Assessing uncertainties related to the use of satellite remote sensing indices to estimate Gross Primary Production

By

Ronny A. Hernández Mora

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Earth and Atmospheric Sciences
University of Alberta

Abstract

Methods to quantify Gross Primary Production (GPP) are classified into two categories: Eddy Covariance techniques (EC) and satellite data-driven. EC techniques can measure carbon fluxes directly, albeit with spatial constraints. Satellite data-driven methods are promising because they overcome spatial constraints associated with EC techniques. However, satellite-driven products have potentially greater uncertainty than EC methods for GPP estimation such as mixed pixels, cloud cover, and the ability of the sensor to retrieve vegetation under saturation conditions in high biomass environments. Therefore, an effort to analyze and quantify the uncertainty of GPP products derived from satellite platforms is needed. This study quantifies the uncertainty of commonly used satellite vegetation indices such as Normalized Difference Vegetation Index (NDVI), Enhance Vegetation Index (EVI), Chlorophyll/Carotenoid Index (CCI), and Near-Infrared Reflectance Index (NIRv) for GPP estimation compared with direct methods such as EC measurements. We conduct this study on three different sites: the University of Michigan Biological Station (USA), the Borden Forest Research Station flux-site (Canada), and Bartlett Experimental Forest (USA) using traditional regression methods and ML approaches.

Preface

This thesis is an original work by Ronny A. Hernández Mora. No part of this thesis has been previously published.

Dedication

If you can keep your leaves when all about you, are losing theirs and blaming it on drought; If you can photosynthesize when others shut their stomates and stop transpiring; yours will be the canopy and all beneath it and -which is more-you'll be a tree, my seed.

– Jennifer L. Funk digression from Rudyard Kipling "If" poem

"No one gets anywhere without the help of someone else."

– Franklin Chang Díaz

Acknowledgements

Firstly, I would like to express my gratitude to Arturo Sánchez for providing me with the opportunity to contribute to the Tropi-dry project in Santa Rosa National Park, Costa Rica. Climbing the towers and delving into the remote sensing field for the first time was both an enjoyable and enriching experience. I am especially thankful for the encouragement to pursue graduate studies and the invaluable assistance in facilitating my journey to Canada, allowing me to become a part of the CEOS lab. Also, I want to thank Richard Fernandes whose guidance has been instrumental in pushing me further and navigating through a research area that has become truly fascinating to me. I appreciate your patience in answering my questions and explaining several concepts. Your mentorship has been incredibly helpful to me.

Also, thanks to the many individuals who supported me before and during this journey. To my friend J. Antonio Guzmán Q. who helped me a lot in my transition from Costa Rica to Canada and for all his advice in an academic career path. To my ixpantia colleagues, who generously took the time to assist me when I encountered coding puzzles, and especially to my friend Frans Van Dunné who took the time to not just explain me technical programming concepts but also took the time for advising me on career path decisions and life. Your support was invaluable. To my CEOS lab partners, thank you for helping me navigate the various aspects of graduate studies and for all your constructive feedback on my research.

Finally, heartfelt thanks to my family, who believe in me more than I do. To María, a constant source of happiness and support despite the distance, navigating with me through all the ups and downs. To my mother, whom I strive to emulate in many aspects of my life. Your resilience in facing life, even in challenging moments, is truly admirable. Thank you for taking care of me, even when I knew you were carrying so much on your shoulders. Special thanks to my grandmother, Abi, who always encouraged me to study. She knew that, regardless of circumstances, education was our path to a better quality of life.

My Master's culmination was made financially possible, thanks to the Canadian Centre for Remote Sensing through its Research Affiliate Program Bursary Agreement, and the Earth and Atmospheric Sciences Department, which supported me through teaching assistantships. Addition- ally, I had the opportunity to participate in the American Geophysical Union Fall Meeting 2022 to present my research, thanks to funding provided by the University of Alberta through the Mary Louise Imrie Graduate Student Award and the Graduate Students' Association Academic Travel Grant.

Table of contents

List of tables

List of figures

List of Abbreviations

Abbreviation	Full phrase
AIC	Akaike Information Criterion
APAR	Absorbed Photosynthetically Active Radiation
BIC	Bayesian Information Criterion
CCI	Chlorophyll/Carotenoid Index
CRI	Chlorophyll Index Red
DT	Day Time method
EC	Eddy Covariance
EO	Earth Observation
EVI	Enhanced Vegetation Index
GAM	Generalized Additive Model
GEE	Google Earth Engine
GPP	Gross Primary Production
LAI	Leaf Area Index
LM	Linear Model
LUE	Light Use Efficiency
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NEE	Net Ecosystem Exchange
NIR	Near Infrared
NIRv	Near-Infrared Reflectance Index
PAR	Photosynthetically Active Radiation
QA	Quality Assurance
QC	Quality Control
RECO	Ecosystem Respiration

RMSE	root mean square error
SIF	Solar-Induced Fluorescence
SWIR	Short wave infrared
VI	Vegetation Index
fAPAR	fraction of Photosynthetically Active Radiation

1. Introduction

The study and assessment of ecosystem dynamics has greatly benefited from the development and advances of the remote sensing field (Running et al. 2004; Baldocchi 2020a) coupled with the growing freely available remote sensing data (Xiao et al. 2019). Elements on the Earth's surface exhibit distinctive spectral signatures arising from their physical and chemical properties and interactions with electromagnetic radiation (Montero et al. 2023). Environmental factors (fires, floods, drought, etc) can affect these interactions modifying the spectral signatures which can provide insights about surface processes when measured by remote sensing instruments (Montero et al. 2023).

The launch in the 1970s of the Earth Resources Technology Satellite (ERTS), later known as Landsat-1 (Donato 1997), initiated a new era in the remote sensing field of the terrestrial biosphere, with a special focus on vegetation monitoring (Montero et al. 2023). With information available about spectral bands, studies about their relations to biophysical characteristics became an object of study to estimate biomass and photosynthetic activity (Myneni et al. 1997).

In fact, one application of remote sensing has been the development of vegetation indices (VI) to study the characteristics of vegetation canopies and their functioning (Houborg, Fisher, and Skidmore 2015). VIs are a function of two or more spectral bands, typically derived from satellite images (Zeng et al. 2022) that can contain information about the status of the vegetation, biochemical characteristics, and structural characteristics such as Leaf Area Index (LAI), among others (Huete et al. 2002).

Early studies about spectral bands relations, led to the development of the Normalized Difference Vegetation Index (NDVI), which showed how the Near Infrared (NIR) and Red ratio had advantages over other bands ratios for monitoring vegetation biomass and structure given surface reflectance measurements in the shortwave spectrum (Tucker 1979; A. J. Richardson and Wiegand 1977). The basis of NDVI rests in the fact that chlorophyll absorption leads to

a decrease in red reflectance with increasing green vegetation relative to the NIR reflectance, that is minimally impacted by chlorophyll absorption (Huete et al. 2010).

The use of a relative indicator serves to reduce the effect of factors such as canopy spatial structure and non-photosynthetic vegetation on relationships between NDVI and biophysical variables such as biomass or absorbed photosynthetically active radiation variances (Sellers 1987; Huete et al. 2010). Nonetheless, NDVI is sensitive to soil background and saturation at high biomass (Huete et al. 2002). To overcome this problem, the Enhance Vegetation Index (EVI) was designed to improve vegetation characterizations such as canopy greenness without soil background sensitivity and aerosol variations by combining the blue, red, and NIR bands (Huete 1988). EVI is also less prone to saturate at high biomass instances (Gao 2000).

The ability to combine reflectance from multiple spectral bands, coupled with advancements and the development of new sensor technologies introducing additional bands or different spectral characteristics, has created opportunities to develop multiple VIs and the total number of indices has grown steadily (Zeng et al. 2022). These equations can function as proxies for various vegetation canopy properties or enhance existing estimations across different ecosystems (Huete et al. 2010).

According to the Awesome Spectral Indices open community catalogue (version 4.0.0), there are 127 spectral indices designed to monitor vegetation. The aim behind the ongoing development of new VIs is to achieve a more accurate mapping of ecosystem variables such as GPP (Montero et al. 2023). These indices can be classified into 3 main categories: biophysical, biochemical and physiological properties (Zeng et al. 2022).

One important variable in understanding ecosystem processes is GPP. At present, VIs derived from satellites cannot directly estimate GPP (Gensheimer et al. 2022) and instead, instruments like eddy covariance have been the traditional method to estimate GPP more directly (Baldocchi 2020a).

GPP represents the total amount of carbon compounds produced by plant photosynthesis in a given period (Ashton et al. 2012). GPP plays a crucial role in the carbon cycle since it quantifies the ability of vegetation to fix carbon from the atmosphere using only solar energy and nutrients (Xiao et al. 2019). GPP is responsive to a range of factors, including abiotic elements such as radiation, temperature, and precipitation (Beer et al. 2010), as well as biotic factors like vegetation type, leaf chemical traits, and species composition (Musavi et al. 2017).

Some of the carbon taken in by GPP is released back to the atmosphere through plant respiration (autotrophic respiration), and the difference between GPP and plant respiration is called Net Primary Productivity (NPP) (Lieth 1975; Xiao et al. 2019). GPP, along with Ecosystem Respiration (ER), which includes both autotrophic and heterotrophic respiration, together determines Net Ecosystem Exchange (NEE) (Xiao et al. 2019). NEE is fundamental to the planetary carbon cycle as it represents the net accumulation or loss of carbon by the ecosystem. Measurement of GPP and NEE is critical to understanding the role of the biosphere in the carbon cycle as well as the status and trends of vegetation productivity (Schimel 1995).

Eddy Covariance (EC) is a method used to measure the exchange of energy and materials between ecosystems and the atmosphere within an area that can range from a few hundred meters to a few kilometers (Pabon-Moreno et al. 2022). Specifically, EC methods can estimate GPP in-situ at a temporal resolution of seconds, by subtracting estimates of respiration from chamber measurements or models from direct measurements of NEE (Reichstein et al. 2005). However, EC estimates are both, costly and limited in spatial coverage (Baldocchi 2020a). Also, uncertainties from this method lie in the typical increase in spatial and temporal variability in GPP over heterogeneous landscapes such as forest boundaries (Reinmann and Hutyra 2017). The prevalent practice of taking measurements primarily from intact forests introduces the risk of potentially misleading estimations for the broader forest being quantified, given the potential differences in dynamics between edge and interior environments (Smith et al. 2018).

Since direct GPP measurements from satellites are not presently available, one of the primary objectives of remote sensing is to enhance the spatial coverage of EC GPP estimates (Xiao et al. 2019). In this context, EC values prove invaluable as information for calibrating VIs values with GPP inferred from flux towers (Huete et al. 2010). The utilization of EC measurements bridges the gap in GPP assessment, providing a practical means to evaluate and refine the reliability of remotely sensed data against ground-based observations (Xiao et al. 2019).

Various methods exist for estimating GPP through satellite measurements. Satellite-derived VIs serve as commonly employed proxies for GPP, whereas quantifying GPP more directly often involves the use of Light Use Efficiency (LUE) models, and process-based models (Xiao et al. 2019). These diverse methodologies cater to different aspects of GPP estimation, with satellite-derived VIs offering a convenient indirect assessment, while LUE models, and

process-based models provide more detailed and direct quantification (Lin et al. 2021).

The concept of a more detailed and direct quantification of GPP achieved through LUE and process-based models can be attributed to the incorporation of a constant maximal light use efficiency, which is a plant functional type dependent parameter that is downregulated by stress factors such as temperature and vapor pressure deficit (Pabon-Moreno et al. 2022). This approach is then integrated with APAR derived from remote sensing data to estimate GPP in the case of LUE models (Running et al. 2004). In contrast, process-based models distinguish themselves by prioritizing a mechanistic description of the photosynthetic biochemical processes at both leaf and canopy scales, incorporating the comprehensive theoretical underpinnings of photosynthesis (Ryu et al. 2011). This methodology not only refines GPP quantification but also offers a more in-depth understanding of the underlying physiological processes governing vegetation productivity.

However, both Light Use Efficiency (LUE) models and process-based models demand a multitude of inputs, some of which may not be readily available for all sites. In such cases, a data-driven approach becomes particularly advantageous, and VIs serve as effective proxies for GPP estimation. Notably, emerging VIs like the Near-Infrared Reflectance Index (NIRv) demonstrate a capability to account for 68% of the monthly variability in GPP across 105 fluxnet sites (Badgley et al. 2019). Additionally, indices like CCI, a pigment-based index designed to capture carotenoid/chlorophyll ratios during seasonal photosynthetic activity in evergreen leaves (Gamon et al. 2016), have demonstrated proficiency as estimators of GPP in certain forest types. These indices offer valuable insights into the dynamics of vegetation productivity.

Despite the notable progress in VIs development and their user-friendly applications there is a consensus that no single VI can be universally applied to address every GPP estimation across species and periods, even for a single biome (Zeng et al. 2022). VI measurement error poses additional challenges to remote sensing applications, with diverse sources affecting satellite-derived data, including atmospheric effects, retrieval errors, cloud contamination, and sensor degradation (Van Leeuwen et al. 2006; Fang et al. 2012). Thus, a quantitative and rigorous understanding of the specific context and characteristics of multiple sites is essential to understanding the potential and limits of VIs for GPP estimation over large extents of space and time (Zeng et al. 2022).

Together with the advancement of new VIs, there has been an upsurge in data analysis meth-

ods, specifically Machine Learning (ML) approaches, has garnered significance in recent years. The capability of ML methods to navigate non-linear relationships proves crucial, particularly when predicting GPP from various predictors (Meyer et al. 2019). However, despite their simplicity and effectiveness, ML approaches encounter challenges, and the resulting fluxes estimates bear various sources of uncertainty. The distribution and quantity of EC sites play a pivotal role in influencing the accuracy and variability of regional flux estimates derived from machine learning methods (Papale et al. 2015).

The goal of this study is to assess the advantages and constraints associated with the utilization of 4 widely employed VIs for estimating GPP. The study is limited to temperate broadleaf forests of the Eastern United States of America and Canada as it required consistent multi-annual EC GPP measurements over multiple sites to quantify uncertainty related to the spatial location of calibration data. The VIs are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS), which has long-standing temporal records and acceptable spatial resolution to study extended forest areas. The analysis employs both traditional regression methods and ML techniques to evaluate the performance of these indices in estimating GPP. This methodological approach seeks to provide a nuanced understanding of the efficacy and limitations of the selected indices in capturing the complexities of ecosystem processes, particularly in the context of temperate broadleaf forests.

1.1. References

Ashton, Mark S., Mary L. Tyrrell, Deborah Spalding, and Bradford Gentry, eds. 2012. *Managing Forest Carbon in a Changing Climate*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2232-3.

Badgley, Grayson, Leander D. L. Anderegg, Joseph A. Berry, and Christopher B. Field. 2019. "Terrestrial Gross Primary Production: Using NIR v to Scale from Site to Globe." *Global Change Biology* 25 (11): 3731–40. https://doi.org/10.1111/gcb.14729.

Baldocchi, Dennis D. 2020. "How Eddy Covariance Flux Measurements Have Contributed to Our Understanding of *Global Change Biology*." *Global Change Biology* 26 (1): 242–60. https://doi.org/10.1111/gcb.14807.

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, et al. 2010. "Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate." *Science* 329 (5993): 834–38. https://doi.org/10.1126/science.1184984.

Donato, David. 1997. "Landsat Data - A Brief History of the Landsat Program," December.

Fang, Hongliang, Shanshan Wei, Chongya Jiang, and Klaus Scipal. 2012. "Theoretical Uncertainty Analysis of Global MODIS, CYCLOPES, and GLOBCARBON LAI Products Using a Triple Collocation Method." *Remote Sensing of Environment* 124 (September): 610–21. https://doi.org/10.1016/j.rse.2012.06.013.

Gamon, John A., K. Fred Huemmrich, Christopher Y. S. Wong, Ingo Ensminger, Steven Garrity, David Y. Hollinger, Asko Noormets, and Josep Peñuelas. 2016. "A Remotely Sensed Pigment Index Reveals Photosynthetic Phenology in Evergreen Conifers." *Proceedings of the National Academy of Sciences* 113 (46): 13087–92. https://doi.org/10.1073/pnas.1606162113.

Gao, X. 2000. "Optical–Biophysical Relationships of Vegetation Spectra Without Background Contamination." *Remote Sensing of Environment* 74 (3): 609–20. https://doi.org/10.1016/S0034-4257(00)00150-4.

Gensheimer, Johannes, Alexander J. Turner, Philipp Köhler, Christian Frankenberg, and Jia Chen. 2022. "A Convolutional Neural Network for Spatial Downscaling of Satellite-Based Solar-Induced Chlorophyll Fluorescence (SIFnet)." *Biogeosciences* 19 (6): 1777–93. https://doi.org/10.5194/bg-19-1777-2022.

Houborg, Rasmus, Joshua B. Fisher, and Andrew K. Skidmore. 2015. "Advances in Remote Sensing of Vegetation Function and Traits." *International Journal of Applied Earth Ob-*

servation and Geoinformation 43 (December): 1–6. https://doi.org/10.1016/j.jag.2015.06.001.

Huete, Alfredo. 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25 (3): 295–309.

Huete, Alfredo, Kamel Didan, Willem Van Leeuwen, Tomoaki Miura, and Ed Glenn. 2010. "MODIS Vegetation Indices." In *Land Remote Sensing and Global Environmental Change*, edited by Bhaskar Ramachandran, Christopher O. Justice, and Michael J. Abrams, 11:579–602. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-6749-7_26.

Huete, Alfredo, K Didan, T Miura, E. P Rodriguez, X Gao, and L. G Ferreira. 2002. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." *Remote Sensing of Environment* 83 (1-2): 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

Lieth, Helmut. 1975. "Modeling the Primary Productivity of the World." *Primary Productivity of the Biosphere*, 237–63.

Lin, Shangrong, Jing Li, Qinhuo Liu, Beniamino Gioli, Eugenie Paul-Limoges, Nina Buchmann, Mana Gharun, et al. 2021. "Improved Global Estimations of Gross Primary Productivity of Natural Vegetation Types by Incorporating Plant Functional Type." *International Journal of Applied Earth Observation and Geoinformation* 100 (August): 102328. https://doi.org/10.1016/j.jag.2021.102328.

Meyer, Hanna, Christoph Reudenbach, Stephan Wöllauer, and Thomas Nauss. 2019. "Importance of Spatial Predictor Variable Selection in Machine Learning Applications – Moving from Data Reproduction to Spatial Prediction." *Ecological Modelling* 411 (November): 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815.

Montero, David, César Aybar, Miguel D. Mahecha, Francesco Martinuzzi, Maximilian Söchting, and Sebastian Wieneke. 2023. "A Standardized Catalogue of Spectral Indices to Advance the Use of Remote Sensing in Earth System Research." *Scientific Data* 10 (1): 197. https://doi.org/10.1038/s41597-023-02096-0.

Musavi, Talie, Mirco Migliavacca, Markus Reichstein, Jens Kattge, Christian Wirth, T. Andrew Black, Ivan Janssens, et al. 2017. "Stand Age and Species Richness Dampen Interannual Variation of Ecosystem-Level Photosynthetic Capacity." *Nature Ecology & Evolution* 1 (2): 0048. https://doi.org/10.1038/s41559-016-0048.

Myneni, R. B., R. Ramakrishna, R. Nemani, and S. W. Running. 1997. "Estimation of Global

Leaf Area Index and Absorbed Par Using Radiative Transfer Models." *IEEE Transactions on Geoscience and Remote Sensing* 35 (6): 1380–93. https://doi.org/10.1109/36.649788.

Pabon-Moreno, Daniel E., Mirco Migliavacca, Markus Reichstein, and Miguel D. Mahecha. 2022. "On the Potential of Sentinel-2 for Estimating Gross Primary Production." *IEEE Transactions on Geoscience and Remote Sensing*, 1–1. https://doi.org/10.1109/TGRS.2022.3152272.

Papale, Dario, T. Andrew Black, Nuno Carvalhais, Alessandro Cescatti, Jiquan Chen, Martin Jung, Gerard Kiely, et al. 2015. "Effect of Spatial Sampling from European Flux Towers for Estimating Carbon and Water Fluxes with Artificial Neural Networks." *Journal of Geophysical Research: Biogeosciences* 120 (10): 1941–57. https://doi.org/10.1002/2015JG002997.

Reichstein, Markus, Eva Falge, Dennis Baldocchi, Dario Papale, Marc Aubinet, Paul Berbigier, Christian Bernhofer, et al. 2005. "On the Separation of Net Ecosystem Exchange into Assimilation and Ecosystem Respiration: Review and Improved Algorithm." *Global Change Biology* 11 (9): 1424–39. https://doi.org/10.1111/j.1365-2486.2005.001002.x.

Reinmann, Andrew B., and Lucy R. Hutyra. 2017. "Edge Effects Enhance Carbon Uptake and Its Vulnerability to Climate Change in Temperate Broadleaf Forests." *Proceedings of the National Academy of Sciences* 114 (1): 107–12. https://doi.org/10.1073/pnas.1612369114.

Richardson, Arthur J, and CL Wiegand. 1977. "Distinguishing Vegetation from Soil Background Information." *Photogrammetric Engineering and Remote Sensing* 43 (12): 1541–52.

Running, Steven W., Ramakrishna R. Nemani, Faith Ann Heinsch, Maosheng Zhao, Matt Reeves, and Hirofumi Hashimoto. 2004. "A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production." *BioScience* 54 (6): 547. https://doi.org/10.1641/0006-3568(2004)054%5B0547:ACSMOG%5D2.0.CO;2.

Ryu, Youngryel, Dennis D. Baldocchi, Hideki Kobayashi, Catharine van Ingen, Jie Li, T. Andy Black, Jason Beringer, et al. 2011. "Integration of MODIS Land and Atmosphere Products with a Coupled-Process Model to Estimate Gross Primary Productivity and Evapotranspiration from 1 Km to Global Scales: GLOBAL GPP AND ET." *Global Biogeochemical Cycles* 25 (4): n/a–. https://doi.org/10.1029/2011GB004053.

Schimel, David S. 1995. "Terrestrial Ecosystems and the Carbon Cycle." *Global Change*

Biology 1 (1): 77–91. https://doi.org/10.1111/j.1365-2486.1995.tb00008.x.

Sellers, P. J. 1987. "Canopy Reflectance, Photosynthesis, and Transpiration, II. The Role of Biophysics in the Linearity of Their Interdependence." *Remote Sensing of Environment* 21 (2): 143–83. https://doi.org/10.1016/0034-4257(87)90051-4.

Smith, Ian A, Lucy R Hutyra, Andrew B Reinmann, Julia K Marrs, and Jonathan R Thompson. 2018. "Piecing Together the Fragments: Elucidating Edge Effects on Forest Carbon Dynamics." *Frontiers in Ecology and the Environment* 16 (4): 213–21. https://doi.org/10.1002/fee.1793.

Tucker, Compton J. 1979. "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." *Remote Sensing of Environment* 8 (2): 127–50.

Van Leeuwen, Willem J. D., Barron J. Orr, Stuart E. Marsh, and Stefanie M. Herrmann. 2006. "Multi-Sensor NDVI Data Continuity: Uncertainties and Implications for Vegetation Monitoring Applications." *Remote Sensing of Environment* 100 (1): 67–81. https://doi.org/10.1016/j.rse.2005.10.002.

Xiao, Jingfeng, Frederic Chevallier, Cecile Gomez, Luis Guanter, Jeffrey A. Hicke, Alfredo R. Huete, Kazuhito Ichii, et al. 2019. "Remote Sensing of the Terrestrial Carbon Cycle: A Review of Advances over 50 Years." *Remote Sensing of Environment* 233 (November): 111383. https://doi.org/10.1016/j.rse.2019.111383.

Zeng, Yelu, Dalei Hao, Alfredo Huete, Benjamin Dechant, Joe Berry, Jing M. Chen, Joanna Joiner, et al. 2022. "Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally." *Nature Reviews Earth & Environment* 3 (7): 477–93. https://doi.org/10.1038/s43017-022-00298-5.

2. Assessing uncertainties related to satellite remote sensing indices to estimate Gross Primary Production

2.1. Introduction

Vegetation Gross Primary Production (GPP) is the total amount of carbon fixation by plants through photosynthesis (Badgley et al. 2019). Quantifying GPP is essential for understanding land-atmosphere carbon exchange (Köhler et al. 2018), ecosystem function, and ecosystem responses to climate change (Guan et al. 2022; Brown et al. 2021; Myneni and Williams 1994). However, terrestrial GPP cannot be directly measured due to the contribution of respiration to land carbon fluxes (Anav et al. 2015). Instead, GPP can be inferred in a non-destructive manner by the net carbon exchange measurements at the ecosystem level, or at broader scales using models that incorporate various assumptions and limitations (Reichstein et al. 2005; M. Jung, Reichstein, and Bondeau 2009).

GPP estimations can be grouped into two broad categories: Eddy Covariance (EC) techniques, and satellite data-driven methods (Guan et al. 2022; Xie et al. 2020). EC is the primary in-situ non-destructive method for measuring terrestrial fluxes, and specifically for quantifying the exchange of CO2 between land and the atmosphere, using advanced field instrumentation (Baldocchi 2020b; Badgley et al. 2019; Ryu, Berry, and Baldocchi 2019; Tramontana et al. 2016). However, EC measurements come with certain limitations, such as their relatively low spatial resolution, typically less than $< 1km^2$, which constrains the accuracy of estimating ecosystem carbon and water fluxes at regional and global scales. (Badgley, Field, and Berry 2017).

Additionally, it's important to note that EC techniques directly measure Net Ecosystem Exchange (NEE), not GPP. Subsequently, GPP must be estimated by both subtracting respiration using models and ancillary measurements while accounting for the removal or deposition of carbon stocks due to natural or anthropogenic transport processes such as water flow, fires, or harvest (Beer et al. 2010; Reichstein et al. 2005).

The satellite data-driven models constitute the second category of methods to estimate GPP. Because of their dependency on earth observation platforms, they are not spatially constrained but can have greater uncertainties than the EC techniques (Ryu, Berry, and Baldocchi 2019; Wang et al. 2011). Satellite data-driven models can be classified into process-based models, Light Use Efficiency models (LUE), and Vegetation Index models (Xie et al. 2020).

