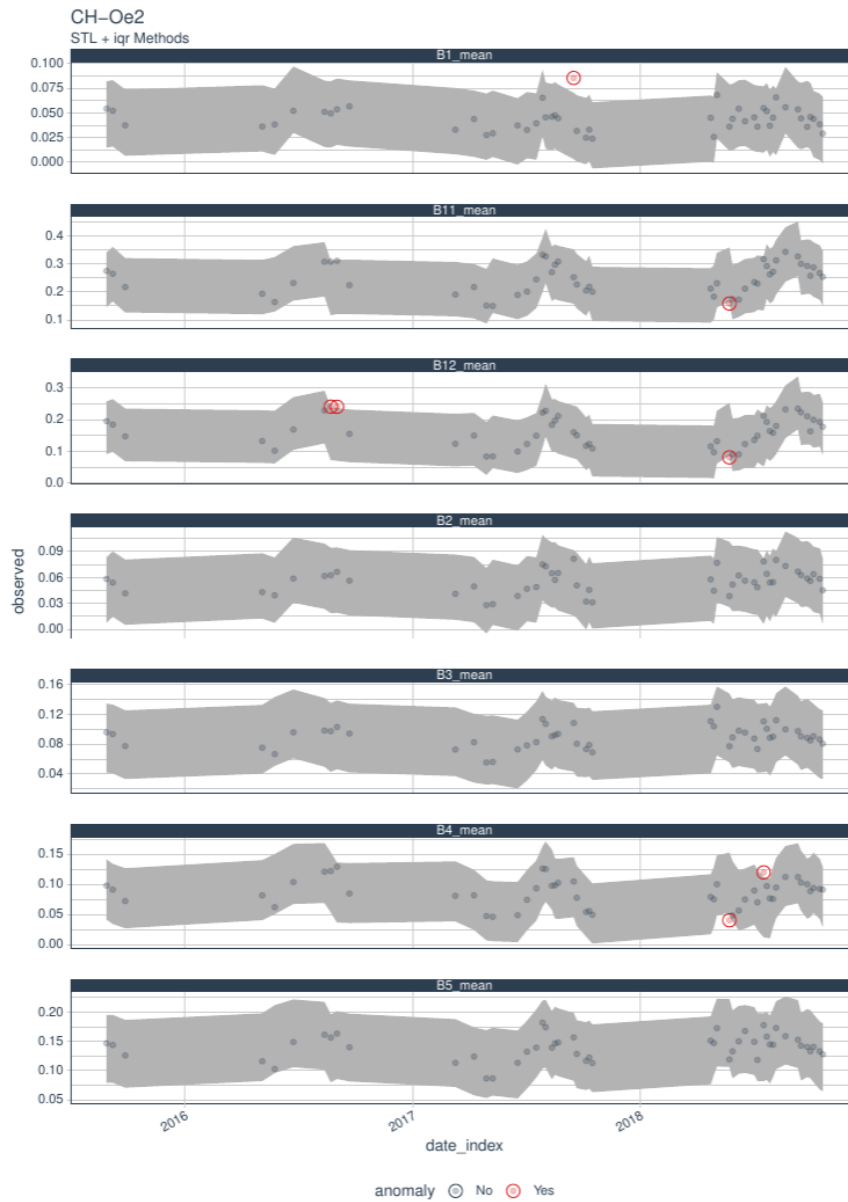


Figure 10: Outliers detection in Sentinel-2 images for the Oensingen crop (CH-Oe2) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

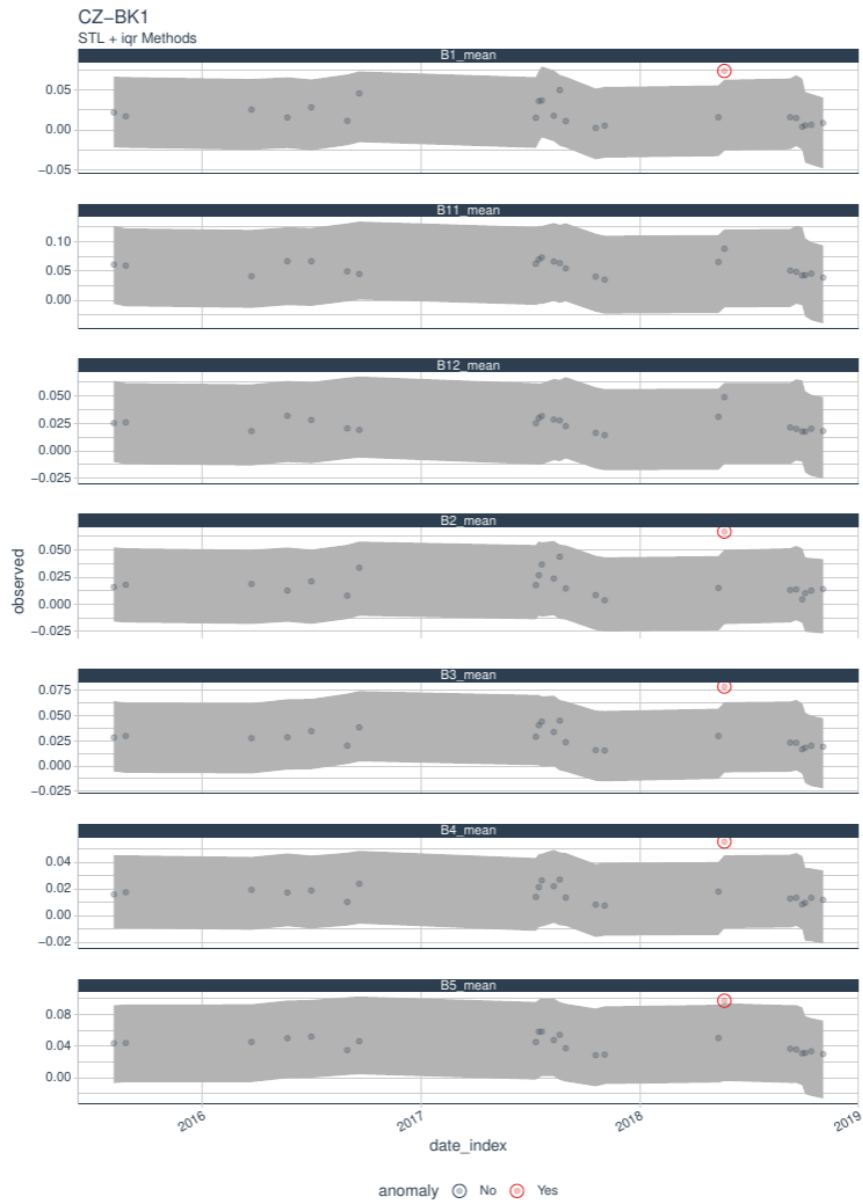


Figure 11: Outliers detection in Sentinel-2 images for the Bily Kriz forest (CZ-BK1) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

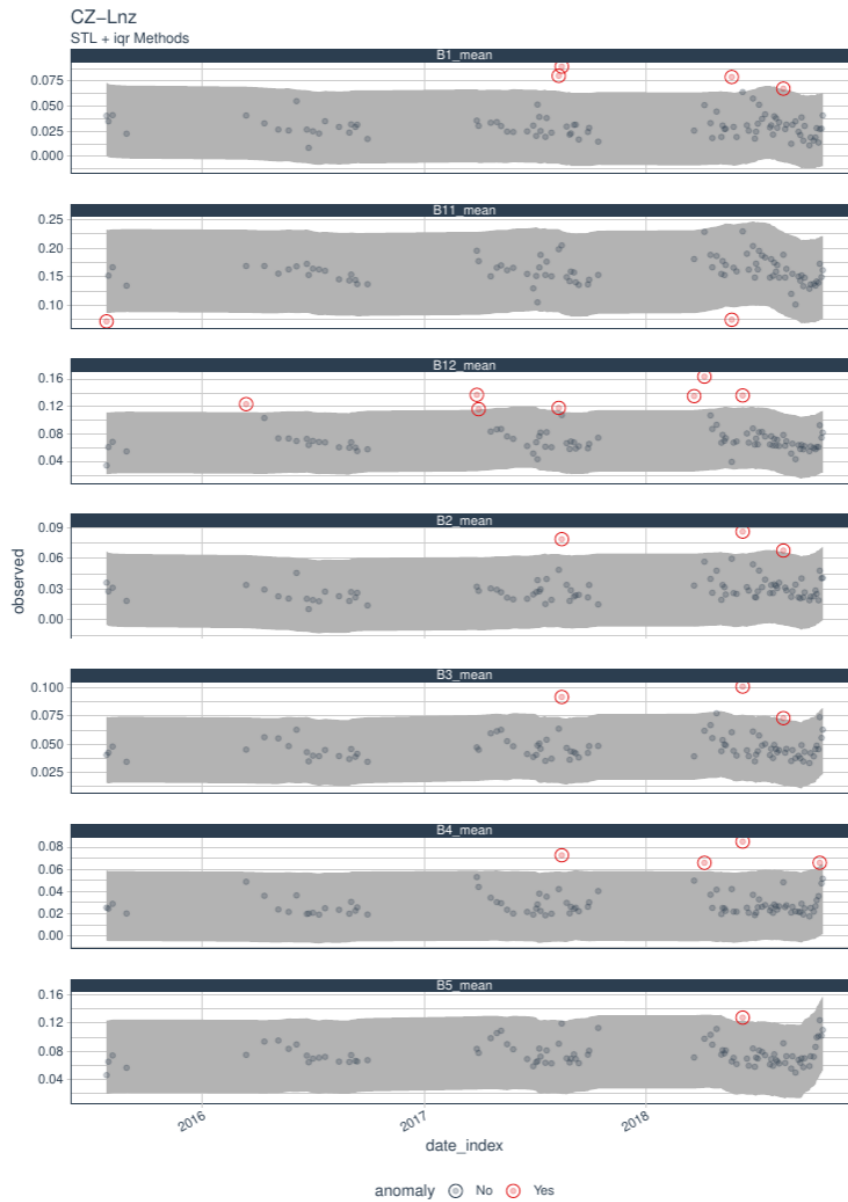


Figure 12: Outliers detection in Sentinel-2 images for the Lanzhot (CZ-Lnz) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

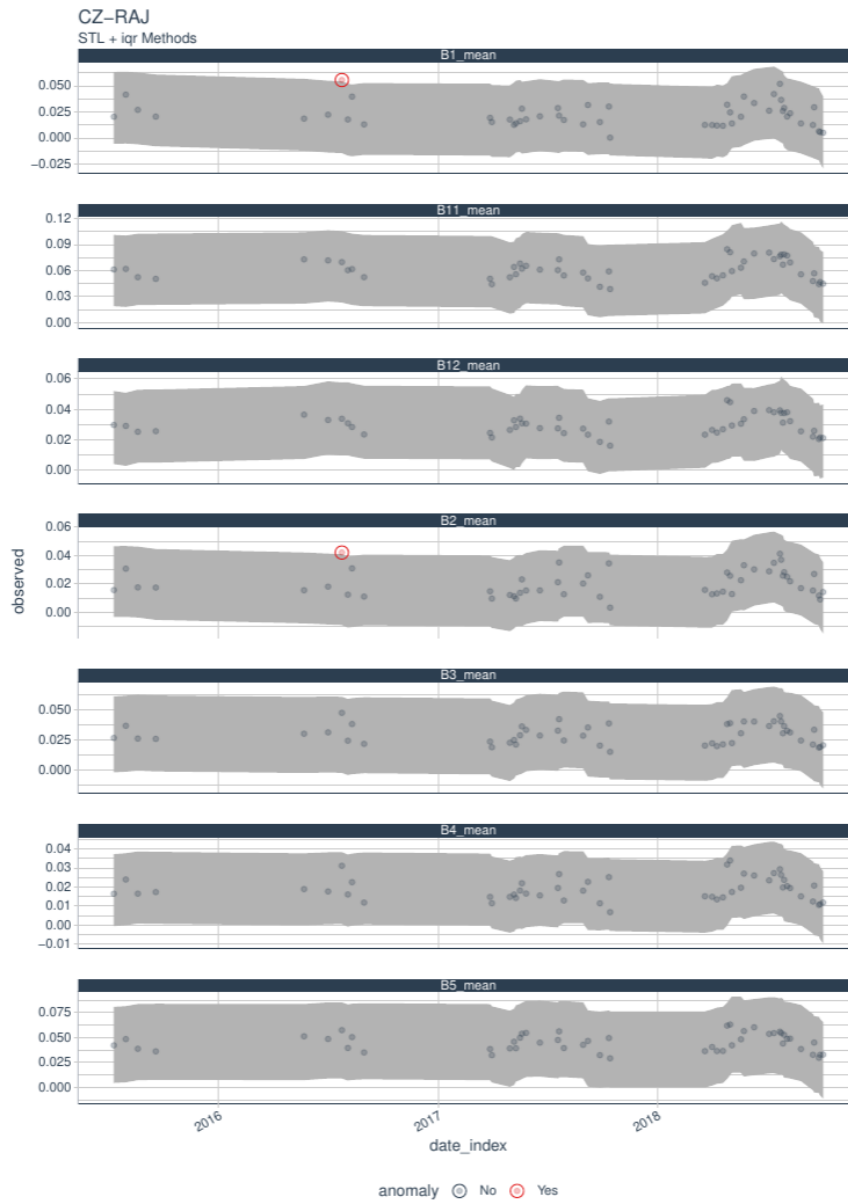


Figure 13: Outliers detection in Sentinel-2 images for the Rajec (CZ-RAJ) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

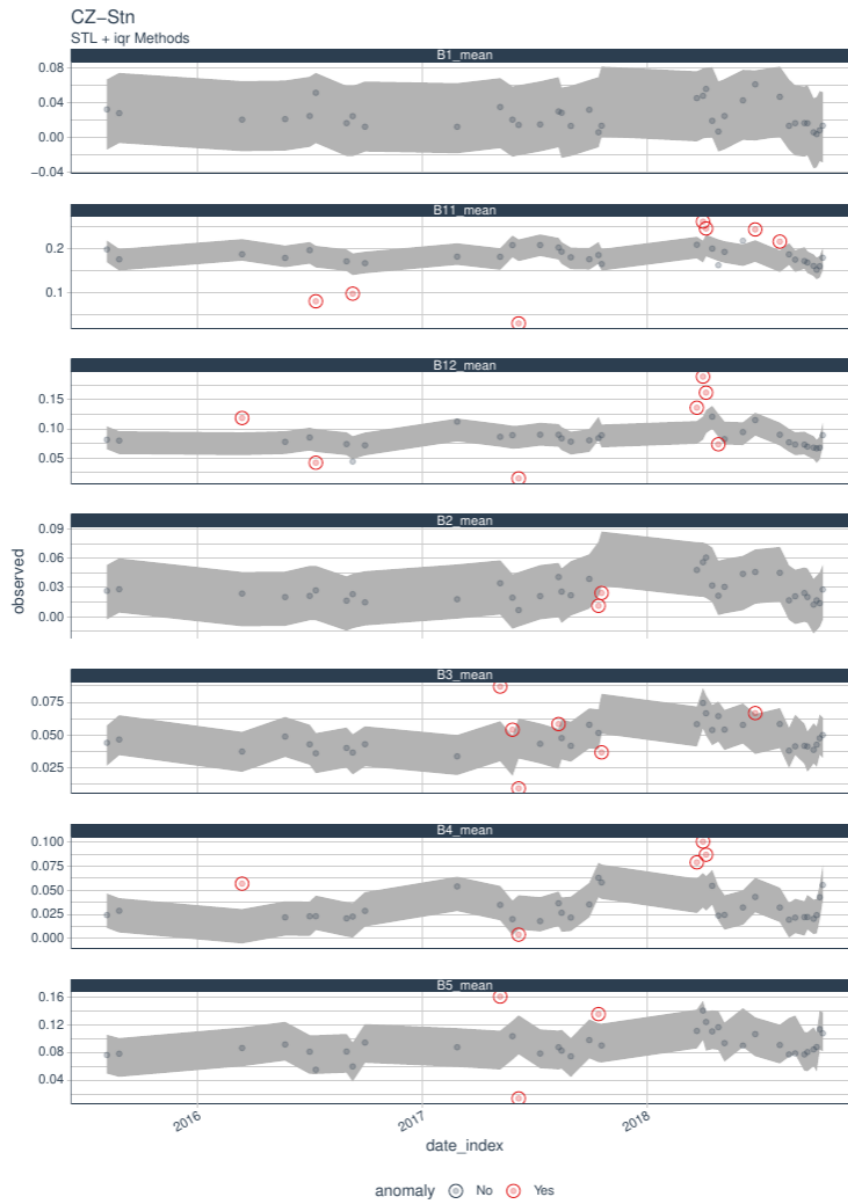


Figure 14: Outliers detection in Sentinel-2 images for the Stitna (CZ-Stn) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

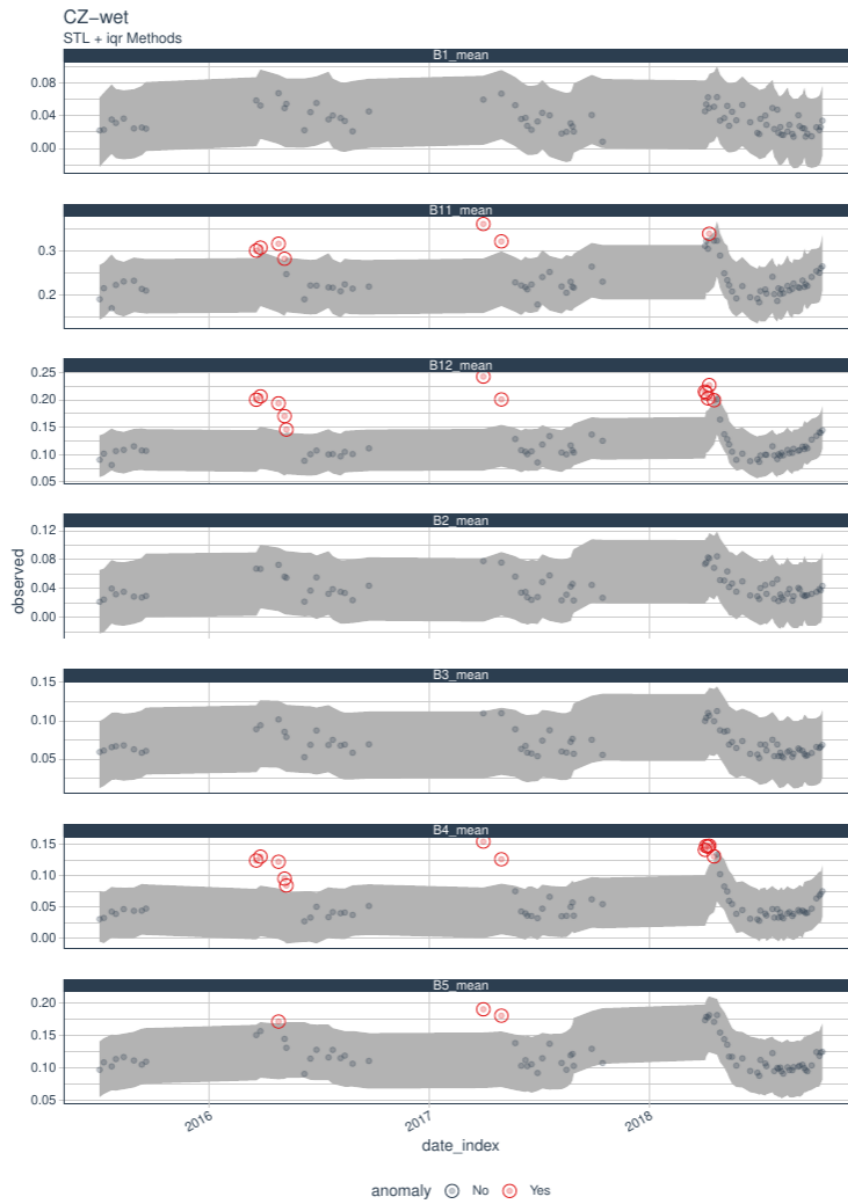


Figure 15: Outliers detection in Sentinel-2 images for the Trebon (CZ-wet) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

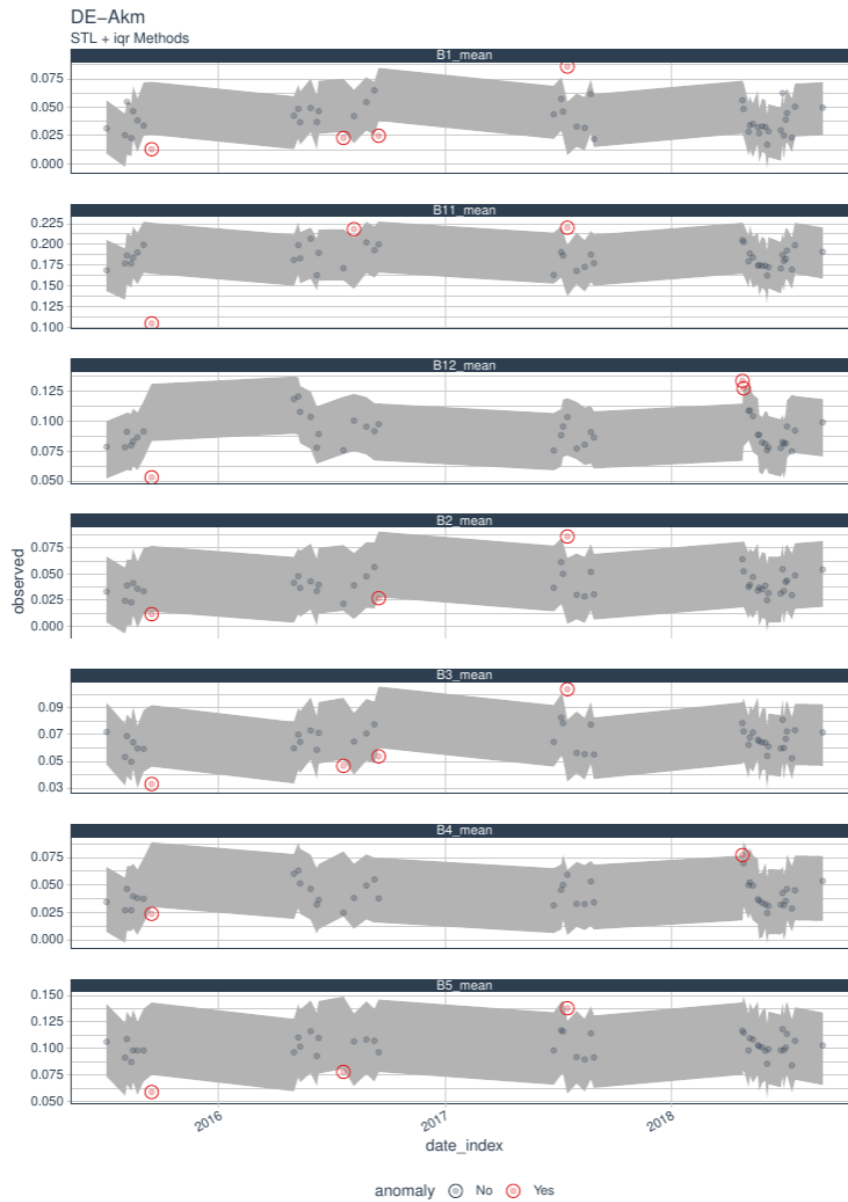


Figure 16: Outliers detection in Sentinel-2 images for the Anklam (DE-Akm) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

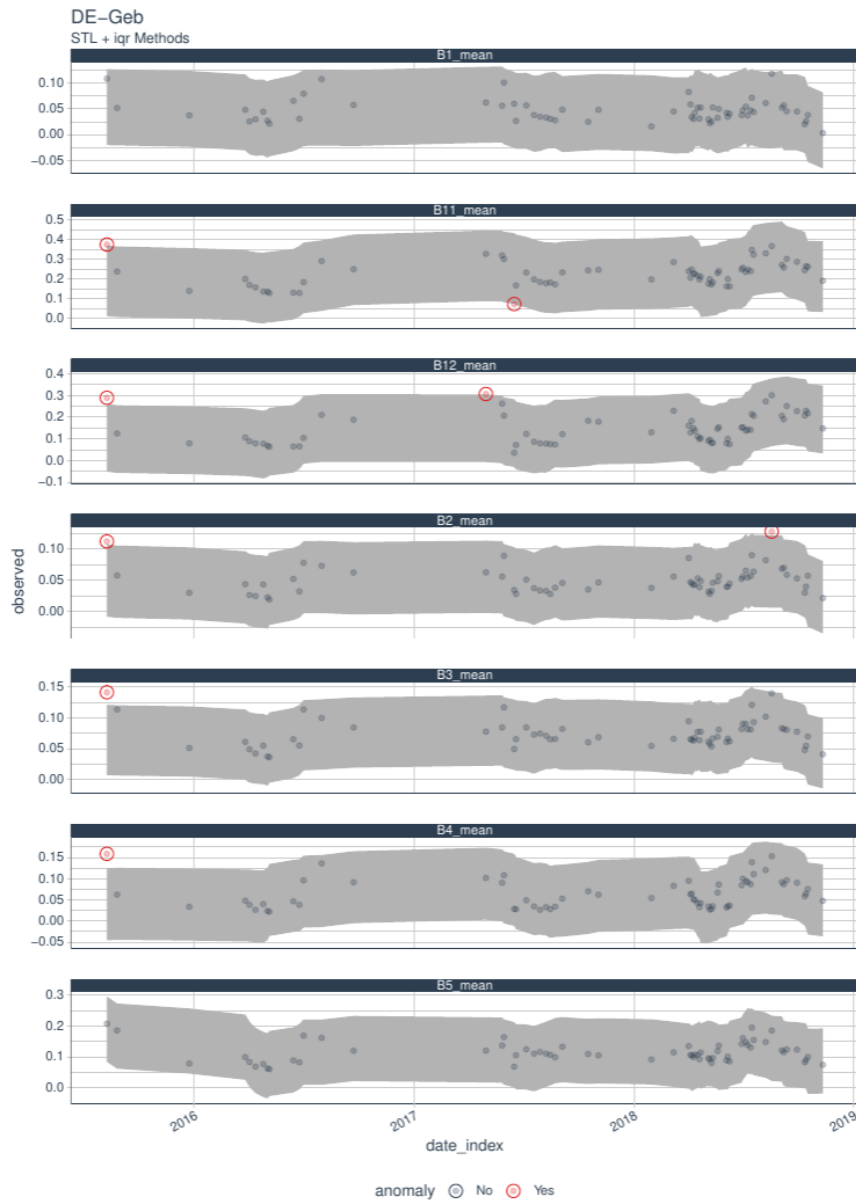


Figure 17: Outliers detection in Sentinel-2 images for the Gebesee (DE-Geb) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