Process-based models integrate climate, canopy, and soil information derived from multiple sources, including satellite EO, into biophysical models of carbon, water, energy, and nutrient cycles with varying levels of detail (Running and Coughlan 1988; Harris et al. 2021). While these models can be scaled globally (Beer et al. 2010; M. Jung, Reichstein, and Bondeau 2009) they require many parameters that may not be readily available in changing landscapes or for fine-scale studies.

LUE models are based on the concept of radiation conversion efficiency and take into consideration ecological processes (Liu et al. 1997; Heinsch et al. 2006). This efficiency signifies the amount of carbon a specific vegetation type can fix per unit of solar radiation (J. L. Monteith 1972). Initially used for NPP estimation (Prince 1991), LUE models were subsequently adapted for GPP and respiration calculations (Goetz et al. 1999). These models explicitly account for the impact of environmental stress on plant physiological responses. The imposition of environmental stressors may lead to a reduction in the rates of daily carbon assimilation, thereby diminishing overall efficiency. (Prince 1991; Running and Coughlan 1988).

The GPP product from the Moderate Resolution Imaging Spectroradiometer (MODIS) employs an algorithm based on the radiation conversion efficiency concept. This algorithm establishes a connection between absorbed photosynthetically active radiation (APAR) and the LUE term (Heinsch et al. 2006) as shown in Equation 2.1.

$$GPP = PAR \times fAPAR \times LUE \tag{2.1}$$

Where PAR is the incident photosynthetically active radiation (John Lennox Monteith and

Unsworth 2013) and fAPAR is the fraction of the PAR that is effectively absorbed by plants (GCOS 2016). The LUE term depends on vegetation type but also physiological conditions are driven by water availability, temperature stress, and vapour pressure deficit (Goetz et al. 1999; Running and Coughlan 1988). Obtaining these variables for every vegetation type on Earth can be challenging, introducing assumptions that amplify uncertainties (Goetz et al. 1999). Nevertheless, under unstressed conditions, LUE remains constant for a given vegetation type, requiring only PAR and fAPAR to assess primary productivity (Running et al. 2004).

VIs are the Satellite data-driven models' third approach. VIs are a summary of non-linear functions of surface bi-directional reflectance spectra (Myneni and Williams 1994) derived from optical sensors that are combined with climate variables to calculate GPP (C. Wu, Chen, and Huang 2011). This is usually done with some form of regression and physical methods that associate interactions between vegetation and incoming radiation (Fernández-Martínez et al. 2019).

VIs have been used to provide inputs to Equation 2.1 related to fAPAR from regional to global extents or to estimate GPP (Sellers et al. 1994; Running et al. 2004). Some of the most common VIs to estimate GPP are the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), the Near-Infrared Reflectance Index (NIRv), or the Chlorophyll/Carotenoid Index (CCI) among others. (Balzarolo, Peñuelas, and Veroustraete 2019; Rahman et al. 2005, 2005; Xie et al. 2020; Badgley et al. 2019; Zhang et al. 2020; Sellers et al. 1994). These VIs are based on a spectral reflectance ratio between the red and near-infrared regions of the electromagnetic spectrum (Glenn et al. 2008) which tracks an integrated impact of fraction of photosynthetically active radiation (fAPAR) and LUE on productivity (Myneni and Williams 1994).

An index such as NDVI is good for detecting structural vegetation changes in seasonal variability, but it becomes saturated with high biomass conditions (Badgley, Field, and Berry 2017). Other indices such as EVI can overcome the soil and atmospheric effects by adding the blue band (Huete 1988) and may be better suited for predicting GPP in large biomass forests (Badgley, Field, and Berry 2017). Nonetheless, it has been evaluated in a narrow range of ecosystems and needs inputs of start and end dates of the growing seasons which can increase uncertainties (Shi et al. 2017). For specific types of ecosystems such as evergreen conifers, CCI can track the seasonality of daily GPP due to its sensibility to the chlorophyll/carotenoid

pigment ratios (Gamon et al. 2016).

Other indices such as the near-infrared reflectance index (NIRv) have been formulated to address the mixed-pixel effect (pixel with vegetated and non-vegetated features) and to determine the vegetation photosyntetic capacity (Badgley et al. 2019). NIRv is defined as the fraction of reflected NIR light that originates from vegetation. NIRv was originally proposed as a replacement for fPAR in LUE models in that the NIR and PAR reflectance of vegetation are correlated and the scaling by NDVI corrects for soil contributions in the signal (Badgley et al. 2019). NIRv has been shown to have a stronger correlation to GPP at flux towers and on a regional basis than fAPAR notwithstanding the fact that GPP is directly related to fAPAR. NIRv has been extended to include weighting with PAR (Dechant et al. 2022) and replacing the NIR reflectance with NIRv radiance (G. Wu et al. 2020). Both of these approaches have been shown to have even stronger correlations with GPP than NIRv as could be expected since they either directly or indirectly weight the NIRv with PAR.

The stronger correlation between the NIRv index to GPP in comparison to the correlation between VIs related to fAPAR and GPP seemingly contradicts the hypotheses in the LUE model that GPP should be linearly related to fAPAR. There are two explanations: i. Many of the reported comparative studies use fAPAR based on VIs and not APAR. We hypothesize that the NIR indices are simply better estimators of APAR than these VI approximations due to lower measurement error for the NIR indices or a stronger physical relationship between them and APAR versus the historical VIs. ii. The strength of the correlation between APAR and GPP depends on LUE having a linear relationship to observed APAR. While this may hold in some circumstances (e.g. early seasonal measurements for vegetation such as crops where leaf chlorophyll concentration increases during the growth phase) it is not the case in general and definitely during stress conditions (J. L. Monteith 1972).

Despite these efforts, relying solely on remote satellite VIs presents a challenge. The assumptions accompanying VI models suggest the importance of systematically quantifying their predictive capacity for GPP to validate and improve their accuracy (Anav et al. 2015; Brown et al. 2021) across various growing seasons and locations. This is particularly crucial because photosynthesis regulation can occur with no major changes in canopy structure or leaf pigments that can undergo without being detected with reflectance data (Pabon-Moreno et al. 2022; Pierrat et al. 2022), implying that temporal aggregation may be critical for VI models. In-situ eddy-covariance flux measurements coupled with locally calibrated models for

respiration (Baldocchi 2020b) represent a suitable reference GPP for validation solution given that they represent site-level observations (Chu et al. 2021).

As such, the main objective of this M.Sc. thesis chapter is to quantify and compare the uncertainty associated with the VI/LUE models to estimate GPP. To control for variability in environmental conditions (E), sites are selected that share the same land cover, climate, and biome characteristics. VIs models are evaluated by comparison to EC based estimates of GPP for multiple growing seasons from the Bartlett Experimental Forest (USA), the Borden Forest Research Station flux-site (Canada), and the University of Michigan Biological Station (USA). Considering the reviewed studies, we hypothesize that the uncertainty of this approach will depend on the nature of the VI and the spatial and temporal aggregation of application. Specifically, we expect that i. the NIRv and CCI indices will consistently demonstrate a stronger correlation and lower prediction uncertainty for GPP compared to NDVI and EVI across the tested sites. ii. larger temporal aggregations will result in improved predictions due to the reduction in observation variability.

2.2. Methods

2.2.1. Eddy Covariance sites

We used three deciduous broadleaf forest sites located in the northern hemisphere (see Figure 2.1) with eddy covariance (EC) data collected by Ameriflux. For each site, we used the daily, weekly, and monthly GPP values (GPP_DT_VUT_REF variable) estimated using the ONEFlux workflow (Pastorello et al. 2020). The ONEFlux processing does the estimation of the CO_2 fluxes into GPP and Ecosystem Respiration (RECO) from Net Ecosystem Exchange (NEE) through two methods known as daytime and nighttime. Here we selected the daytime method (DT) which uses daytime and nighttime to parameterize a model with two components: one based on light response curve and vapour pressure deficit and a second one using a respiration-temperature relationship to estimate RECO which in turn is used to obtain the difference with NEE and provide GPP (Pastorello et al. 2020).

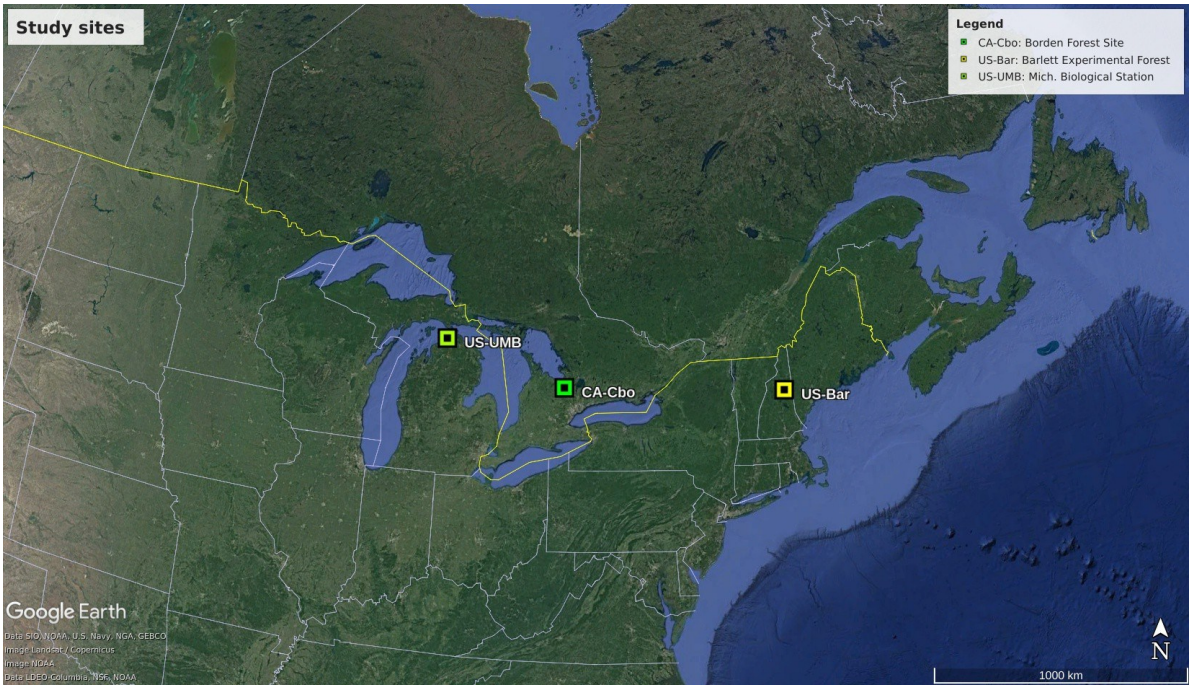

Figure 2.1.: Sites locations

Figure 2.2 displays the GPP trends for the University of Michigan Biological Station, Bartlett Experimental Forest, and the Borden Forest Research Station. Additional details regarding the characteristics of these datasets can be found in Table Table 2.1.

Table 2.1.: ONEFlux sites datasets description

Site	Data range available	Dataset name	Reference
Bartlett	Jan 2015 to Dec 2017	US-Bar: Barlett Experimental Forest (version: beta-3)	(Staebler 2019)
Borden	Jan 2015 to Jan 2022	CA-Cbo: Ontario - Mixed Deciduous, Borden Forest Site	(A. Richardson and Hollinger 2016)
Michigan	Jan 2015 to Jan 2018	US-UMB: Univ. of Mich. Biological Station (version: beta-4)	(C. Gough, Bohrer, and Curtis 2016)

The Bartlett experimental forest is located in New Hampshire, USA (44°06 N, 71°3 W). This site is characterized by a forest with an average canopy height ranging from 20 to 22 meters with a mean annual temperature of 6°C. Despite events such as a hurricane in 1938 and small scale forest management, the forest's mean stand age is around 120-125 years (Ouimette et al. 2018).

The second flux site is Borden Forest Research Station located in Ontario (44°19 N, 79°56 W), Canada. This is one of the largest patches of forest in Southern Ontario which has been collecting EC data since 1996 (Rogers et al. 2020). This site has a forest cover of over 60% with a height of approximately 22 m. It's a deciduous broadleaf natural re-growth forest since 1916 dominated by woody vegetation (Lee et al. 1999).

The third site is the University of Michigan Biological Station which is located in northern Michigan, USA (45°350 N 84°430 W). The site has a forest with different succesional stages, with an average stand age of 90 years (C. M. Gough et al. 2010), and a mean height of 22m. The mean annual temperature is around 5.5°C (C. M. Gough et al. 2021).

The three sites exhibit similar characteristics, indicating their representation of a specific ecosystem type. This uniformity enables meaningful comparisons and offers valuable insights into the relationship between GPP and VIs. A tabular summary of site characteristics, guided by insights detailed in Teets (2022), is presented in Table 2.2.

Table 2.2.: ONEFlux Site characteristics overview

	Site		
Variable	Bartlett	Michigan	Borden
Mean annual temperature (°C)	6	5.5	7.4
Mean annual precipitation (mm)	1246	803	784
Elevation (m)	272	234	209
Dominant genera	Acer, Fagus, Betula	Populus, Quercus, Pinus	Acer, Pinus, Populus
Climate Koeppen[1]	Dfb	Dfb	Dfb

[1]**D** stands for the warm-summer continental or hemiboreal climate. **f** indicates that this climate has significant precipitation in all seasons. **b** indicates that the warmest month has an average temperature between 22°C and 28°C.

2.2.2. Satellite imagery

We used data from the Terra Moderate Resolution Imaging Spectroradiometer (MODIS), specifically the collection MOD09GA Version 6.1 product (MODIS/Terra Surface Reflectance Daily L2G Global 1 km and 500 m SIN Grid) chosen for its daily sampling and broad temporal coverage. Data retrieval was performed using Google Earth Engine (GEE). For each site, a square polygon with an area of 3 km surrounding the EC tower was defined, and the pixel values within this polygon were extracted for comprehensive analysis.

The MODIS contains the surface spectral reflectance from bands 1 through 7 with a spatial resolution of 500m, with corrections for atmospheric conditions such as aerosols, gasses, and Rayleigh scattering (Vermote 2021), and validation of cloud-free pixels as well. Bands used to derive the vegetation indices are shown in Table 2.3

Table 2.3.: MODIS (MOD09GA.061 product) bands used to calculate the VIs

Name	**Description**	**Resolution**	**Wavelength**
sur_refl_01	Red	500 meters	620-670nm
sur_refl_02	NIR	500 meters	841-876nm
sur_refl_03	Blue	500 meters	459-479nm
sur_refl_04	Green	500 meters	545-565nm

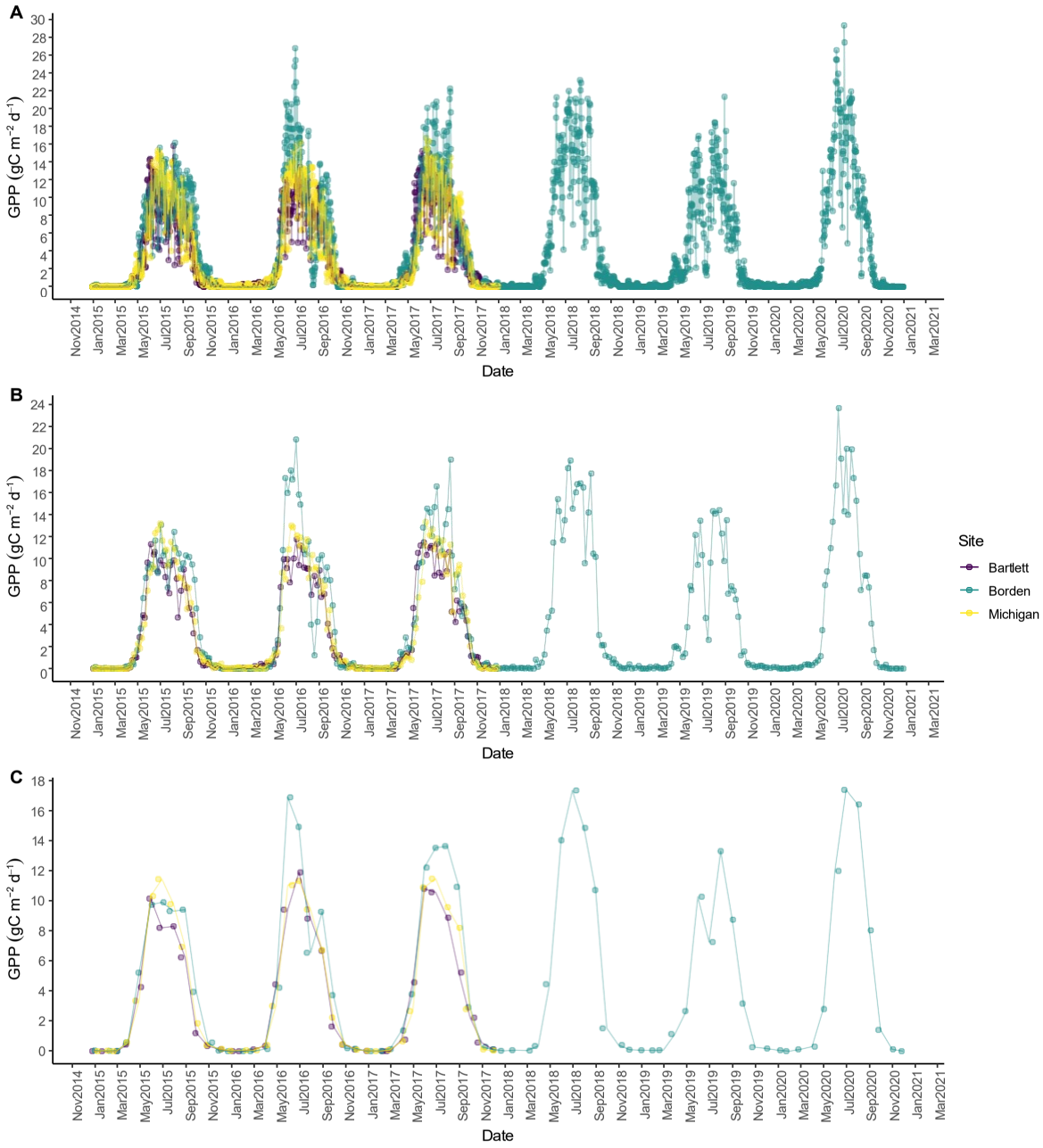

Figure 2.2.: Reference in-situ GPP time series from the study sites on a daily (a), weekly (b), and monthly (c) basis for the University of Michigan Biological Station, Bartlett experimental forest, and the Borden Forest Research Station

The data processing involved three main steps: selecting high-quality pixels, scaling band values, and calculating vegetation indices. MODIS product data have 4 bit-encoded variables which provide information about the observation quality. From those variables, only the 1km Reflectance Data State QA (state_1km) and Surface Reflectance 500m Quality Assurance (qc_500m) variables were used along with each of the band's bits quality indicators, as 250m scan value information (q_scan) was not informative and Geolocation flags (g_flags) had the same value for all observations. The bit-encoded variables were transformed into categorical strings, and only the categories indicating the best quality were selected to filter the pixels (Figure 2.3). The specific bit strings selected for state_1km are shown in Table A.1 and for qc_500m in Table A.2. Subsequently, the surface reflectance for each filtered pixel was determined by scaling the digital number recorded by 0.0001.

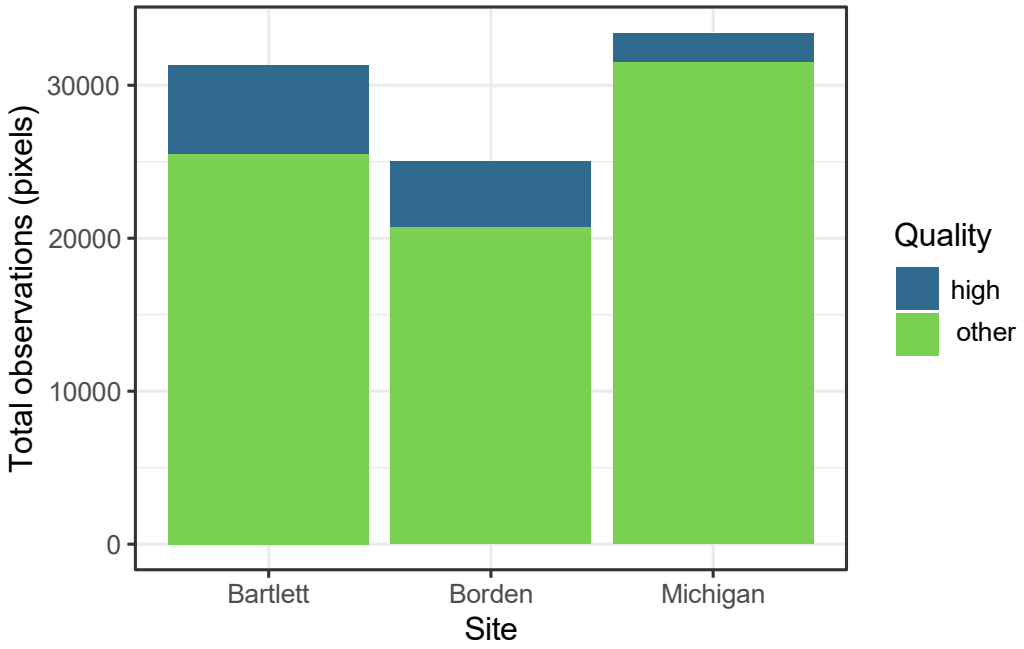

Figure 2.3.: Total number of observations (pixels) from MODIS classified as high quality (used in the analysis) or other quality (filtered out from the analysis) per site.

Following the MODIS Collection 6.1 (C61) LSR Product User Guide (Vermote 2021), any scaled value that fell outside the range of 0 to 1 was considered a fill value or uncorrected Level 1B data and was subsequently discarded. These values were deemed unreliable or lacking meaningful information for the analysis. The number of high-quality surface reflectance observations was summarized on a monthly basis for each site (Figure 2.4)

The VIs NDVI (Equation 2.2), NIRv (Equation 2.3), EVI (Equation 2.4), and CCI (Equation 2.5) were calculated and then matched with the corresponding date in the flux datasets (Green,

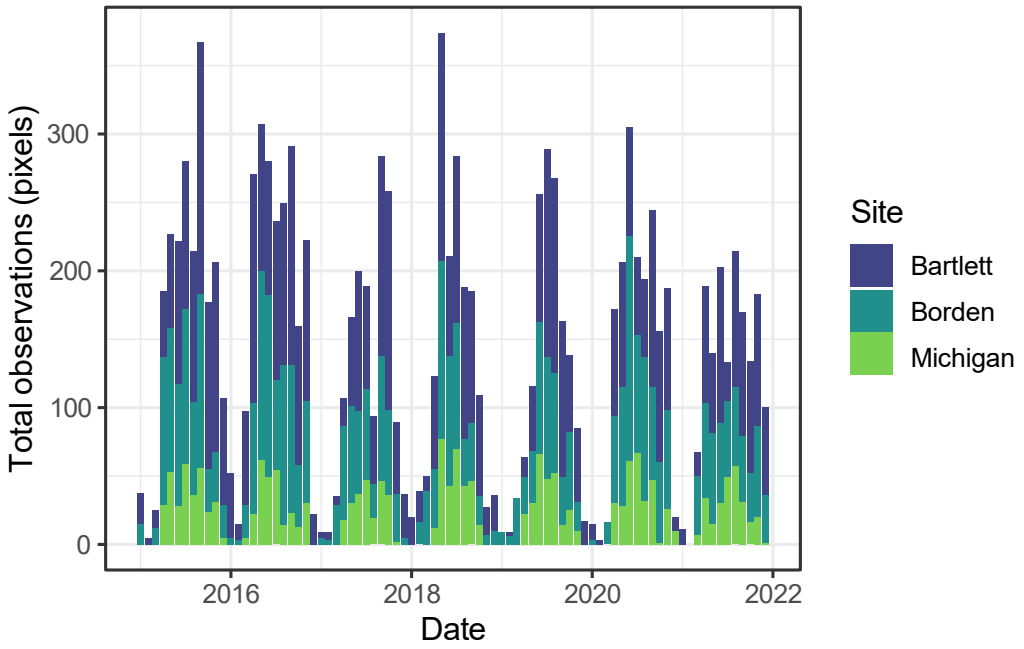

Figure 2.4.: Total number of observations (pixels) from MODIS classified as high quality (used in the analysis) or other quality (filtered out from the analysis)

Red, and NIR correspond to the bands defined in Table 2.3).

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{2.2}$$

$$NIRv = NIR \times \frac{NIR - Red}{NIR + Red} \tag{2.3}$$

$$EVI = 2.5 \times \frac{\text{NIR} - \text{Red}}{(\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1)} \tag{2.4}$$

$$CCI = \frac{Green - Red}{Green + Red} \tag{2.5}$$

2.2.3. Data Preparation

Three datasets were prepared for each site: a daily, a weekly, and a monthly dataset. These datasets were generated from the satellite imagery data with the selected high-quality pixels and the ONEFluxprocess data in order to capture variations in vegetation indices (VI), band values, and GPP over different time scales.

The daily dataset included the VI values, band values only from the high-quality pixels, and GPP measurements derived from the ONEFlux process collected on a daily basis for the time period of available GPP at each site Table 2.1. This dataset provided a high-resolution representation of the variables, allowing for a detailed analysis of their daily fluctuations.

The weekly and monthly datasets were derived from the corresponding daily dataset. These datasets contained summarized values of the VIs and band values, aggregated over the weekly and monthly time frames, respectively. The aggregation process involved calculating an average for the VIs and band values within each week or month.

For the GPP values in the weekly and monthly datasets, rather than summarizing the daily GPP values, the GPP measurements for the weekly and monthly time frames were obtained directly from the ONEFlux process, which provided a reliable estimation of GPP for these longer time intervals.

By creating these three datasets (daily, weekly, and monthly), the study allowed for a comprehensive analysis of the VIs, band values, and GPP at different temporal resolutions. This approach provided insights into the temporal dynamics and patterns of the variables, enabling a more thorough understanding of the processes and relationships under investigation.

For each site and across all time scales (daily, weekly, and monthly), GPP values below 1 gC m^{-2} d^{-1} were excluded. These values were deemed either below the detection limit or insufficient to make a significant contribution to the overall analysis, particularly in representing meaningful vegetation productivity. This process aimed to refine the dataset and focus on values within a range considered more pertinent to the growing season. The final number of observations per site after selecting the high-quality pixels and GPP observations are shown in Figure 2.5.

2.2.4. Data Analysis

To address the potential non-linearity of the lowest uncertainty data-driven model across different temporal aggregation scales or for specific VIs, we employed two distinct modeling approaches: a Linear Model (LM) and a Generalized Additive Model (GAM). GAM models allow a better fit for those cases where the distribution and variability observed in the data is greater, due to the varying temporal scales. We aimed to evaluate whether this variance could be better explained by this type of model, which might not be optimal to capture with a linear

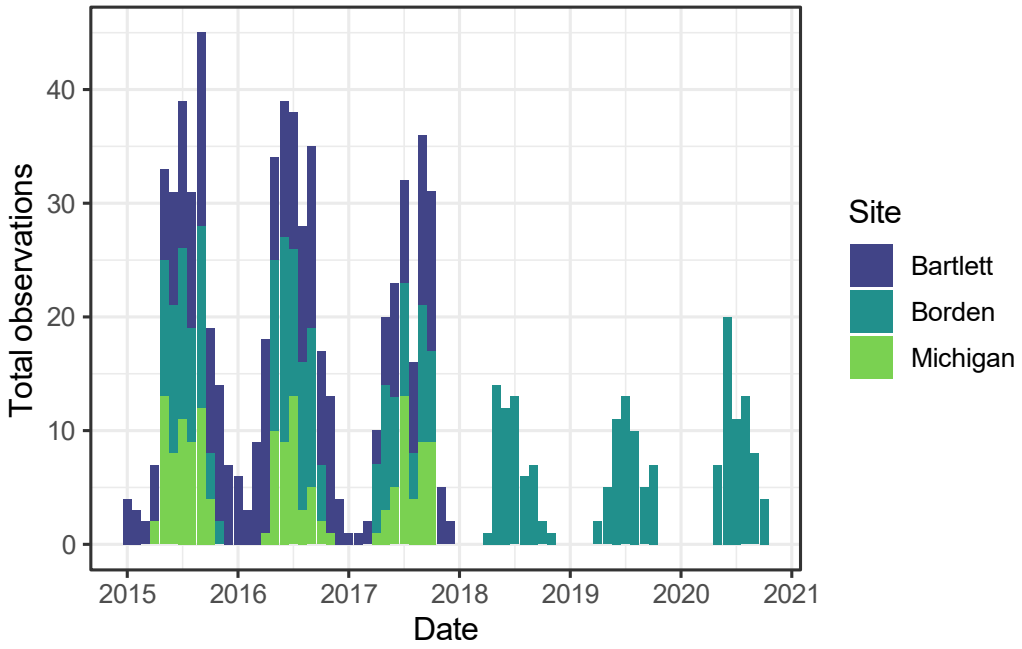

Figure 2.5.: Monthly high-quality MODIS observations after joining with flux observations containing Gross Primary Productivity (GPP) values higher than 1.

relationship, potentially leading to greater residuals. Both models' approaches were applied individually to test our hypothesis regarding VI predictive uncertainty. Additionally, a single model was applied to assess whether a combination of VIs could improve GPP estimation uncertainty.

This facilitated a comparative analysis of how each individual index independently explains the variation in GPP per site. Additionally, the performance of each VI was also evaluated using models calibrated using all sites to assess their robustness to spatial variability within the selected temperate broadleaf biome across various temporal scales. Another step of the analysis was the examination of the relationship between GPP and all indices functioning as covariates, both on an individual site basis and without site distinction.

In total, per each site plus the category of all-sites we created 5 linear models and 5 GAM models for every timescale (daily, weekly, and monthly): one per each index (NDVI, CCI, EVI, NIRv) and one with all the indices as covariates (NDVI + CCI + EVI + NIRv), resulting in a total of 120 models. To assess and compare the models' performance, we calculated the coefficient of determination (R^2) to measure the correlation between the actual observations and predictions. Additionally, we used the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) as indicators of the model estimation error.

2.3. Results

2.3.1. Analysis of GPP-Vegetation Index Relationships Using Linear Models

The Table 2.4 provides a summary of linear models used for GPP estimation at each site, employing the vegetation indices as predictors. For each site and predictor, the table includes the relevant model summary statistics, such as R^2, MAE, and RMSE as indicators of the statistical significance of the model fit. All models add p-values < 0.05 More metrics such as the p-value, adjusted r-squared, the Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) are displayed in the Table A.3 for monthly results, Table A.4 for weekly results, and the Table A.5 for daily outputs. A residuals distribution for each of the models is in Figure A.1

Our findings show that when using all the indices as covariates within the linear model, the model's performance demonstrates better outcomes compared to using any single index alone across all scenarios. In the assessment of individual VI performance on a monthly basis, CCI tend to perform better than EVI, NDVI, and NIRv. Although all models were statistically significant ($p < 0.05$), it is important to note that for the Bartlett and Michigan sites, while CCI shows favourable predictive accuracy, its advantage over the NIRv and EVI is slightly better and mostly due to the MAE (0.86 $gC\ m^{-2}\ d^{-1}$ for Bartlett and 1.02 $gC\ m^{-2}\ d^{-1}$ for Michigan) and RMSE (1.05 $gC\ m^{-2}\ d^{-1}$ for Bartlett and 1.29 $gC\ m^{-2}\ d^{-1}$ for Michigan) results.

The NDVI displays relatively diminished performance, indicating a 9% and 15% reduction in the R^2 compared to CCI in the Bartlett and Michigan sites, respectively. In contrast, EVI exhibits less favourable predictive results ($R^2 = 0.75$, MAE = 1.92 $gC\ m^{-2}\ d^{-1}$, RMSE = 2.54 $gC\ m^{-2}\ d^{-1}$) for the Borden site and the aggregated sites category, although in the specific context of the Borden site, its performance aligns closely with NDVI and NIRv. Among the individual sites, Bartlett has the most favourable predictive outcomes in terms of R^2, MAE, and RMSE for every individual VI model, while the aggregated sites category yields the least favourable predictive results in the same scenarios.

In the case of the weekly models, NDVI records the least favourable results in terms of R^2, MAE, and RMSE when assessed at the three individual sites, while EVI demonstrated its weakest predictive capabilities when all sites were treated as a single entity. However, for this case, EVI shows marginal differences for individual indices such as NDVI and NIRv (R^2 =

0.01 variance explained and no more than 0.02 $gC\ m^{-2}\ d^{-1}$ in RMSE). The best performant individual indices were CCI in terms of R^2, MAE, and RMSE for Borden and the aggregated sites, NIRv for Michigan and EVI for Bartlett. Nonetheless, those superior performances are subtle when compared with the other individual indices. On a weekly basis, differences in variability explained between the best and least performing models range from $R^2 = 0.11$ to $R^2 = 0.2$. Bartlett and Michigan sites consistently yield the most accurate predictive models.

On a daily basis, CCI outperforms the rest of the individual indices in 3 cases: for Bartlett by $R^2 = 0.02$ in variance explanation compared with EVI, for Borden by $R^2 = 0.01$ compared with EVI, and for the combined sites dataset by $R^2 = 0.1$. Nonetheless is worth mentioning that the variance explained in any of the models by individual indices in Borden or the combined site is less than $R^2 = 0.5$ in most of the cases and the error scores are the highest. In the case of Michigan, the best performing individual index was EVI which outperformed the next best-performing individual index NIRv by $R^2 = 0.03$. Generally, the Bartlett and Michigan sites consistently yielded the most accurate predictive models across various configurations. Conversely, the Borden site consistently exhibited the poorest model performance across all scenarios.

Overall, when comparing individual indices, CCI consistently performed better across different timeframes in terms of variance explainability and error metrics. Nonetheless, those differences are subtle compared to EVI and NIRv. In contrast, NDVI consistently performs less mentioning across all evaluated timeframes. Notably, models based on monthly values consistently exhibit better performance than those based on weekly or daily values as ilustrated in Figure 2.6.

2.3.2. Analysis of GPP-Vegetation Index Relationships Using GAM Models

In Table 2.5, we present a summary of the results obtained from the GAM models used for GPP estimation at each site, employing the vegetation indices as predictors. To compare between the models, we include in the table relevant model summary statistics, such as R^2, MAE, and RMSE. Furthermore, additional metrics such as the p-value, the F statistic (f), effective degrees of freedom (edf), and the Akaike Information Criterion (AIC), are displayed in Table A.10, and Table A.11 for daily models, Table A.8, and Table A.9 for weekly outputs, and for monthly results the Table A.6, and Table A.7. A residuals distribution for each of the models is in Figure A.2

Table 2.4.: Summary of Linear models for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis. MAE and RMSE metrics units are gC m- ² d- ¹

(a)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Bartlett	0.89	0.92	1.08	0.80	1.23	1.44	0.89	0.89	1.08	0.89	0.86	1.05	0.92	0.72	0.93
Michigan	0.86	1.05	1.35	0.72	1.62	1.92	0.86	1.05	1.33	0.87	1.02	1.29	0.94	0.72	0.90
Borden	0.75	1.92	2.54	0.76	1.95	2.46	0.75	1.90	2.50	0.79	1.82	2.32	0.80	1.72	2.24
All	0.57	2.21	2.87	0.64	1.95	2.62	0.59	2.18	2.81	0.74	1.69	2.26	0.78	1.63	2.04

(b)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Bartlett	0.79	1.16	1.54	0.64	1.62	2.03	0.77	1.21	1.61	0.79	1.23	1.56	0.82	1.06	1.43
Michigan	0.78	1.50	1.79	0.61	1.99	2.39	0.78	1.50	1.78	0.71	1.72	2.09	0.81	1.43	1.67
Borden	0.57	2.62	3.54	0.50	2.91	3.83	0.57	2.64	3.58	0.58	2.62	3.52	0.61	2.53	3.39
All	0.45	2.65	3.53	0.46	2.64	3.50	0.46	2.61	3.51	0.56	2.35	3.16	0.56	2.34	3.15

(c)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Bartlett	0.81	1.42	1.98	0.66	2.01	2.68	0.80	1.44	2.05	0.83	1.42	1.88	0.87	1.20	1.67
Michigan	0.70	1.87	2.33	0.52	2.31	2.96	0.67	1.92	2.48	0.61	2.06	2.67	0.72	1.81	2.27
Borden	0.43	3.42	4.48	0.28	3.88	5.04	0.38	3.57	4.66	0.44	3.31	4.42	0.53	3.07	4.08
All	0.49	3.07	4.12	0.45	3.22	4.30	0.48	3.14	4.19	0.59	2.71	3.70	0.61	2.65	3.63

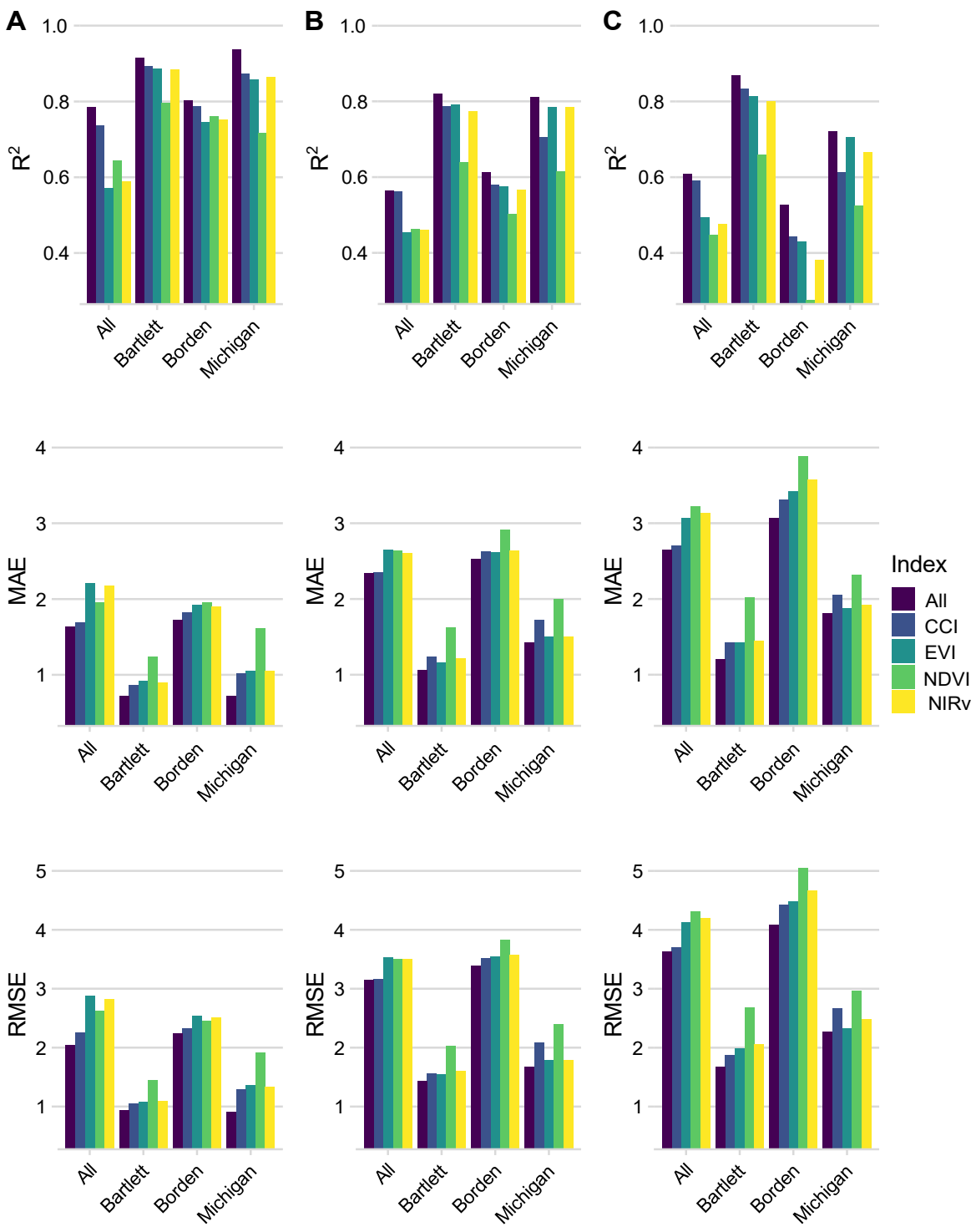

Figure 2.6.: Summary of Linear models for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis. MAE and RMSE metrics units are gC m ² d ¹

When using GAM models on a monthly basis, no single VI demonstrates consistent superiority over the others. For the all sites category, CCI has an $R^2 = 0.03$ better variance explanation than NDVI, which is the second best model with an $R^2 = 0.75$. In the case of Michigan EVI and NIRv were the best individual indices with an $R^2 = 0.96$ of the variance in GPP, but EVI had slightly lower error metrics values in MAE ($0.01\ gC\ m^{-2}\ d^{-1}$) and RMSE ($0.02\ gC\ m^{-2}\ d^{-1}$). Bartlett had a better model performance when using CCI with an $R^2 = 0.91$ of variance explained with the lowest error metrics among all the GAM monthly models. It's important to highlight that for Michigan and Bartlett, implementing a GAM model using all the VIs as covariates posed challenges due to limited observations for the model parameters, raising concerns of potential overfitting.

Finally, it's noteworthy that NDVI on a monthly basis displayed suboptimal performance in two of the sites (Michigan and Bartlett) with a difference of $R^2 = 0.1$ and $R^2 = 0.13$ with the best performing models respectively, while EVI exhibited less favourable results in just the Borden site with MAE $1.92\ gC\ m^{-2}\ d^{-1}$) and RMSE $2.54\ gC\ m^{-2}\ d^{-1}$).

On a weekly basis, models incorporating all VIs as covariates consistently obtained better performance compared to any individual VI, irrespective of the site. Specifically when evaluating the all sites category, the inclusion of all indices as covariates yielded an $R^2 = 0.07$ increase in variance explanation compared with the best individual VI result CCI. Further, for Michigan, this improvement amounted to $R^2 = 0.02$ compared to EVI, $R^2 = 0.01$ for Bartlett in contrast to EVI, and an $R^2 = 0.02$ enhancement for Borden when compared with EVI.

Conversely, NDVI showed a diminished performance as an individual index when compared with all the other individual VIs. In the context of all sites, NDVI yielded $R^2 = 0.07$ less variance explanation than CCI. Notably, for Michigan, NDVI's performance lagged by $R^2 = 0.17$ compared to EVI, Bartlett an $R^2 = 0.18$ reduction relative to EVI, and Borden exhibited an $R^2 = 0.09$ deficit when contrasted with EVI.

On a daily basis, when considering the all sites category, the model with all the VIs as covariates explained $R^2 = 0.04$ more variance in GPP when compared with the best performing individual VI CCI. In the case of Bartlett, the increase was also $R^2 = 0.04$ but in this case, the best individual performing VI was EVI. For Michigan, the model using all VIs as covariates outperformed EVI by $R^2 = 0.05$ in GPP prediction, and for Borden, it was an $R^2 = 0.09$ improvement when compared with NIRv.

Table 2.5.: Summary of GAM models for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis. MAE and RMSE metrics units are gC m- ² d- ¹

(a)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
All	0.66	1.91	2.50	0.72	1.72	2.27	0.67	1.89	2.47	0.75	1.60	2.16	0.79	1.47	1.91
Michigan	0.96	0.55	0.64	0.86	0.91	1.23	0.96	0.56	0.66	0.91	0.79	0.97	NA	NA	NA
Bartlett	0.88	0.92	1.08	0.78	1.23	1.44	0.88	0.89	1.08	0.91	0.67	0.87	NA	NA	NA
Borden	0.74	1.92	2.54	0.77	1.83	2.37	0.75	1.90	2.50	0.78	1.80	2.30	0.78	1.72	2.24

(b)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
All	0.53	2.35	3.25	0.51	2.40	3.34	0.52	2.37	3.29	0.58	2.17	3.06	0.59	2.17	3.02
Michigan	0.84	1.20	1.49	0.67	1.61	2.16	0.82	1.30	1.61	0.75	1.46	1.88	0.86	1.05	1.33
Bartlett	0.81	1.03	1.45	0.63	1.62	2.03	0.79	1.10	1.51	0.79	1.13	1.50	0.82	0.98	1.39
Borden	0.59	2.51	3.41	0.50	2.90	3.82	0.58	2.57	3.49	0.58	2.62	3.52	0.61	2.47	3.35

(c)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
All	0.58	2.70	3.73	0.55	2.79	3.87	0.56	2.74	3.83	0.61	2.55	3.60	0.65	2.39	3.37
Bartlett	0.85	1.22	1.79	0.74	1.63	2.30	0.83	1.25	1.85	0.85	1.26	1.76	0.89	1.04	1.51
Michigan	0.74	1.67	2.13	0.58	2.01	2.73	0.67	1.84	2.43	0.64	1.92	2.55	0.79	1.45	1.92
Borden	0.50	3.22	4.17	0.42	3.41	4.47	0.46	3.33	4.33	0.45	3.27	4.36	0.55	2.94	3.92

Among the individual VIs, NDVI consistently demonstrated the poorest performance across all four cases. As an individual VI, EVI performed better in Bartlett, Michigan, and Borden. However is worth noting that for Borden the variance explained was limited to $R^2 = 0.5$.

In summary, among the three individual sites, Bartlett consistently produced the most favourable results in terms of models explaining variance and yielding lower residuals, followed by Michigan. In the case of Borden, when employing individual indices, the models struggled to achieve a variance explanation exceeding $R^2 = 0.5$.

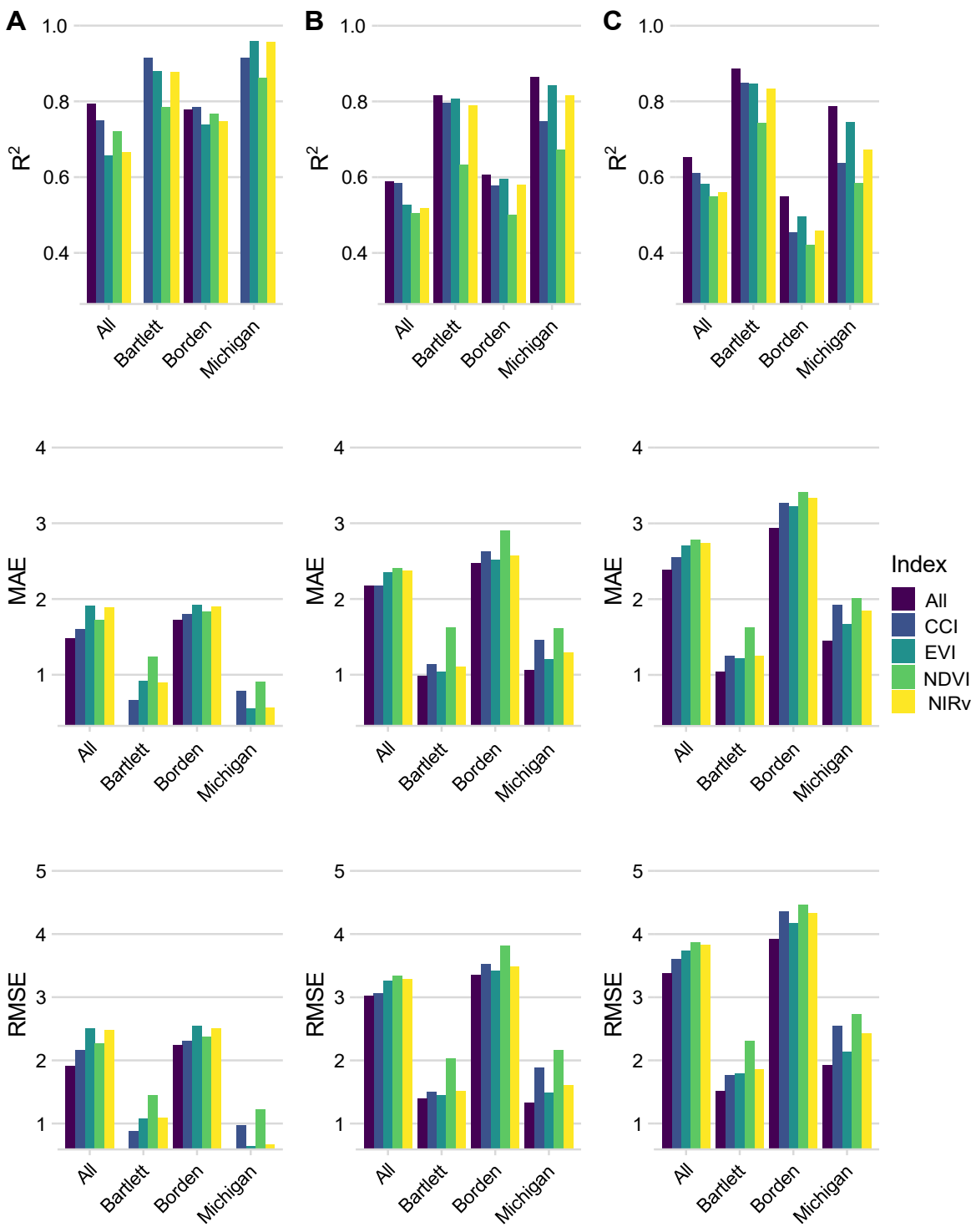

Figure 2.7.: Summary of GAM models for GPP (gC m-² d-¹) estimation using the vegetation indices. Column A represents the metrics for the monthly models, B the weekly, and C the daily metrics. MAE and RMSE metrics units are gC m ² d ¹

2.3.3. LM vs GAM

Monthly GAM applied in the context of Michigan exhibited a better performance when employing EVI and NIRv in comparison to LMs with all vegetation indices as covariates. Regarding weekly GAM models, they demonstrated improved performance compared to LM across all sites, except for Borden, where both models yielded equivalent R^2 values. Notably, despite similar R^2 values, the GAM model consistently displayed lower error metrics, suggesting its capacity to better accommodate potential nonlinear relationships and produce more accurate predictions. This adaptability was particularly evident when addressing the daily, weekly, or monthly variations in GPP. Furthermore, it is crucial to highlight that monthly models, both in LM and GAM frameworks, consistently exhibited superior metrics compared to their weekly and daily counterparts, as illustrated in Figure 2.6 and Figure 2.7. This can be an effect of reducing variance when summarizing values as it is shown in the daily Figure 2.8, weekly Figure 2.9, and monthly Figure 2.10

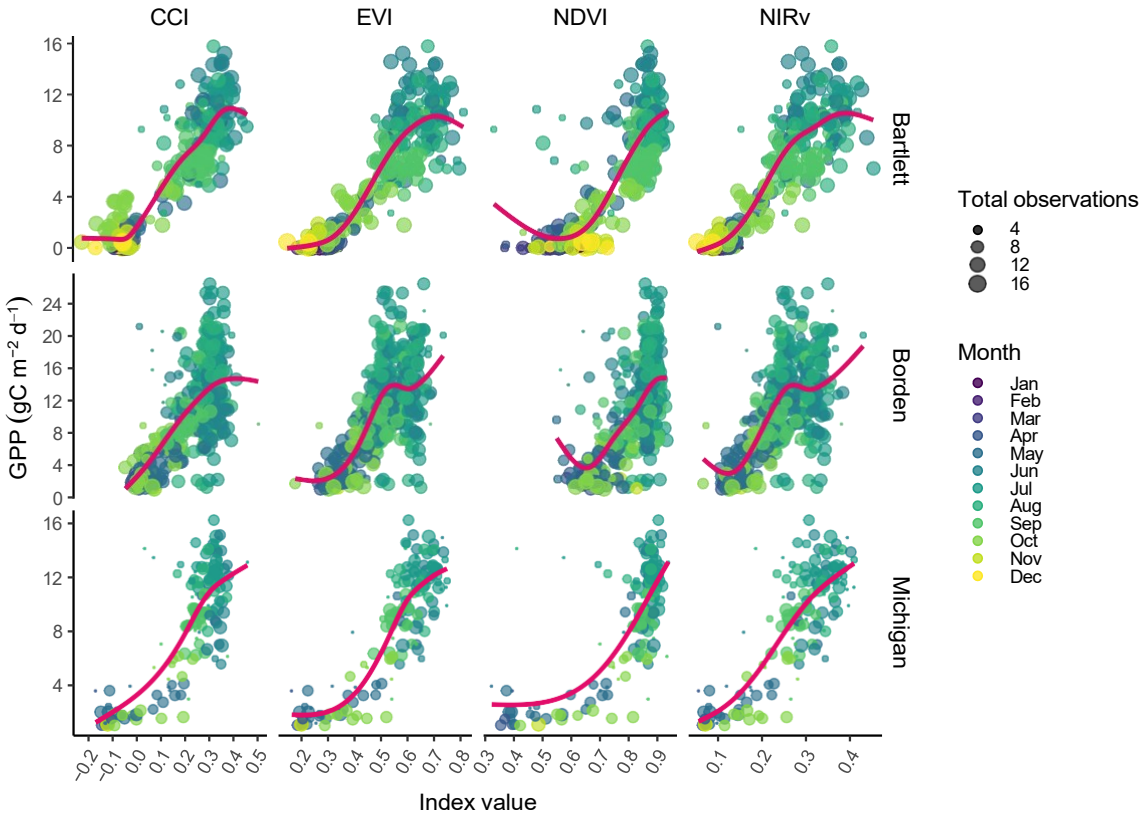

Figure 2.8.: Scatterplot of MODIS 500m derived VIs and GPP with daily values. Every observation corresponds to the observed GPP from a flux tower site. Total observations corresponds to the number of observations used to obtain the mean of the vegetation index (NDVI, NIRv, CCI and EVI). The red line indicates the GAM fit.