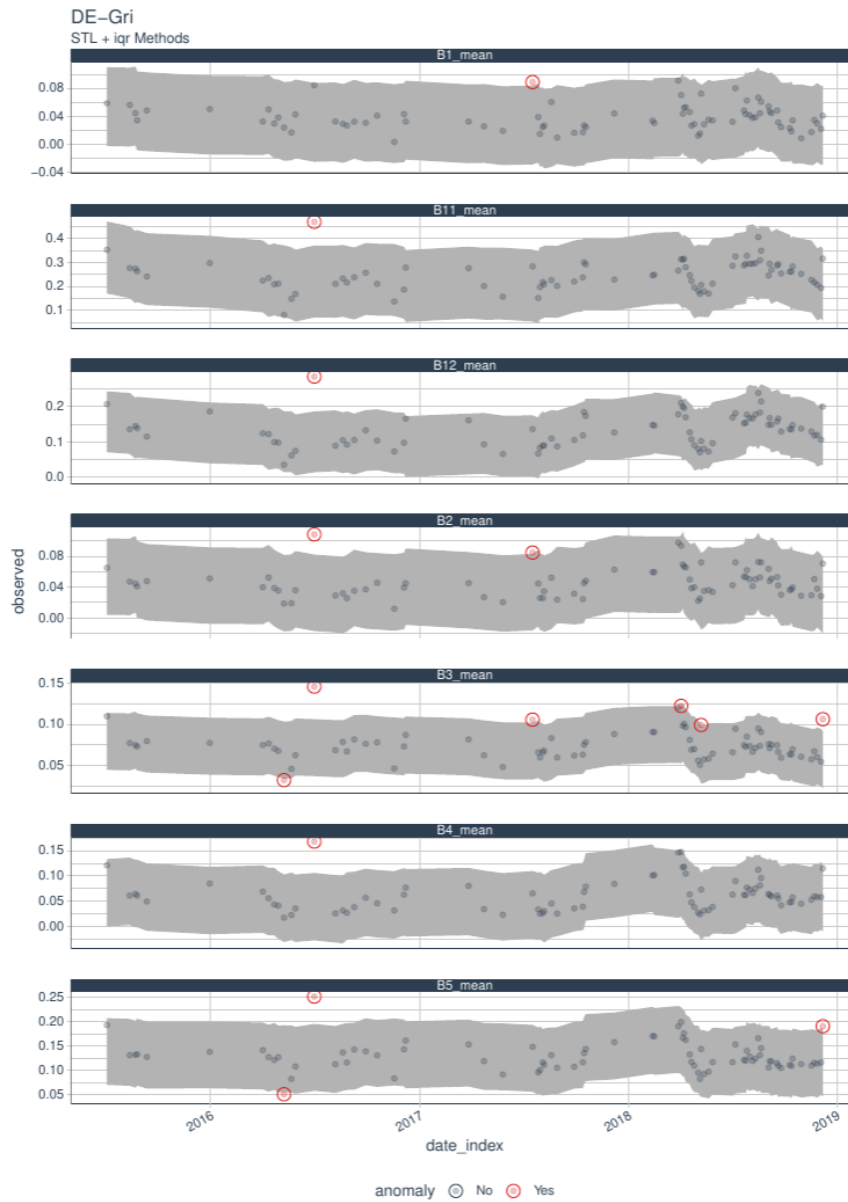


Figure 18: Outliers detection in Sentinel-2 images for the Grillenburg (DE-Gri) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

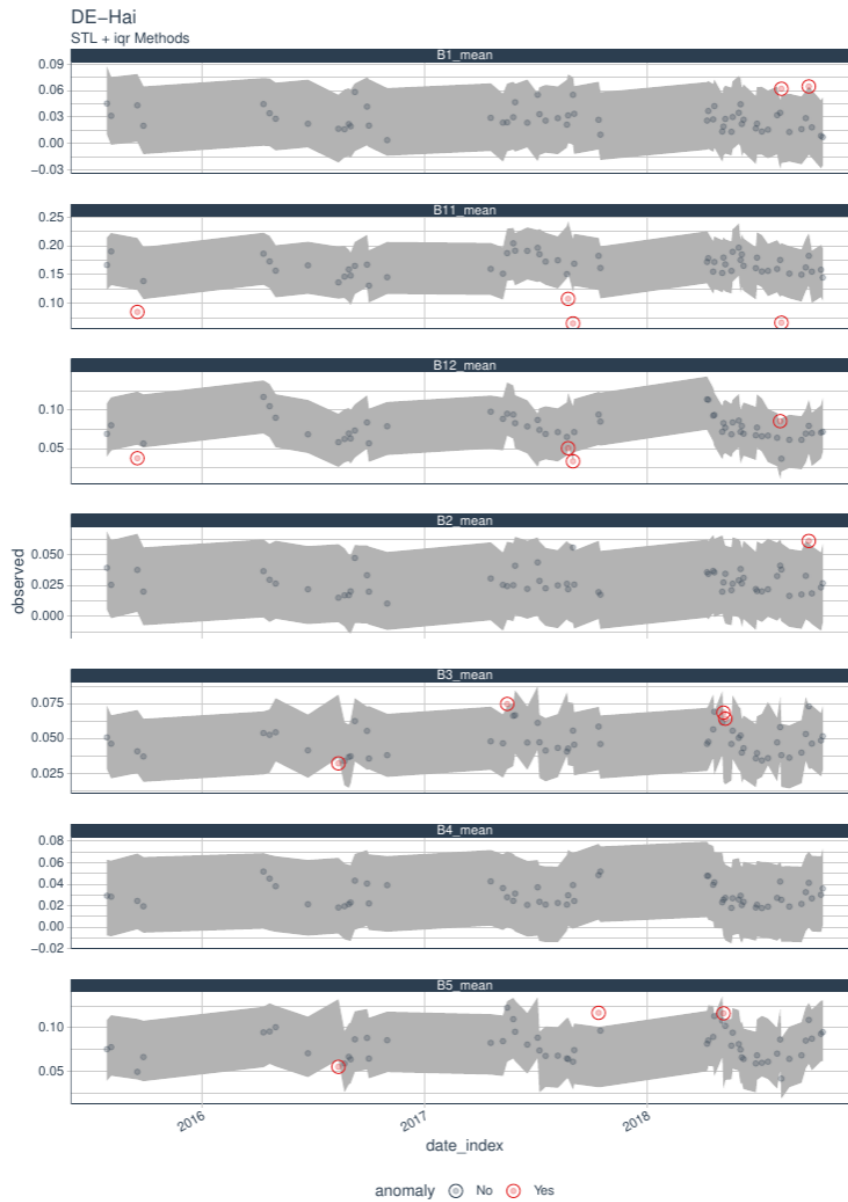


Figure 19: Outliers detection in Sentinel-2 images for the Hainich (DE-Hai) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

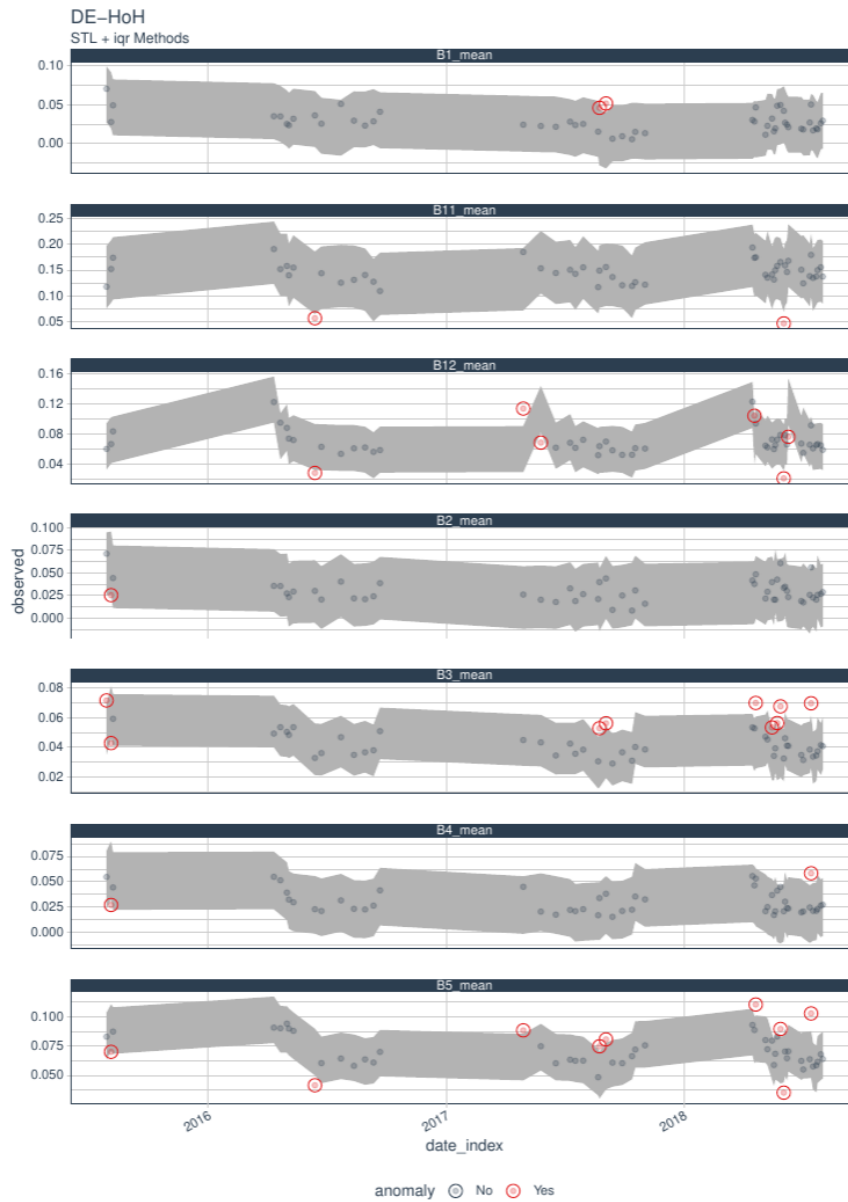


Figure 20: Outliers detection in Sentinel-2 images for the Hohes Holz (DE-HoH) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

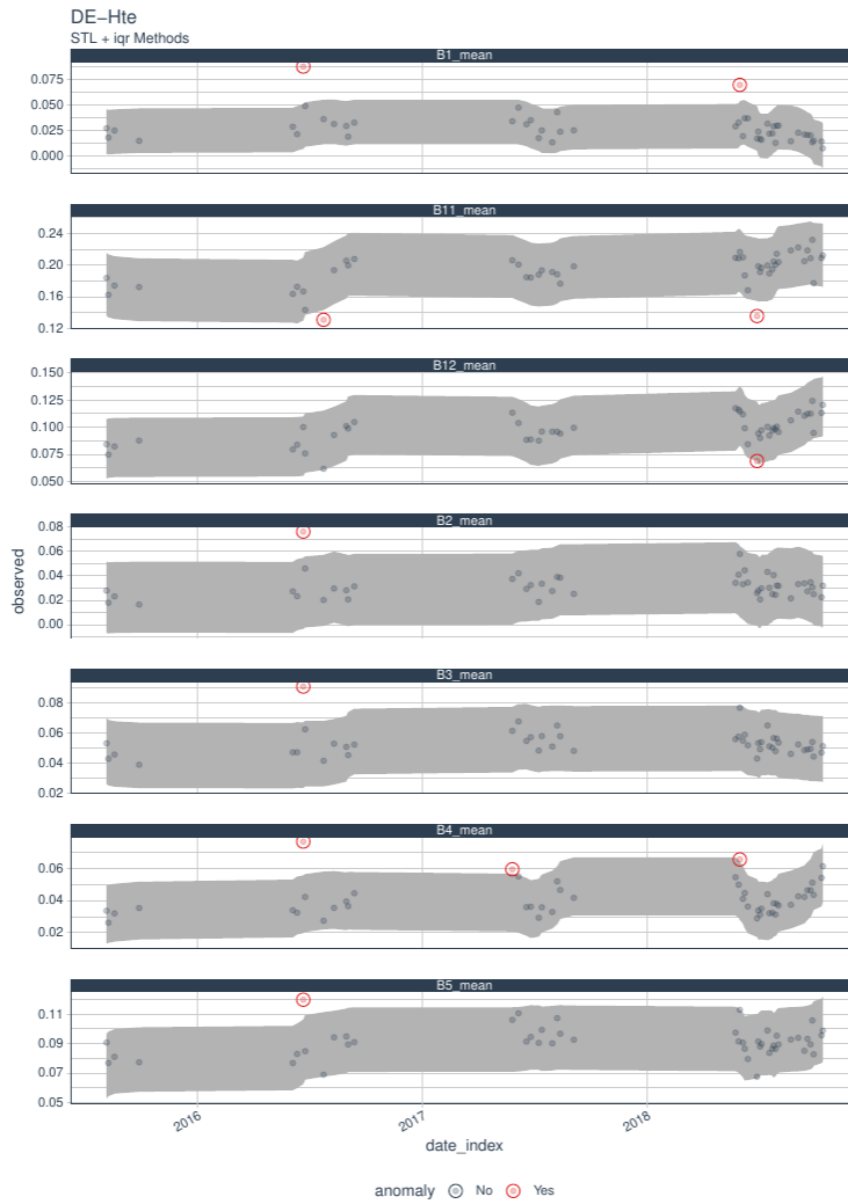


Figure 21: Outliers detection in Sentinel-2 images for the Huetelmoor (DE-Hte) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

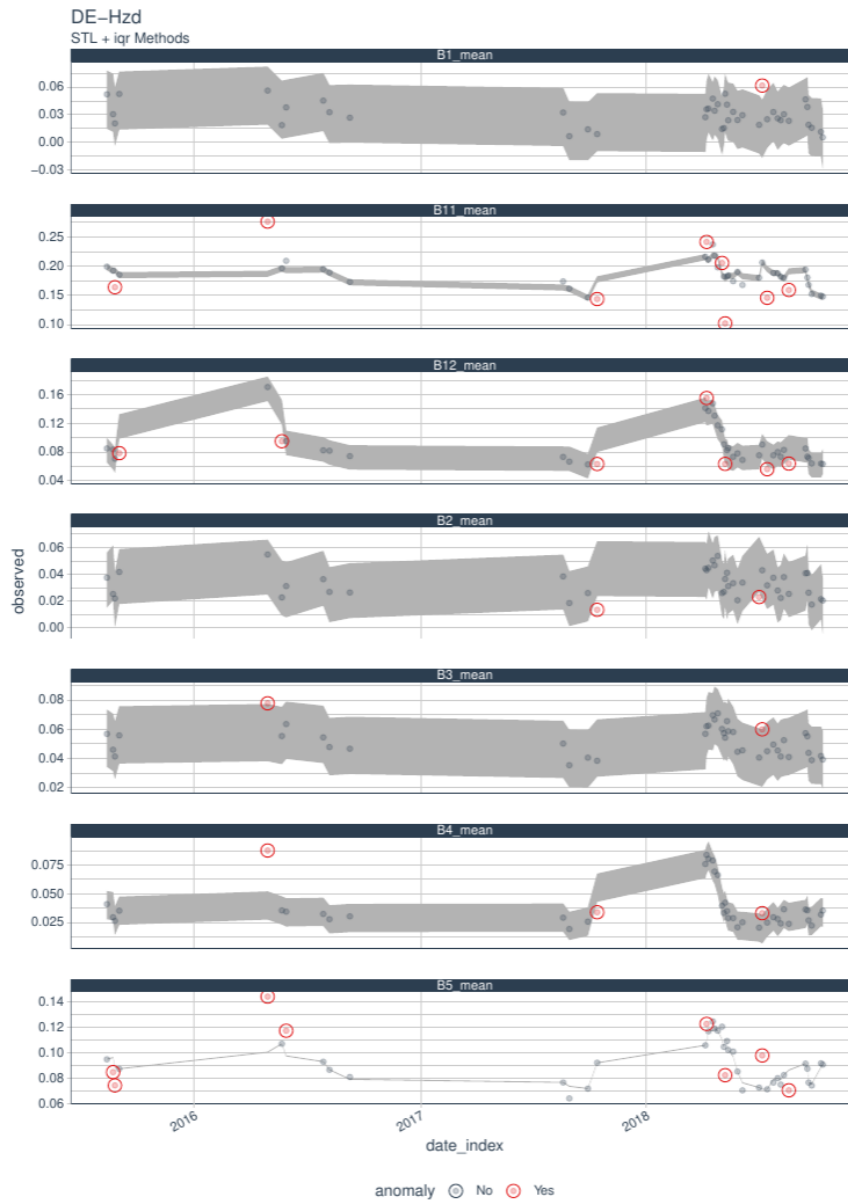


Figure 22: Outliers detection in Sentinel-2 images for the Hetzdorf (DE-Hzd) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

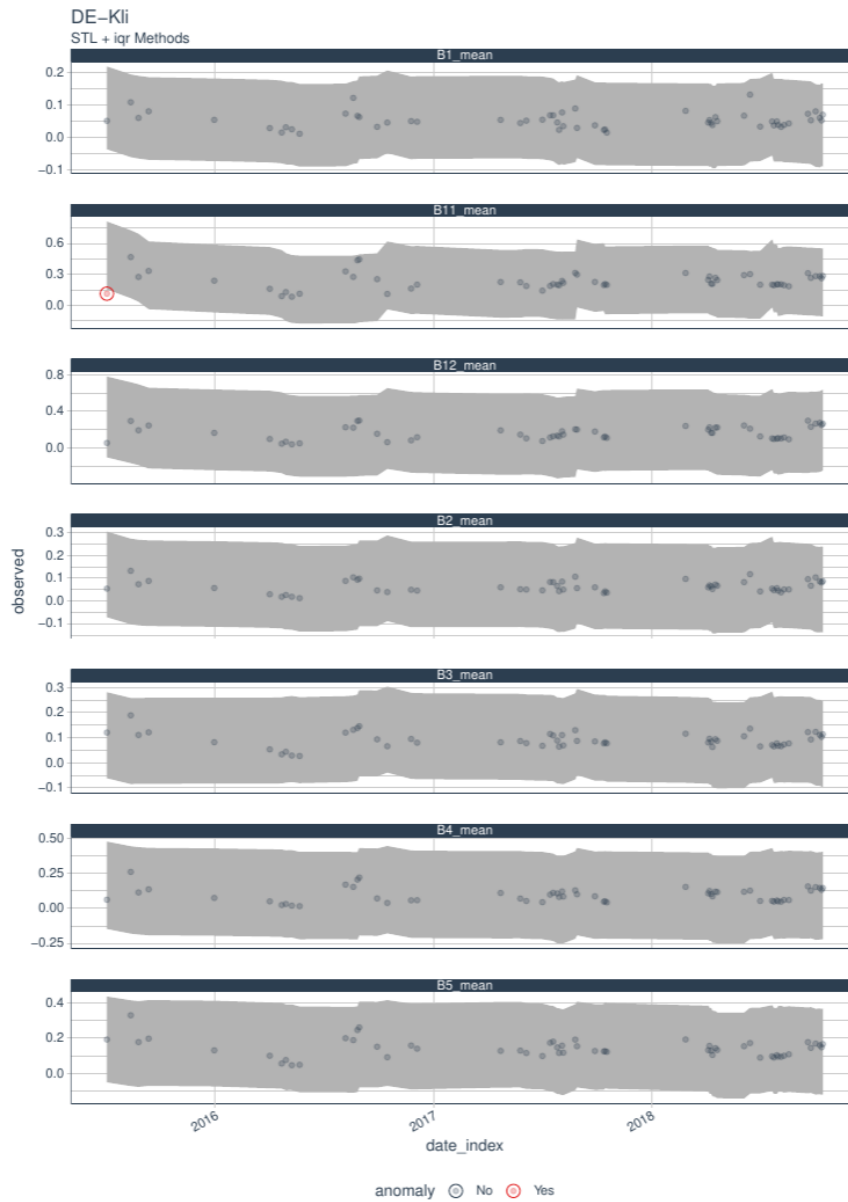


Figure 23: Outliers detection in Sentinel-2 images for the Klingenberg (DE-Kli) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

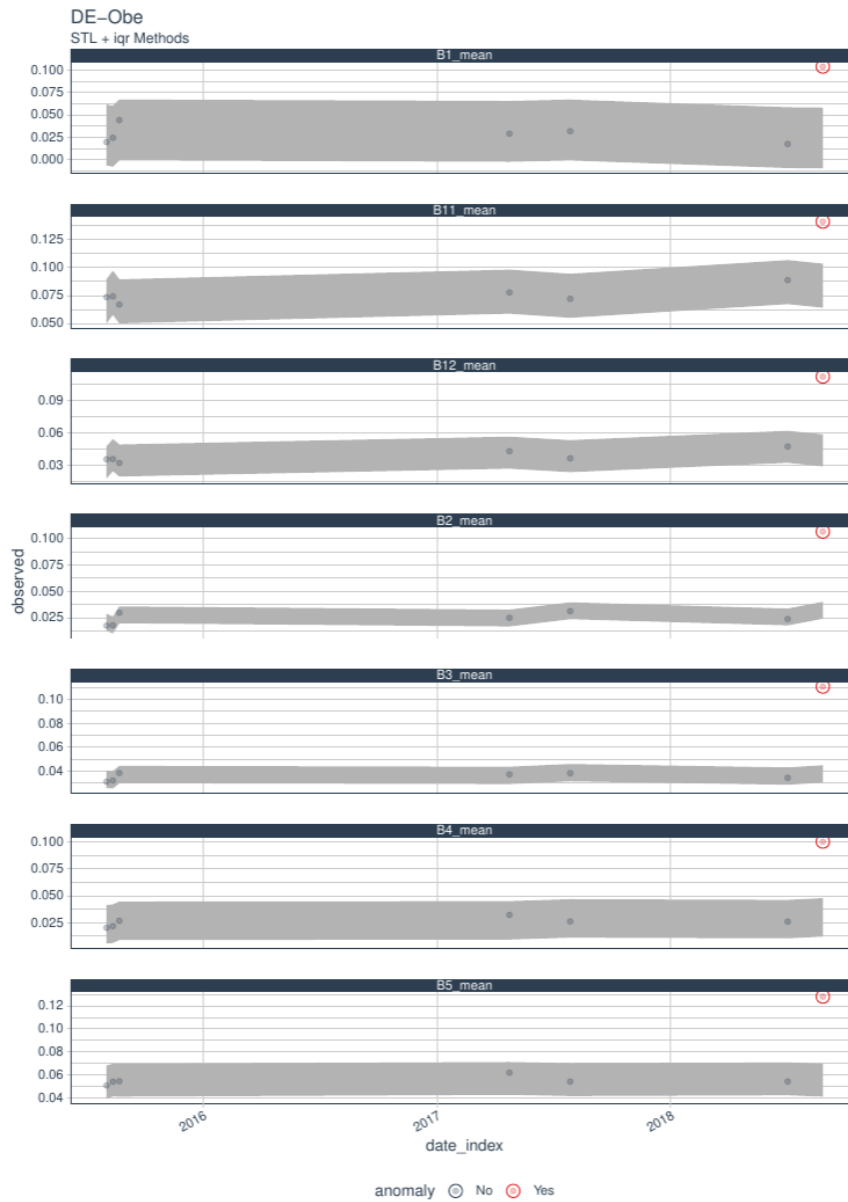


Figure 24: Outliers detection in Sentinel-2 images for the Oberbärenburg (DE-Obe) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