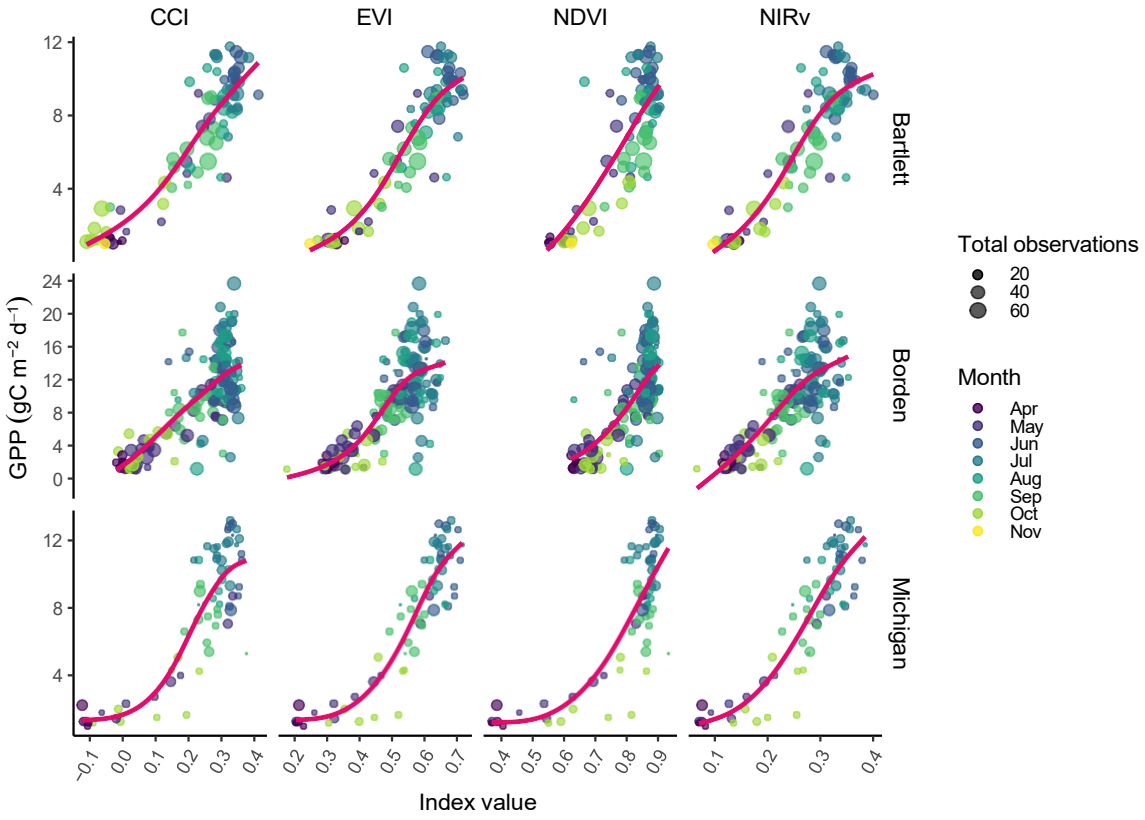

Figure 2.9.: Scatterplot of MODIS 500m derived VIs and GPP with weekly values. Every observation corresponds to the observed GPP from a flux tower site. Total observations corresponds to the number of observations used to obtain the mean of the vegetation index (NDVI, NIRv, CCI and EVI). The red line indicates the GAM fit.

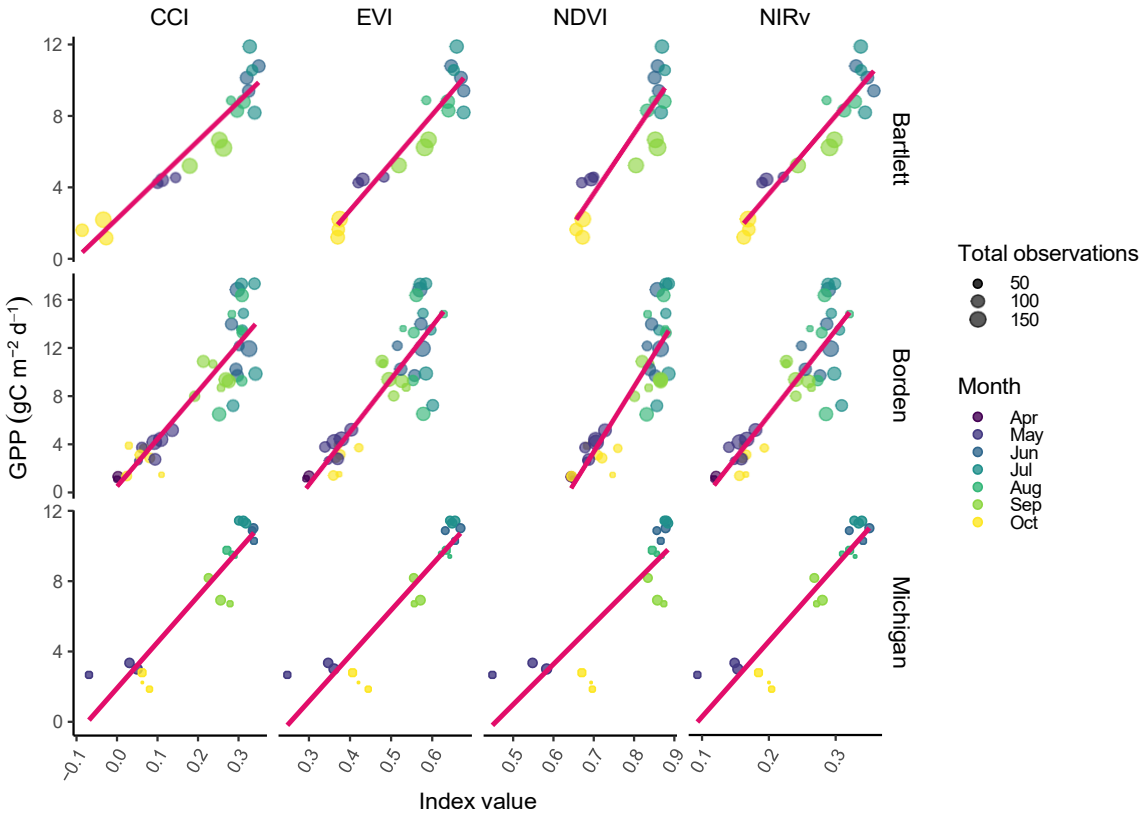

Figure 2.10.: Scatterplot of MODIS 500m derived VIs and GPP with monthly values. Every observation corresponds to the observed GPP from a flux tower site. Total observations corresponds to the number of observations used to obtain the mean of the vegetation index (NDVI, NIRv, CCI and EVI). The red line indicates the GAM fit.

2.4. Discussion

In summary, our LM results demonstrate that incorporating all VIs as covariates in the model enhances predictive accuracy for GPP compared to models using single VIs. This holds true across the three temporal scales and for each of the sites individually or when considered as part of the same biotype. This observation suggests that the relationship is really non-linear, and combining different VIs allows to use of an LM to capture non-linear patterns throughout the entire study duration, contrasting with the predictive power of any individual VI. Additionally, VIs exhibit different noise sensitivities (Zeng et al. 2022), which explains why using a single VI is insufficient to capture nuanced variations in GPP.

Conversely, when aggregating all study sites indiscriminately to represent a unified ecosystem type, model metrics exhibit diminished performance compared to employing distinct models for each individual site. This decline in performance may be attributed to the introduction of variability inherent in the Borden site. The Borden site presents larger ranges of GPP values (see Figure Figure 2.2) than the Bartlett and Michigan sites. Incorporating the Borden values together with those from the Bartlett and Michigan sites to represent a single ecosystem type in a model introduces higher GPP values that contribute to variability. This variability cannot be accurately tracked by the vegetation indices, resulting in a reduction in RMSE and R2.

If such discrepancies arise when constructing models with identical specifications for various sites categorized under the same biotype ecosystem, it may indicate that, for global GPP models, exploring the use of diverse VIs is crucial. An approach like this one could open the possibility for a more accurate estimation of GPP by leveraging specific indices that might yield superior results for each distinct type of ecosystem potentially contributing to more robust and accurate global GPP estimations. For example, the study by (Lin et al. 2019) showed that Chlorophyll Index Red (CRI) had a betterThis hypothesis could be further elucidated through standardized processes for calculating GPP with in-situ data across all sites with flux towers available; a task currently in development by FLUXNET (Pastorello et al. 2020). Standardized GPP values obtained through this process could then be used to train models and assess which VIs have the lowest errors when predicting GPP.

Considering the availability of data with daily, weekly, and monthly values, a pertinent question arises about the temporal scale's impact on GPP estimation. Our results indicate that models based on monthly data demonstrated better model fit for each model and smaller resid-

uals metrics. This superiority may be attributed to reduced data variation, as daily values are aggregated into monthly summaries, leading to values more centralized around a mean which can reduce the measurement noise and concurrently mitigate the saturation effect of VIs for daily peaks. The performance of model fits and residuals metrics declined when transitioning from monthly to weekly and, subsequently, to daily values (See Figure A.1 for LM residuals and Figure A.2 for GAM residuals). Although the decrease in model fit was not substantial, the increase in residuals metrics suggests a potential for larger errors in predictions.

While NDVI is one of the most used VIs in EO (Pabon-Moreno et al. 2022), its performance, as measured by the GPP RMSE, was comparatively less favorable on two out of the three sites when compared to other indices. However, the performance variation among the sites was not significant, with the Bartlett experimental forest showing slightly better results in all GPP ~ VIs relationships compared to the other sites. These findings suggest that, although NDVI may not be the most optimal vegetation index for GPP estimation at these specific sites, variations in performance across different locations are still noteworthy.

When examining the data trends for GPP at each site and its relationship with VIs, it becomes evident that at higher GPP values, the dispersion of the data is more pronounced. This observed pattern may be ascribed to the inherent limitation of these indices, which predominantly track the presence of green leaves. It is important to note, however, that the mere presence of green leaves does not consistently signify active photosynthesis. This discrepancy can arise for two reasons. Firstly, the photosynthetic process may undergo temporary suspension without any manifest changes in chlorophyll content or leaf abscission (Camps-Valls et al. 2021) Secondly, VIs relying on the NIR band face challenges in detecting photosynthesis in instances when a higher amount of healthy plant biomass is present. The increased biomass leads to greater scattering and reflection of NIR radiation, resulting in a saturation effect (Camps-Valls et al. 2021).

This observation holds significance as, despite VIs serving as indicators of vegetation capacity rather than vegetation physiology, historical data records are predominantly derived from sensors equipped with bands for generating these VIs. Conversely, newer sensors capable of capturing additional bands to derive indices such as Solar-Induced Fluorescence (SIF) lack comprehensive historical records, hindering the feasibility of long-term studies that extend back to years preceding the 2000s.

The application of both LM and GAM models to the dataset revealed some differences in

model performance. The GAM model exhibited a slightly superior performance in capturing the underlying patterns within the weekly and daily data compared to the LM counterpart, but not in the monthly data. This was evident from assessments of model fit and residuals metrics. The GAM model's ability to flexibly capture non-linear relationships allowed for a more accurate representation of the complex structure inherent in the data. Consequently, the results suggest that the GAM framework may be more suitable for capturing the nuances present in the dataset, emphasizing the importance of considering model flexibility when analyzing non-linear relationships.

It is worth noticing that using a GAM model with only CCI as a predictor, performs almost as well as the GAM model with all the VIs as covariates. This could imply that the limitation in using a single VI with a LM arises from its inability to capture the non-linear nature of the relationship. A selection of a proficient VI with a non-linear model could better estimate GPP without the need to create models with multiple predictors. In this scenario, the GAM model with CCI to estimate GPP emerges as a potentially adequate choice among generalized linear fits. However, this assessment does not address the question of whether combining diverse models could offer more robustness, as such an approach might implicitly incorporate spatial variability to a certain degree.

2.5. Conclusions

In conclusion, our analysis demonstrates that incorporating all VIs as covariates in our models yields a substantial improvement in predictive accuracy for GPP compared to using any single VI. Additionally, our research highlights the impact of time aggregation on prediction accuracy across different models. Monthly LM models exhibit the best performance metrics, while weekly and daily LM models present lower metrics, attributable to higher variability in observations that makes tracking GPP challenging. A similar pattern is observed when using GAM; however, weekly and daily GAM models outperform LM models.

Notably, no single VI emerges as the universal best predictor for every site or time aggregation, but CCI with a GAM model emerges as a potentially adequate choice among generalized linear fits. Moreover, the impact of large variations in GPP ranges within a site is evident in the quality of predictions. These variations introduce complexities and uncertainties, emphasizing the necessity of accounting for local site characteristics and inherent heterogeneity when

aiming for accurate predictions of GPP. Growing seasons exhibit the most pronounced variability, posing challenges for vegetation indices that saturate at high biomass concentrations, making it more difficult to track changes in GPP.

2.6. References

Anav, Alessandro, Pierre Friedlingstein, Christian Beer, Philippe Ciais, Anna Harper, Chris Jones, Guillermo Murray-Tortarolo, et al. 2015. "Spatiotemporal Patterns of Terrestrial Gross Primary Production: A Review: GPP Spatiotemporal Patterns." *Reviews of Geophysics* 53 (3): 785–818. https://doi.org/10.1002/2015RG000483.

Badgley, Grayson, Leander D. L. Anderegg, Joseph A. Berry, and Christopher B. Field. 2019. "Terrestrial Gross Primary Production: Using NIR v to Scale from Site to Globe." *Global Change Biology* 25 (11): 3731–40. https://doi.org/10.1111/gcb.14729.

Badgley, Grayson, Christopher B. Field, and Joseph A. Berry. 2017. "Canopy Near-Infrared Reflectance and Terrestrial Photosynthesis." *Science Advances* 3 (3): e1602244. https://doi.org/10.1126/sciadv.1602244.

Baldocchi, Dennis D. 2020. "How Eddy Covariance Flux Measurements Have Contributed to Our Understanding of *Global Change Biology*." *Global Change Biology* 26 (1): 242–60. https://doi.org/10.1111/gcb.14807.

Balzarolo, Manuela, Josep Peñuelas, and Frank Veroustraete. 2019. "Influence of Landscape Heterogeneity and Spatial Resolution in Multi-Temporal In Situ and MODIS NDVI Data Proxies for Seasonal GPP Dynamics." *Remote Sensing* 11 (14): 1656. https://doi.org/10.3390/rs11141656.

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, et al. 2010. "Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate." *Science* 329 (5993): 834–38. https://doi.org/10.1126/science.1184984.

Brown, Luke A., Fernando Camacho, Vicente García-Santos, Niall Origo, Beatriz Fuster, Harry Morris, Julio Pastor-Guzman, et al. 2021. "Fiducial Reference Measurements for Vegetation Bio-Geophysical Variables: An End-to-End Uncertainty Evaluation Framework." *Remote Sensing* 13 (16): 3194. https://doi.org/10.3390/rs13163194.

Camps-Valls, Gustau, Manuel Campos-Taberner, Álvaro Moreno-Martínez, Sophia Walther, Grégory Duveiller, Alessandro Cescatti, Miguel D. Mahecha, et al. 2021. "A Unified Vegetation Index for Quantifying the Terrestrial Biosphere." *Science Advances* 7 (9): eabc7447. https://doi.org/10.1126/sciadv.abc7447.

Chu, Housen, Xiangzhong Luo, Zutao Ouyang, W. Stephen Chan, Sigrid Dengel, Sébastien C. Biraud, Margaret S. Torn, et al. 2021. "Representativeness of Eddy-Covariance Flux Footprints for Areas Surrounding AmeriFlux Sites." *Agricultural and Forest Meteorology*

301-302 (May): 108350. https://doi.org/10.1016/j.agrformet.2021.108350.

Dechant, Benjamin, Youngryel Ryu, Grayson Badgley, Philipp Köhler, Uwe Rascher, Mirco Migliavacca, Yongguang Zhang, et al. 2022. "NIRVP: A Robust Structural Proxy for Sun-Induced Chlorophyll Fluorescence and Photosynthesis Across Scales." *Remote Sensing of Environment* 268 (January): 112763. https://doi.org/10.1016/j.rse.2021.112763.

Fernández-Martínez, Marcos, Rong Yu, John Gamon, Gabriel Hmimina, Iolanda Filella, Manuela Balzarolo, Benjamin Stocker, and Josep Peñuelas. 2019. "Monitoring Spatial and Temporal Variabilities of Gross Primary Production Using MAIAC MODIS Data." *Remote Sensing* 11 (7): 874. https://doi.org/10.3390/rs11070874.

Gamon, John A., K. Fred Huemmrich, Christopher Y. S. Wong, Ingo Ensminger, Steven Garrity, David Y. Hollinger, Asko Noormets, and Josep Peñuelas. 2016. "A Remotely Sensed Pigment Index Reveals Photosynthetic Phenology in Evergreen Conifers." *Proceedings of the National Academy of Sciences* 113 (46): 13087–92. https://doi.org/10.1073/pnas.1606162113.

GCOS. 2016. "The Global Observing System for Climate Implementation Needs." GCOS 200. Geneva: World Meteorological Organization (WMO); United Nations Educational, Scientific; Cultural Organization (UNESCO); Intergovernmental Oceanographic Commission (IOC); United Nations Environment Programme (UNEP) ; International Council of Scientific Unions (ICSU). https://library.wmo.int/idurl/4/55469.

Glenn, Edward, Alfredo Huete, Pamela Nagler, and Stephen Nelson. 2008. "Relationship Between Remotely-Sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape." *Sensors* 8 (4): 2136–60. https://doi.org/10.3390/s8042136.

Goetz, Scott J, Stephen D Prince, Samuel N Goward, Michelle M Thawley, Jennifer Small, and Andrew Johnston. 1999. "Mapping Net Primary Production and Related Biophysical Variables with Remote Sensing: Application to the BOREAS Region." *Journal of Geophysical Research: Atmospheres* 104 (D22): 27719–34. https://doi.org/10.1029/1999JD900269.

Gough, Christopher M., Gil Bohrer, Brady S. Hardiman, Lucas E. Nave, Christoph S. Vogel, Jeff W. Atkins, Ben Bond-Lamberty, et al. 2021. "Disturbance-accelerated Succession Increases the Production of a Temperate Forest." *Ecological Applications* 31 (7). https://doi.org/10.1002/eap.2417.

Gough, Christopher M., Christoph S. Vogel, Brady Hardiman, and Peter S. Curtis. 2010.

"Wood Net Primary Production Resilience in an Unmanaged Forest Transitioning from Early to Middle Succession." *Forest Ecology and Management* 260 (1): 36–41. https://doi.org/10.1016/j.foreco.2010.03.027.

Gough, Christopher, Gil Bohrer, and Peter Curtis. 2016. "AmeriFlux AmeriFlux US-UMB Univ. Of Mich. Biological Station." Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….

Guan, Xiaobin, Jing M. Chen, Huanfeng Shen, Xinyao Xie, and Jianbo Tan. 2022. "Comparison of Big-Leaf and Two-Leaf Light Use Efficiency Models for GPP Simulation After Considering a Radiation Scalar." *Agricultural and Forest Meteorology* 313 (February): 108761. https://doi.org/10.1016/j.agrformet.2021.108761.

Harris, Nancy L., David A. Gibbs, Alessandro Baccini, Richard A. Birdsey, Sytze De Bruin, Mary Farina, Lola Fatoyinbo, et al. 2021. "Global Maps of Twenty-First Century Forest Carbon Fluxes." *Nature Climate Change* 11 (3): 234–40. https://doi.org/10.1038/s41558-020-00976-6.

Heinsch, F. A., Maosheng Zhao, S. W. Running, J. S. Kimball, R. R. Nemani, K. J. Davis, P. V. Bolstad, et al. 2006. "Evaluation of Remote Sensing Based Terrestrial Productivity from MODIS Using Regional Tower Eddy Flux Network Observations." *IEEE Transactions on Geoscience and Remote Sensing* 44 (7): 1908–25. https://doi.org/10.1109/TGRS.2005.853936.

Huete, Alfredo. 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25 (3): 295–309.

Jung, M, M Reichstein, and A Bondeau. 2009. "Towards Global Empirical Upscaling of FLUXNET Eddy Covariance Observations: Validation of a Model Tree Ensemble Approach Using a Biosphere Model."

Köhler, Philipp, Christian Frankenberg, Troy S. Magney, Luis Guanter, Joanna Joiner, and Jochen Landgraf. 2018. "Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2." *Geophysical Research Letters* 45 (19): 10, 456–10, 463. https://doi.org/10.1029/2018GL079031.

Lee, Xuhui, Jose D. Fuentes, Ralf M. Staebler, and Harold H. Neumann. 1999. "Long-Term Observation of the Atmospheric Exchange of CO ₂ with a Temperate Deciduous Forest in Southern Ontario, Canada." *Journal of Geophysical Research: Atmospheres* 104 (D13): 15975–84. https://doi.org/10.1029/1999JD900227.

Lin, Shangrong, Jing Li, Qinhuo Liu, Longhui Li, Jing Zhao, and Wentao Yu. 2019. "Evalu-

ating the Effectiveness of Using Vegetation Indices Based on Red-Edge Reflectance from Sentinel-2 to Estimate Gross Primary Productivity." *Remote Sensing* 11 (11): 1303. https://doi.org/10.3390/rs11111303.

Liu, J, JM Chen, J Cihlar, and WM Park. 1997. "A Process-Based Boreal Ecosystem Productivity Simulator Using Remote Sensing Inputs." *Remote Sensing of Environment* 62 (2): 158–75.

Monteith, J. L. 1972. "Solar Radiation and Productivity in Tropical Ecosystems." *The Journal of Applied Ecology* 9 (3): 747. https://doi.org/10.2307/2401901.

Monteith, John Lennox, and M. H. Unsworth. 2013. *Principles of Environmental Physics: Plants, Animals, and the Atmosphere*. Elsevier Science.

Myneni, R. B., and D. L. Williams. 1994. "On the Relationship Between FAPAR and NDVI." *Remote Sensing of Environment* 49 (3): 200–211. https://doi.org/10.1016/0034-4257(94)90016-7.

Ouimette, Andrew P., Scott V. Ollinger, Andrew D. Richardson, David Y. Hollinger, Trevor F. Keenan, Lucie C. Lepine, and Matthew A. Vadeboncoeur. 2018. "Carbon Fluxes and Interannual Drivers in a Temperate Forest Ecosystem Assessed Through Comparison of Top-down and Bottom-up Approaches." *Agricultural and Forest Meteorology* 256-257 (June): 420–30. https://doi.org/10.1016/j.agrformet.2018.03.017.

Pabon-Moreno, Daniel E., Mirco Migliavacca, Markus Reichstein, and Miguel D. Mahecha. 2022. "On the Potential of Sentinel-2 for Estimating Gross Primary Production." *IEEE Transactions on Geoscience and Remote Sensing*, 1–1. https://doi.org/10.1109/TGRS.2022.3152272.

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, et al. 2020. "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." *Scientific Data* 7 (1): 1–27.

Pierrat, Zoe, Troy Magney, Nicholas C. Parazoo, Katja Grossmann, David R. Bowling, Ulli Seibt, Bruce Johnson, et al. 2022. "Diurnal and Seasonal Dynamics of Solar-Induced Chlorophyll Fluorescence, Vegetation Indices, and Gross Primary Productivity in the Boreal Forest." *Journal of Geophysical Research: Biogeosciences* 127 (2). https://doi.org/10.1029/2021JG006588.

Prince, S. D. 1991. "A Model of Regional Primary Production for Use with Coarse Resolution Satellite Data." *International Journal of Remote Sensing* 12 (6): 1313–30. https://doi.org/10.1080/01431169108929728.

Rahman, A. F., D. A. Sims, V. D. Cordova, and B. Z. El-Masri. 2005. "Potential of MODIS EVI and Surface Temperature for Directly Estimating Per-Pixel Ecosystem C Fluxes: MODIS EVI FOR ECOSYSTEM C FLUX." *Geophysical Research Letters* 32 (19): n/a–. https://doi.org/10.1029/2005GL024127.

Reichstein, Markus, Eva Falge, Dennis Baldocchi, Dario Papale, Marc Aubinet, Paul Berbigier, Christian Bernhofer, et al. 2005. "On the Separation of Net Ecosystem Exchange into Assimilation and Ecosystem Respiration: Review and Improved Algorithm." *Global Change Biology* 11 (9): 1424–39. https://doi.org/10.1111/j.1365-2486.2005.001002.x.

Richardson, Andrew, and David Hollinger. 2016. "AmeriFlux AmeriFlux US-Bar Bartlett Experimental Forest." Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….

Rogers, Cheryl A., Jing M. Chen, Ting Zheng, Holly Croft, Alemu Gonsamo, Xiangzhong Luo, and Ralf M. Staebler. 2020. "The Response of Spectral Vegetation Indices and Solar-Induced Fluorescence to Changes in Illumination Intensity and Geometry in the Days Surrounding the 2017 North American Solar Eclipse." *Journal of Geophysical Research: Biogeosciences* 125 (10). https://doi.org/10.1029/2020JG005774.

Running, Steven W., and Joseph C. Coughlan. 1988. "A General Model of Forest Ecosystem Processes for Regional Applications I. Hydrologic Balance, Canopy Gas Exchange and Primary Production Processes." *Ecological Modelling* 42 (2): 125–54. https://doi.org/10.1016/0304-3800(88)90112-3.

Running, Steven W., Ramakrishna R. Nemani, Faith Ann Heinsch, Maosheng Zhao, Matt Reeves, and Hirofumi Hashimoto. 2004. "A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production." *BioScience* 54 (6): 547. https://doi.org/10.1641/0006-3568(2004)054%5B0547:ACSMOG%5D2.0.CO;2.

Ryu, Youngryel, Joseph A. Berry, and Dennis D. Baldocchi. 2019. "What Is Global Photosynthesis? History, Uncertainties and Opportunities." *Remote Sensing of Environment* 223 (March): 95–114. https://doi.org/10.1016/j.rse.2019.01.016.

Sellers, P. J., C. J. Tucker, G. J. Collatz, S. O. Los, C. O. Justice, D. A. Dazlich, and D. A. Randall. 1994. "A Global 1° by 1° NDVI Data Set for Climate Studies. Part 2: The Generation of Global Fields of Terrestrial Biophysical Parameters from the NDVI." *International Journal of Remote Sensing* 15 (17): 3519–45. https://doi.org/10.1080/01431169408954343.

Shi, Hao, Longhui Li, Derek Eamus, Alfredo Huete, James Cleverly, Xin Tian, Qiang Yu, et al. 2017. "Assessing the Ability of MODIS EVI to Estimate Terrestrial Ecosystem Gross Primary Production of Multiple Land Cover Types." *Ecological Indicators* 72 (January): 153–64. https://doi.org/10.1016/j.ecolind.2016.08.022.

Staebler, Ralf. 2019. "AmeriFlux AmeriFlux CA-Cbo Ontario-Mixed Deciduous, Borden Forest Site." Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….

Teets, Aaron, David J. P. Moore, M. Ross Alexander, Peter D. Blanken, Gil Bohrer, Sean P. Burns, Mariah S. Carbone, et al. 2022. "Coupling of Tree Growth and Photosynthetic Carbon Uptake Across Six North American Forests." *Journal of Geophysical Research: Biogeosciences* 127 (4): e2021JG006690. https://doi.org/10.1029/2021JG006690.

Tramontana, Gianluca, Martin Jung, Christopher R. Schwalm, Kazuhito Ichii, Gustau Camps-Valls, Botond Ráduly, Markus Reichstein, et al. 2016. "Predicting Carbon Dioxide and Energy Fluxes Across Global FLUXNET Sites Withregression Algorithms." *Biogeosciences* 13 (14): 4291–4313. https://doi.org/10.5194/bg-13-4291-2016.

Vermote, Wolfe, E. 2021. "MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061." *NASA EOSDIS Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MOD09GA.061.

Wang, Weile, Jennifer Dungan, Hirofumi Hashimoto, Andrew R. Michaelis, Cristina Milesi, Kazuhito Ichii, and Ramakrishna R. Nemani. 2011. "Diagnosing and Assessing Uncertainties of Terrestrial Ecosystem Models in a Multimodel Ensemble Experiment: 1. Primary Production: ENSEMBLE MODEL UNCERTAINTIES: GPP/NPP." *Global Change Biology* 17 (3): 1350–66. https://doi.org/10.1111/j.1365-2486.2010.02309.x.

Wu, Chaoyang, Jing M. Chen, and Ni Huang. 2011. "Predicting Gross Primary Production from the Enhanced Vegetation Index and Photosynthetically Active Radiation: Evaluation and Calibration." *Remote Sensing of Environment* 115 (12): 3424–35. https://doi.org/10.1016/j.rse.2011.08.006.

Wu, Genghong, Kaiyu Guan, Chongya Jiang, Bin Peng, Hyungsuk Kimm, Min Chen, Xi Yang, et al. 2020. "Radiance-Based NIR $_{v}$ as a Proxy for GPP of Corn and Soybean." *Environmental Research Letters* 15 (3): 034009. https://doi.org/10.1088/1748-9326/ab65cc.

Xie, Xinyao, Ainong Li, Jianbo Tan, Huaan Jin, Xi Nan, Zhengjian Zhang, Jinhu Bian, and Guangbin Lei. 2020. "Assessments of Gross Primary Productivity Estimations with Satellite Data-Driven Models Using Eddy Covariance Observation Sites over the Northern Hemisphere." *Agricultural and Forest Meteorology* 280 (January): 107771. https://doi.

org/10.1016/j.agrformet.2019.107771.

Zeng, Yelu, Dalei Hao, Alfredo Huete, Benjamin Dechant, Joe Berry, Jing M. Chen, Joanna Joiner, et al. 2022. "Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally." *Nature Reviews Earth & Environment* 3 (7): 477–93. https://doi.org/10.1038/s43017-022-00298-5.

Zhang, Zhaoying, Yongguang Zhang, Yao Zhang, Nadine Gobron, Christian Frankenberg, Songhan Wang, and Zhaohui Li. 2020. "The Potential of Satellite FPAR Product for GPP Estimation: An Indirect Evaluation Using Solar-Induced Chlorophyll Fluorescence." *Remote Sensing of Environment* 240 (April): 111686. https://doi.org/10.1016/j.rse.2020.111686.

3. A data driven approach to predict GPP from VIs through machine learning methods

3.1. Introduction

The field of Earth Science has witnessed a transformative shift with the integration of Machine Learning (ML) methods, which has led to a deeper understanding of our planet's complex ecosystems and processes (Reichstein et al. 2019). ML methods are now well established in environmental sciences (Lary et al. 2016), including their use within studies aimed at mapping and quantifying vegetation characteristics and ecological processes such as vegetation cover, structure, and disturbances, among others (Lehnert et al. 2015; Jochem Verrelst et al. 2012).

The increasing number of EC sites (Tramontana et al. 2016) coupled with the continuously growing amount of Earth system data surpassing dozens of petabytes (Reichstein et al. 2019), has led to an emergence of purely data-driven methodologies for quantifying ecosystem status and fluxes. These approaches have shown promise for the quantification of global terrestrial photosynthesis (Martin Jung et al. 2011; Tramontana et al. 2016) and have resulted in good progress in the estimation of biogeo-physical parameters using remotely sensed reflectance data, both at local and global scales (Coops et al. 2003; J. Verrelst et al. 2012).

Furthermore, these data-driven approaches have contributed significantly to the scientific community by providing spatial, seasonal, and interannual variations in predicted fluxes. These predictions, generated through machine learning methodologies, are now serving as important benchmarks for evaluating the performance of physical land-surface and climate models (Martin Jung et al. 2010; Bonan et al. 2011; Anav et al. 2015).

Some of the differences between data driven models and process-based methods are the inherent observational character of data driven models and that functional relationships emerge

from the patterns found in the data, rather than being stated before (Tramontana et al. 2016). So functional relationships between in-situ measured fluxes with the explanatory variables can emerge (Tramontana et al. 2016). This paradigm shift toward data-driven modeling to extract patterns represents an opportunity to come up with new ideas and question established theories in earth system models.

In contrast to process based models, data driven models inherently possess an observational nature where functional relationships emerge from the patterns found in the data, rather than being predefined, such as the relationships between in-situ measured fluxes and the explanatory variables. (Tramontana et al. 2016). For example, the application of spatially explicit global data driven methods, has unveiled discrepancies in the estimation of photosynthesis within tropical rainforests when compared to climate models (Beer et al. 2010). This overestimation, has led to the creation of hypothesis for a better understanding of radiative transfer in vegetation canopies (Bonan et al. 2011) which can result in better photosynthesis estimates. This paradigm shift toward data-driven modeling to extract patterns, represents an opportunity to explore novel ideas and question established theories in earth system models (Reichstein et al. 2019).

Predicting dynamics in the biosphere is challenging due to biologically mediated processes (Reichstein et al. 2019). The term "prediction" should not be confused with forecasting, as most models are not aiming at predicting into the future. Instead, the focus of these data-driven models is to improve historical estimates or enable the use of reflectance values to predict GPP when no in-situ data is available in the present times (Meyer et al. 2018).

These predictions can include many forms of uncertainty (Reichstein et al. 2019). One form is that individual ML methods can have different responses, especially when these models are applied beyond the conditions presented in the training dataset. (M. Jung, Reichstein, and Bondeau 2009; Papale et al. 2015). Another form is related to the explanatory variables used in ML methods derived from satellite remote sensing, which are partial in providing information about the vegetation state (Tramontana et al. 2016). Consequently, they lack the information required to explain the complete variability in fluxes (Tramontana et al. 2015). For instance, if a model is created using only reflectance data to estimate GPP without meteorological data, phenomena such as drought could lead to predicted values with large errors, given that stomata closure during water deficit have an immediate effect on the fluxes that can not be detected by the reflectance values only until later when the stress conditions

persist (Tramontana et al. 2015).

This chapter evaluates data drive approaches for estimating GPP given MODIS surface reflectance data for temperate broadleaf forests in North America. Two state-of-the-art approaches previously used for GPP estimation are tested: regression random forests (James et al. 2022) and AutoML (LeDell and Poirier 2020) to understand how the ML algorithm impacts prediction uncertainty. Both approaches were tested locally and using pooled data from three sites to quantify their ability to be applied over larger spatial extents.

3.2. Methods

3.2.1. Eddy Covariance sites

For this study, we selected three deciduous broadleaf forest forests sites: University of Michigan Biological Station located in northern Michigan, USA (45°350 N 84°430 W), Bartlett experimental forest in New Hampshire, USA (44°06 N, 71°3 W), and the Borden Forest Research Station (44°19 N, 79°56 W) in Ontario, Canada. These sites were selected to ensure they represented a single ecosystem type, characterized by shared environmental features. This approach allowed us to treat the dataset as a representation of a specific vegetation type terrestrial ecosystem.

In-situ data such as GPP was obtained utilizing the ONEFlux estimation processing by Ameriflux. Here, we selected GPP estimation done by the daytime method (Pastorello et al. 2020) on a daily, weekly, and monthly basis.

To capture seasonal variations and long-term trends, GPP data was collected over a minimum of 2 years. Specifically, University of Michigan Biological Station collected data spanned from January 2015 to January 2018, Bartlett experimental forest data ranges from January 2015 to December 2018, and Borden Forest Research Station from January 2015 to January 2022.

3.2.2. Satellite imagery

We used Google Earth Engine (GEE) to retrieve the Terra Moderate Resolution Imaging Spectroradiometer (MODIS), specifically the collection MOD09GA Version 6.1 daily 500m resolution surface reflectance products (MODIS/Terra Surface Reflectance Daily L2G Global 1 km and 500 m SIN Grid). A square polygon with an area of 3km surrounding the EC tower was defined for each of the study sites, and all daily data pixel values within this polygon were extracted for analysis.

MOD09GA contains the surface spectral reflectance from bands 1 through 7 (Table 3.1) with a spatial resolution of 500m, with corrections for atmospheric conditions such as aerosols, gasses, and Rayleigh scattering (Vermote 2021).

We selected the highest quality pixels according to the 1km Reflectance Data State QA (state_1km) (Table A.1) and Surface Reflectance 500m Quality Assurance (qc_500m) (Table A.2) variables. Once we had just the highest quality pixels, all the band values were scaled

by a factor of 0.0001. If any value fell outside the range of 0 to 1 after the scaling, it was discarded.

Once all the band values were scaled, we calculated 4 Vegetation Indices: NDVI (Equation 2.2), NIRv (Equation 2.3), EVI (Equation 2.4), and CCI (Equation 2.5). Then all the MODIS bands values and VIs were summarized on a daily, weekly, and monthly basis to be merged with the GPP values from ONEFlux.

Table 3.1.: MODIS (MOD09GA.061 product) bands used for ML methods

Name	Description	Resolution	Wavelength
sur_refl_01	Red	500 meters	620-670nm
sur_refl_02	NIR	500 meters	841-876nm
sur_refl_03	Blue	500 meters	459-479nm
sur_refl_04	Green	500 meters	545-565nm
sur_refl_05	Red Edge	500 meters	1230-1250nm
sur_refl_06	SWIR 1	500 meters	1628-1652nm
sur_refl_07	SWIR 2	500 meters	2105-2155nm

3.2.3. Random Forests

A random forest is an ensemble learning technique that leverages the power of multiple decision trees to improve predictive accuracy and robustness. It combines regression and classification trees, constructing each tree from random subsets of both trained data and features (James et al. 2022). The final prediction is obtained by aggregating individual tree predictions, which are then averaged to produce the ultimate estimate. Additionally, at each split, the best predictor from the random subset is selected to effectively partition the data and in the case of the regression random forest, the process is adapted to predict a continuous numeric outcome(James et al. 2022; Meyer et al. 2018).

Regression random forests were used as an approach to predict GPP as a function of all available MOD09GA bands values (from B01 to B07) and the calculated VIs. Three distinct models were developed, each one tailored for a specific time scale (daily, weekly, and monthly). This approach allowed us to assess the prediction performance of GPP at different time scales.

Each model was calibrated using a random data splitting procedure, dividing the data into a training set with 70% of the observations and a test set with the remaining 30%. Due to the varying number of observations for each site, to avoid an imbalanced training dataset, we employed a stratified data split to ensure a proportional representation of each site category in both sets. To ensure reproducibility, we used a consistent random number generator state throughout the process.

To implement the RF models, we use the ranger package in R (Wright and Ziegler 2017), utilizing 1000 trees within the forest ensemble. The models were trained using the bootstrap resampling technique with 100 folds, which helps to improve the robustness and accuracy of the predictions.

We calculated the variable of importance (VIP) to understand which MODIS bands or VIs are driving the predictions in each of the regression random forest models. To measure the influence of each feature on the overall model's predictive performance, we quantify how this performance deteriorates when a particular variable is permuted while keeping others constant.

To understand which features contributed the most on average to a particular GPP prediction in different coalitions (Molnar 2020), we calculated the Shapley values (Lundberg and Lee 2017). These values were computed with the DALEX package in R (Biecek 2018) for both low GPP and high GPP scenarios within each of the temporally aggregated models.

3.2.4. AutoML

The AutoML approach is designed to identify the most optimal ML pipeline for a specific problem and available training data by evaluating different combinations of data processing steps, ML models, and hyperparameters settings (Gaber et al. 2023). To create a stacked ensemble of models and evaluate the performance, we utilized the H2O AutoML framework (LeDell and Poirier 2020).

For the data training and evaluation, we randomly split the dataset into a training and a test set, allocating 80% and 20% of the observations, respectively. We excluded the total observations and site variables from both datasets, while the remaining features (bands and VIs), were considered predictors. The response variable for prediction was GPP. We transformed the data into an H2O frame to ensure compatibility with the AutoML process.

After generating machine learning models and hyperparameter configurations, we imposed a time constraint of a maximum of 2 minutes for model training. The best performing models were then selected in terms of lower predictions erros and used to form an ensemble. This trained ensemble model was used to generate predictions on the test set. Model performance was evaluated using metrics such as the coefficient of determination and the RMSE.

To asses the variable importance, we determined the influence of each variable on the model's predictions. The results of this analysis were visually represented as heatmaps for each of the models within the ensemble.

3.3. Results

3.3.1. Data-Driven GPP Prediction: A Regression Random Forest Approach

The best performing model, based on R^2 and RMSE was the monthly aggregated model, achieving a R^2 of 0.81 and an RMSE of 2.03 $gC\ m^{-2}\ d^{-1}$ (see Figure 3.1). The daily model exhibited the second best performance model with a RMSE of 3.20 $gC\ m^{-2}\ d^{-1}$ and a R^2 of 0.69 (see Figure 3.2). However, the weekly model displayed comparatively lower performance, with an R^2 of 0.56 and an RMSE of 3.23 $gC\ m^{-2}\ d^{-1}$ (see Figure 3.3)

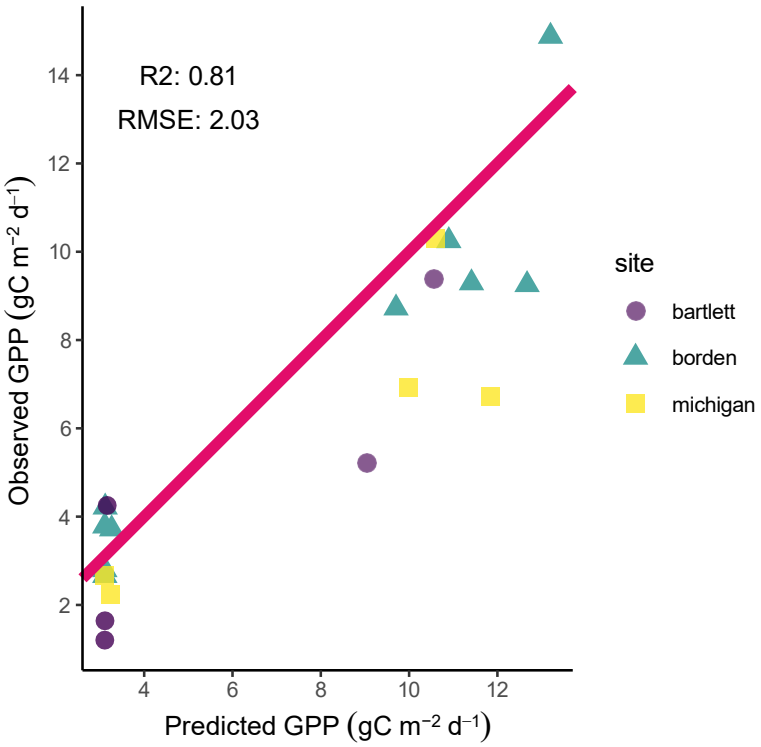

Figure 3.1.: GPP observed and predicted values from the Random Forest for all the sites at a monthly basis. The red line represents a 1:1 relation. Metrics units are gC m^{2} d^{1}

Moreover, in assessing the importance of predictor variables within each model, VIs variables held top positions across all models, surpassing the importance of any other spectral bands alone. Specifically, the CCI contributed most significantly to the predictive performance in two of the models, the monthly and weekly models (Figure 3.4 and Figure 3.5). In the case of the daily model, CCI emerged as the second most influential predictor variable (as shown in Figure 3.6).

Despite NDVI being ranked as the second variable in importance for the monthly model, it

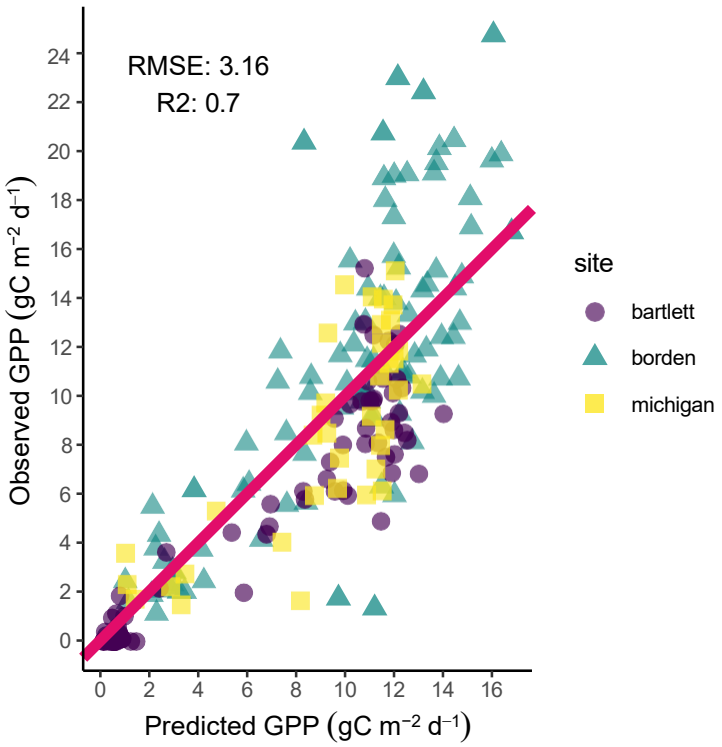

Figure 3.2.: GPP observed and predicted values from the Random Forest model for all the sites at a daily basis. The red line represents a 1:1 relation. Metrics units are gC m ² d ¹

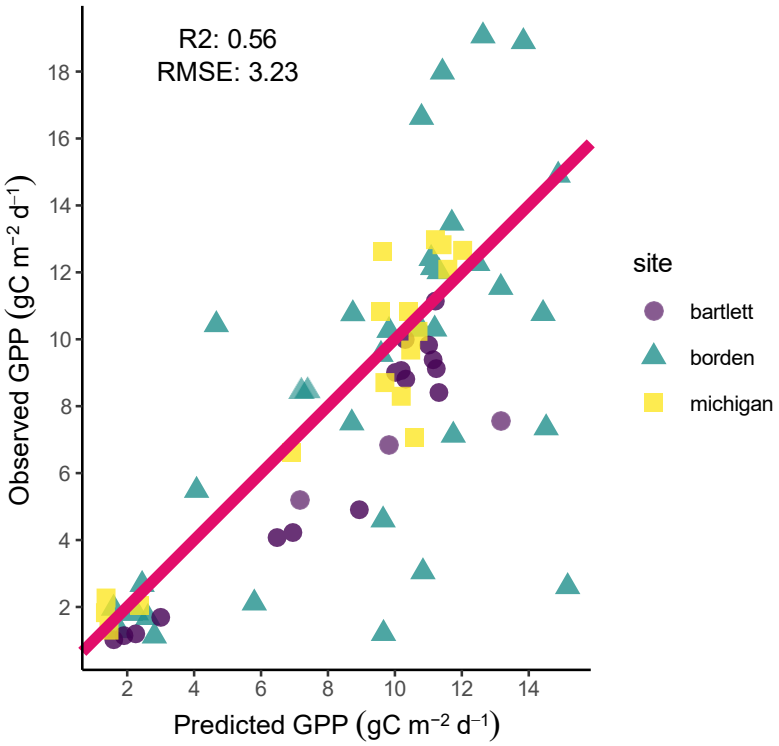

Figure 3.3.: GPP observed and predicted values from the Random Forest for all the sites at a weekly basis. The red line represents a 1:1 relation. Metrics units are gC m ² d ¹

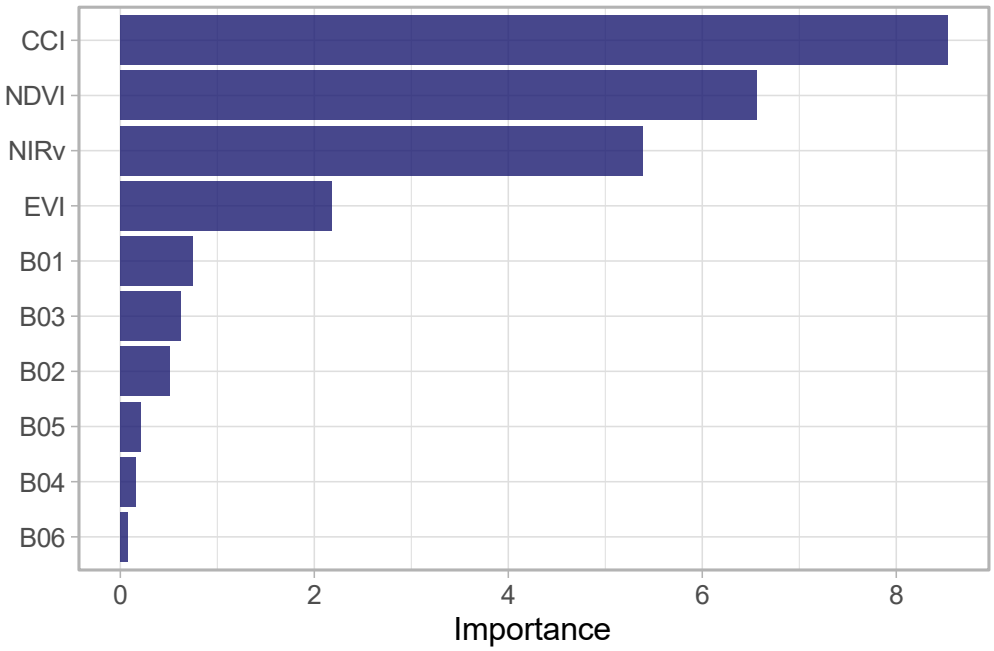

Figure 3.4.: Variable of importance derived from the Random forest model for the monthly values at 500 m spatial resolution model.

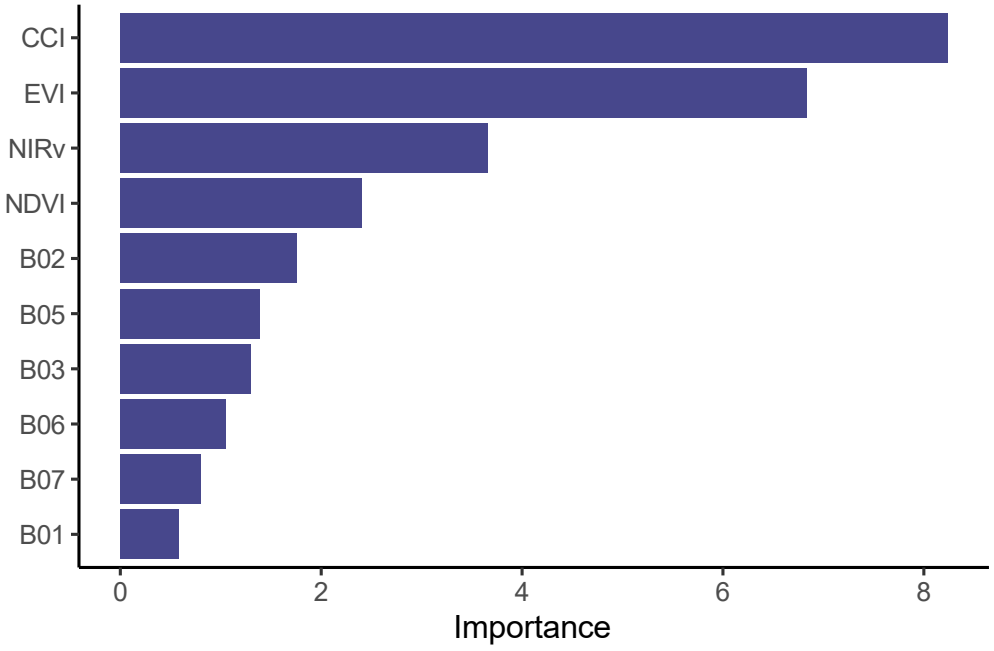

Figure 3.5.: Variable of importance derived from the Random forest model for the weekly values at 500 m spatial resolution model.