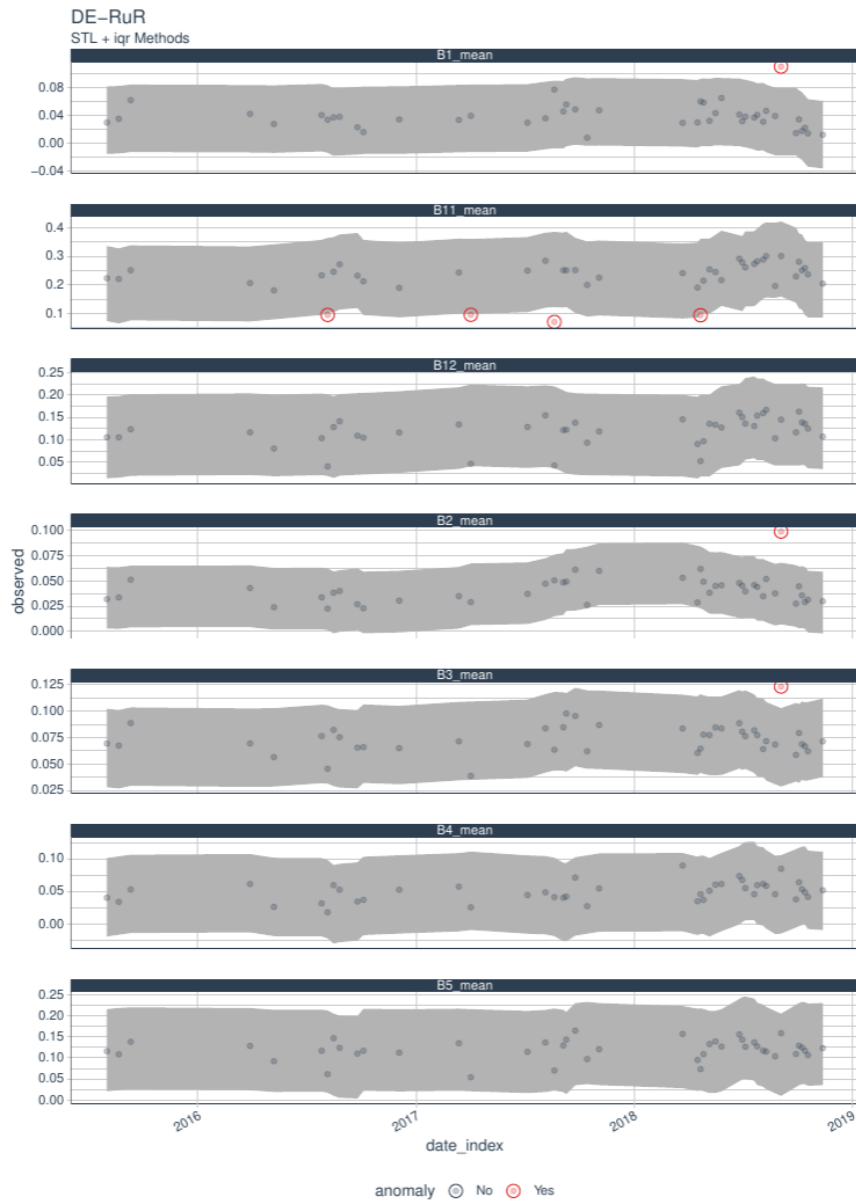


Figure 25: Outliers detection in Sentinel-2 images for the Rollesbroich (DE-RuR) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

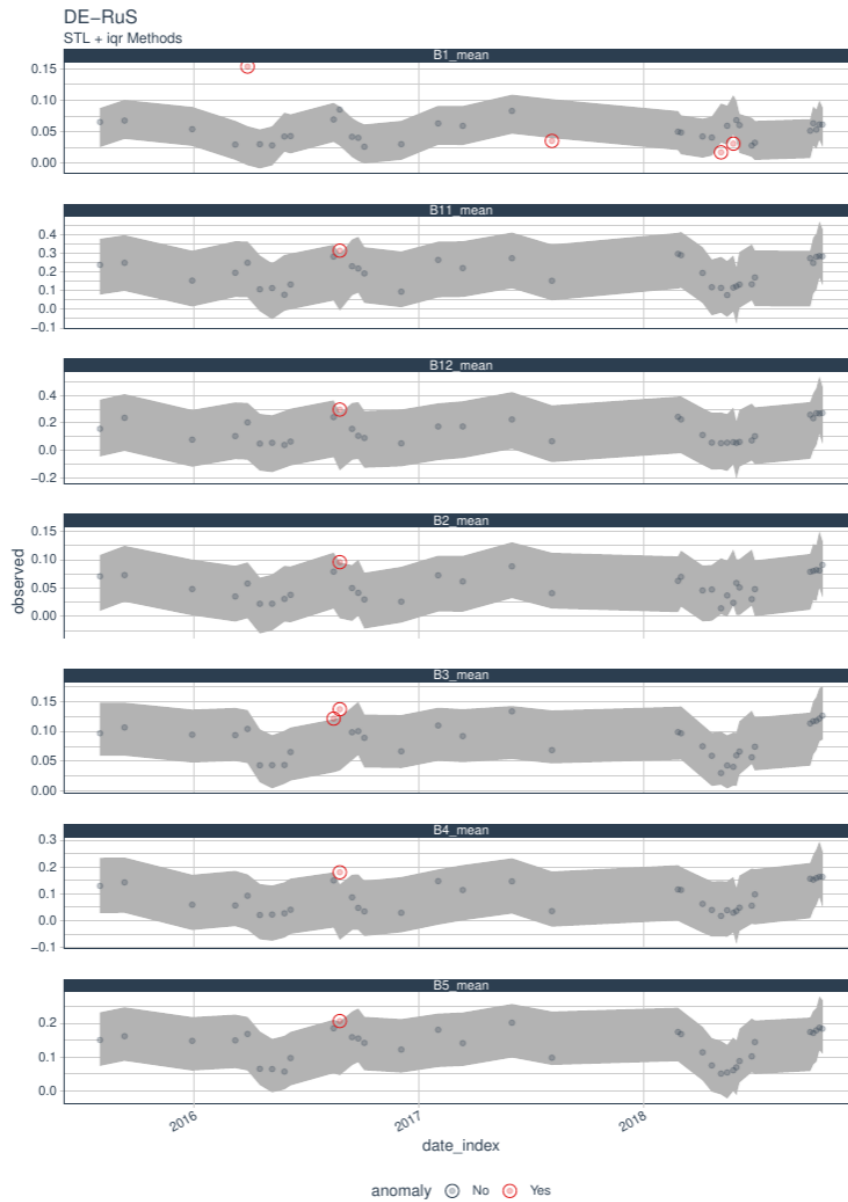


Figure 26: Outliers detection in Sentinel-2 images for the Selhausen Juelich (DE-Rus) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

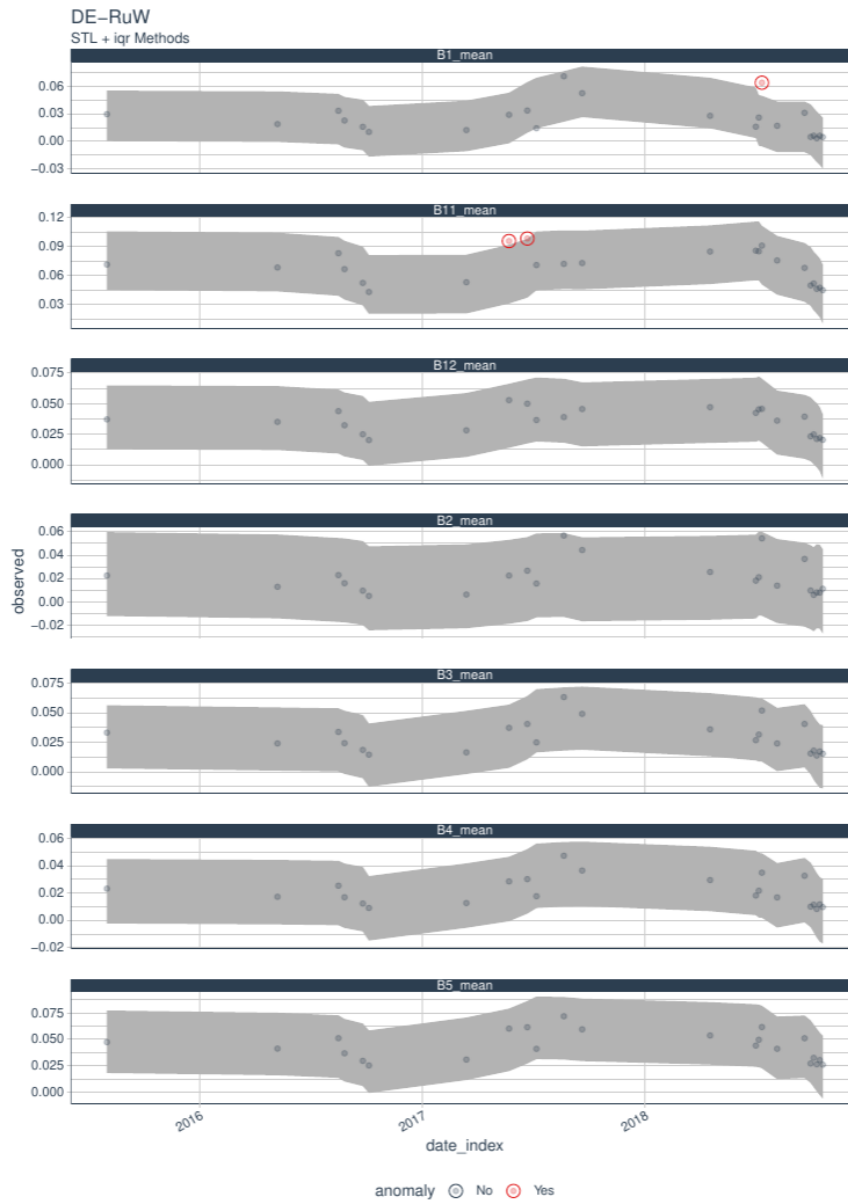


Figure 27: Outliers detection in Sentinel-2 images for the Wustebach (DE-RuW) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

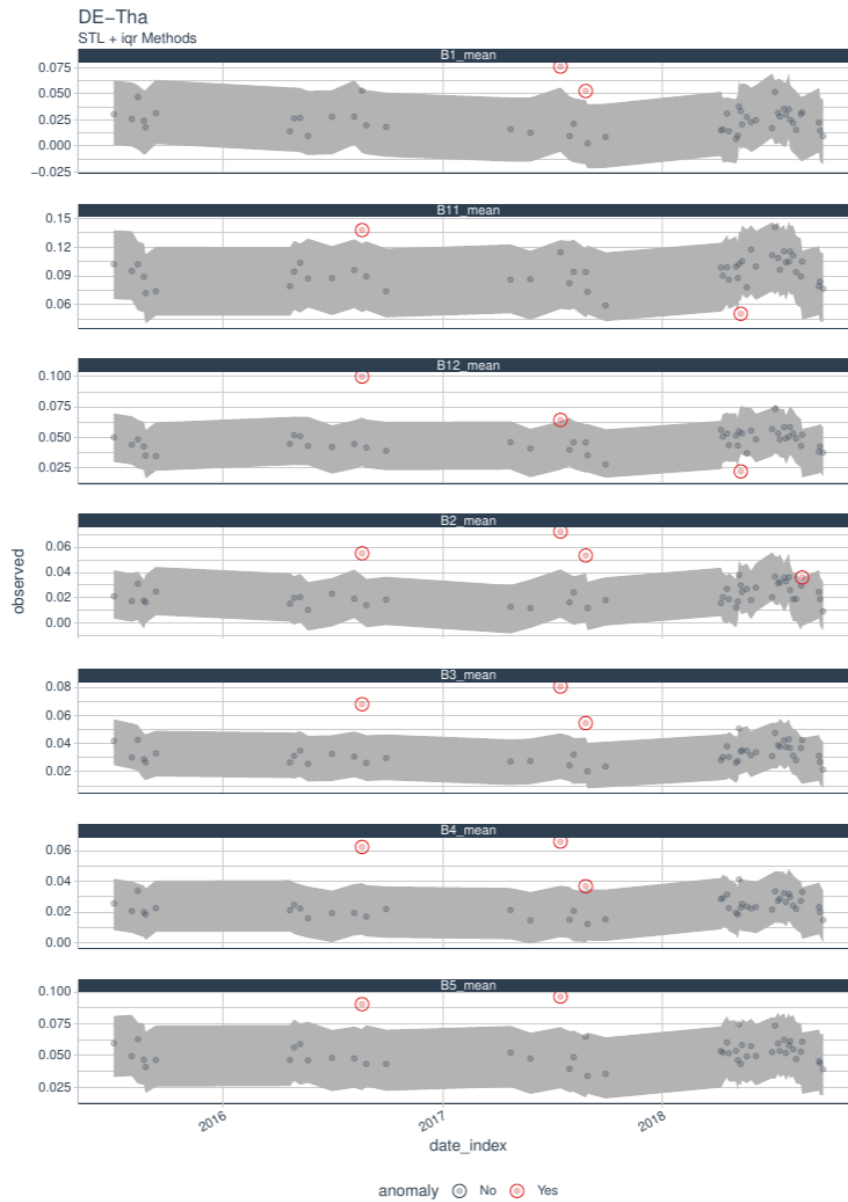


Figure 28: Outliers detection in Sentinel-2 images for the Tharandt (DE-Tha) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

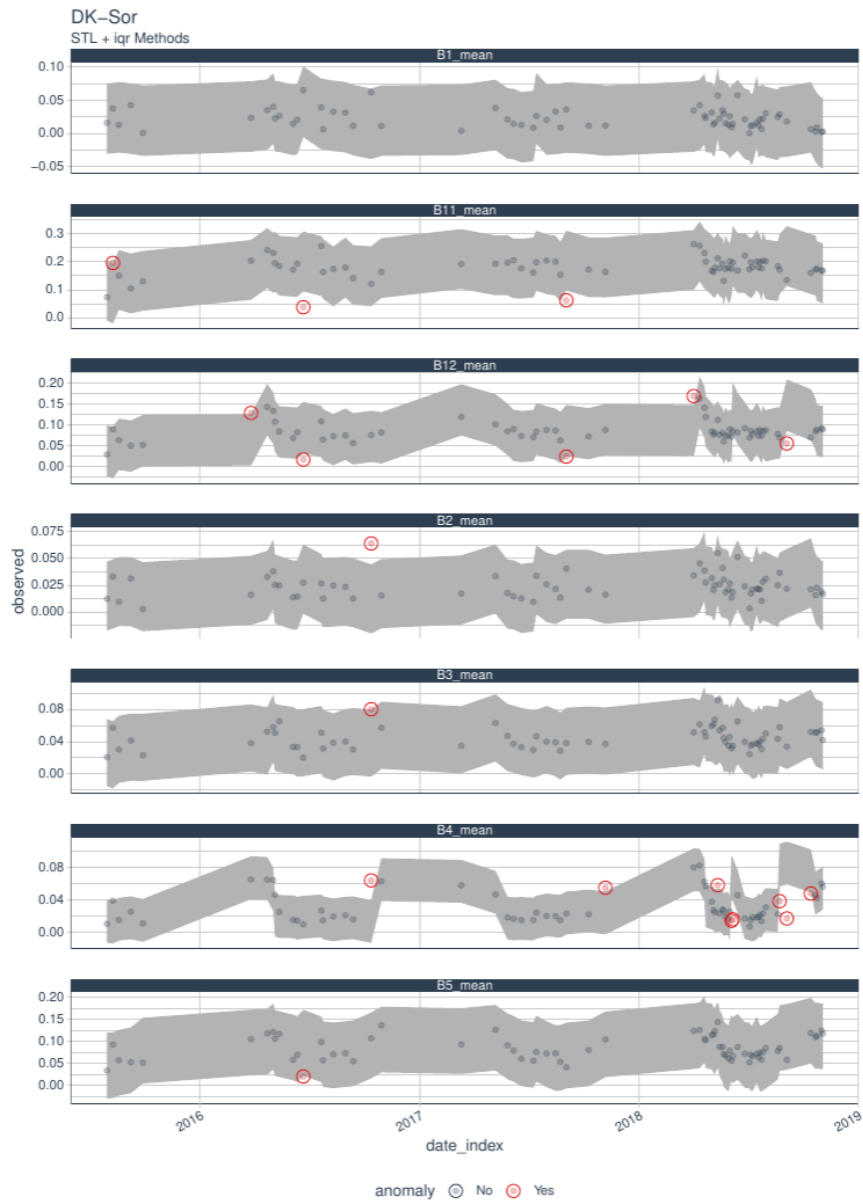


Figure 29: Outliers detection in Sentinel-2 images for the Soroe (DK-Sor) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

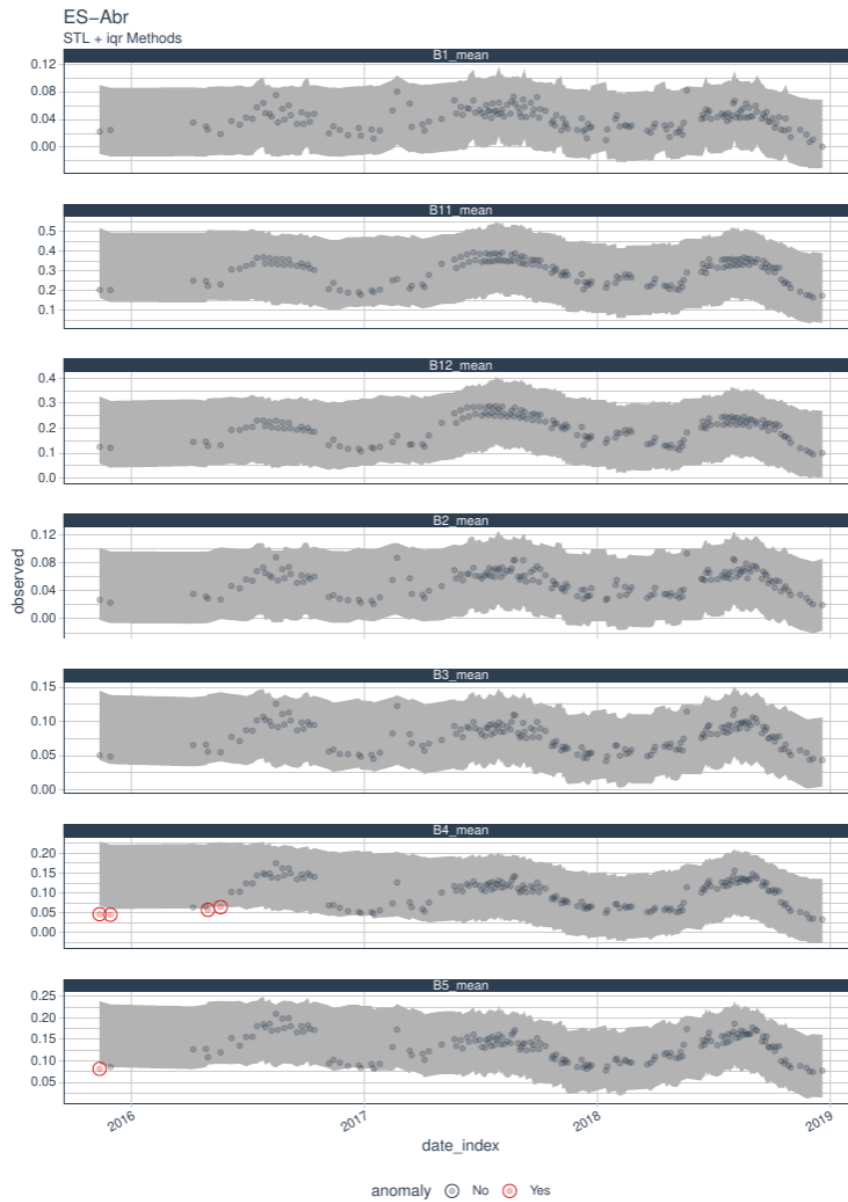


Figure 30: Outliers detection in Sentinel-2 images for the Albuera (ES-Abr) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

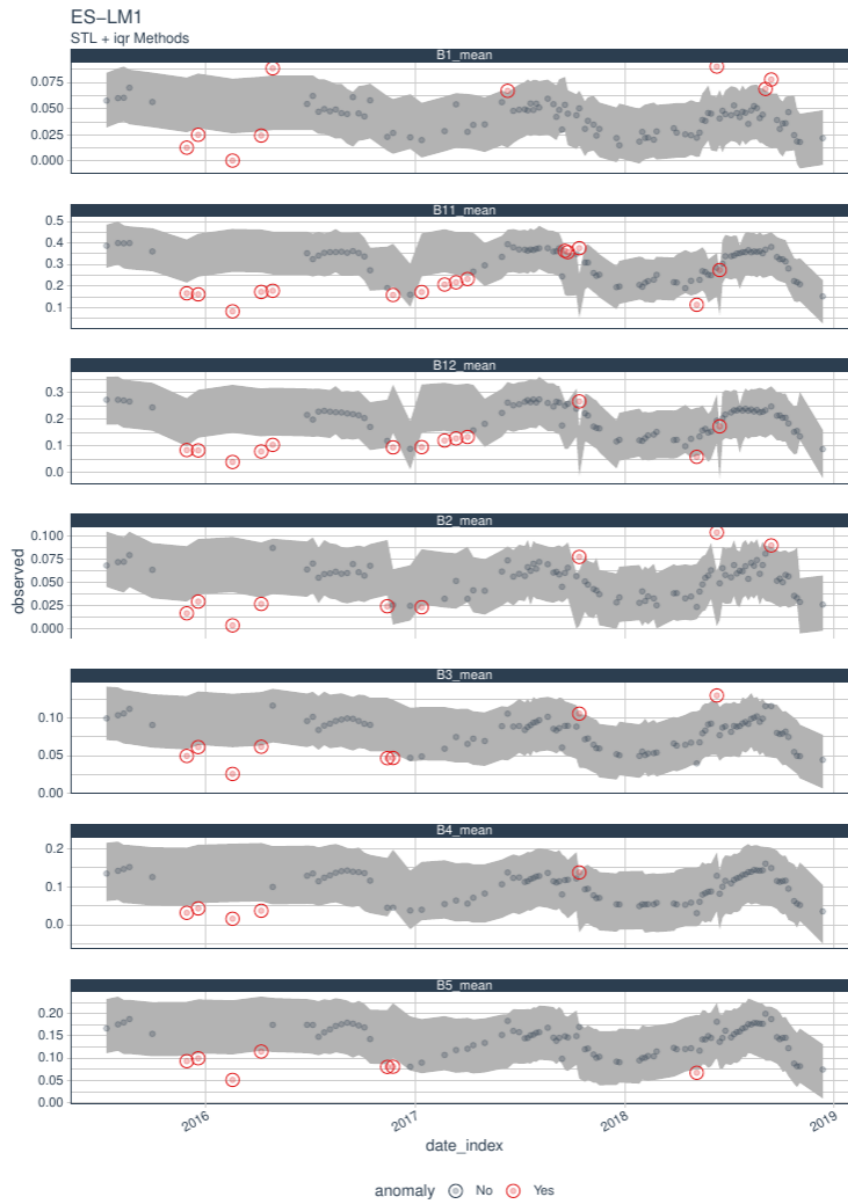


Figure 31: Outliers detection in Sentinel-2 images for the Majadas del Tietar North (ES-LM1) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

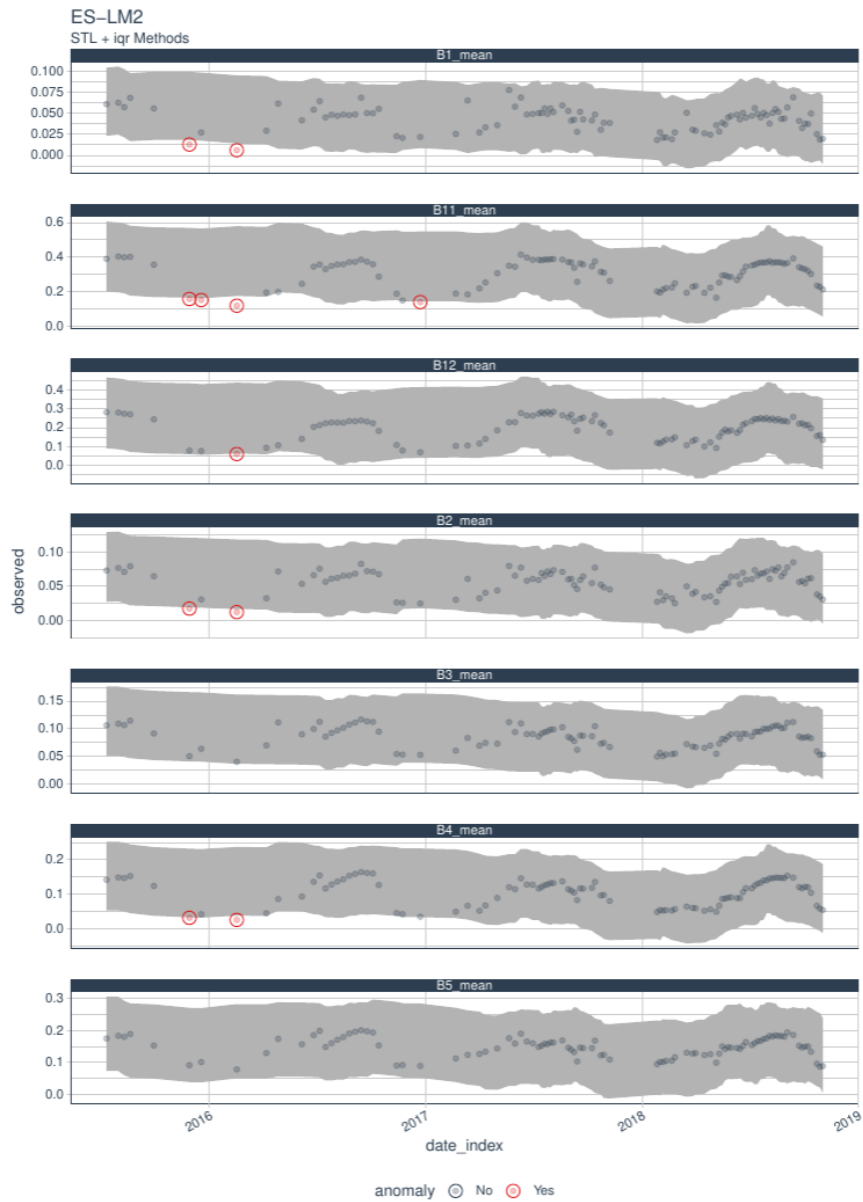


Figure 32: Outliers detection in Sentinel-2 images for the Majadas del Tietar South (ES-LM2) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