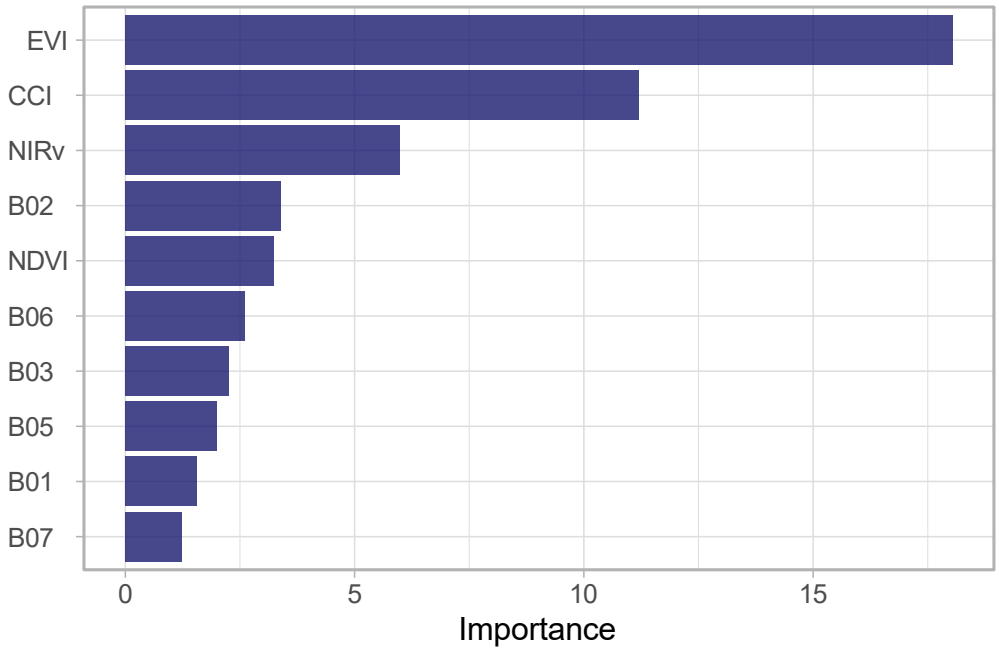

Figure 3.6.: Variable of importance derived from the Random forest model for the daily values at 500 m spatial resolution model.

was the fifth variable in importance for the daily model and the fourth one for the weekly model. Meanwhile, the NIRv consistently held the third most influential position across all models. EVI was found to be most important for the daily model, the second most important variable in the weekly model, but fourth in the monthly model.

To evaluate how the predictor variables had an impact on the model's predictions, we employed the Shapley values. For each of the models created, we calculated the contributions to each variable in predicting GPP for two cases: for a known high GPP value and a known low GPP value from the test data set.

In the case of the daily model, we chose a low GPP value from the test dataset, specifically 0.01 $gC\ m^{-2}\ d^{-1}$and obtained a model prediction of 0.60 $gC\ m^{-2}\ d^{-1}$. Our Shapley value analysis showed that EVI, CCI and NIRv were the most influential attributes affecting the model prediction (See Figure 3.7 **A**). These VIs contribute negatively to the prediction, reducing the predicted GPP.

Conversely, the high GPP value selected (16.21 $gC\ m^{-2}\ d^{-1}$) from the test dataset, had a model prediction of 12.70 $gC\ m^{-2}\ d^{-1}$. In this case the Shapley values analysis indicated that the most influential variables were CCI and NIRv (See Figure 3.7 **B**). These variables contributed positively to the prediction.

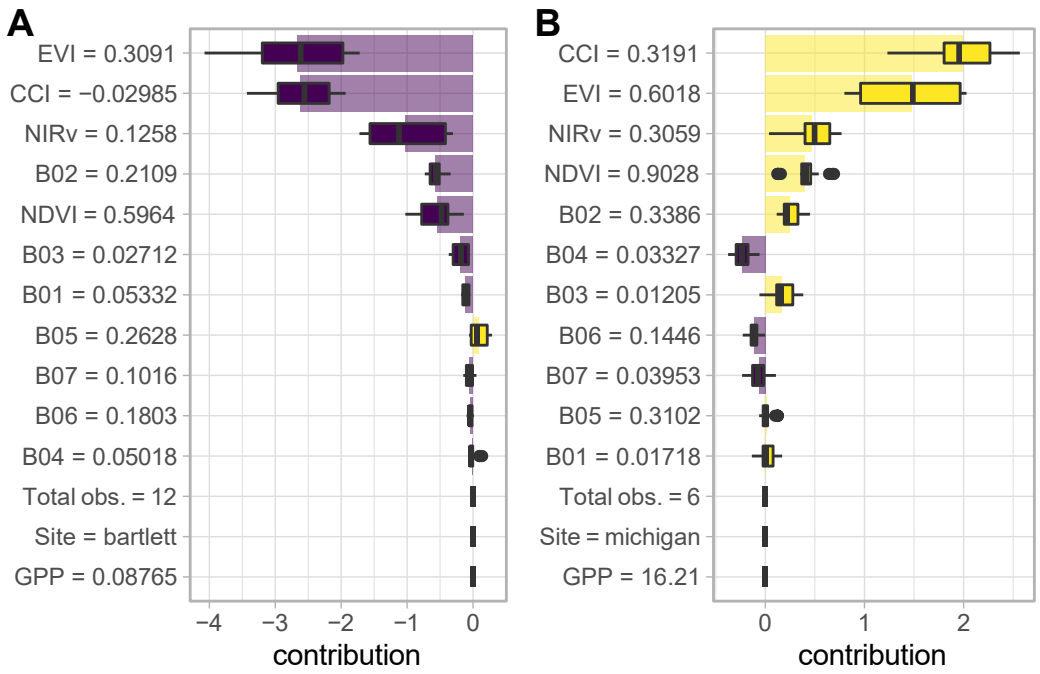

Figure 3.7.: Shapley values derived from the Random forest model for the daily values at 500 m spatial resolution model. Predicted value for the low GPP value is 0.59 (A) and 12.7 for the selected high GPP value (B).

For the weekly model, we selected high and low GPP values of 1.21 gC m^{-2} d^{-1} and 13.17 gC m^{-2} d^{-1}, respectively, resulting in predictions of 1.60 gC m^{-2} d^{-1} and 11.0 gC m^{-2} d^{-1}. For the low GPP value, the most influential variables were CCI, EVI, NIRv, and NDVI, all contributing negatively to the prediction (see Figure 3.8 **A**). Conversely, for the high GPP value, CCI, EVI, NIRv, B02 and NDVI had the most influence on the prediction positively (refer to Figure 3.8 **B**).

In the case of the monthly model, the low GPP value selected was 1.86 gC m^{-2} d^{-1} and the high GPP value was 11.87 gC m^{-2} d^{-1}, leading to predictions of 3.31 gC m^{-2} d^{-1} and 10.60 gC m^{-2} d^{-1} respectively. In both scenarios, CCI, NDVI, and NIRv emerged as the most influential variables, albeit contributing negatively to the low GPP value prediction (See Figure 3.9 **A**) and positively to the high GPP value prediction (See Figure 3.9 **B**).

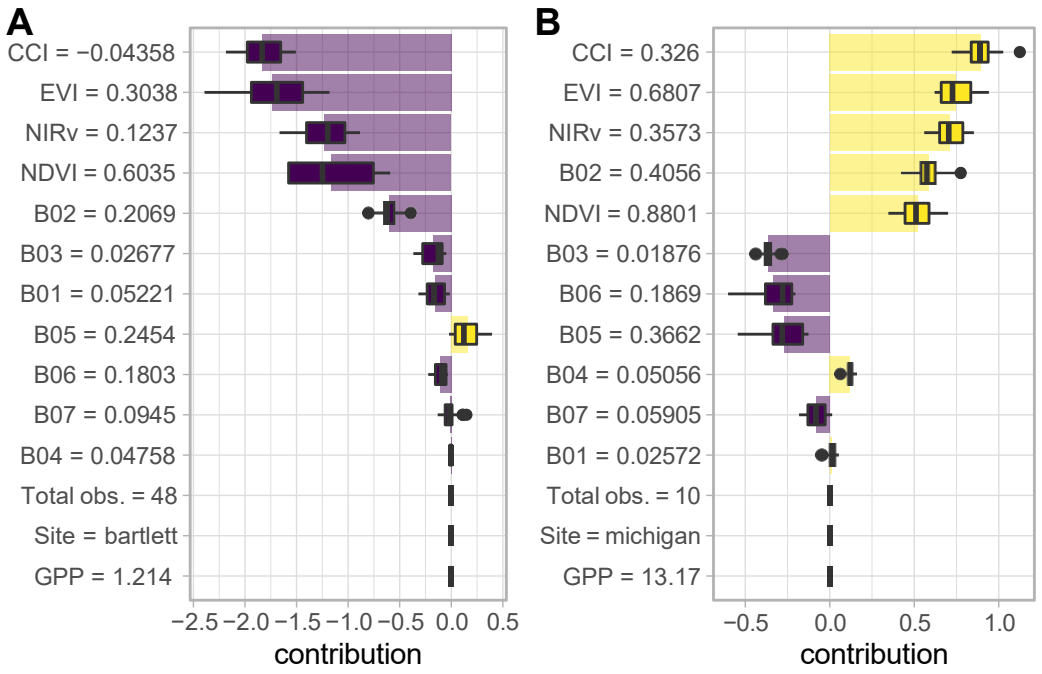

Figure 3.8.: Shapley values derived from the Random forest model for the weekly values at 500 m spatial resolution model. Predicted value for the low GPP value is 1.59 (A) and 11.0 for the selected high GPP value (B).

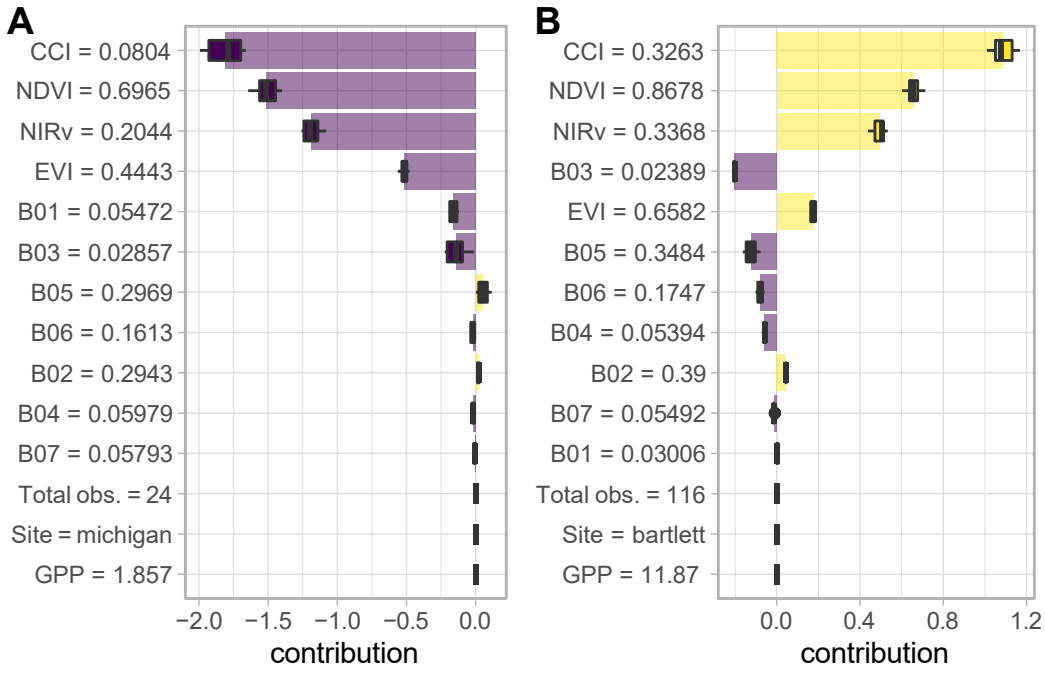

Figure 3.9.: Shapley values derived from the Random forest model for the monthly values at 500 m spatial resolution model. Predicted value for the low GPP value is 3.31 (A) and 10.6 for the selected high GPP value (B).

3.3.2. The potential of AutoML approaches for GPP predictions

The AutoML approach yielded varying performance outcomes across different temporal scales as measured by R^2 and RMSE. The monthly prediction model (Figure 3.10) emerged as the top performer, with an R^2 of 0.76 and a low RMSE of 1.84 gC m^{-2} d^{-1}. Following this, the weekly model (Figure 3.11) demonstrated the second-highest explanatory power, capturing R^2 = 0.72 of the GPP variability, albeit with a slightly elevated RMSE of 3.08 gC m^{-2} d^{-1}. Conversely, the daily model (Figure 3.12) exhibited comparatively diminished performance, explaining 0.67 of the variability with the highest RMSE of 3.11 gC m^{-2} d^{-1}

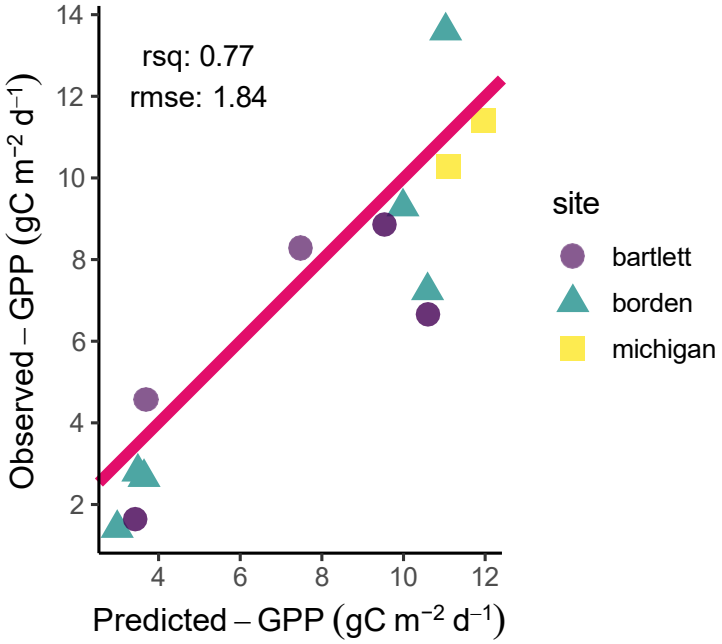

Figure 3.10.: GPP observed and predicted values from the autoML for all the sites at a monthly basis. The red line represents a 1:1 relation. Metrics units are gC m^{-2} d^{-1}

An examination of variable importance in the AutoML model revealed distinctive patterns in significant contributors to GPP prediction across temporal scales. In the daily prediction model, EVI and CCI emerged as the most important (Figure 3.13). Moving to a weekly time frame, the importance of EVI and CCI persisted, with the addition of Band 02 (NIR band) (Figure 3.14). In the monthly prediction model, EVI once again took precedence, accompanied by Band 02 and NIRv (Figure 3.15), highlighting the enduring importance of these variables.

Additionally, a comparative assessment between the regression random forest model and the

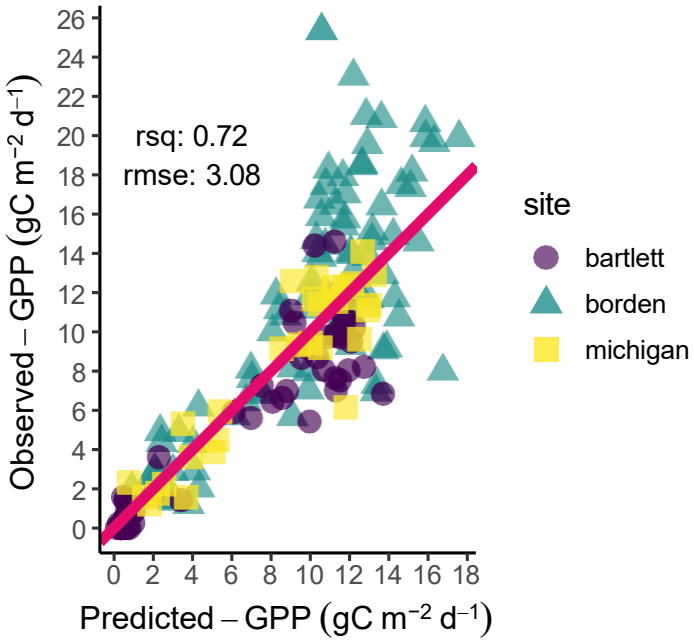

Figure 3.11.: GPP observed and predicted values from the autoML for all the sites at a weekly basis. The red line represents a 1:1 relation. Metrics units are gC m ² d ¹

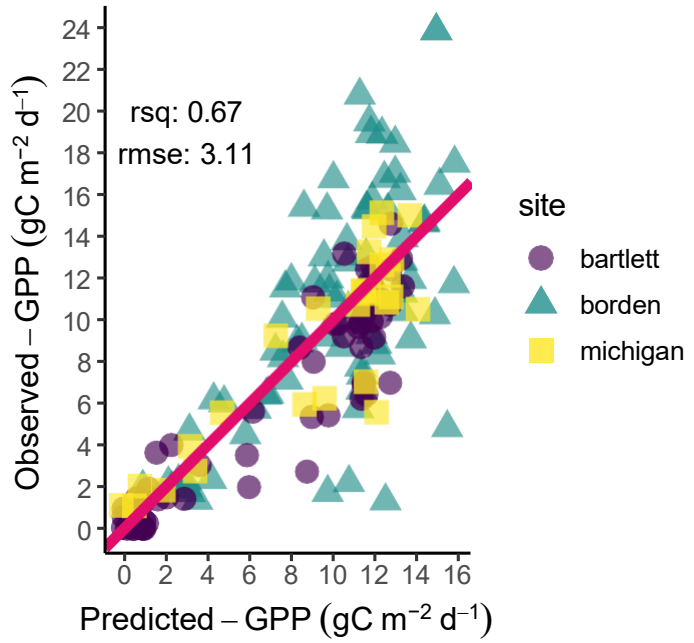

Figure 3.12.: GPP observed and predicted values from the autoML for all the sites at a daily basis. The red line represents a 1:1 relation. Metrics units are gC m ² d ¹

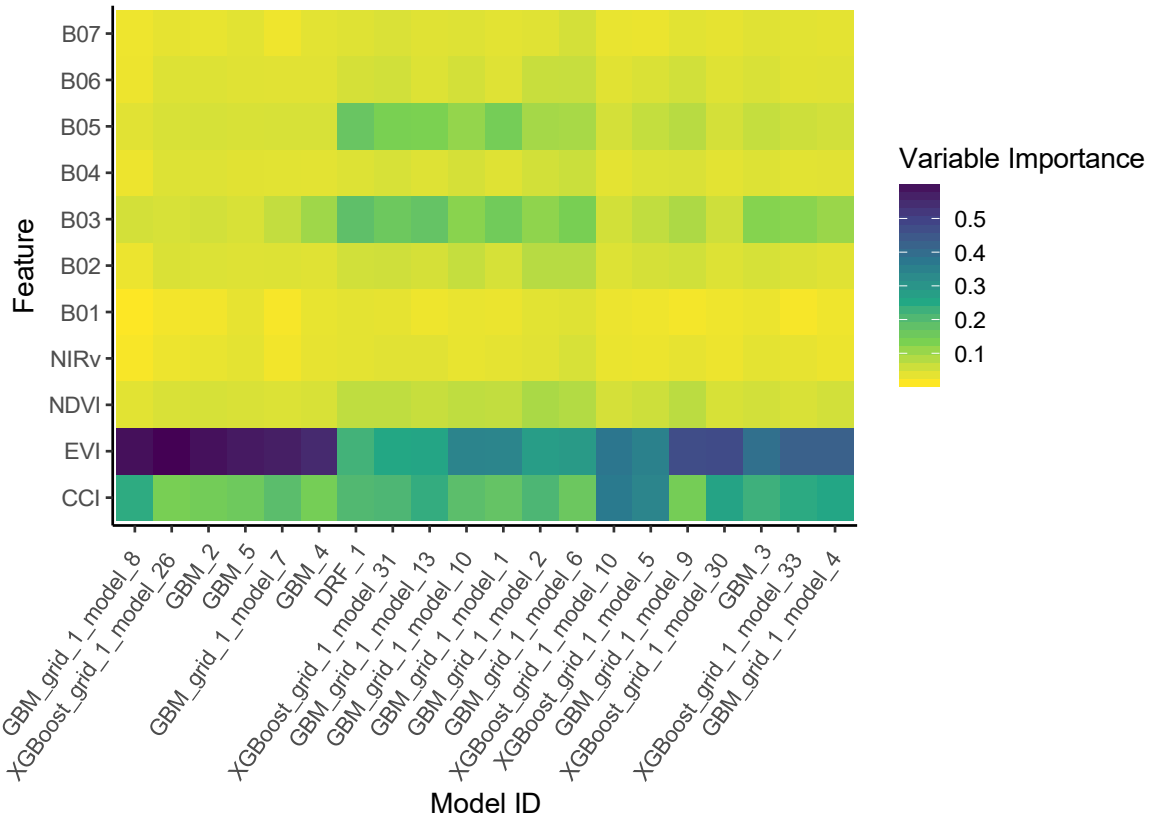

Figure 3.13.: Variable of importance derived from the autoML model for the daily values at 500 m spatial resolution model.

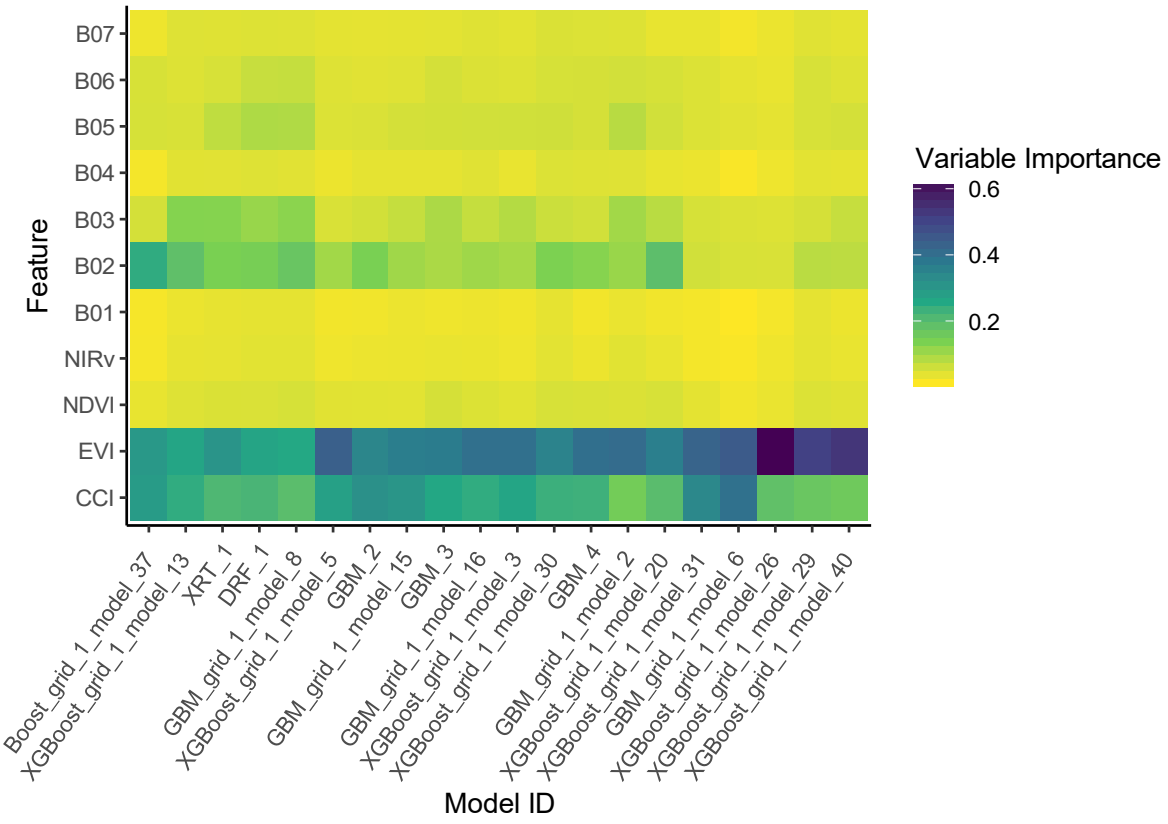

Figure 3.14.: Variable of importance derived from the autoML model for the weekly values at 500 m spatial resolution model.

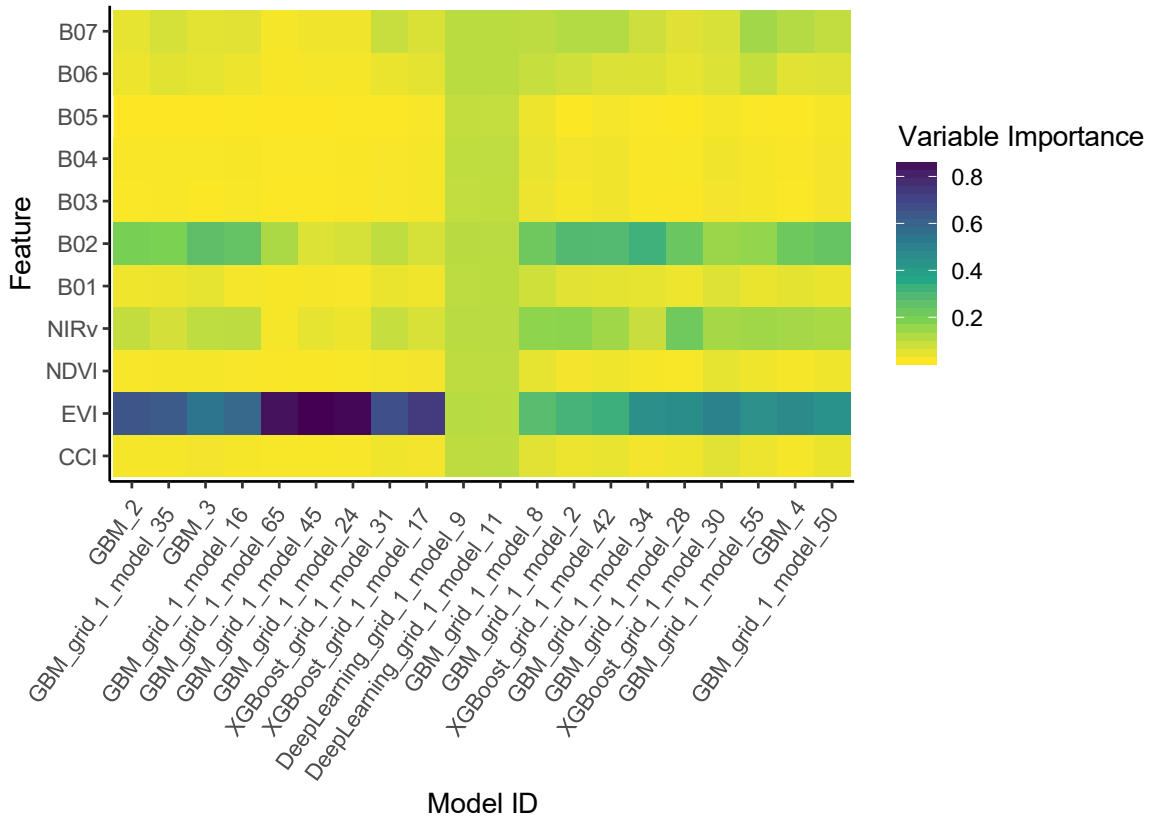

Figure 3.15.: Variable of importance derived from the autoML model for the monthly values at 500 m spatial resolution model.

AutoML model, both applied to the same datasets, revealed nuanced differences in their predictive performance Table 3.2. While the regression random forest model exhibited superior R^2 values in daily and monthly predictions, indicating a better overall fit to the data, the AutoML model demonstrated lower RMSE. Conversely, for weekly predictions, the AutoML model outperformed in both metrics. These findings underscore the importance of considering multiple metrics and temporal scales when evaluating and selecting models for GPP predictions.

Table 3.2.: Summary ML metrics in gC m^{-2} d^{-1}

Variable	RF R2	RF RMSE	automl R2	automl RMSE
Daily	0.70	3.17	0.67	3.11
Weekly	0.55	3.23	0.72	3.08
Monthly	0.81	2.03	0.76	1.84

3.4. Discussion

These results underscore the temporal variations in model performance, with the monthly models having the best performance metrics for both, the regression random forest and the autoML. The lower RMSE of the AutoML model indicates it can potentially result in lower uncertainty for monthly predictions than RF. Given the higher R^2, regression RF captures a larger proportion of the variance in the dependent variable than the autoML. This discrepancy underscores the importance of considering multiple evaluation metrics when assessing model performance. The choice between these models may depend on the specific goals of the analysis, weighing the trade-off between explaining variability and achieving precision in predictions.

Upon examining the six specific predictions analyzed through Shapley values for the Random Forest models across all timeframes, it became evident that high-predicted values of Gross Primary Productivity (GPP) were consistently underestimated, while low GPP values were overestimated. It is crucial to clarify that these particular predictions and Shapley values are specific to each instance and do not represent the entirety of potential predictions. However, when scrutinizing the graphs of predicted GPP values against observed GPP values, especially in the case of the daily model, it is noteworthy that the maximum predicted values hover around 16 $gC\,m^{-2}\,d^{-1}$, whereas observed values can reach up to 24 $gC\,m^{-2}\,d^{-1}$ (See Figure 3.2). Importantly, these values all belong to the Borden site, where the range of observed GPP is notably higher compared to the other two sites.

The discrepancy between the observed values and the predicted values may arise from a potential lag between the change in photosynthetic rate and the concentration of photosynthetic pigments, particularly the change in chlorophyll. Since the predicted GPP values for the constructed models are solely based on reflectance values, the primary changes related to GPP that they can capture are APAR and, to some extent, chlorophyll concentration (Pabon-Moreno et al. 2022). Given that photosynthesis can change rapidly without significant alterations in pigment concentrations, there might be an overestimation of predicted values compared to observed and estimated GPP values at the site mediating EC. In situations of stress, the dynamics of GPP could shift unnoticed solely based on satellite reflectance values.

Conversely, underestimation may be attributed to the well-known phenomenon of satura-

tion. When using indices created with the NIR band for GPP estimation, challenges arise in scenarios where there is a substantial increase in vegetation biomass, which happens during summers in the study sites. In such situations, the dense biomass leads to increased scattering and reflection of radiation. While the NIR band is sensitive to changes in vegetation structure and density, it encounters limitations as the amount of biomass intensifies. This results in a phenomenon known as saturation, where the sensor reaches its maximum capacity to detect changes in reflectance values.

Saturation occurs because the dense vegetation causes a greater proportion of the incoming radiation to be scattered or reflected, particularly in the NIR spectrum (Camps-Valls et al. 2021). As a consequence, the sensor becomes saturated, meaning that further increases in biomass or productivity do not translate proportionally into higher measured reflectance values. This limits the sensor's ability to capture and differentiate changes in the productivity of vegetation beyond a certain threshold.

While, for the monthly model, AutoML demonstrates an acceptable percentage of explained variability, with Random Forest surpassing it by 0.05, for the weekly and daily models, it appears that capturing the inherent variability in the data is challenging solely with the utilized indices and the entirety of MODIS bands. There seems to be a necessity to incorporate additional variables that possibly could further enhance predictive capabilities.

Regarding the question of whether it is preferable to employ a Random Forest (RF) model or an AutoML model, the advantages revealed by the results of this study do not exhibit marked distinctions. The performances are comparable, with the only instance of a notable increase, a $R^2 = 0.17$ improvement in explaining variance, observed in the weekly model by the AutoML model. In the monthly model, Random Forest demonstrates a $R^2 = 0.05$ superiority in terms of explained variance compared to AutoML, although AutoML achieves a marginal improvement in error reduction by 0.198. These nuanced differences suggest that the choice between RF and AutoML may depend on specific considerations, emphasizing the importance of assessing both explanatory power and error metrics for comprehensive model evaluation.

3.5. Conclusions

In summary, the monthly models generated using both methods (RF and autoML) exhibited superior performance based on various metrics, surpassing the outcomes from weekly or daily temporal aggregations. Specifically, in the context of monthly modelling, although RF demonstrated a slightly higher variance explanation compared to autoML (by 0.05), autoML showcased a lower RMSE, signifying more accurate predictions with minimal error. When incorporating VIs alongside all bands into the GPP prediction models (both RF and autoML), the VIs consistently demonstrated their significance as the most influential variables in the predictions, with CCI and EVI consistently having pivotal roles. Notably, among the bands, B02 (NIR) emerged as the most crucial for predictions, surpassing the importance of the remaining bands.

Analyzing the RF model revealed interesting insights through Shapley values, indicating a tendency to underestimate high GPP values and overestimate low GPP values. This observation was consistent across scatterplots depicting predicted versus actual values for diverse temporal aggregations. When comparing both methods (RF and autoML), no definitive superiority emerges, as both exhibit nuanced distinctions, except for the weekly models where autoML outperforms with a 0.17 better variance explanation and a lower RMSE.

3.6. References

Anav, Alessandro, Pierre Friedlingstein, Christian Beer, Philippe Ciais, Anna Harper, Chris Jones, Guillermo Murray-Tortarolo, et al. 2015. "Spatiotemporal Patterns of Terrestrial Gross Primary Production: A Review: GPP Spatiotemporal Patterns." *Reviews of Geophysics* 53 (3): 785–818. https://doi.org/10.1002/2015RG000483.

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, et al. 2010. "Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate." *Science* 329 (5993): 834–38. https://doi.org/10.1126/science.1184984.

Biecek, Przemyslaw. 2018. "DALEX: Explainers for Complex Predictive Models." https://arxiv.org/abs/1806.08915.

Bonan, Gordon B., Peter J. Lawrence, Keith W. Oleson, Samuel Levis, Martin Jung, Markus Reichstein, David M. Lawrence, and Sean C. Swenson. 2011. "Improving Canopy Processes in the Community Land Model Version 4 (CLM4) Using Global Flux Fields Empirically Inferred from FLUXNET Data." *Journal of Geophysical Research* 116 (G2): G02014. https://doi.org/10.1029/2010JG001593.

Camps-Valls, Gustau, Manuel Campos-Taberner, Álvaro Moreno-Martínez, Sophia Walther, Grégory Duveiller, Alessandro Cescatti, Miguel D. Mahecha, et al. 2021. "A Unified Vegetation Index for Quantifying the Terrestrial Biosphere." *Science Advances* 7 (9): eabc7447. https://doi.org/10.1126/sciadv.abc7447.

Coops, N. C., M.-L. Smith, M. E. Martin, and S. V. Ollinger. 2003. "Prediction of Eucalypt Foliage Nitrogen Content from Satellite-Derived Hyperspectral Data." *IEEE Transactions on Geoscience and Remote Sensing* 41 (6): 1338–46. https://doi.org/10.1109/TGRS.2003.813135.

Gaber, Max, Yanghui Kang, Guy Schurgers, and Trevor Keenan. 2023. "Using Automated Machine Learning for the Upscaling of Gross Primary Productivity." *Biogeosciences Discussions* 2023: 1–26.

James, Gareth, Daniela Witten, Trevor Hastie, and Rob Tibshirani. 2022. "An Introduction to Statistical Learning with Applications in R: By Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, eISBN: 978-1-4614-7137-7." *Statistical Theory and Related Fields* 6 (1): 87–87. https://doi.org/10.1080/24754269.2021.1980261.

Jung, Martin, Markus Reichstein, Philippe Ciais, Sonia I. Seneviratne, Justin Sheffield,

Michael L. Goulden, Gordon Bonan, et al. 2010. "Recent Decline in the Global Land Evapotranspiration Trend Due to Limited Moisture Supply." *Nature* 467 (7318): 951–54. https://doi.org/10.1038/nature09396.

Jung, Martin, Markus Reichstein, Hank A. Margolis, Alessandro Cescatti, Andrew D. Richardson, M. Altaf Arain, Almut Arneth, et al. 2011. "Global Patterns of Land-Atmosphere Fluxes of Carbon Dioxide, Latent Heat, and Sensible Heat Derived from Eddy Covariance, Satellite, and Meteorological Observations." *Journal of Geophysical Research* 116 (September): G00J07. https://doi.org/10.1029/2010JG001566.

Jung, M, M Reichstein, and A Bondeau. 2009. "Towards Global Empirical Upscaling of FLUXNET Eddy Covariance Observations: Validation of a Model Tree Ensemble Approach Using a Biosphere Model."

Lary, David J., Amir H. Alavi, Amir H. Gandomi, and Annette L. Walker. 2016. "Machine Learning in Geosciences and Remote Sensing." *Geoscience Frontiers* 7 (1): 3–10. https://doi.org/10.1016/j.gsf.2015.07.003.

LeDell, Erin, and Sebastien Poirier. 2020. "H2o Automl: Scalable Automatic Machine Learning." In *Proceedings of the AutoML Workshop at ICML*. Vol. 2020. ICML.

Lehnert, Lukas W., Hanna Meyer, Yun Wang, Georg Miehe, Boris Thies, Christoph Reudenbach, and Jörg Bendix. 2015. "Retrieval of Grassland Plant Coverage on the Tibetan Plateau Based on a Multi-Scale, Multi-Sensor and Multi-Method Approach." *Remote Sensing of Environment* 164 (July): 197–207. https://doi.org/10.1016/j.rse.2015.04.020.

Lundberg, Scott M, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*.

Meyer, Hanna, Christoph Reudenbach, Tomislav Hengl, Marwan Katurji, and Thomas Nauss. 2018. "Improving Performance of Spatio-Temporal Machine Learning Models Using Forward Feature Selection and Target-Oriented Validation." *Environmental Modelling & Software* 101 (March): 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

Molnar, Christoph. 2020. *Interpretable Machine Learning*. Lulu. com.

Pabon-Moreno, Daniel E., Mirco Migliavacca, Markus Reichstein, and Miguel D. Mahecha. 2022. "On the Potential of Sentinel-2 for Estimating Gross Primary Production." *IEEE Transactions on Geoscience and Remote Sensing*, 1–1. https://doi.org/10.1109/TGRS.2022.3152272.

Papale, Dario, T. Andrew Black, Nuno Carvalhais, Alessandro Cescatti, Jiquan Chen, Martin Jung, Gerard Kiely, et al. 2015. "Effect of Spatial Sampling from European Flux

Towers for Estimating Carbon and Water Fluxes with Artificial Neural Networks." *Journal of Geophysical Research: Biogeosciences* 120 (10): 1941–57. https://doi.org/10.1002/2015JG002997.

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, et al. 2020. "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." *Scientific Data* 7 (1): 1–27.

Reichstein, Markus, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566 (7743): 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Tramontana, Gianluca, Kazuito Ichii, Gustau Camps-Valls, Enrico Tomelleri, and Dario Papale. 2015. "Uncertainty Analysis of Gross Primary Production Upscaling Using Random Forests, Remote Sensing and Eddy Covariance Data." *Remote Sensing of Environment* 168 (October): 360–73. https://doi.org/10.1016/j.rse.2015.07.015.

Tramontana, Gianluca, Martin Jung, Christopher R. Schwalm, Kazuhito Ichii, Gustau Camps-Valls, Botond Ráduly, Markus Reichstein, et al. 2016. "Predicting Carbon Dioxide and Energy Fluxes Across Global FLUXNET Sites Withregression Algorithms." *Biogeosciences* 13 (14): 4291–4313. https://doi.org/10.5194/bg-13-4291-2016.

Vermote, Wolfe, E. 2021. "MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061." *NASA EOSDIS Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MOD09GA.061.

Verrelst, J., L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno. 2012. "Retrieval of Vegetation Biophysical Parameters Using Gaussian Process Techniques." *IEEE Transactions on Geoscience and Remote Sensing* 50 (5): 1832–43. https://doi.org/10.1109/TGRS.2011.2168962.

Verrelst, Jochem, Jordi Muñoz, Luis Alonso, Jesús Delegido, Juan Pablo Rivera, Gustavo Camps-Valls, and José Moreno. 2012. "Machine Learning Regression Algorithms for Biophysical Parameter Retrieval: Opportunities for Sentinel-2 and -3." *Remote Sensing of Environment* 118 (March): 127–39. https://doi.org/10.1016/j.rse.2011.11.002.

Wright, Marvin N., and Andreas Ziegler. 2017. "**Ranger** : A Fast Implementation of Random Forests for High Dimensional Data in *C++* and *r*." *Journal of Statistical Software* 77 (1). https://doi.org/10.18637/jss.v077.i01.

4. Conclusions and Future Work

4.1. Synthesis

This research evaluated the advantages and limitations associated with estimating GPP using traditional VIs like NDVI, EVI, and recent ones like CCI and NIRv. Specifically, the study focused on VIs derived from MODIS, chosen for its extensive time records. To assess GPP estimates through the VIs, we employed traditional methods such as linear regression and GAM, as well as ML techniques.

Traditional methods, such as LM and GAM, offer the advantage of being more straightforward to interpret in relation to their predictions and associated errors. In contrast, ML techniques like regression RF and autoML pursue predictions through methods that are harder to elucidate, a phenomenon referred to as "black model boxes." Despite their complexity, these ML methods have the potential to achieve superior accuracy in their predictions.

To evaluate the capacity of VIs to estimate GPP,Productivity (GPP), our analysis spanned three temporal aggregations: daily, weekly, and monthly. This assessment considered the utilization of VIs individually or collectively as covariates in the model. For ML methods, the same temporal aggregations were employed, with an assessment involving all indices and the entire set of MODIS bands as predictors.

The results indicate that, across all models, monthly temporal aggregation consistently yields better outcomes in comparison to weekly and daily aggregations. This superiority can be attributed to the reduced variation present in monthly aggregated data, resulting in lower errors in the models. According to LM and GAM, incorporating all indices as covariates in the models enhances predictive capability, suggesting that a single VI in isolation may fall short of capturing the entire variation in the data.

Examining the outcomes of models utilizing individual VIs reveals minimal differences. Nonetheless, VIs consistently displaying a slightly superior performance in terms of explained variation and lower error are typically EVI and CCI, whereas NDVI tends to exhibit lower performance. This trend is also evident in ML models, with EVI and CCI frequently occupying top positions in variable of importance rankings. In ML models, the inclusion of all bands as predictors highlights that B02 (NIR) often ranks among the top variables in importance over other bands.

In comparing the various models and their configurations (LM and GAM with VIs as co-variates; RF and autoML with VIs and complete bands as predictors), it becomes evident that ML methods exhibit a better performance across different time aggregations. For daily models, assessed by R^2 and RMSE metrics, RF surpassed autoML, GAM, and LM, in that sequence. For the weekly models, autoML led, followed by GAM, LM, and RF. Regarding the monthly models, RF had better performance, followed by LM, autoML, and GAM.

4.2. Limitations

The uncertainties in VIs, in this case, for GPP estimation, may be due to their inability to capture variations in GPP or inherent limitations falling into two categories: artifacts or external factors. Artifacts originate from the sensors used for deriving VIs, encompassing calibration, quality control, or sensor degradation (Zeng et al. 2022). In this study, we addressed these effects by excluding pixels with suboptimal quality using data from quality assurance (QA) and quality control (QC). These variables offer insights into the quality of each pixel, aiding in the selection of pixels potentially harboring erroneous values. However, the accuracy of these data may be compromised by the presence of pixels with poor quality not identified as such by postprocessing algorithms, or influenced by external factors such as clouds, cloud shadows, inaccuracies in snow detection, atmospheric pollution, or technical issues in image acquisition (Pesquer, Domingo-Marimon, and Pons 2019).

Another potential source of uncertainty in GPP estimates is related to the representation of the flux tower footprint within the pixel area. EC towers exhibit a flux footprint that can vary over time. The footprint is defined as the extent to which measurements taken at a specific time and location accurately reflect the actual flows in time and specific area. The monthly climatologies of the footprint can exhibit variations, usually falling within the range

of 100 m to 450 m. (Chu et al. 2021). The heterogeneity of the landscape beyond this footprint may affect flow conditions differently than those measured within the footprint. In our methodology, we established a square area of 3 km² around the EC tower to ensure that the flux footprint climatology aligns with this region, considering the homogeneity in land cover type. However, the resolution of MODIS might not be adequate for capturing nuanced fine-scale spatial variations (Robinson et al. 2018).

The limitations of the study include the unavailability of data from a sensor like Sentinel-2, which offers better spatial resolution compared to the MODIS product employed. This data could have allowed us to explore whether, despite estimating in areas with homogeneous land cover, it is feasible to capture nuanced fine-scale spatial variations. Such an exploration could potentially lead to improved GPP estimates, thereby reducing uncertainty. The restriction in using Sentinel-2 data stems from the fact that calibrated data is accessible only from 2017 onward. However, the GPP data processed by ONEFlux for each site commenced in 2015, concluding for Bartlett and Michigan in 2017 and 2018, respectively, thus resulting in the exclusion of a significant portion of the data.

Finally, it is important to acknowledge that the sites selection can constitute a potential limitation in this study. The focus on deciduous broadleaf forest sites in the northern hemisphere, encompassing only three locations, restricts the generalization of the findings to this specific ecosystem type. Expanding the scope to include more sites representing other various ecosystem types could further contribute to understanding which indices have limitations or can perform better in estimating GPP across a broader range of scenarios.

4.3. Future work

As uncertainties persist in GPP estimation models relying on VIs as predictors, the potential for improvement lies in models that can consider critical stress factors influencing photosynthesis (Rogers et al. 2020; Ryu, Berry, and Baldocchi 2019; Xiao et al. 2019). While incorporating meteorological data can address this, it poses challenges in various regions around the globe. In such circumstances, a more advantageous approach involves the utilization and validation of VIs that demonstrate sensitivity to stress factors affecting photosynthetic activity, not just photosynthetic capacity. Remote sensing provides a means to achieve this by integrating data products such as thermal information, which can indirectly address water deficit

in ecosystems, for example (Pabon-Moreno et al. 2022). Despite potential uncertainties associated with these products, there is a prospect for enhanced estimations by incorporating considerations beyond traditional reflectance values.

Addressing the inherent variability across ecosystems, it is essential to recognize that no single VI will be the best, and their representation of vegetation functioning has limitations (Zeng et al. 2022). Nonetheless, the continual introduction of new VIs derived from both traditional and advanced sensors (Montero et al. 2023) presents an opportunity to enhance the estimation of GPP on a global level. A potentially effective strategy would involve establishing a standardized guide of VIs or combinations tailored to different ecosystem types, providing researchers with a reliable foundation for accurate assessments. Furthermore, the field could benefit from adopting a qualitative rating of model performance, such as the one used by for the FAO model AquaCrop (Raes et al. 2022). This practice involves standardizing model evaluations based on their performance within specific research domains. Such standardized benchmarks can provide valuable guidance for researchers seeking optimal model selection and interpretation within their respective contexts.

Given the growing reliance on ML methods to enhance GPP predictions in recent years, a notable challenge arises from their tendency to obscure the understanding of the underlying mechanisms explaining the relationship between VIs and GPP (Molnar 2020). Hence, it becomes crucial for new studies to articulate their research objectives with clarity and consistency. Specifically, researchers should explicitly state whether their primary focus is to explore mechanistic relationships between VIs or if their aim is to improve GPP predictions. This distinction is vital, whether the goal is to predict GPP in sites lacking validation data (flux towers) or to advance forecasting capabilities (Meyer et al. 2019).

The implementation of explainable ML methods could enhance our understanding of the variables influencing predictions (Molnar 2020). Additionally, it is well-known that a prediction model's ability to generalize the rules it has learned from the training data set to a new (unseen) data set is poor. Diverse scenarios, such as droughts, intense rainfall, and vegetation mortality, among others, can affect predictive capacity. Therefore, consistent model predictive power assessment over time and retraining as necessary are crucial (Kuhn and Silge 2022). This underscores the significance of having flux towers and extending their presence to underrepresented sites to acquire validation data. These efforts contribute to refining future global GPP estimation models on a broader scale (Meyer et al. 2019).

4.4. References

Chu, Housen, Xiangzhong Luo, Zutao Ouyang, W. Stephen Chan, Sigrid Dengel, Sébastien C. Biraud, Margaret S. Torn, et al. 2021. "Representativeness of Eddy-Covariance Flux Footprints for Areas Surrounding AmeriFlux Sites." *Agricultural and Forest Meteorology* 301-302 (May): 108350. https://doi.org/10.1016/j.agrformet.2021.108350.

Kuhn, Max, and Julia Silge. 2022. *Tidy Modeling with r*. " O'Reilly Media, Inc.".

Meyer, Hanna, Christoph Reudenbach, Stephan Wöllauer, and Thomas Nauss. 2019. "Importance of Spatial Predictor Variable Selection in Machine Learning Applications – Moving from Data Reproduction to Spatial Prediction." *Ecological Modelling* 411 (November): 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815.

Molnar, Christoph. 2020. *Interpretable Machine Learning*. Lulu. com.

Montero, David, César Aybar, Miguel D. Mahecha, Francesco Martinuzzi, Maximilian Söchting, and Sebastian Wieneke. 2023. "A Standardized Catalogue of Spectral Indices to Advance the Use of Remote Sensing in Earth System Research." *Scientific Data* 10 (1): 197. https://doi.org/10.1038/s41597-023-02096-0.

Pabon-Moreno, Daniel E., Mirco Migliavacca, Markus Reichstein, and Miguel D. Mahecha. 2022. "On the Potential of Sentinel-2 for Estimating Gross Primary Production." *IEEE Transactions on Geoscience and Remote Sensing*, 1–1. https://doi.org/10.1109/TGRS.2022.3152272.

Pesquer, Lluís, Cristina Domingo-Marimon, and Xavier Pons. 2019. "Spatial and Spectral Pattern Identification for the Automatic Selection of High-Quality MODIS Images." *Journal of Applied Remote Sensing* 13 (01): 1. https://doi.org/10.1117/1.JRS.13.014510.

Raes, Dirl, Pasquale Steduto, Theodore Hsiao, and Elias Fereres. 2022. "Reference Manual, Chapter 2 – AquaCrop, Version 7.0 - August 2022."

Robinson, Nathaniel P., Brady W. Allred, William K. Smith, Matthew O. Jones, Alvaro Moreno, Tyler A. Erickson, David E. Naugle, and Steven W. Running. 2018. "Terrestrial Primary Production for the Conterminous United States Derived from Landsat 30 m and <Span Style="font-Variant:small-Caps;">MODIS</Span> 250 m." Edited by Nathalie Pettorelli and Jose Paruelo. *Remote Sensing in Ecology and Conservation* 4 (3): 264–80. https://doi.org/10.1002/rse2.74.

Rogers, Cheryl A., Jing M. Chen, Ting Zheng, Holly Croft, Alemu Gonsamo, Xiangzhong Luo, and Ralf M. Staebler. 2020. "The Response of Spectral Vegetation Indices and Solar-Induced Fluorescence to Changes in Illumination Intensity and Geometry in the Days

Surrounding the 2017 North American Solar Eclipse." *Journal of Geophysical Research: Biogeosciences* 125 (10). https://doi.org/10.1029/2020JG005774.

Ryu, Youngryel, Joseph A. Berry, and Dennis D. Baldocchi. 2019. "What Is Global Photosynthesis? History, Uncertainties and Opportunities." *Remote Sensing of Environment* 223 (March): 95–114. https://doi.org/10.1016/j.rse.2019.01.016.

Xiao, Jingfeng, Frederic Chevallier, Cecile Gomez, Luis Guanter, Jeffrey A. Hicke, Alfredo R. Huete, Kazuhito Ichii, et al. 2019. "Remote Sensing of the Terrestrial Carbon Cycle: A Review of Advances over 50 Years." *Remote Sensing of Environment* 233 (November): 111383. https://doi.org/10.1016/j.rse.2019.111383.

Zeng, Yelu, Dalei Hao, Alfredo Huete, Benjamin Dechant, Joe Berry, Jing M. Chen, Joanna Joiner, et al. 2022. "Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally." *Nature Reviews Earth & Environment* 3 (7): 477–93. https://doi.org/10.1038/s43017-022-00298-5.

References

Anav, Alessandro, Pierre Friedlingstein, Christian Beer, Philippe Ciais, Anna Harper, Chris Jones, Guillermo Murray-Tortarolo, et al. 2015. "Spatiotemporal Patterns of Terrestrial Gross Primary Production: A Review: GPP Spatiotemporal Patterns." *Reviews of Geophysics* 53 (3): 785–818. https://doi.org/10.1002/2015RG000483.

Ashton, Mark S., Mary L. Tyrrell, Deborah Spalding, and Bradford Gentry, eds. 2012. *Managing Forest Carbon in a Changing Climate*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2232-3.

Badgley, Grayson, Leander D. L. Anderegg, Joseph A. Berry, and Christopher B. Field. 2019. "Terrestrial Gross Primary Production: Using NIR v to Scale from Site to Globe." *Global Change Biology* 25 (11): 3731–40. https://doi.org/10.1111/gcb.14729.

Badgley, Grayson, Christopher B. Field, and Joseph A. Berry. 2017. "Canopy Near-Infrared Reflectance and Terrestrial Photosynthesis." *Science Advances* 3 (3): e1602244. https://doi.org/10.1126/sciadv.1602244.