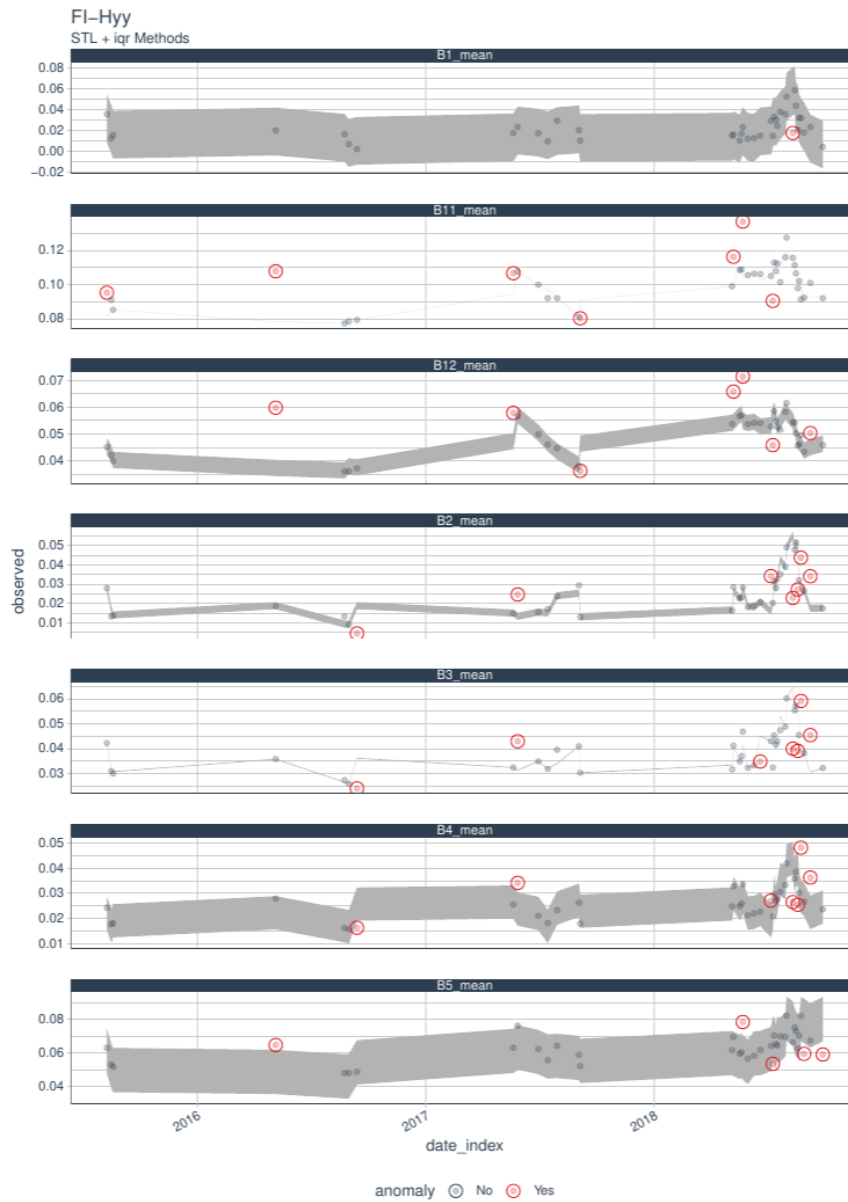


Figure 33: Outliers detection in Sentinel-2 images for the Hyytiala (FI-Hyy) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

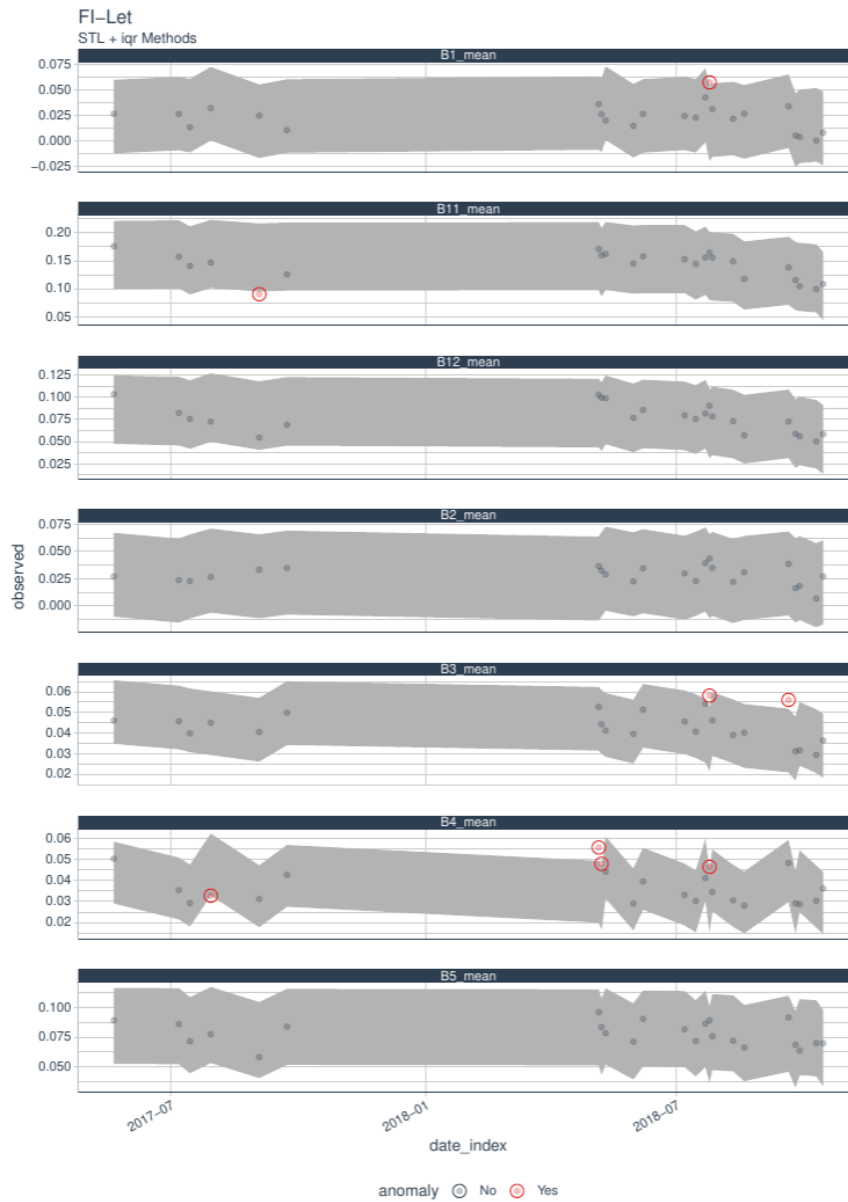


Figure 34: Outliers detection in Sentinel-2 images for the Lettosuo (FI-Let) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

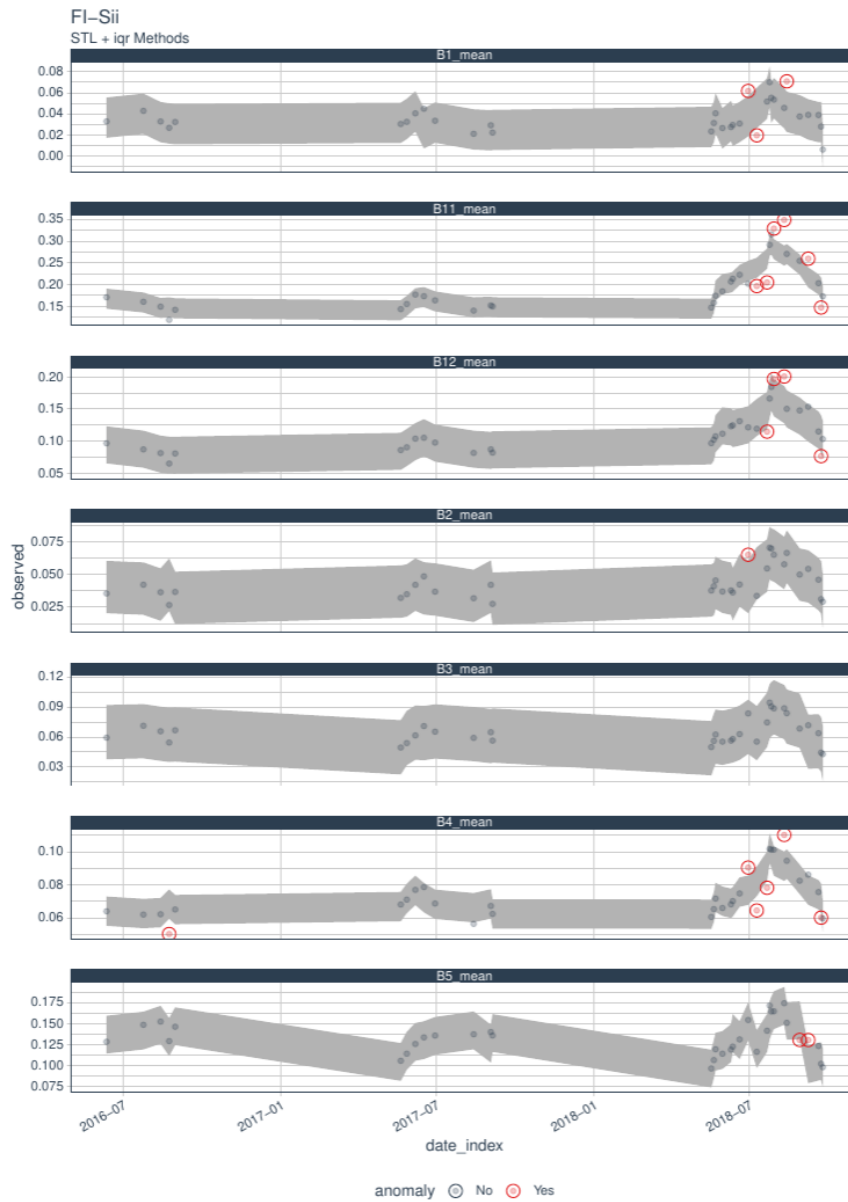


Figure 35: Outliers detection in Sentinel-2 images for the Siikaneva (FI-Sii) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

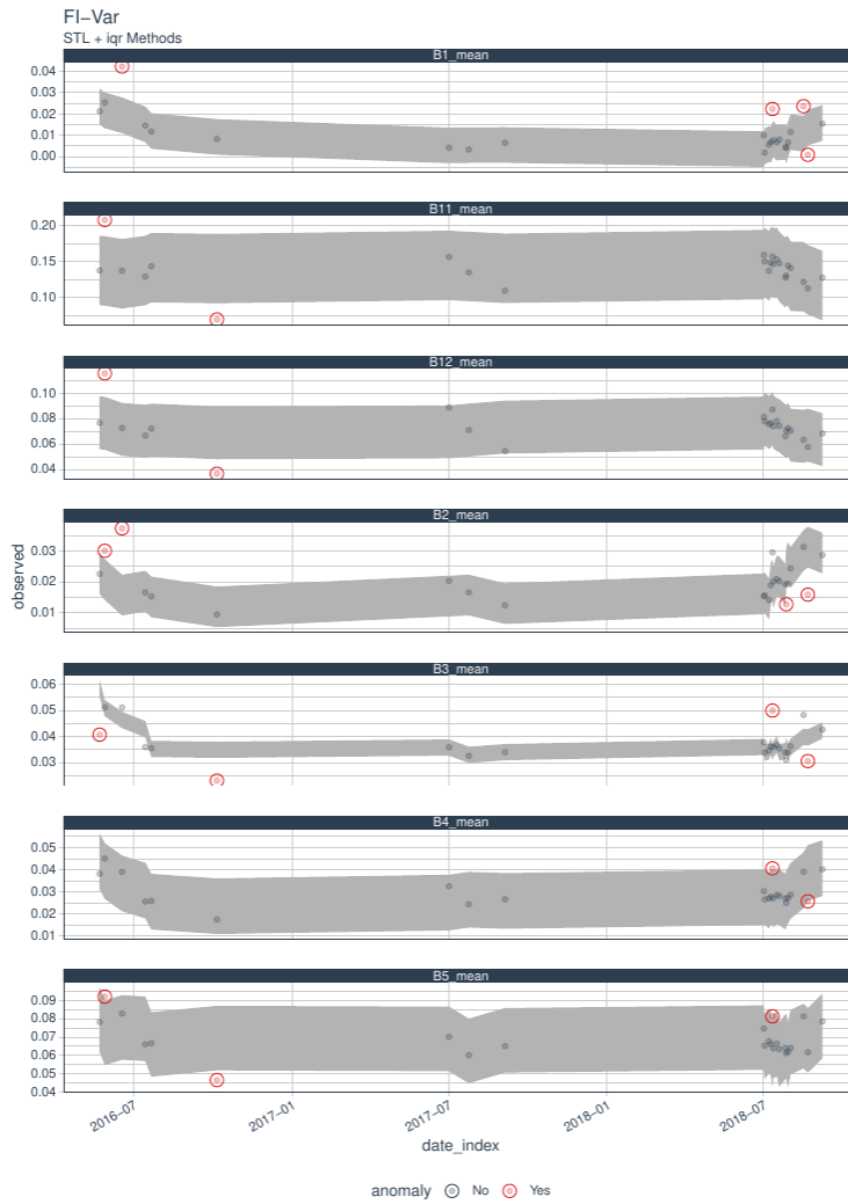


Figure 36: Outliers detection in Sentinel-2 images for the Varrío (FI-Var) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

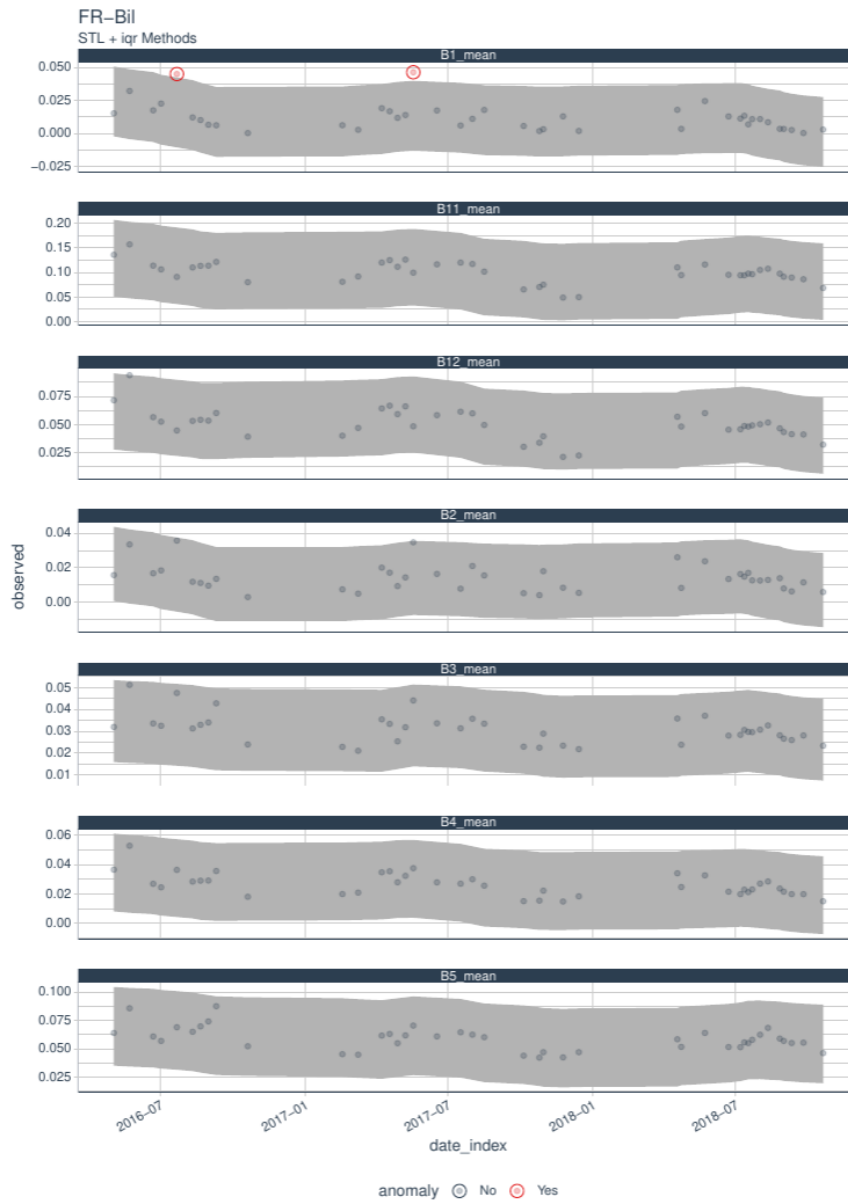


Figure 37: Outliers detection in Sentinel-2 images for the Bilos (FR-Bil) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

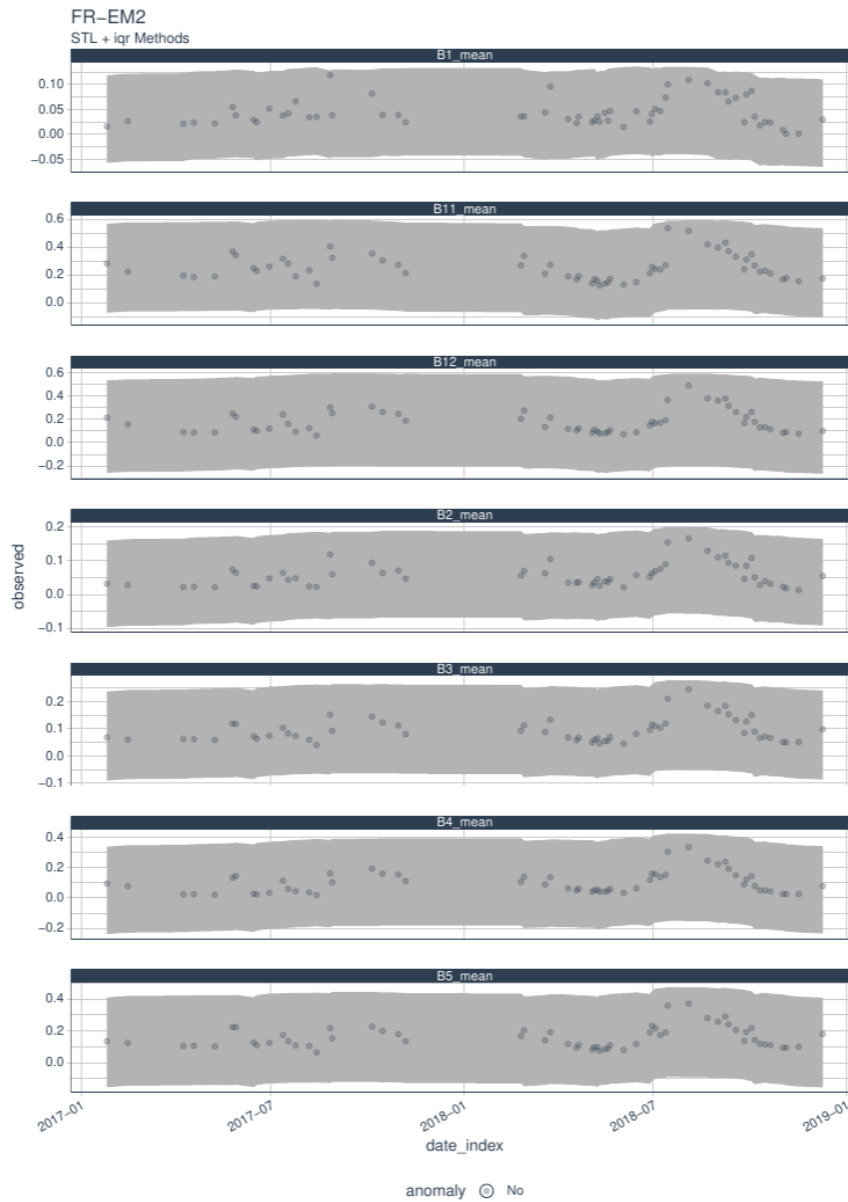


Figure 38: Outliers detection in Sentinel-2 images for the Estrees-Mons A28 (FR-EM2) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

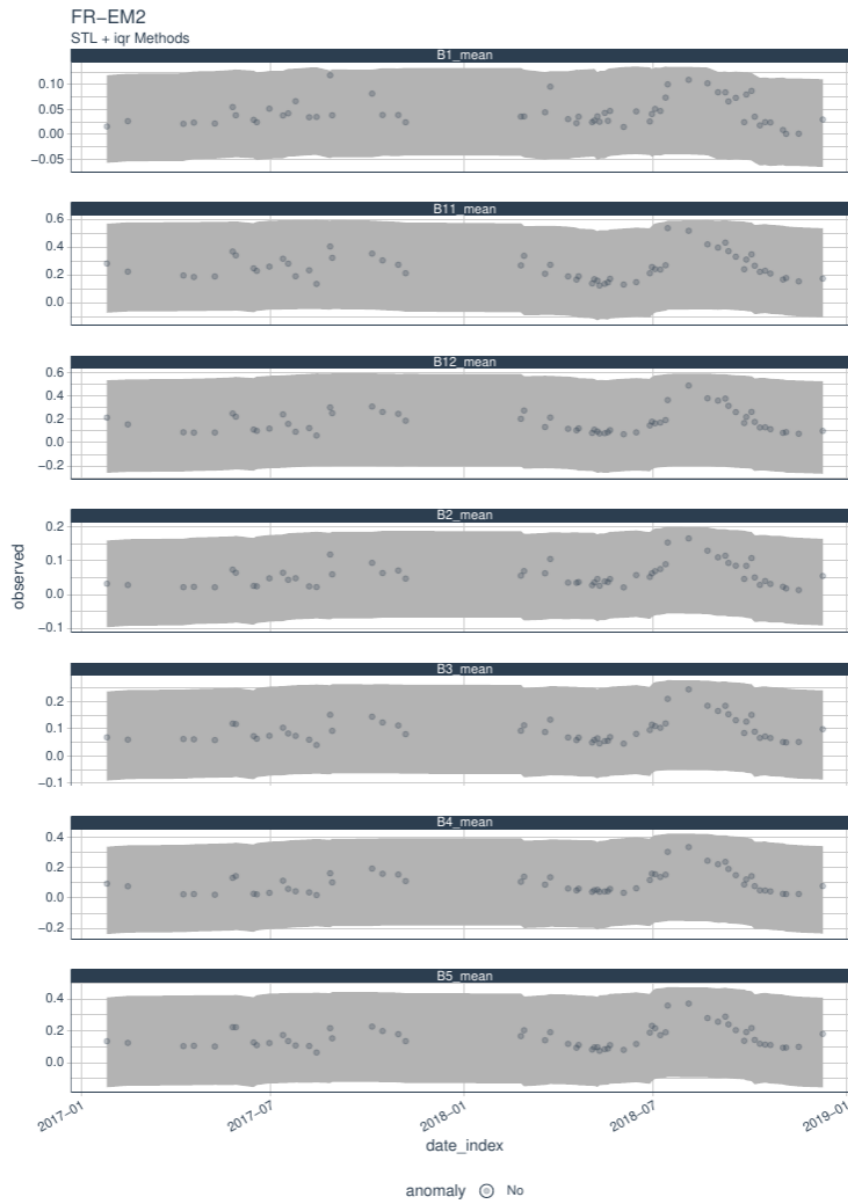


Figure 39: Outliers detection in Sentinel-2 images for the Hesse (FR-Hes) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

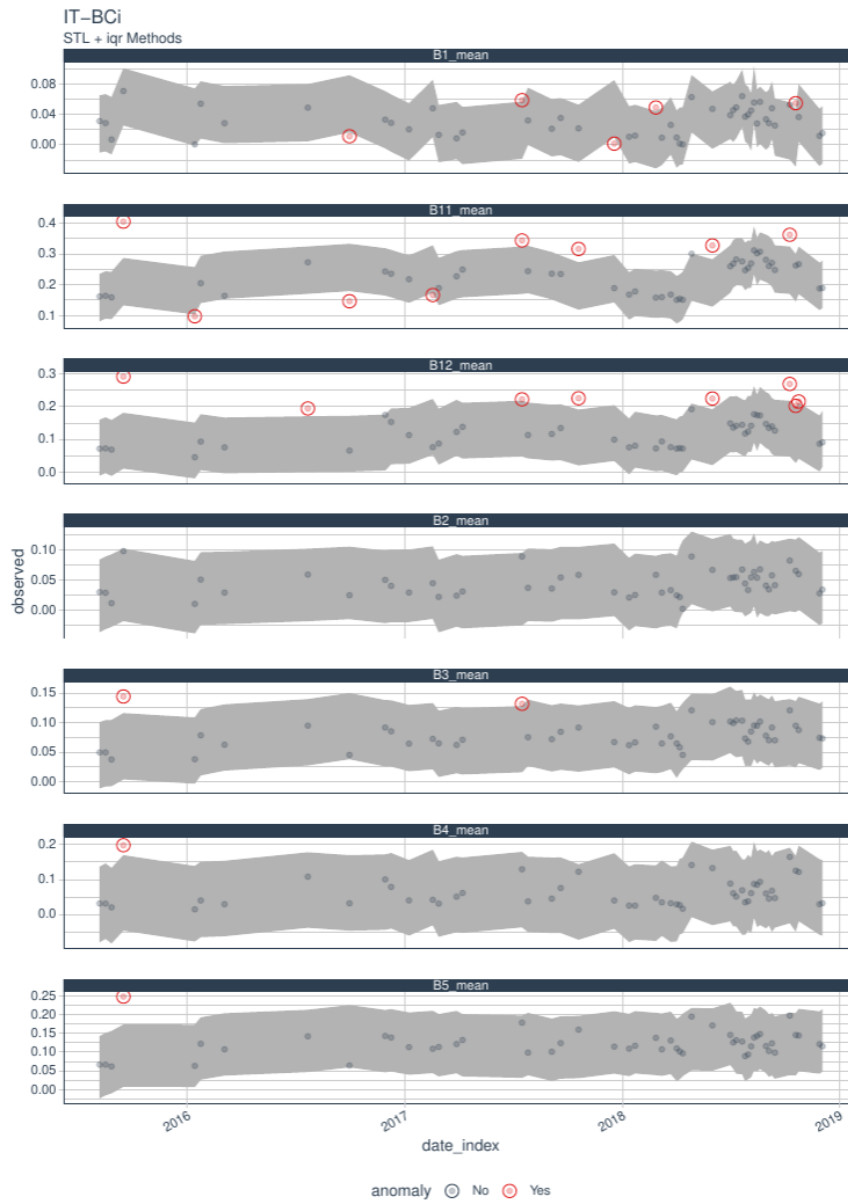


Figure 40: Outliers detection in Sentinel-2 images for the Borgo Cioffi (IT-BCi) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

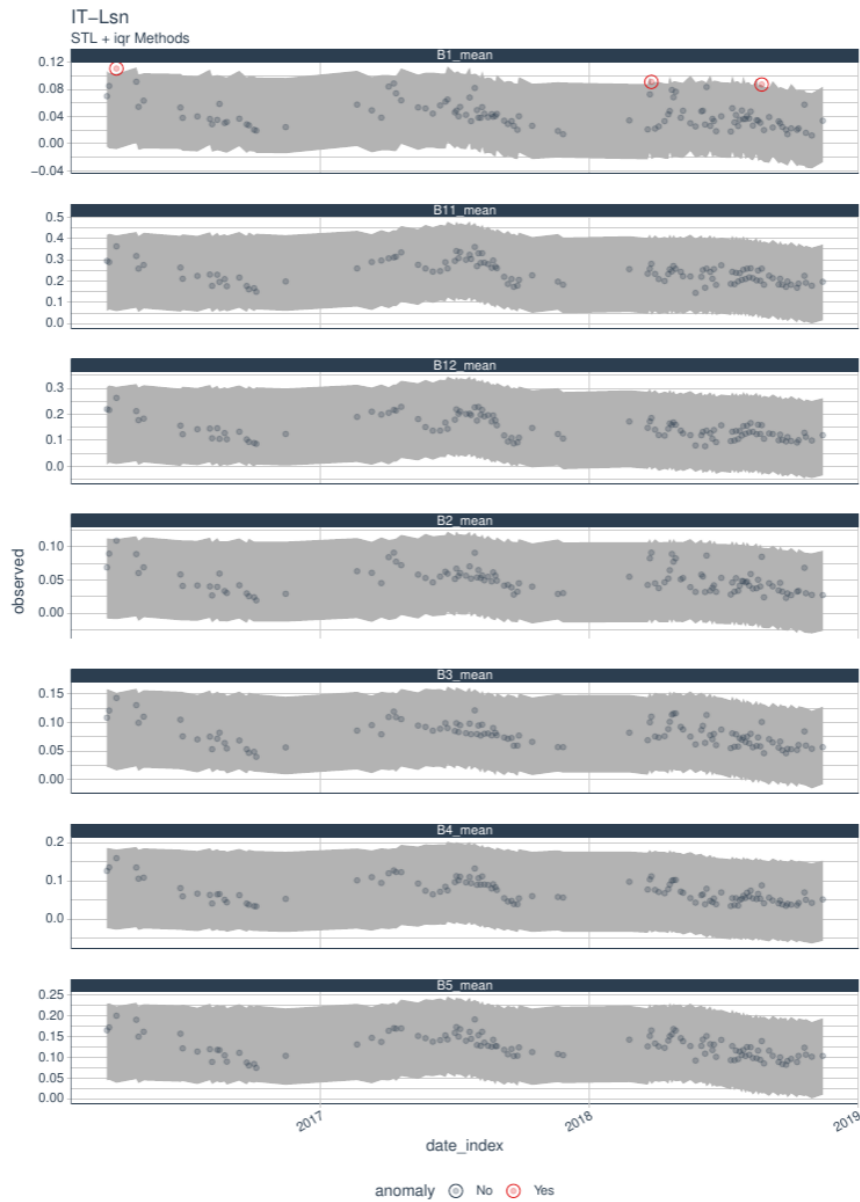


Figure 41: Outliers detection in Sentinel-2 images for the Lison (IT-Lsn) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

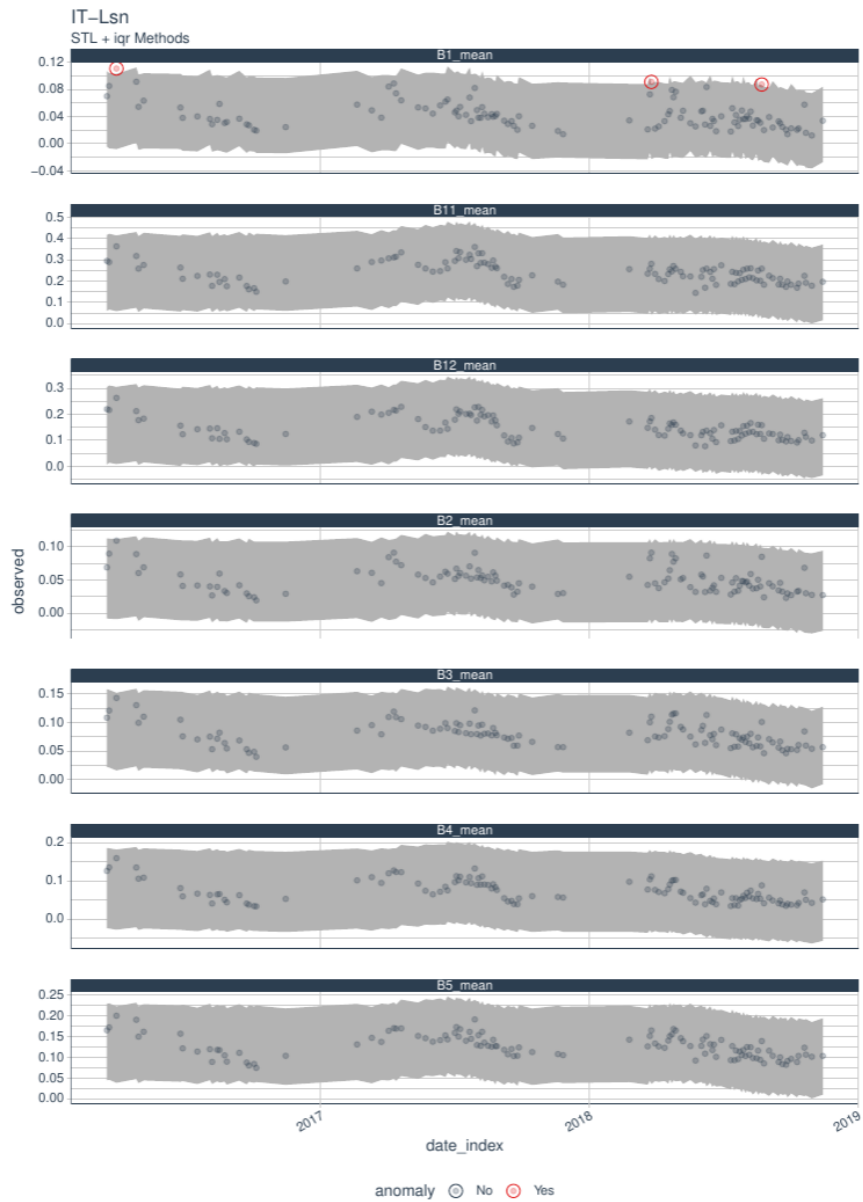


Figure 42: Outliers detection in Sentinel-2 images for the Lison (IT-Lsn) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

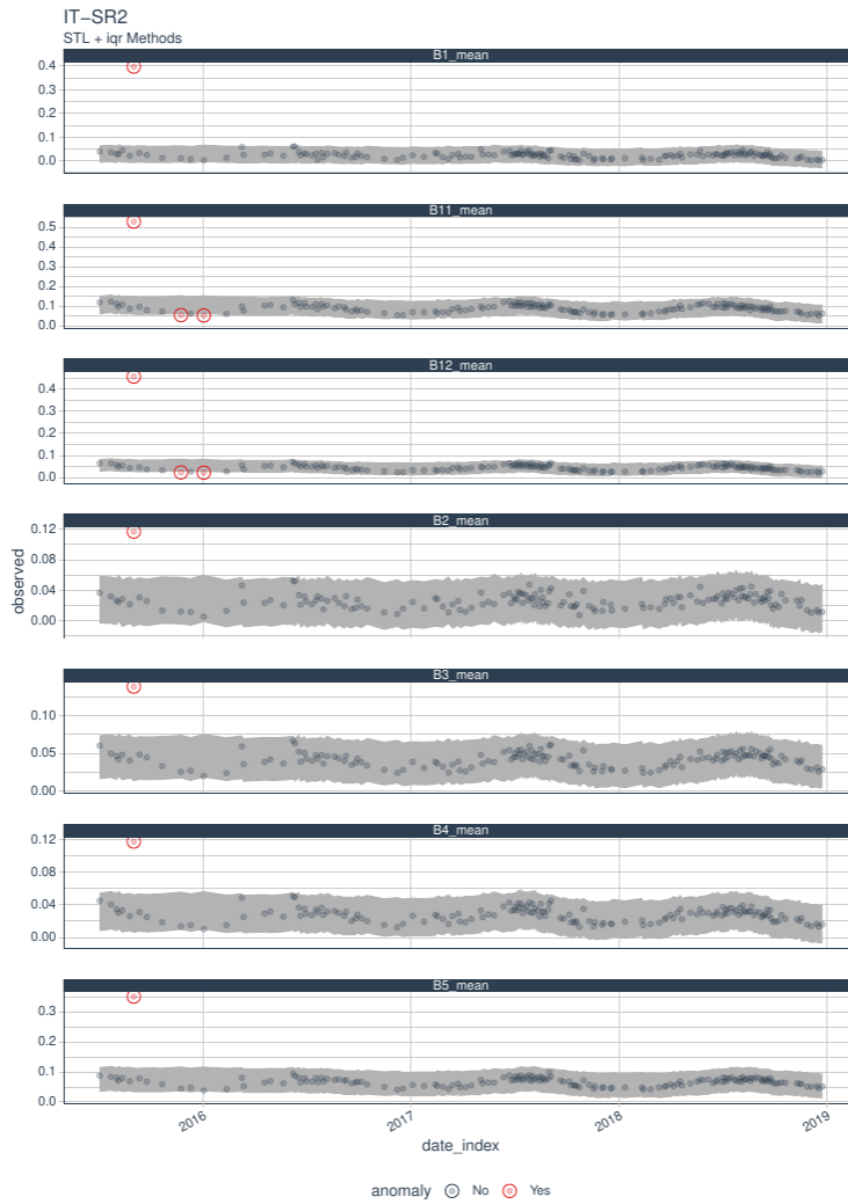


Figure 43: Outliers detection in Sentinel-2 images for the San Rossore 2 (IT-SR2) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

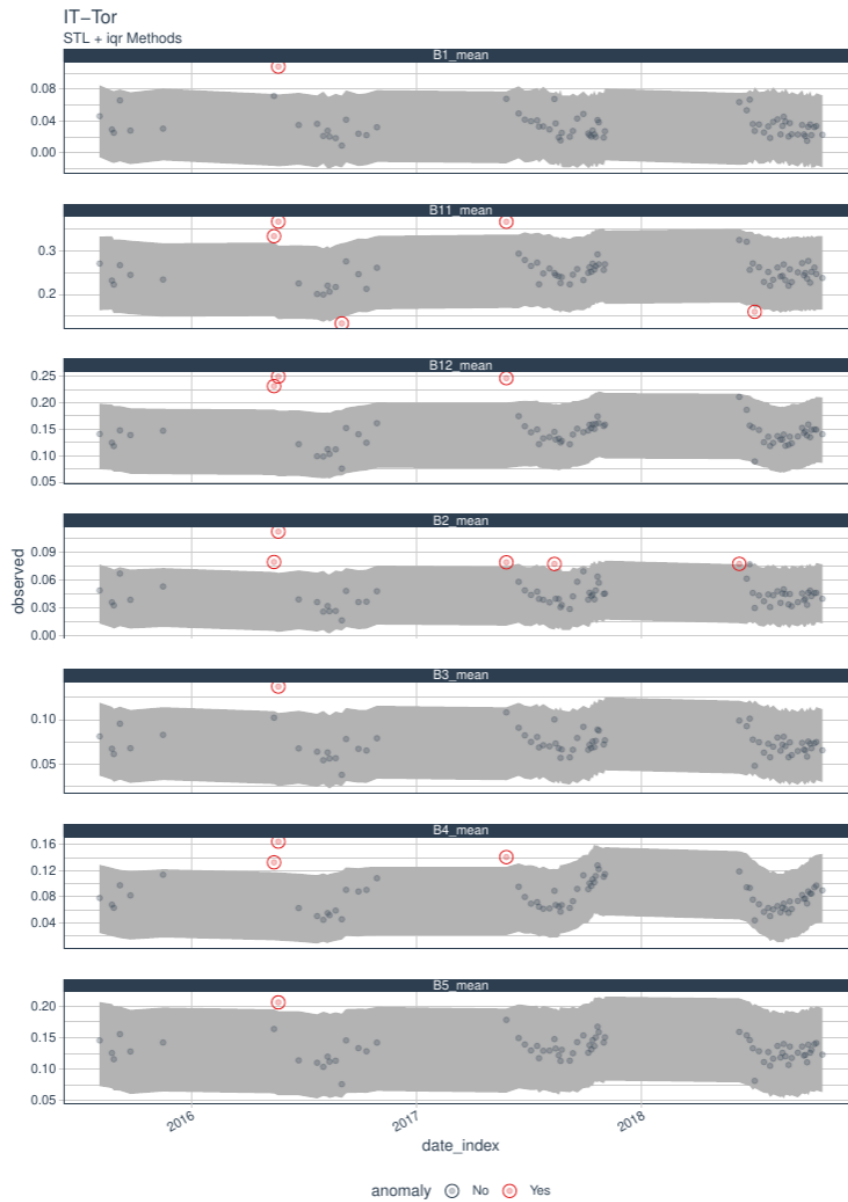


Figure 44: Outliers detection in Sentinel-2 images for the Torgnon (IT-Tor) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

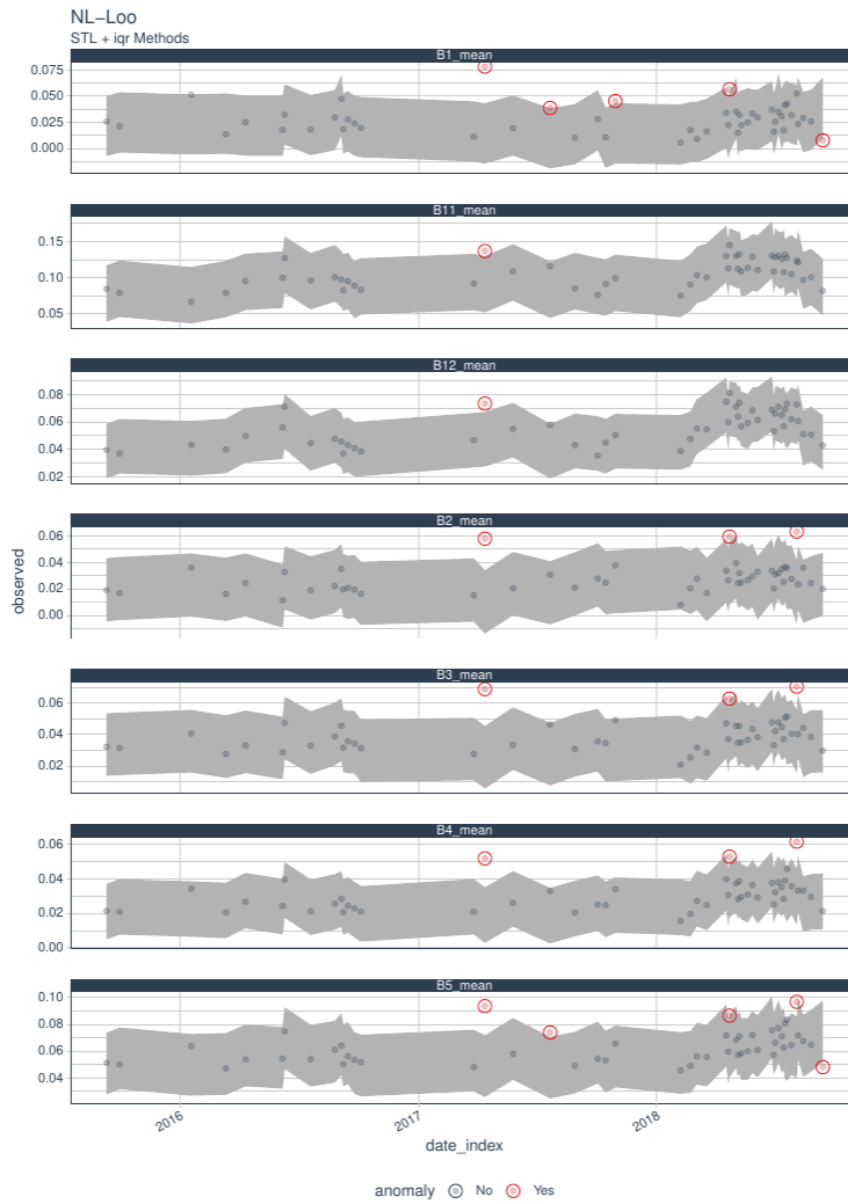


Figure 45: Outliers detection in Sentinel-2 images for the Loobos (NL-Loo) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

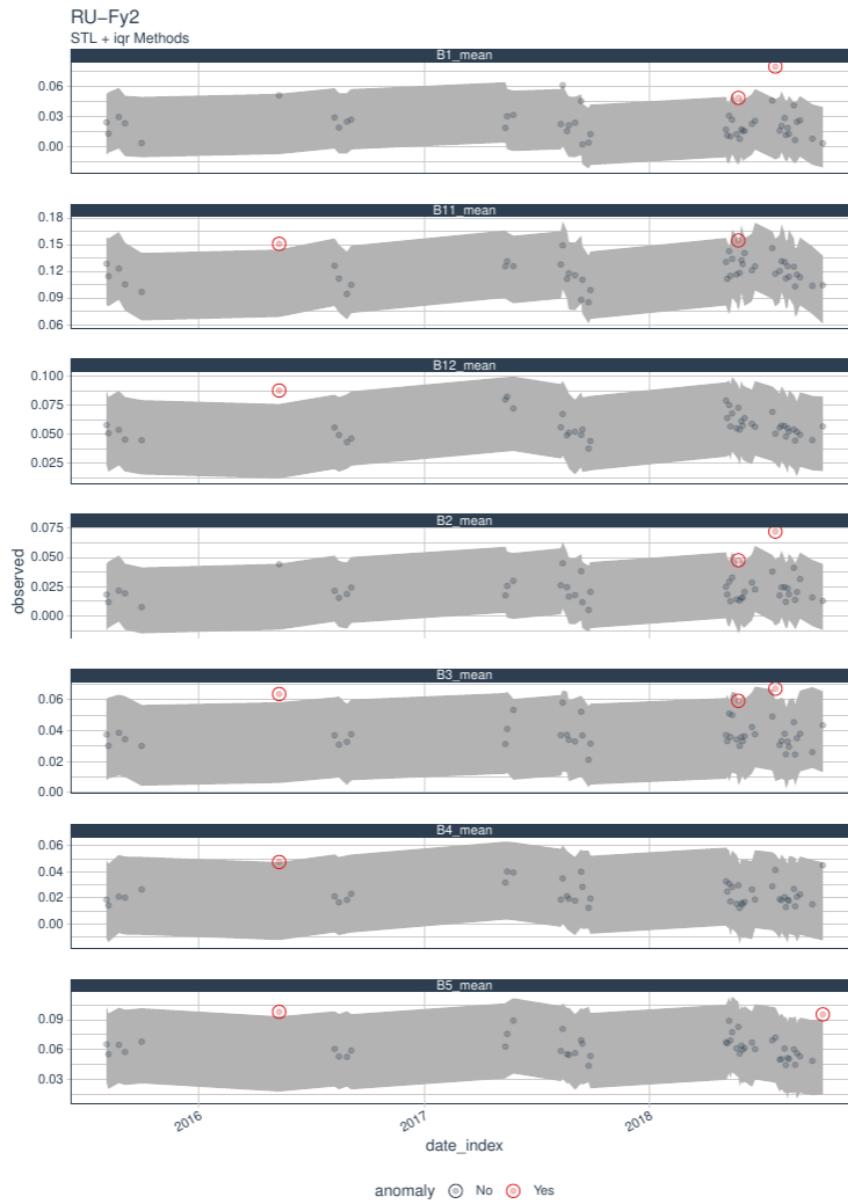


Figure 46: Outliers detection in Sentinel-2 images for the Fyodorovskoye dry spruce stand (RU-Fy2) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

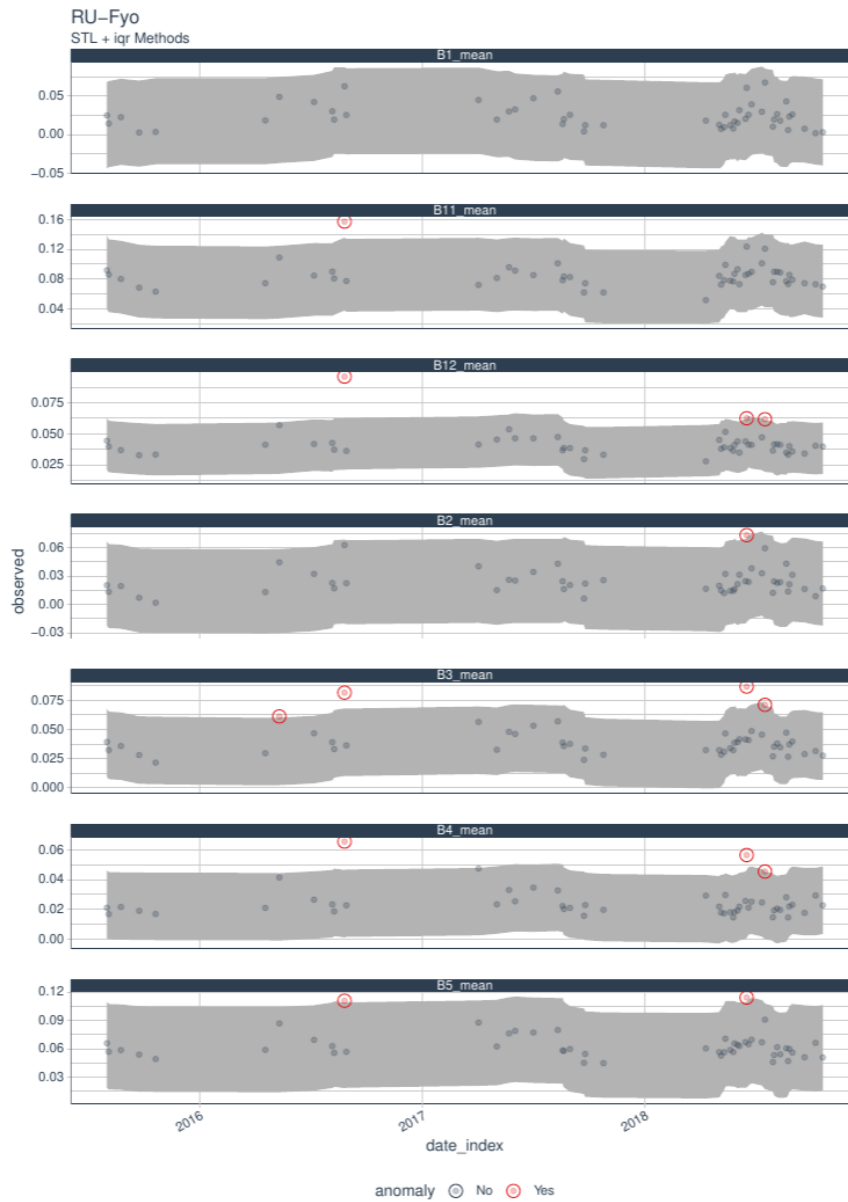


Figure 47: Outliers detection in Sentinel-2 images for the Fyodorovskoye (RU-Fyo) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

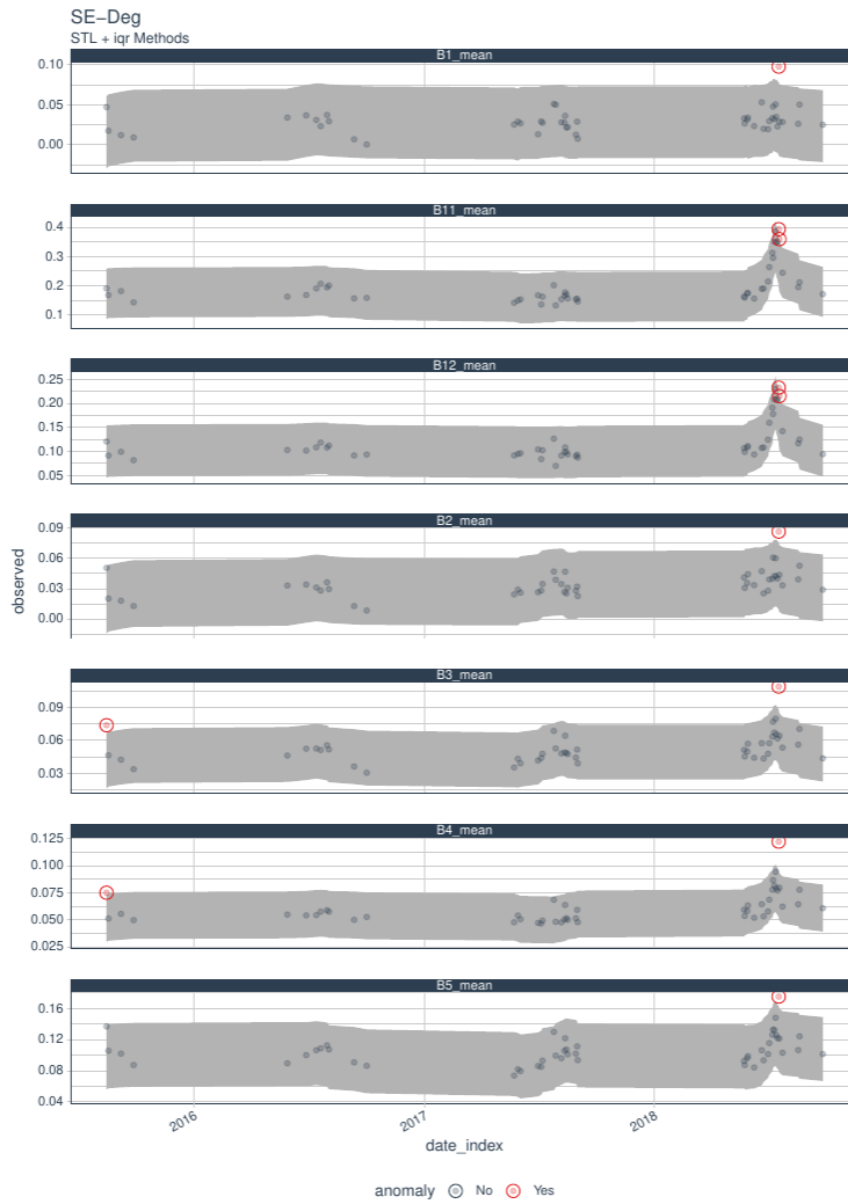


Figure 48: Outliers detection in Sentinel-2 images for the Degero (SE-Deg) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

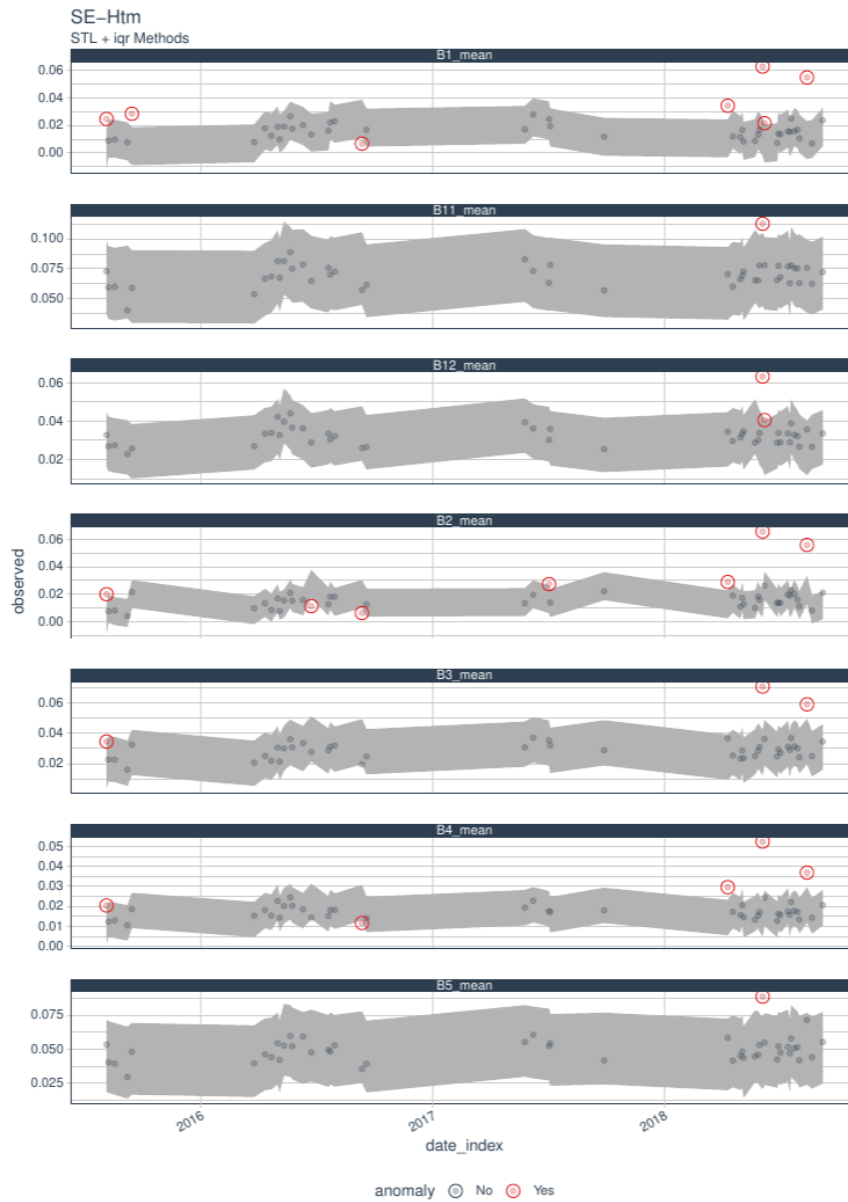


Figure 49: Outliers detection in Sentinel-2 images for the Hyltemossa (SE-Htm) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

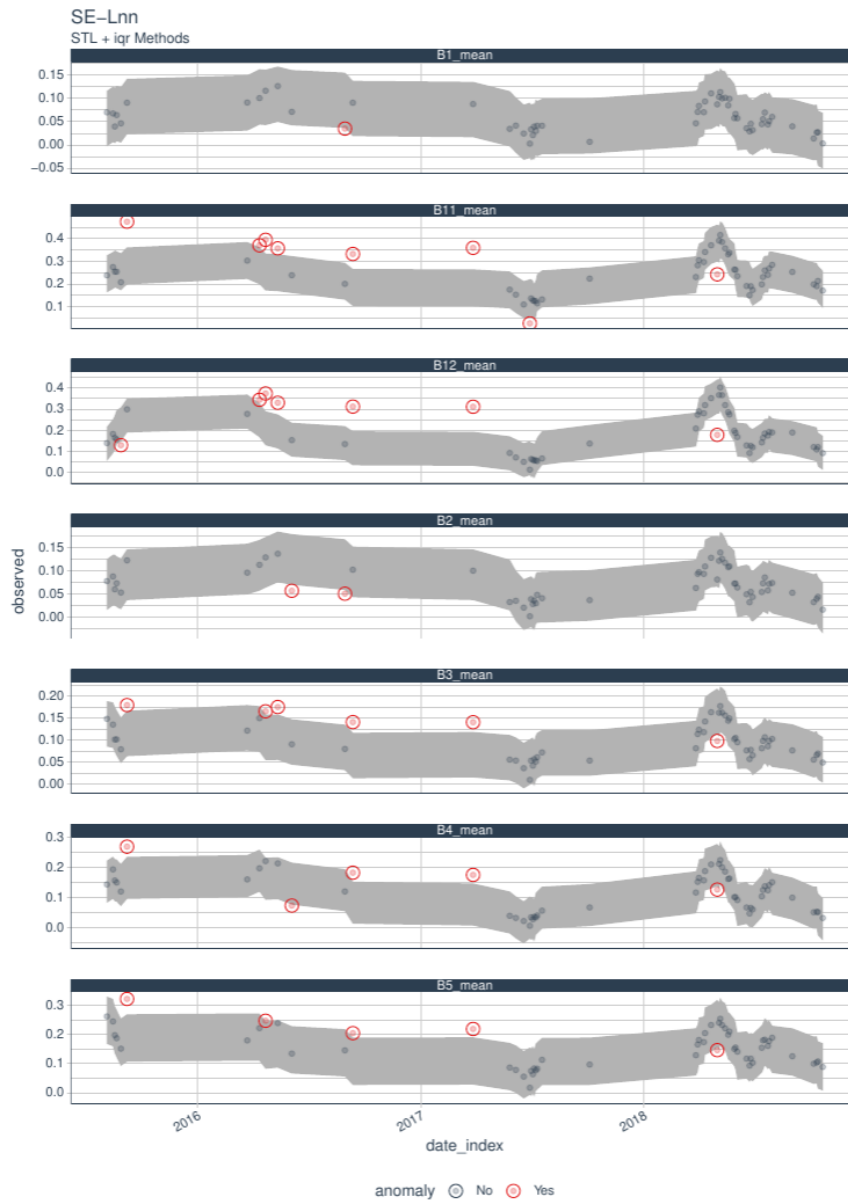


Figure 50: Outliers detection in Sentinel-2 images for the Lanna (SE-Lnn) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

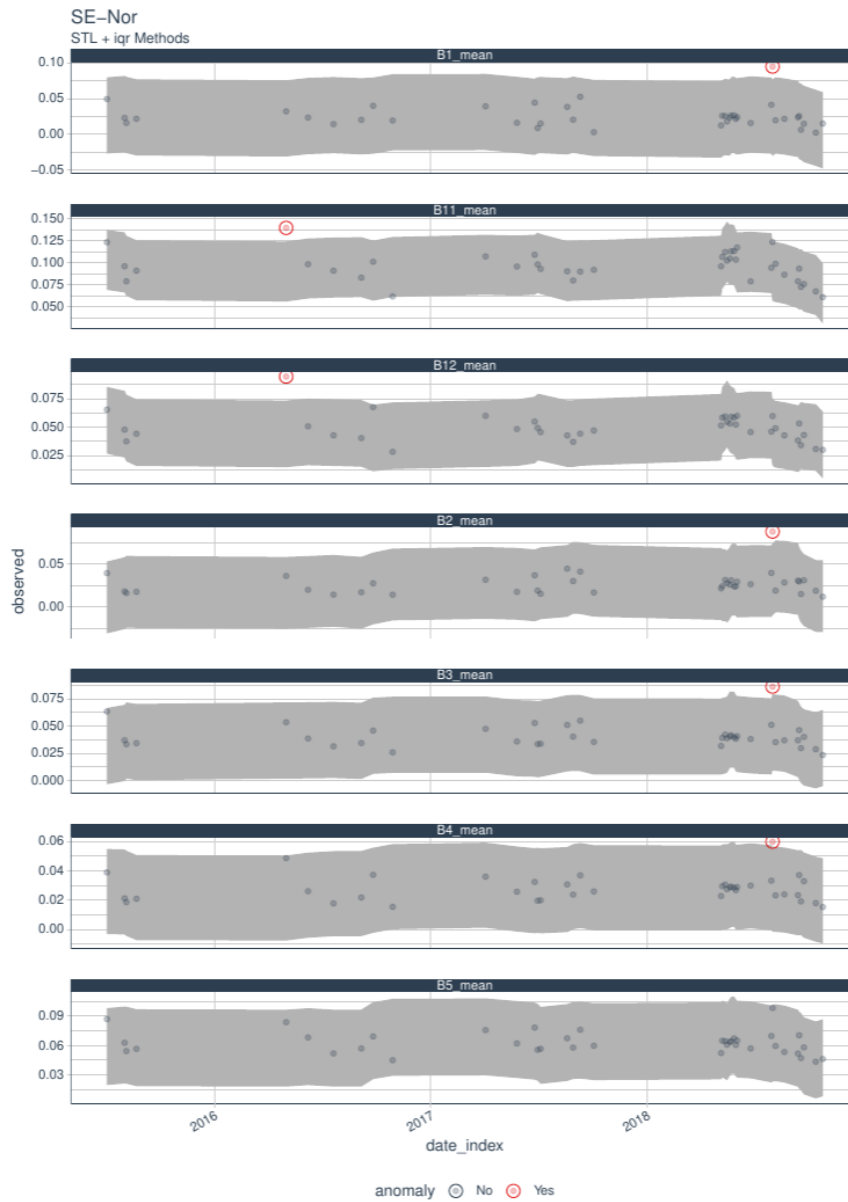


Figure 51: Outliers detection in Sentinel-2 images for the Norunda (SE-Nor) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

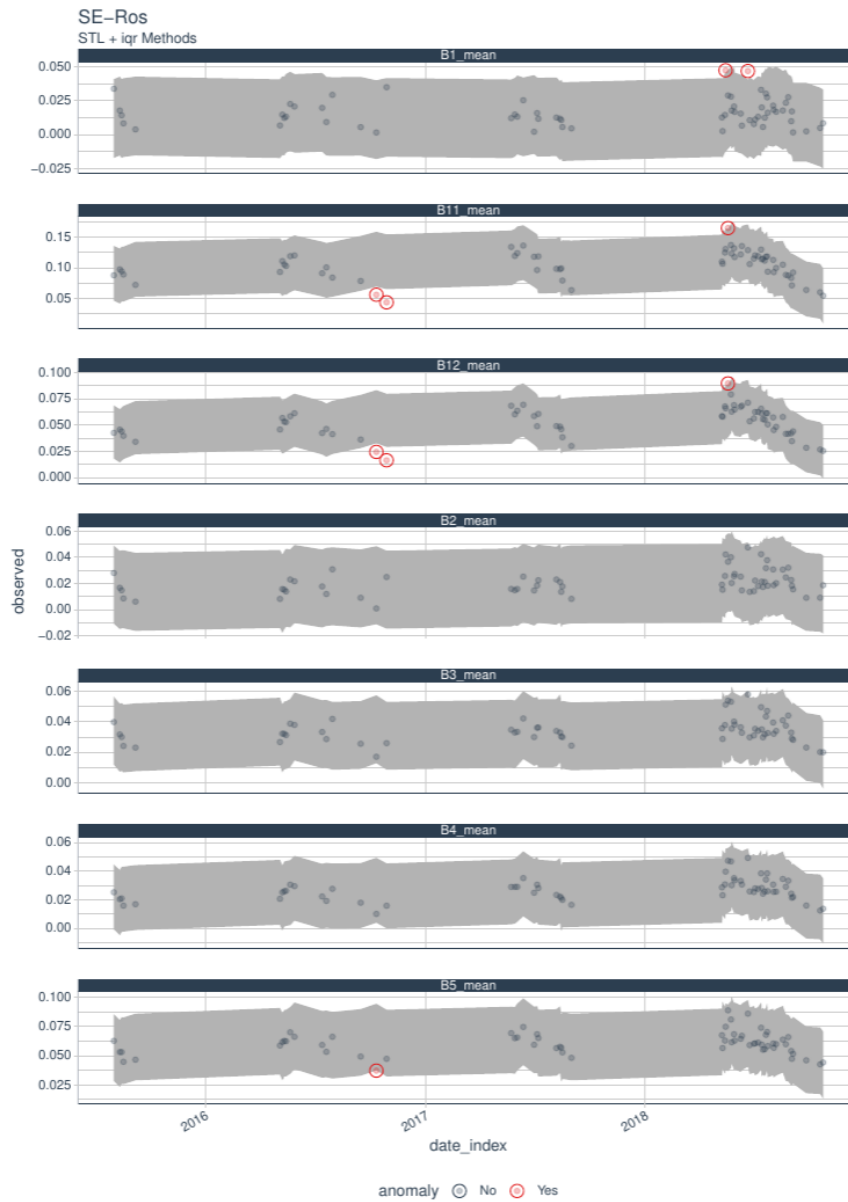


Figure 52: Outliers detection in Sentinel-2 images for the Rosinedal-3 (SE-Ros) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

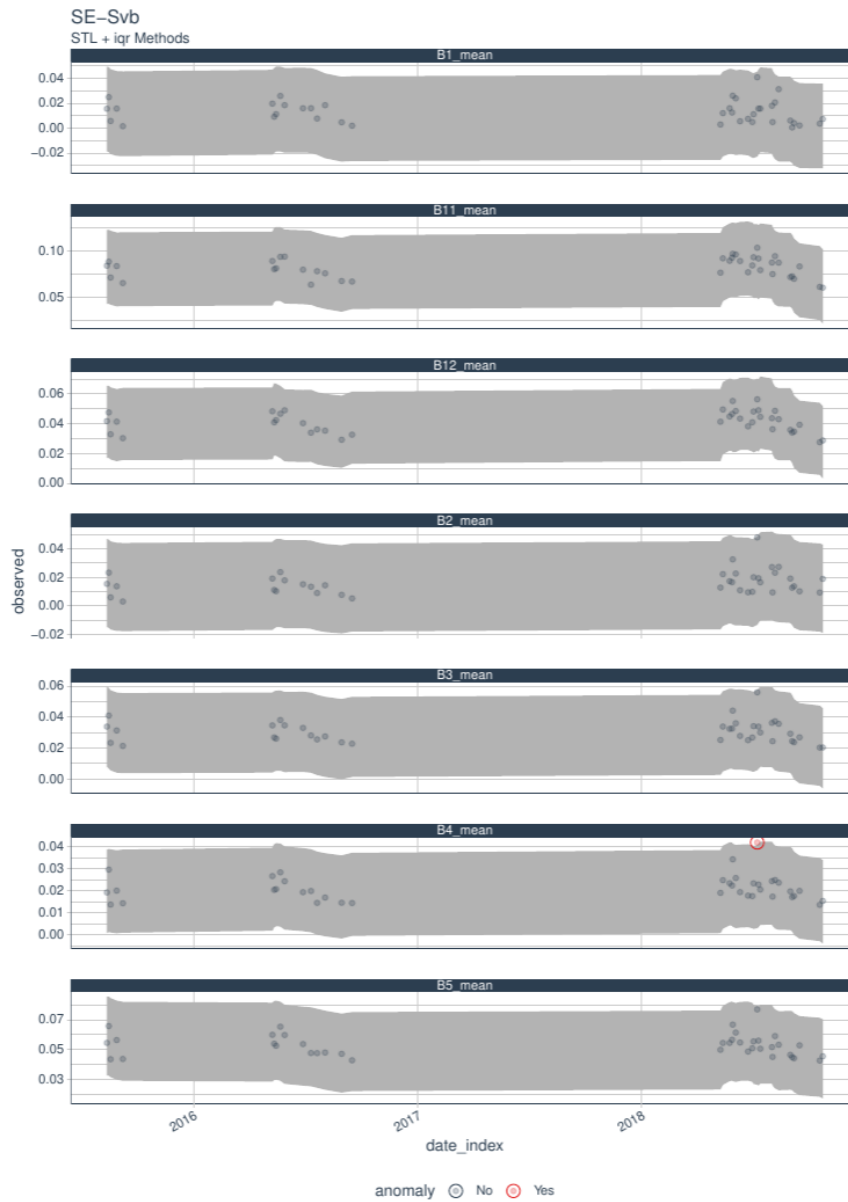


Figure 53: Outliers detection in Sentinel-2 images for the Svartberget (SE-Svb) ICOS site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

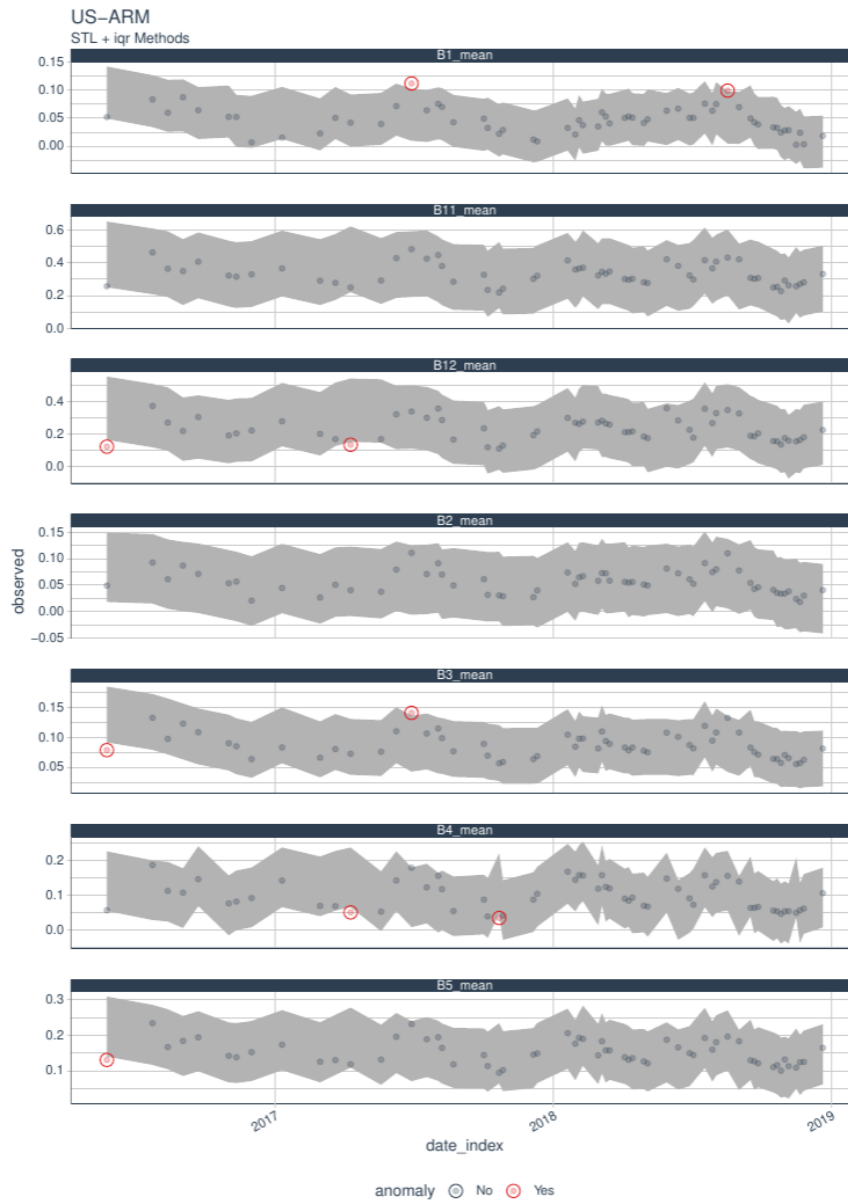


Figure 54: Outliers detection in Sentinel-2 images for the ARM Southern Great Plains site- Lamont (US-ARM) AmeriFlux site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

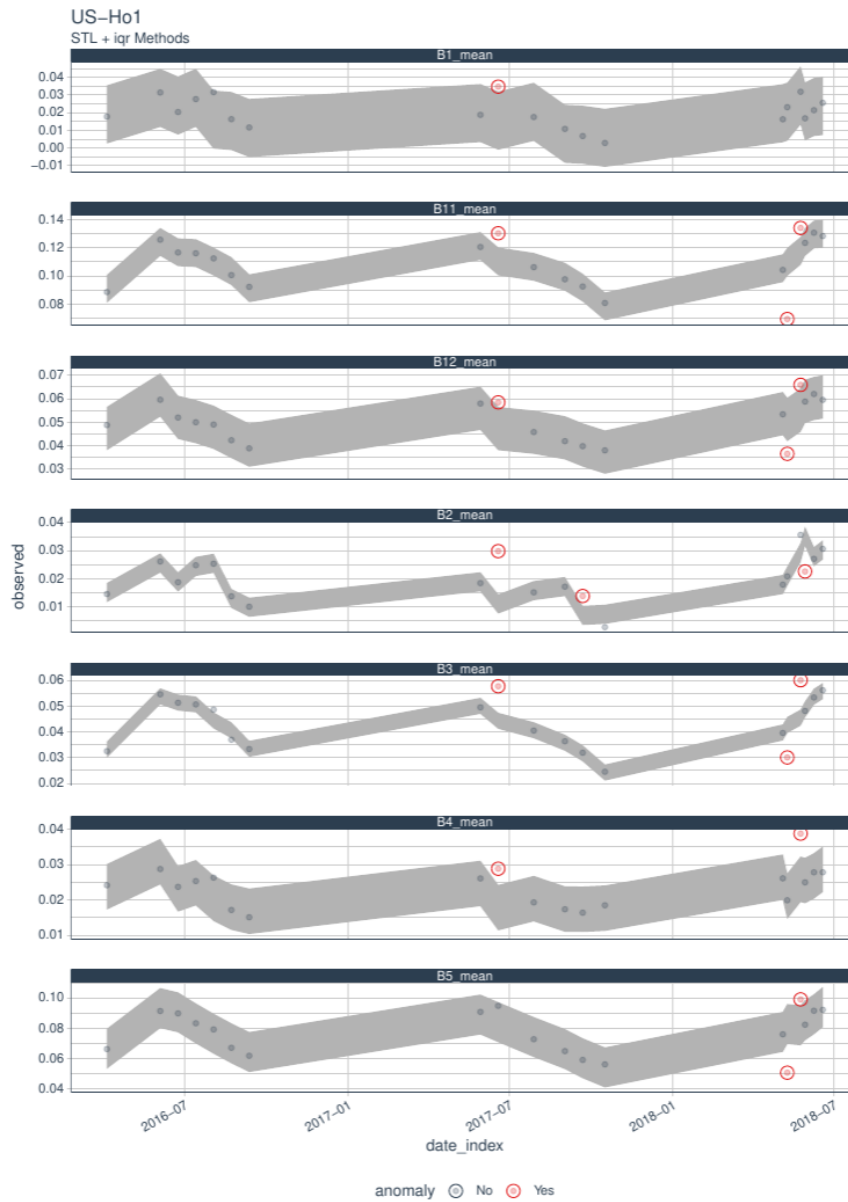


Figure 55: Outliers detection in Sentinel-2 images for the Howland Forest (main tower) (US-Ho1) AmeriFlux site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

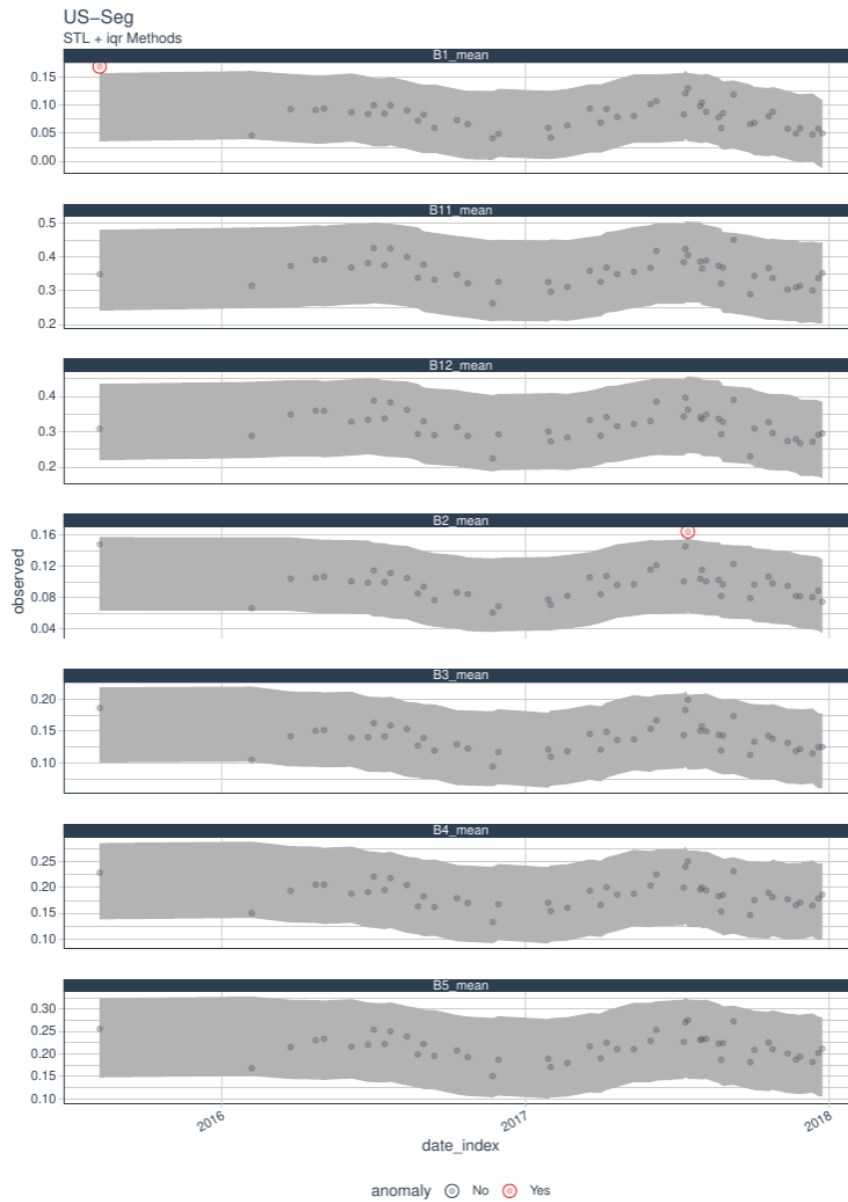


Figure 56: Outliers detection in Sentinel-2 images for the Sevilleta grassland (US-Seg) AmeriFlux site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

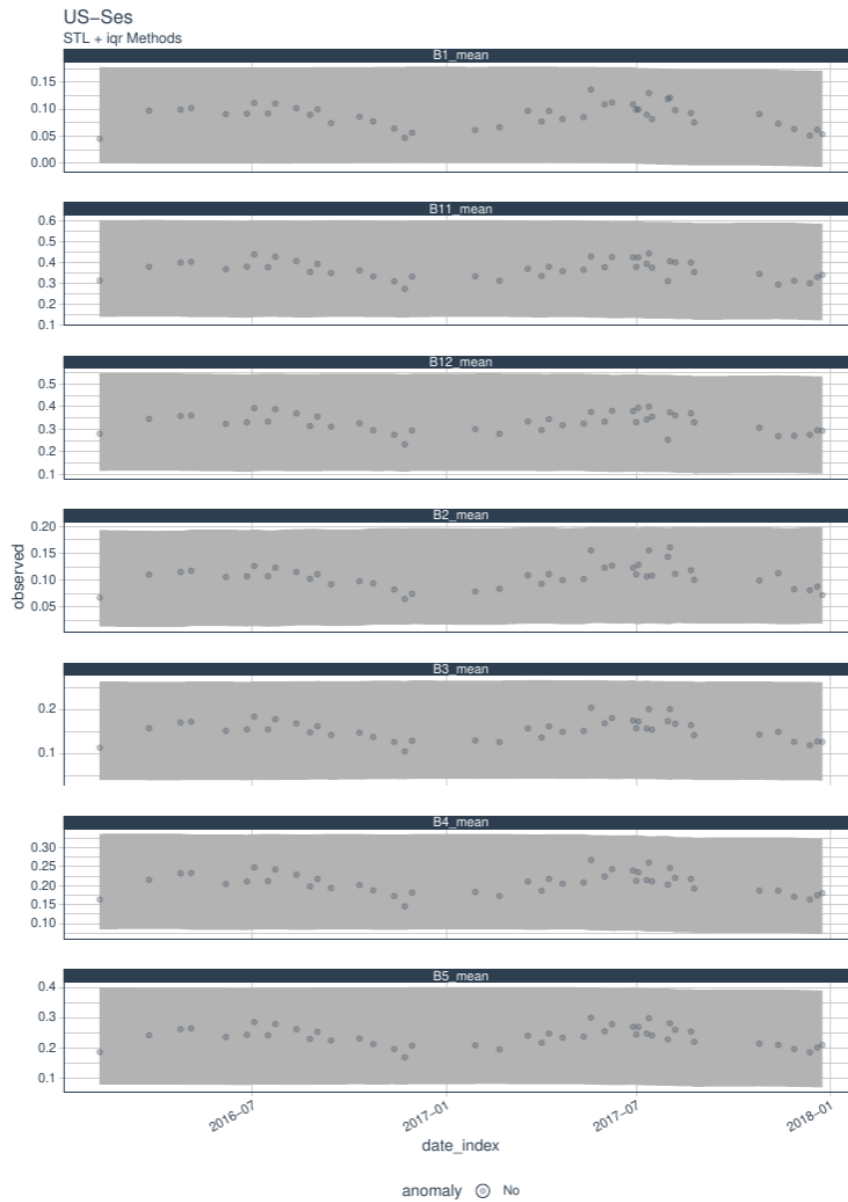


Figure 57: Outliers detection in Sentinel-2 images for the Sevillea shrubland (US-SeS) AmeriFlux site from 2015 to 2018. Bands B1, B11, B12, B2, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

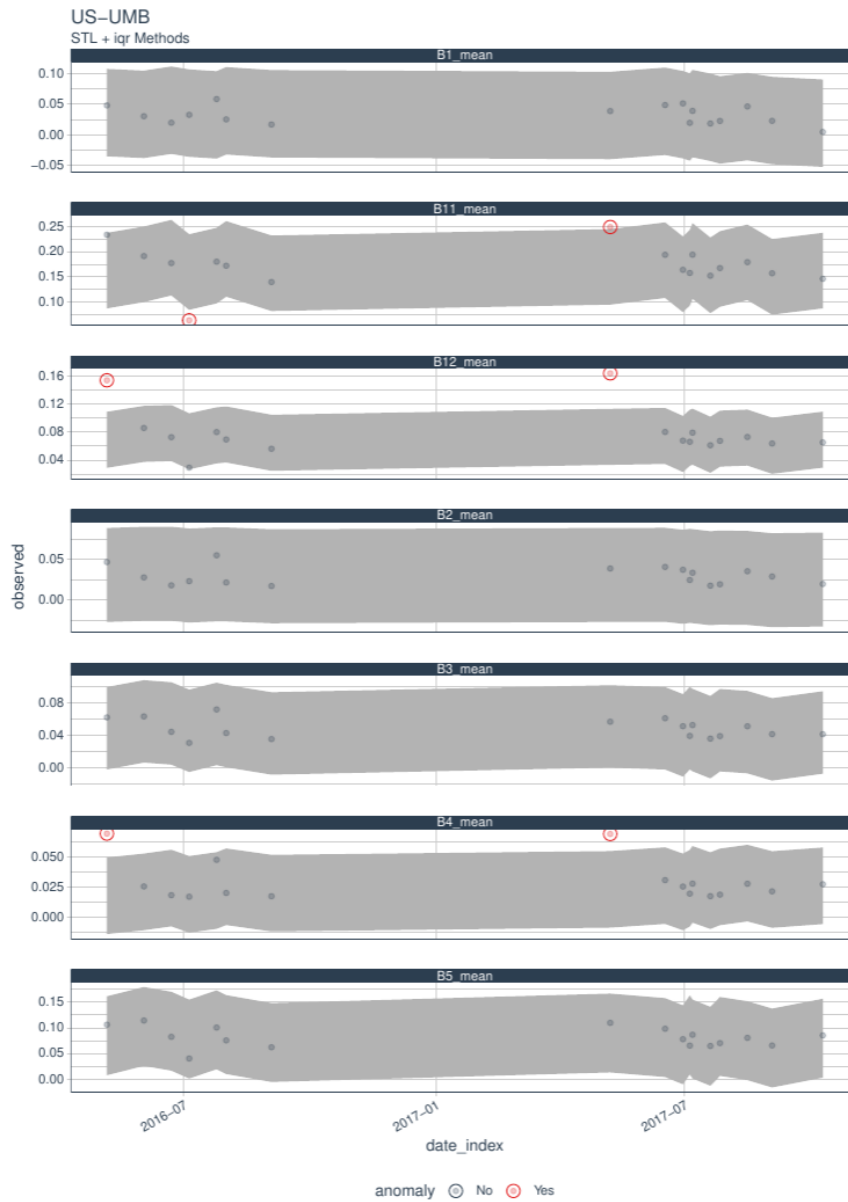


Figure 58: Outliers detection in Sentinel-2 images for the Univ. of Mich. Biological Station (US-UMB) AmeriFlux site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

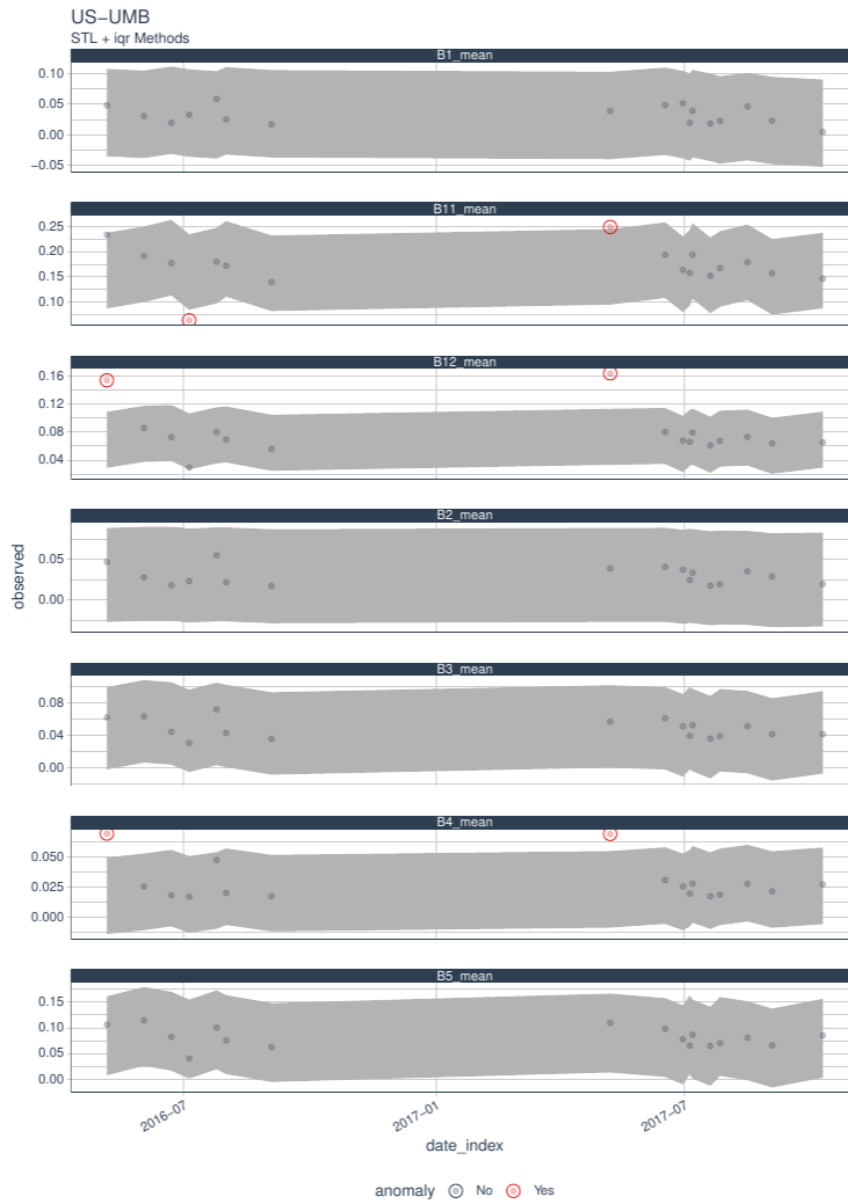


Figure 59: Outliers detection in Sentinel-2 images for the Willard Juniper Savannah (US-Wjs) AmeriFlux site from 2015 to 2018. Bands B1, B11, B12, B2, B3, B4, and B5 are shown. The gray ribbon represent the interquartile range used to classify outliers. Outliers are shown in red.

Supplement Material 4
Prediction of GPP using linear regressions, and balanced datasets

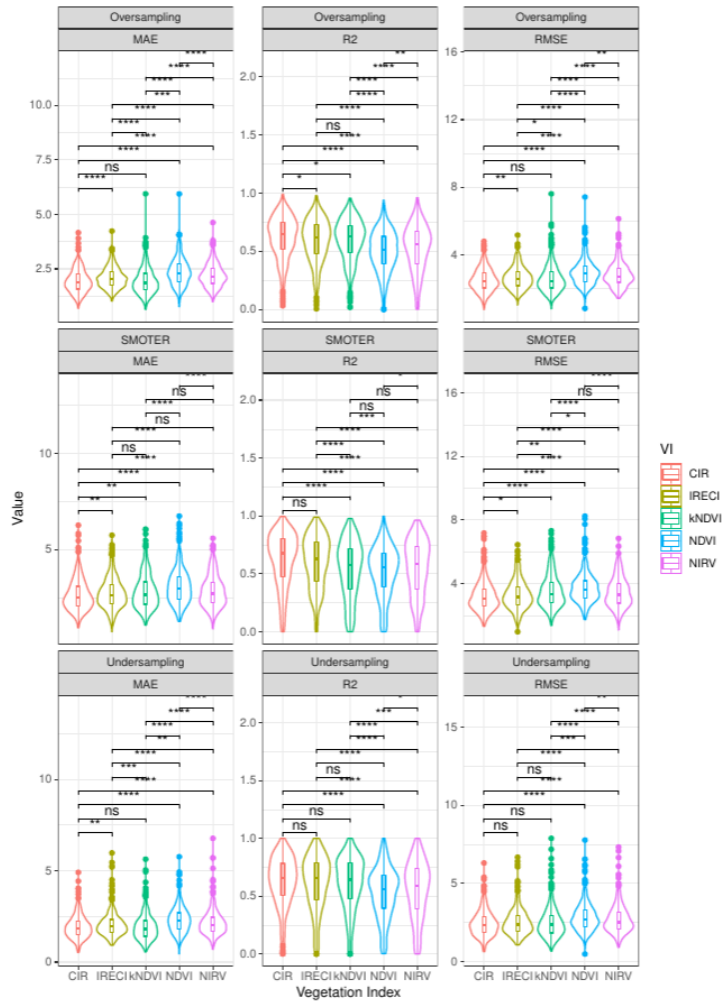


Figure 1: Prediction of GPP using linear regression and different vegetation indices as predictors. Each row of the plot show a balanced dataset. Each column correspond to the evaluation metric. The name of each vegetation index is presented in the X-axis. CIR: Chlorophyll index red. IRECI: Inverted Red-Edge Chlorophyll Index. NIRv: Near Infrared vegetation. NDVI: Normalized difference vegetation index. The Y-axis correspond to the value of each metric. The vertical lines correspond to the results of the Wilcoxon test in pairs, where ns: Non-significative differences; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$.

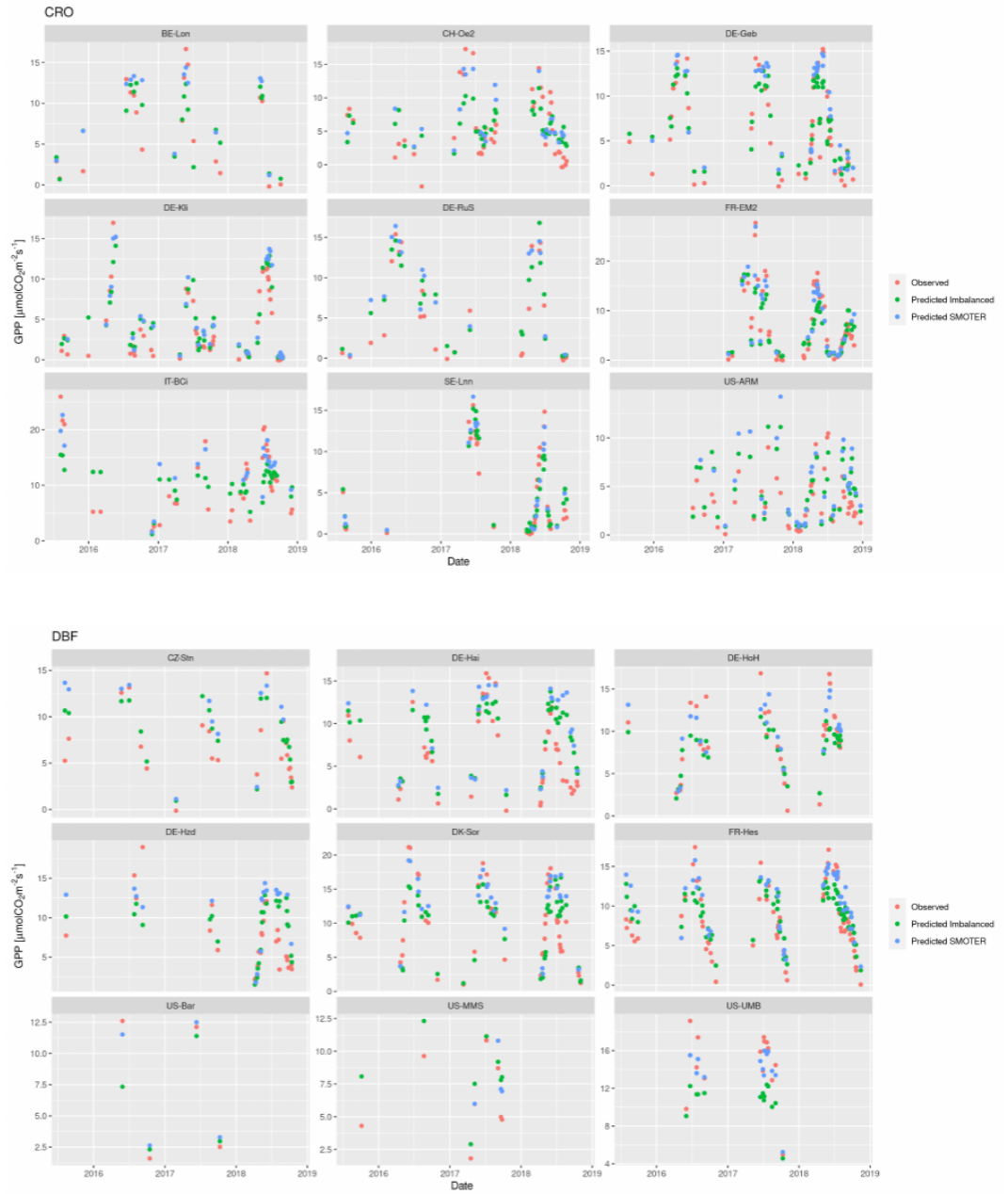
Supplement Material 5

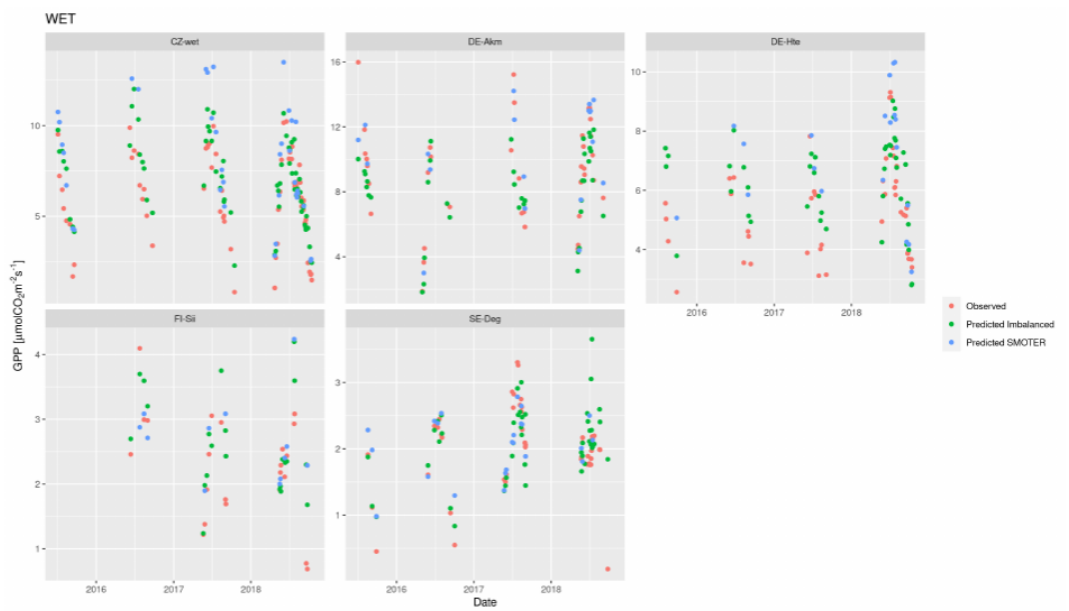
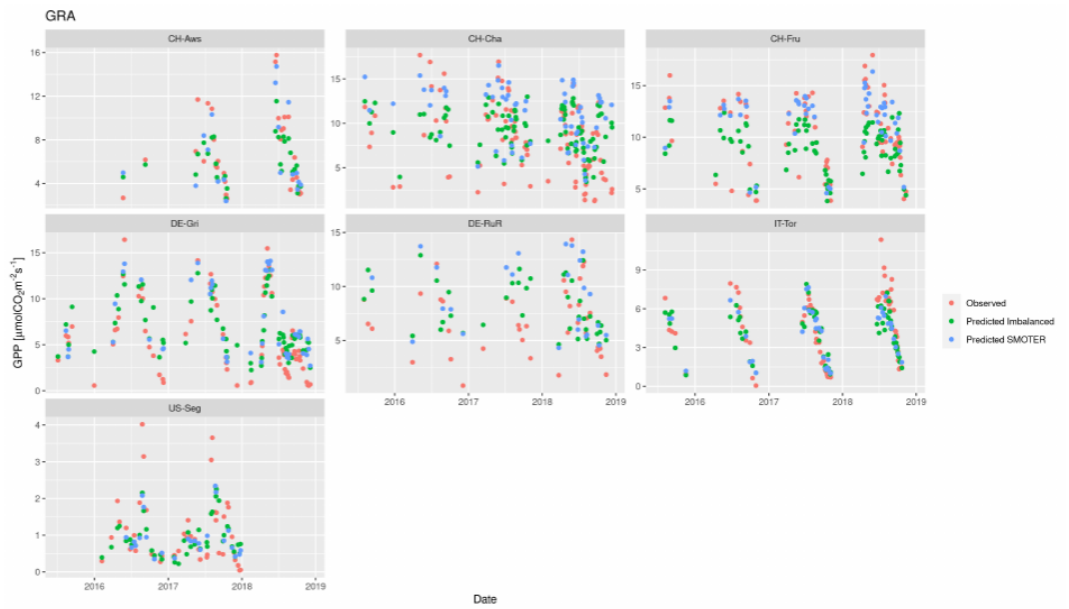
GPP prediction using random forest and 10-folds spatio-temporal cross validation

GPP prediction using random forest and 10-folds spatio-temporal cross validation. The plots are grouped by vegetation type. Evergreen needleleaf forests (ENF), Croplands (CRO), Deciduous broadleaf forests (DBF), Grasslands (GRA), Wetlands (WET), Mixed Forest (MF), Savannas (SAV), Open shrublands (OSH). The site name is presented in the upper part of each plot.

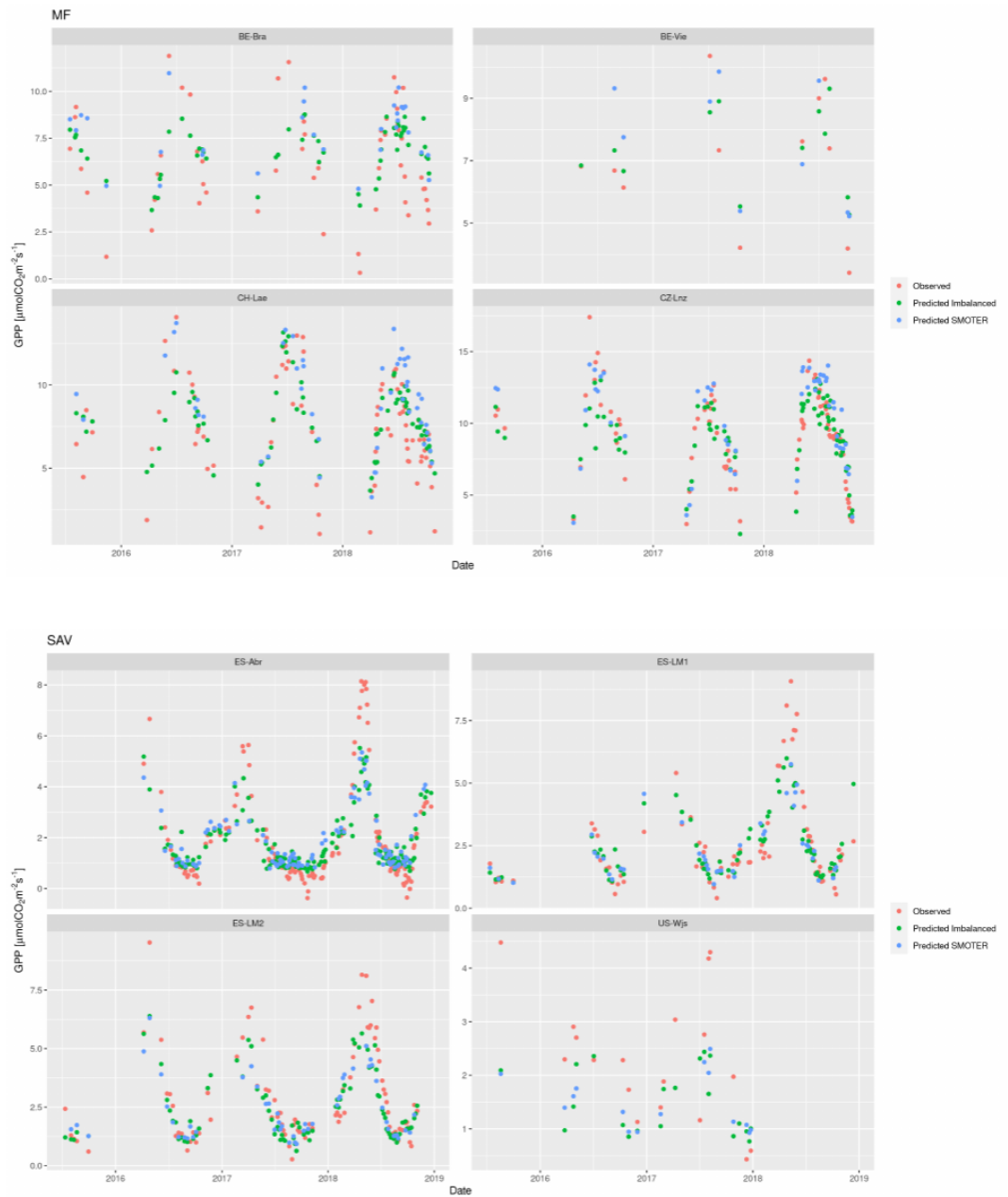


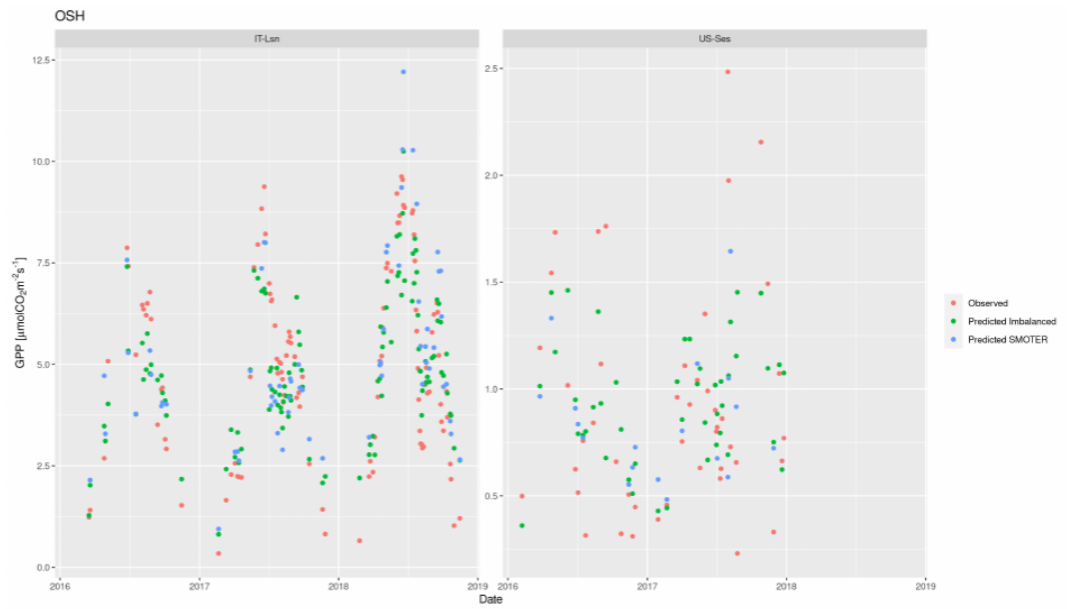
Appendix C. Supplementary information: On the Potential of Sentinel-2 for Estimating Gross Primary Production





Appendix C. Supplementary information: On the Potential of Sentinel-2 for Estimating Gross Primary Production





Supplement Material 8 Sentinel-2 Bands

SENTINEL-2 SPECTRAL BANDS

Band	Central Wavelength (μm)	Resolution (m)
B1	0.443	60
B2	0.490	10
B3	0.560	10
B4	0.665	10
B5	0.705	20
B6	0.740	20
B7	0.783	20
B8	0.842	10
B8A	0.865	20
B9	0.945	60
B11	1.610	20
B12	2.190	20

Supplement Material 9
Vegetation Indices

Vegetation indices calculated from Sentinel-2 images.

Vegetation index	Equation using Sentinel-2 bands	Reference
Atmospherically Resistant Vegetation Index (ARVI)	$ARVI = \frac{B8-rb}{B8+rb}$ $rb = B4 - gamma * (B2 - B4)$ $\gamma = 1$	[108]
Difference Vegetation Index (DVI)	$DVI = B8 - B4$	[109]
Global Environmental Monitoring Index (GEMI)	$GEMI = \frac{\eta(1-0.25\eta) - (B4-0.125)}{1-B4}, \eta = \frac{2(B8A^2 - B4^2)}{B8A + B4 + 0.5}$	[110]
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{B7-B3}{B7+B3}$	[111]
Infrared Percentage Vegetation Index (IPVI)	$IPVI = \frac{B8}{B8-B4}$	[112]
Inverted Red-Edge Chlorophyll Index (IRECI)	$IRECI = \frac{B7-B4}{\frac{B5}{B6}}$	[39]
Modified Chlorophyll Absorption Ratio Index (MCARI)	$MCARI = \frac{B5}{B4} [(B5 - B4) - 0.2 * (B5 - B3)]$	[113]
Second Modified Soil Adjusted Vegetation Index (MSAVI2)	$MSAVI2 = \frac{(2*B8+1) - \sqrt{(2*B8+1)^2 - 8*(B8-B4)}}{2}$	[114]
Modified Soil Adjusted Vegetation Index (MSAVI)	$MSAVI = \frac{(1+L) \frac{B8-B4}{B8+B4+L}}{s} = 1 - 2s(NDVI)(WDVI)$ $s = 0.5$	[114, p.199]
Meris Terrestrial Chlorophyll Index (MTCI)	$MTCI = \frac{B6-B5}{B5-B4}$	[115]
Normalized Difference Index 45 (NDI45)	$NDI45 = \frac{B5-B4}{B5+B4}$	[116]
Pigment Specific Simple Ratio (PSSRA)	$PSSRA = \frac{B7}{B4}$	[117]
Perpendicular Vegetation Index (PVI)	$PVI = \sin(\alpha)B8 - \cos(\alpha)B4$ $\alpha = 45$	[118]
Ratio Vegetation Index (RVI)	$RVI = \frac{B8}{B4}$	[119]
Sentinel-2 Red-Edge Position Index (S2REP)	$S2REP = 705 + 35 \frac{B7+B3 - B5}{B6 - B5}$	[120]
Soil Adjusted Vegetation Index (SAVI)	$SAVI = \frac{(1+L) \frac{B8-B4}{B8+B4+L}}{L} = 0.5$	[121]
Transformed Normalized Difference Vegetation Index (TNDVI)	$TNDVI = \sqrt{\frac{B8-B4}{B8+B4} + 0.5}$	[122]

Transformed Soil Adjusted Vegetation Index (TSAVI)	$TSAVI = s * \frac{B8 - sB4 - a}{sB8 + B4 - sa + X(1 + s^2)}$ $a = 0.5$ $s = 0.5$ $X = 0.08$	[123]
Weighted Difference Vegetation Index (WDVI)	$WDVI = B8 - gB4; g = 0.5$	[124]
Red-edge chlorophyll index (CIR)	$CIR = \frac{B7}{B5} - 1$	[125]
Green Chlorophyll index (CIG)	$CIG = \frac{B7}{B3} - 1$	[62, 126]
Normalized difference vegetation index (NDVI)	$NDVI = \frac{B8 - B4}{B8 + B4}$	[127]
Near-infrared vegetation index (NIRv)	$NIRv = \frac{B8 - B4}{B8 + B4} B8$	[32]
Kernel NDVI (KNDVI)	$kNDVI = \tanh\left(\left(\frac{B8 - B4}{2\sigma}\right)^2\right)$ $\sigma = 0.5(B8 + B4)$	[34]

Supplement Material 10

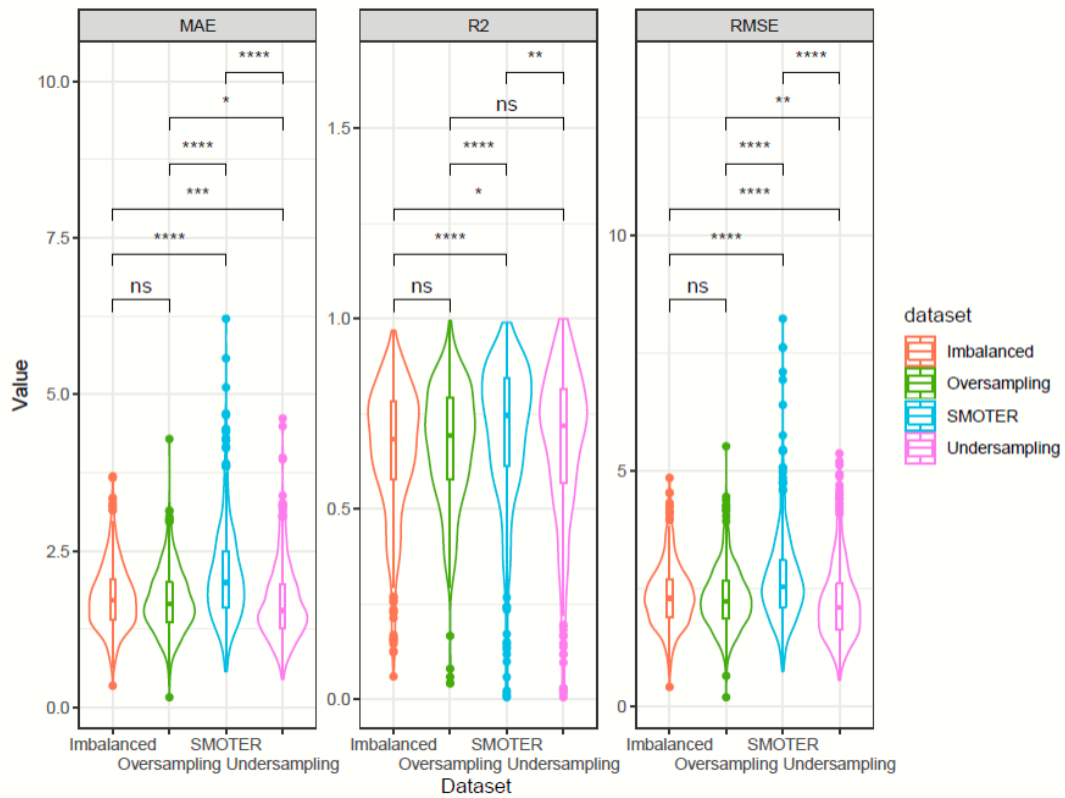


FIG. E1: Comparison of the GPP prediction using different imbalanced and balanced datasets. The y-axis corresponds to the value of each metric. The vertical lines correspond to the results of the Wilcoxon test in pairs, where ns is the nonsignificant differences, *: $p < 0.05$, and **: $p < 0.01$.

APPENDIX D

Declaration Forms

FORM 2⁴

Manuscript No. 1

Short reference Pabon-Moreno et al. (2022), In prep.

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles):

Figure(s) # 1:4	<input checked="" type="checkbox"/> 100% (the data presented in this figure come entirely from experimental work carried out by the candidate)
	<input type="checkbox"/> 0% (the data presented in this figure are based exclusively on the work of other co-authors)
	<input type="checkbox"/> Approximate contribution of the doctoral candidate to the figure: _____% Brief description of the contribution: (e.g. "Figure parts a, d and f" or "Evaluation of the data" etc.)

Signature candidate

Signature supervisor (member of the Faculty)

⁴ The signatures must be original only in the completed form to be submitted separately to the Dean's Office. The signatures and signature fields are not necessarily required in the version included in the dissertation.

FORM 2⁵

Manuscript No. 2

Short reference Pabon-Moreno et al. (2020), Biogeosciences.

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles):

Figure(s) # 1:7	<input checked="" type="checkbox"/> 100% (the data presented in this figure come entirely from experimental work carried out by the candidate)
	<input type="checkbox"/> 0% (the data presented in this figure are based exclusively on the work of other co-authors)
	<input type="checkbox"/> Approximate contribution of the doctoral candidate to the figure: _____% Brief description of the contribution: (e.g. "Figure parts a, d and f" or "Evaluation of the data" etc.)

Signature candidate

Signature supervisor (member of the Faculty)

⁵ The signatures must be original only in the completed form to be submitted separately to the Dean's Office. The signatures and signature fields are not necessarily required in the version included in the dissertation.

FORM 2⁶

Manuscript No. 3

Short reference Pabon-Moreno et al (2022), IEEE Trans Geosci Remote Sens.

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles):

Figure(s) # 1:5	<input checked="" type="checkbox"/> 100% (the data presented in this figure come entirely from experimental work carried out by the candidate)
	<input type="checkbox"/> 0% (the data presented in this figure are based exclusively on the work of other co-authors)
	<input type="checkbox"/> Approximate contribution of the doctoral candidate to the figure: _____% Brief description of the contribution: (e.g. "Figure parts a, d and f" or "Evaluation of the data" etc.)

Signature candidate

Signature supervisor (member of the Faculty)

⁶ The signatures must be original only in the completed form to be submitted separately to the Dean's Office. The signatures and signature fields are not necessarily required in the version included in the dissertation.

APPENDIX E

Curriculum Vitae

Daniel E. Pabon-Moreno

Personal Information

Name: Daniel Ernesto

Surname: Pabon Moreno

Year of birth: 1993

Nationality: Colombian

Institutional Email: dpabon@bgc-jena.mpg.de

Personal Email: daniel.pabon@protonmail.com

Github: dpabon

Education

2017 - PhD Student. Max Planck Institute For Biogeochemistry / Friedrich-Schiller-Universität Jena.

Department Biogeochemical Integration.

Research group: Empirical Inference of the Earth System.

Supervisors: Prof. Dr. Miguel Mahecha, Dr. Mirco Migliavacca, Prof. Dr. Christine Römermann, Prof. Dr. Markus Reichstein.

Early Stage Researcher. Training on Remote Sensing for Ecosystem Modelling. H2020-MSCA-ITN-2016. TRUSTEE Network.

2011 - 2017 Biologist. Industrial University of Santander. Bucaramanga, Colombia. (Official duration: 5 years).

Thesis: “Vulnerability of the Colombian vascular plant flora to elevational upslope shifts in a warming world”.

2005 - 2010 High School Graduate with Emphasis on Education. Escuela Normal Superior Francisco de Paula Santander. Málaga, Colombia.

Publications

- 2022 Pabon-Moreno, D. E.**, Migliavacca, M., Reichstein, M., & Mahecha, M. D. (2022). On the Potential of Sentinel-2 for Estimating Gross Primary Production. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 112. <https://doi.org/10.1109/TGRS.2022.3152272>
- 2021** Migliavacca, M., Musavi, T., Mahecha, M. D., Nelson, J. A., Knauer, J., Baldocchi, D. D., ..., **Pabon-Moreno, D. E.**, ..., Reichstein, M. (2021). The three major axes of terrestrial ecosystem function. *Nature*, 598(7881), 468472. <https://doi.org/10.1038/s41586-021-03939-9>
- 2021** Estupinan-Suarez, L. M., Gans, F., Brenning, A., Gutierrez-Velez, V. H., Londono, M. C., **Pabon-Moreno, D. E.**, Poveda, G., Reichstein, M., Reu, B., Sierra, C. A., Weber, U., & Mahecha, M. D. (2021). A Regional Earth System Data Lab for Understanding Ecosystem Dynamics: An Example from Tropical South America. *Frontiers in Earth Science*, 9. <https://www.frontiersin.org/articles/10.3389/feart.2021.613395>
- 2020 Pabon-Moreno, D. E.**, Musavi, T., Migliavacca, M., Reichstein, M., Römermann, C., & Mahecha, M. D. (2020). Ecosystem physio-phenology revealed using circular statistics. *Biogeosciences*, 17(15), 39914006. <https://doi.org/10.5194/bg-17-3991-2020>
- 2019** Ma, X., Mahecha, M. D., Migliavacca, M., van der Plas, F., Benavides, R., Ratcliffe, S., ..., **Pabon-Moreno, D. E.**, ..., Wirth, C. (2019). Inferring plant functional diversity from space: The potential of Sentinel-2. *Remote Sensing of Environment*, 233, 111368. <https://doi.org/10.1016/j.rse.2019.111368>

Complementary training

IMPRS-gBGC: International Max-Planck Research School for Global Biogeochemical Cycles.

TRUSTEE: Training on Remote Sensing for Ecosystem Modelling. H2020-MSCA-ITN-2016.

MPI-BGC: Max-Planck Institute for Biogeochemistry.

2021, January 15-22. Atmosphere, ocean and land. IMPRS-gBGC. Jena, Germany.

2020, January 23-24. Course on Transferable Skills. TRUSTEE Training School. Fondazione Edmund Mach. San Michele all'Adige (Trento), Italy.

2019, October 1 - November 30. Research stay at the University of Twente, Department of Water resources under the supervision of Dr. Christiaan van der Tol. Enschede, The Netherlands.

2019, Wintersemester. Statistical analysis of biological data. Friedrich-Schiller-Universität Jena. Jena, Germany.

2019, May 6-9. How to write a competitive proposal. TRUSTEE Training School. KU Leuven and VITO. Leuven, Belgium.

2019, May 2-3. Remote sensing for traits mapping in agriculture. TRUSTEE Training School. Forschungszentrum Jülich. Jülich, Germany

2019, March 27-28. Academic Writing. IMPRS-gBGC. Jena, Germany.

2019, February 21-22. How to use GPU. IMPRS-gBGC. Jena, Germany.

2019, January 25. Rules for good scientific practice. IMPRS-gBGC. Jena, Germany.

2019, January 16-18. The Julia Programming Language. IMPRS-gBGC. Jena, Germany.

-
- 2018, Wintersemester.** Integration of Remote Sensing, ecosystem observation and biogeochemical modelling. Friedrich-Schiller-Universität Jena. Jena, Germany.
- 2018, September 21.** Web Design for Landscape data dissemination and exploitation.
- 2018, September 12-20.** Training Course on Sentinel-2 for agriculture monitoring activities and remote sensing for plant traits and ecosystem functional properties. TRUSTEE Training School. University of Milano-Bicocca. Milan, Italy.
- 2018, September 3-7.** Earth Observation Techniques. IMPRS-gBGC. Jena, Germany.
- 2018, March 11-18.** Summer School on Field Experimental Design and Data Acquisition. MPI-BGC and Remote Sensing and Field Spectroscopy Laboratory, CSIC. Navalmoral de la Mata- Caceres, Spain.
- 2018, January 22-26.** Advanced statistics & machine learning for data analysis. IMPRS-gBGC. Jena, Germany.
- 2017, November 27 - December 6.** Biogeochemical Cycles in the Earth System - an Overview. IMPRS-gBGC. Jena, Germany.
- 2017, October 30 - November 2.** Drone field operations for environmental monitoring. TRUSTEE Training School. DroneLab. University of Exeter. Falmouth, UK.
- 2017, September 11-13.** Applied statistics & data analysis Course (1. Basic Statistics). IMPRS-gBGC. Jena, Germany.
- 2017, August 21 - September 6.** Terrestrial Biosphere Course. IMPRS-gBGC. Jena, Germany.

2017, June 19-30. Remote Sensing and Ecosystem Modelling. Open Software for Data Analysis. TRUSTEE Training School. University of Milano-Bicocca. Milan, Italy.

2016, September 26-30. Autumn School: Dynamics of natural (eco)systems: theory and applications. Michael Stifel Center Jena for Data-Driven and Simulation Science. Max Planck Institute for Biogeochemistry. Jena, Germany.

2015, June 17-19. Computer Enzymatic Catalysis Course. Theoretical Biochemistry Group. Industrial University of Santander. Bucaramanga, Colombia.

2014, November 24-28. Synthesis and Characterization of Antimicrobial Peptides. Biochemistry and Microbiology Research Group. Industrial University of Santander. Bucaramanga, Colombia.

Academic Events

2019, December 9-13. Fall meeting of the American Geophysical Union 2019. Poster presentation: "Explaining the timing of maximum GPP using circular statistics". San Francisco, USA.

2018, September 24-28. 10th International Conference on Ecological Informatics. Oral presentation: "Potential Gross Primary Production: A phenological approach". Jena, Germany.

2017, November 21-23. BACI Progress and review meeting. Presentation: "Global estimates and exploration of photosynthetic capacity". Talie Musavi & Daniel Pabon-Moreno. Jena, Germany.

2017, October 24-26 Potsdam Greenhouse Gas Workshop From Photosystems to Ecosystems. Assistant. Potsdam, Germany.

2015, March 26 I Symposium on Biotechnology and Nanobiotechnology. Assistant. Bucaramanga, Colombia.

2014, December 11-12 I International Symposium on Biotechnology and Agroindustry. Co-Organizer. Bucaramanga, Colombia.

2013, July 24-26 IV Symposium Colombian network of evolutionary biology. Assistant. Bogota, Colombia.

2012, September 5-7 II Seminar on Biodiversity and Conservation of Endangered Species. Assistant. Bucaramanga, Colombia.

Past Projects

- Pabon-Moreno, Daniel., Reu, Björn. (2017). Vulnerability of the Colombian vascular plant flora to elevational upslope shifts in a warming world. Thesis repository.
- Pimiento-Quiroga, Nataly., Pabon-Moreno, Daniel., Reu, Björn. (2016). Useful plants of Colombia: A data-base about the use of the Colombian tree flora. Manuscript in preparation.

Experience

2020 Lecturer. R course: Advanced modules. International Max Planck Research School for Global Biogeochemical Cycles. September 14-16, 2020.

2016 Web-Admin Socio-Ecological Landscapes Research Group. November 2016.

2014 Digitalization of biological collections database from the school of Biology of the Industrial University of Santander. August 2014.

Software

- **ecofunr** <https://github.com/dpabon/ecofunr>
An R package to derive **E**cosystem **F**unctional Properties.

APPENDIX F

Declaration of honor / Ehrenwörtliche Erklärung

Declaration of honor / Ehrenwörtliche Erklärung

I hereby confirm that I am familiar with the valid doctoral examination regulations; that I produced this doctoral thesis project myself and I did not use any text passage from third parties nor their own previous final theses without citing them; that I have cited the tools, personal communication, and sources that I have used; that I did not receive any assistance from specialised consultants and that any third party did not receive either direct or indirect financial benefits from me for work connected to the doctoral thesis submitted; and that I have not submitted this doctoral thesis project as the final thesis for a state examination or other scientific examination. I declare that all my supervisors: Prof. Dr. Miguel Mahecha, Dr. Mirco Migliavacca, Prof. Dr. Markus Reichstein, and Prof. Dr. Christine Römmermann have supported me in the writing of this document. Other coauthors, stated in the authorship list of each manuscript, have contributed with the writing of each manuscripts.

Hiermit erkläre ich: dass mir die Promotionsordnung der Fakultät bekannt ist; dass ich die Dissertation selbst angefertigt habe, keine Textabschnitte oder Ergebnisse eines Dritten oder eigenen Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen in meiner Arbeit angegeben habe; dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen; und dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts hat mich Prof. Dr. Miguel Mahecha, Dr. Mirco Migliavacca, Prof. Dr. Markus Reichstein, und Prof. Dr. Christine Römmermann. Andere Co-Autoren, die in der Autorenliste jedes Manuskripts angegeben sind, haben bei der Erstellung des jeweiligen Manuskripts mitgewirkt.

Jena, 14 November 2022

Daniel Ernesto Pabon Moreno