Baldocchi, Dennis D. 2020b. "How Eddy Covariance Flux Measurements Have Contributed to Our Understanding of *Global Change Biology*." *Global Change Biology* 26 (1): 242–60. https://doi.org/10.1111/gcb.14807.

———. 2020a. "How Eddy Covariance Flux Measurements Have Contributed to Our Understanding of *Global Change Biology*." *Global Change Biology* 26 (1): 242–60. https://doi.org/10.1111/gcb.14807.

Balzarolo, Manuela, Josep Peñuelas, and Frank Veroustraete. 2019. "Influence of Landscape Heterogeneity and Spatial Resolution in Multi-Temporal In Situ and MODIS NDVI Data Proxies for Seasonal GPP Dynamics." *Remote Sensing* 11 (14): 1656. https://doi.org/10.3390/rs11141656.

Beer, Christian, Markus Reichstein, Enrico Tomelleri, Philippe Ciais, Martin Jung, Nuno Carvalhais, Christian Rödenbeck, et al. 2010. "Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate." *Science* 329 (5993): 834–38. https://doi.org/10.1126/science.1184984.

Biecek, Przemyslaw. 2018. "DALEX: Explainers for Complex Predictive Models." https://arxiv.org/abs/1806.08915.

Bonan, Gordon B., Peter J. Lawrence, Keith W. Oleson, Samuel Levis, Martin Jung, Markus

Reichstein, David M. Lawrence, and Sean C. Swenson. 2011. "Improving Canopy Processes in the Community Land Model Version 4 (CLM4) Using Global Flux Fields Empirically Inferred from FLUXNET Data." *Journal of Geophysical Research* 116 (G2): G02014. https://doi.org/10.1029/2010JG001593.

Brown, Luke A., Fernando Camacho, Vicente García-Santos, Niall Origo, Beatriz Fuster, Harry Morris, Julio Pastor-Guzman, et al. 2021. "Fiducial Reference Measurements for Vegetation Bio-Geophysical Variables: An End-to-End Uncertainty Evaluation Framework." *Remote Sensing* 13 (16): 3194. https://doi.org/10.3390/rs13163194.

Camps-Valls, Gustau, Manuel Campos-Taberner, Álvaro Moreno-Martínez, Sophia Walther, Grégory Duveiller, Alessandro Cescatti, Miguel D. Mahecha, et al. 2021. "A Unified Vegetation Index for Quantifying the Terrestrial Biosphere." *Science Advances* 7 (9): eabc7447. https://doi.org/10.1126/sciadv.abc7447.

Chu, Housen, Xiangzhong Luo, Zutao Ouyang, W. Stephen Chan, Sigrid Dengel, Sébastien C. Biraud, Margaret S. Torn, et al. 2021. "Representativeness of Eddy-Covariance Flux Footprints for Areas Surrounding AmeriFlux Sites." *Agricultural and Forest Meteorology* 301-302 (May): 108350. https://doi.org/10.1016/j.agrformet.2021.108350.

Coops, N. C., M.-L. Smith, M. E. Martin, and S. V. Ollinger. 2003. "Prediction of Eucalypt Foliage Nitrogen Content from Satellite-Derived Hyperspectral Data." *IEEE Transactions on Geoscience and Remote Sensing* 41 (6): 1338–46. https://doi.org/10.1109/TGRS.2003.813135.

Dechant, Benjamin, Youngryel Ryu, Grayson Badgley, Philipp Köhler, Uwe Rascher, Mirco Migliavacca, Yongguang Zhang, et al. 2022. "NIRVP: A Robust Structural Proxy for Sun-Induced Chlorophyll Fluorescence and Photosynthesis Across Scales." *Remote Sensing of Environment* 268 (January): 112763. https://doi.org/10.1016/j.rse.2021.112763.

Donato, David. 1997. "Landsat Data - A Brief History of the Landsat Program," December.

Fang, Hongliang, Shanshan Wei, Chongya Jiang, and Klaus Scipal. 2012. "Theoretical Uncertainty Analysis of Global MODIS, CYCLOPES, and GLOBCARBON LAI Products Using a Triple Collocation Method." *Remote Sensing of Environment* 124 (September): 610–21. https://doi.org/10.1016/j.rse.2012.06.013.

Fernández-Martínez, Marcos, Rong Yu, John Gamon, Gabriel Hmimina, Iolanda Filella, Manuela Balzarolo, Benjamin Stocker, and Josep Peñuelas. 2019. "Monitoring Spatial and Temporal Variabilities of Gross Primary Production Using MAIAC MODIS Data." *Remote Sensing* 11 (7): 874. https://doi.org/10.3390/rs11070874.

Gaber, Max, Yanghui Kang, Guy Schurgers, and Trevor Keenan. 2023. "Using Automated

Machine Learning for the Upscaling of Gross Primary Productivity." *Biogeosciences Discussions* 2023: 1–26.

Gamon, John A., K. Fred Huemmrich, Christopher Y. S. Wong, Ingo Ensminger, Steven Garrity, David Y. Hollinger, Asko Noormets, and Josep Peñuelas. 2016. "A Remotely Sensed Pigment Index Reveals Photosynthetic Phenology in Evergreen Conifers." *Proceedings of the National Academy of Sciences* 113 (46): 13087–92. https://doi.org/10.1073/pnas.1606162113.

Gao, X. 2000. "Optical–Biophysical Relationships of Vegetation Spectra Without Background Contamination." *Remote Sensing of Environment* 74 (3): 609–20. https://doi.org/10.1016/S0034-4257(00)00150-4.

GCOS. 2016. "The Global Observing System for Climate Implementation Needs." GCOS 200. Geneva: World Meteorological Organization (WMO); United Nations Educational, Scientific; Cultural Organization (UNESCO); Intergovernmental Oceanographic Commission (IOC); United Nations Environment Programme (UNEP) ; International Council of Scientific Unions (ICSU). https://library.wmo.int/idurl/4/55469.

Gensheimer, Johannes, Alexander J. Turner, Philipp Köhler, Christian Frankenberg, and Jia Chen. 2022. "A Convolutional Neural Network for Spatial Downscaling of Satellite-Based Solar-Induced Chlorophyll Fluorescence (SIFnet)." *Biogeosciences* 19 (6): 1777–93. https://doi.org/10.5194/bg-19-1777-2022.

Glenn, Edward, Alfredo Huete, Pamela Nagler, and Stephen Nelson. 2008. "Relationship Between Remotely-Sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape." *Sensors* 8 (4): 2136–60. https://doi.org/10.3390/s8042136.

Goetz, Scott J, Stephen D Prince, Samuel N Goward, Michelle M Thawley, Jennifer Small, and Andrew Johnston. 1999. "Mapping Net Primary Production and Related Biophysical Variables with Remote Sensing: Application to the BOREAS Region." *Journal of Geophysical Research: Atmospheres* 104 (D22): 27719–34. https://doi.org/10.1029/1999JD900269.

Gough, Christopher M., Gil Bohrer, Brady S. Hardiman, Lucas E. Nave, Christoph S. Vogel, Jeff W. Atkins, Ben Bond-Lamberty, et al. 2021. "Disturbance-accelerated Succession Increases the Production of a Temperate Forest." *Ecological Applications* 31 (7). https://doi.org/10.1002/eap.2417.

Gough, Christopher M., Christoph S. Vogel, Brady Hardiman, and Peter S. Curtis. 2010. "Wood Net Primary Production Resilience in an Unmanaged Forest Transitioning from

Early to Middle Succession." *Forest Ecology and Management* 260 (1): 36–41. https://doi.org/10.1016/j.foreco.2010.03.027.

Gough, Christopher, Gil Bohrer, and Peter Curtis. 2016. "AmeriFlux AmeriFlux US-UMB Univ. Of Mich. Biological Station." Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….

Guan, Xiaobin, Jing M. Chen, Huanfeng Shen, Xinyao Xie, and Jianbo Tan. 2022. "Comparison of Big-Leaf and Two-Leaf Light Use Efficiency Models for GPP Simulation After Considering a Radiation Scalar." *Agricultural and Forest Meteorology* 313 (February): 108761. https://doi.org/10.1016/j.agrformet.2021.108761.

Harris, Nancy L., David A. Gibbs, Alessandro Baccini, Richard A. Birdsey, Sytze De Bruin, Mary Farina, Lola Fatoyinbo, et al. 2021. "Global Maps of Twenty-First Century Forest Carbon Fluxes." *Nature Climate Change* 11 (3): 234–40. https://doi.org/10.1038/s41558-020-00976-6.

Heinsch, F. A., Maosheng Zhao, S. W. Running, J. S. Kimball, R. R. Nemani, K. J. Davis, P. V. Bolstad, et al. 2006. "Evaluation of Remote Sensing Based Terrestrial Productivity from MODIS Using Regional Tower Eddy Flux Network Observations." *IEEE Transactions on Geoscience and Remote Sensing* 44 (7): 1908–25. https://doi.org/10.1109/TGRS.2005.853936.

Houborg, Rasmus, Joshua B. Fisher, and Andrew K. Skidmore. 2015. "Advances in Remote Sensing of Vegetation Function and Traits." *International Journal of Applied Earth Observation and Geoinformation* 43 (December): 1–6. https://doi.org/10.1016/j.jag.2015.06.001.

Huete, Alfredo. 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25 (3): 295–309.

Huete, Alfredo, Kamel Didan, Willem Van Leeuwen, Tomoaki Miura, and Ed Glenn. 2010. "MODIS Vegetation Indices." In *Land Remote Sensing and Global Environmental Change*, edited by Bhaskar Ramachandran, Christopher O. Justice, and Michael J. Abrams, 11:579–602. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-6749-7_26.

Huete, Alfredo, K Didan, T Miura, E. P Rodriguez, X Gao, and L. G Ferreira. 2002. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." *Remote Sensing of Environment* 83 (1-2): 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

James, Gareth, Daniela Witten, Trevor Hastie, and Rob Tibshirani. 2022. "An Introduction

to Statistical Learning with Applications in R: By Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, eISBN: 978-1-4614-7137-7." *Statistical Theory and Related Fields* 6 (1): 87–87. https://doi.org/10.1080/24754269.2021.1980261.

Jung, Martin, Markus Reichstein, Philippe Ciais, Sonia I. Seneviratne, Justin Sheffield, Michael L. Goulden, Gordon Bonan, et al. 2010. "Recent Decline in the Global Land Evapotranspiration Trend Due to Limited Moisture Supply." *Nature* 467 (7318): 951–54. https://doi.org/10.1038/nature09396.

Jung, Martin, Markus Reichstein, Hank A. Margolis, Alessandro Cescatti, Andrew D. Richardson, M. Altaf Arain, Almut Arneth, et al. 2011. "Global Patterns of Land-Atmosphere Fluxes of Carbon Dioxide, Latent Heat, and Sensible Heat Derived from Eddy Covariance, Satellite, and Meteorological Observations." *Journal of Geophysical Research* 116 (September): G00J07. https://doi.org/10.1029/2010JG001566.

Jung, M, M Reichstein, and A Bondeau. 2009. "Towards Global Empirical Upscaling of FLUXNET Eddy Covariance Observations: Validation of a Model Tree Ensemble Approach Using a Biosphere Model."

Köhler, Philipp, Christian Frankenberg, Troy S. Magney, Luis Guanter, Joanna Joiner, and Jochen Landgraf. 2018. "Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2." *Geophysical Research Letters* 45 (19): 10, 456–10, 463. https://doi.org/10.1029/2018GL079031.

Kuhn, Max, and Julia Silge. 2022. *Tidy Modeling with r*. " O'Reilly Media, Inc.".

Lary, David J., Amir H. Alavi, Amir H. Gandomi, and Annette L. Walker. 2016. "Machine Learning in Geosciences and Remote Sensing." *Geoscience Frontiers* 7 (1): 3–10. https://doi.org/10.1016/j.gsf.2015.07.003.

LeDell, Erin, and Sebastien Poirier. 2020. "H2o Automl: Scalable Automatic Machine Learning." In *Proceedings of the AutoML Workshop at ICML*. Vol. 2020. ICML.

Lee, Xuhui, Jose D. Fuentes, Ralf M. Staebler, and Harold H. Neumann. 1999. "Long-Term Observation of the Atmospheric Exchange of CO ₂ with a Temperate Deciduous Forest in Southern Ontario, Canada." *Journal of Geophysical Research: Atmospheres* 104 (D13): 15975–84. https://doi.org/10.1029/1999JD900227.

Lehnert, Lukas W., Hanna Meyer, Yun Wang, Georg Miehe, Boris Thies, Christoph Reudenbach, and Jörg Bendix. 2015. "Retrieval of Grassland Plant Coverage on the Tibetan Plateau Based on a Multi-Scale, Multi-Sensor and Multi-Method Approach." *Remote*

Sensing of Environment 164 (July): 197–207. https://doi.org/10.1016/j.rse.2015.04.020.

Lieth, Helmut. 1975. "Modeling the Primary Productivity of the World." *Primary Productivity of the Biosphere*, 237–63.

Lin, Shangrong, Jing Li, Qinhuo Liu, Beniamino Gioli, Eugenie Paul-Limoges, Nina Buchmann, Mana Gharun, et al. 2021. "Improved Global Estimations of Gross Primary Productivity of Natural Vegetation Types by Incorporating Plant Functional Type." *International Journal of Applied Earth Observation and Geoinformation* 100 (August): 102328. https://doi.org/10.1016/j.jag.2021.102328.

Lin, Shangrong, Jing Li, Qinhuo Liu, Longhui Li, Jing Zhao, and Wentao Yu. 2019. "Evaluating the Effectiveness of Using Vegetation Indices Based on Red-Edge Reflectance from Sentinel-2 to Estimate Gross Primary Productivity." *Remote Sensing* 11 (11): 1303. https://doi.org/10.3390/rs11111303.

Liu, J, JM Chen, J Cihlar, and WM Park. 1997. "A Process-Based Boreal Ecosystem Productivity Simulator Using Remote Sensing Inputs." *Remote Sensing of Environment* 62 (2): 158–75.

Lundberg, Scott M, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*.

Meyer, Hanna, Christoph Reudenbach, Tomislav Hengl, Marwan Katurji, and Thomas Nauss. 2018. "Improving Performance of Spatio-Temporal Machine Learning Models Using Forward Feature Selection and Target-Oriented Validation." *Environmental Modelling & Software* 101 (March): 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

Meyer, Hanna, Christoph Reudenbach, Stephan Wöllauer, and Thomas Nauss. 2019. "Importance of Spatial Predictor Variable Selection in Machine Learning Applications – Moving from Data Reproduction to Spatial Prediction." *Ecological Modelling* 411 (November): 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815.

Molnar, Christoph. 2020. *Interpretable Machine Learning*. Lulu. com.

Monteith, J. L. 1972. "Solar Radiation and Productivity in Tropical Ecosystems." *The Journal of Applied Ecology* 9 (3): 747. https://doi.org/10.2307/2401901.

Monteith, John Lennox, and M. H. Unsworth. 2013. *Principles of Environmental Physics: Plants, Animals, and the Atmosphere*. Elsevier Science.

Montero, David, César Aybar, Miguel D. Mahecha, Francesco Martinuzzi, Maximilian Söchting, and Sebastian Wieneke. 2023. "A Standardized Catalogue of Spectral Indices to Advance the Use of Remote Sensing in Earth System Research." *Scientific Data* 10 (1):

197. https://doi.org/10.1038/s41597-023-02096-0.

Musavi, Talie, Mirco Migliavacca, Markus Reichstein, Jens Kattge, Christian Wirth, T. Andrew Black, Ivan Janssens, et al. 2017. "Stand Age and Species Richness Dampen Interannual Variation of Ecosystem-Level Photosynthetic Capacity." *Nature Ecology & Evolution* 1 (2): 0048. https://doi.org/10.1038/s41559-016-0048.

Myneni, R. B., R. Ramakrishna, R. Nemani, and S. W. Running. 1997. "Estimation of Global Leaf Area Index and Absorbed Par Using Radiative Transfer Models." *IEEE Transactions on Geoscience and Remote Sensing* 35 (6): 1380–93. https://doi.org/10.1109/36.649788.

Myneni, R. B., and D. L. Williams. 1994. "On the Relationship Between FAPAR and NDVI." *Remote Sensing of Environment* 49 (3): 200–211. https://doi.org/10.1016/0034-4257(94)90016-7.

Ouimette, Andrew P., Scott V. Ollinger, Andrew D. Richardson, David Y. Hollinger, Trevor F. Keenan, Lucie C. Lepine, and Matthew A. Vadeboncoeur. 2018. "Carbon Fluxes and Interannual Drivers in a Temperate Forest Ecosystem Assessed Through Comparison of Top-down and Bottom-up Approaches." *Agricultural and Forest Meteorology* 256-257 (June): 420–30. https://doi.org/10.1016/j.agrformet.2018.03.017.

Pabon-Moreno, Daniel E., Mirco Migliavacca, Markus Reichstein, and Miguel D. Mahecha. 2022. "On the Potential of Sentinel-2 for Estimating Gross Primary Production." *IEEE Transactions on Geoscience and Remote Sensing*, 1–1. https://doi.org/10.1109/TGRS.2022.3152272.

Papale, Dario, T. Andrew Black, Nuno Carvalhais, Alessandro Cescatti, Jiquan Chen, Martin Jung, Gerard Kiely, et al. 2015. "Effect of Spatial Sampling from European Flux Towers for Estimating Carbon and Water Fluxes with Artificial Neural Networks." *Journal of Geophysical Research: Biogeosciences* 120 (10): 1941–57. https://doi.org/10.1002/2015JG002997.

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, et al. 2020. "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." *Scientific Data* 7 (1): 1–27.

Pesquer, Lluís, Cristina Domingo-Marimon, and Xavier Pons. 2019. "Spatial and Spectral Pattern Identification for the Automatic Selection of High-Quality MODIS Images." *Journal of Applied Remote Sensing* 13 (01): 1. https://doi.org/10.1117/1.JRS.13.014510.

Pierrat, Zoe, Troy Magney, Nicholas C. Parazoo, Katja Grossmann, David R. Bowling, Ulli Seibt, Bruce Johnson, et al. 2022. "Diurnal and Seasonal Dynamics of Solar-Induced

Chlorophyll Fluorescence, Vegetation Indices, and Gross Primary Productivity in the Boreal Forest." *Journal of Geophysical Research: Biogeosciences* 127 (2). https://doi.org/10.1029/2021JG006588.

Prince, S. D. 1991. "A Model of Regional Primary Production for Use with Coarse Resolution Satellite Data." *International Journal of Remote Sensing* 12 (6): 1313–30. https://doi.org/10.1080/01431169108929728.

Raes, Dirl, Pasquale Steduto, Theodore Hsiao, and Elias Fereres. 2022. "Reference Manual, Chapter 2 – AquaCrop, Version 7.0 - August 2022."

Rahman, A. F., D. A. Sims, V. D. Cordova, and B. Z. El-Masri. 2005. "Potential of MODIS EVI and Surface Temperature for Directly Estimating Per-Pixel Ecosystem C Fluxes: MODIS EVI FOR ECOSYSTEM C FLUX." *Geophysical Research Letters* 32 (19): n/a–. https://doi.org/10.1029/2005GL024127.

Reichstein, Markus, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566 (7743): 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Reichstein, Markus, Eva Falge, Dennis Baldocchi, Dario Papale, Marc Aubinet, Paul Berbigier, Christian Bernhofer, et al. 2005. "On the Separation of Net Ecosystem Exchange into Assimilation and Ecosystem Respiration: Review and Improved Algorithm." *Global Change Biology* 11 (9): 1424–39. https://doi.org/10.1111/j.1365-2486.2005.001002.x.

Reinmann, Andrew B., and Lucy R. Hutyra. 2017. "Edge Effects Enhance Carbon Uptake and Its Vulnerability to Climate Change in Temperate Broadleaf Forests." *Proceedings of the National Academy of Sciences* 114 (1): 107–12. https://doi.org/10.1073/pnas.1612369114.

Richardson, Andrew, and David Hollinger. 2016. "AmeriFlux AmeriFlux US-Bar Bartlett Experimental Forest." Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….

Richardson, Arthur J, and CL Wiegand. 1977. "Distinguishing Vegetation from Soil Background Information." *Photogrammetric Engineering and Remote Sensing* 43 (12): 1541–52.

Robinson, Nathaniel P., Brady W. Allred, William K. Smith, Matthew O. Jones, Alvaro Moreno, Tyler A. Erickson, David E. Naugle, and Steven W. Running. 2018. "Terrestrial

Primary Production for the Conterminous United States Derived from Landsat 30 m and MODIS 250 m." Edited by Nathalie Pettorelli and Jose Paruelo. *Remote Sensing in Ecology and Conservation* 4 (3): 264–80. https://doi.org/10.1002/rse2.74.

Rogers, Cheryl A., Jing M. Chen, Ting Zheng, Holly Croft, Alemu Gonsamo, Xiangzhong Luo, and Ralf M. Staebler. 2020. "The Response of Spectral Vegetation Indices and Solar-Induced Fluorescence to Changes in Illumination Intensity and Geometry in the Days Surrounding the 2017 North American Solar Eclipse." *Journal of Geophysical Research: Biogeosciences* 125 (10). https://doi.org/10.1029/2020JG005774.

Running, Steven W., and Joseph C. Coughlan. 1988. "A General Model of Forest Ecosystem Processes for Regional Applications I. Hydrologic Balance, Canopy Gas Exchange and Primary Production Processes." *Ecological Modelling* 42 (2): 125–54. https://doi.org/10.1016/0304-3800(88)90112-3.

Running, Steven W., Ramakrishna R. Nemani, Faith Ann Heinsch, Maosheng Zhao, Matt Reeves, and Hirofumi Hashimoto. 2004. "A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production." *BioScience* 54 (6): 547. https://doi.org/10.1641/0006-3568(2004)054%5B0547:ACSMOG%5D2.0.CO;2.

Ryu, Youngryel, Dennis D. Baldocchi, Hideki Kobayashi, Catharine van Ingen, Jie Li, T. Andy Black, Jason Beringer, et al. 2011. "Integration of MODIS Land and Atmosphere Products with a Coupled-Process Model to Estimate Gross Primary Productivity and Evapotranspiration from 1 Km to Global Scales: GLOBAL GPP AND ET." *Global Biogeochemical Cycles* 25 (4): n/a–. https://doi.org/10.1029/2011GB004053.

Ryu, Youngryel, Joseph A. Berry, and Dennis D. Baldocchi. 2019. "What Is Global Photosynthesis? History, Uncertainties and Opportunities." *Remote Sensing of Environment* 223 (March): 95–114. https://doi.org/10.1016/j.rse.2019.01.016.

Schimel, David S. 1995. "Terrestrial Ecosystems and the Carbon Cycle." *Global Change Biology* 1 (1): 77–91. https://doi.org/10.1111/j.1365-2486.1995.tb00008.x.

Sellers, P. J. 1987. "Canopy Reflectance, Photosynthesis, and Transpiration, II. The Role of Biophysics in the Linearity of Their Interdependence." *Remote Sensing of Environment* 21 (2): 143–83. https://doi.org/10.1016/0034-4257(87)90051-4.

Sellers, P. J., C. J. Tucker, G. J. Collatz, S. O. Los, C. O. Justice, D. A. Dazlich, and D. A. Randall. 1994. "A Global 1° by 1° NDVI Data Set for Climate Studies. Part 2: The Generation of Global Fields of Terrestrial Biophysical Parameters from the NDVI."

International Journal of Remote Sensing 15 (17): 3519–45. https://doi.org/10.1080/01431169408954343.

Shi, Hao, Longhui Li, Derek Eamus, Alfredo Huete, James Cleverly, Xin Tian, Qiang Yu, et al. 2017. "Assessing the Ability of MODIS EVI to Estimate Terrestrial Ecosystem Gross Primary Production of Multiple Land Cover Types." *Ecological Indicators* 72 (January): 153–64. https://doi.org/10.1016/j.ecolind.2016.08.022.

Smith, Ian A, Lucy R Hutyra, Andrew B Reinmann, Julia K Marrs, and Jonathan R Thompson. 2018. "Piecing Together the Fragments: Elucidating Edge Effects on Forest Carbon Dynamics." *Frontiers in Ecology and the Environment* 16 (4): 213–21. https://doi.org/10.1002/fee.1793.

Staebler, Ralf. 2019. "AmeriFlux AmeriFlux CA-Cbo Ontario-Mixed Deciduous, Borden Forest Site." Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….

Teets, Aaron, David J. P. Moore, M. Ross Alexander, Peter D. Blanken, Gil Bohrer, Sean P. Burns, Mariah S. Carbone, et al. 2022. "Coupling of Tree Growth and Photosynthetic Carbon Uptake Across Six North American Forests." *Journal of Geophysical Research: Biogeosciences* 127 (4): e2021JG006690. https://doi.org/10.1029/2021JG006690.

Tramontana, Gianluca, Kazuito Ichii, Gustau Camps-Valls, Enrico Tomelleri, and Dario Papale. 2015. "Uncertainty Analysis of Gross Primary Production Upscaling Using Random Forests, Remote Sensing and Eddy Covariance Data." *Remote Sensing of Environment* 168 (October): 360–73. https://doi.org/10.1016/j.rse.2015.07.015.

Tramontana, Gianluca, Martin Jung, Christopher R. Schwalm, Kazuhito Ichii, Gustau Camps-Valls, Botond Ráduly, Markus Reichstein, et al. 2016. "Predicting Carbon Dioxide and Energy Fluxes Across Global FLUXNET Sites Withregression Algorithms." *Biogeosciences* 13 (14): 4291–4313. https://doi.org/10.5194/bg-13-4291-2016.

Tucker, Compton J. 1979. "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." *Remote Sensing of Environment* 8 (2): 127–50.

Van Leeuwen, Willem J. D., Barron J. Orr, Stuart E. Marsh, and Stefanie M. Herrmann. 2006. "Multi-Sensor NDVI Data Continuity: Uncertainties and Implications for Vegetation Monitoring Applications." *Remote Sensing of Environment* 100 (1): 67–81. https://doi.org/10.1016/j.rse.2005.10.002.

Vermote, Wolfe, E. 2021. "MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061." *NASA EOSDIS Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MOD09GA.061.

Verrelst, J., L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno. 2012. "Retrieval of Vegetation Biophysical Parameters Using Gaussian Process Techniques." *IEEE Transactions on Geoscience and Remote Sensing* 50 (5): 1832–43. https://doi.org/10.1109/TGRS.2011.2168962.

Verrelst, Jochem, Jordi Muñoz, Luis Alonso, Jesús Delegido, Juan Pablo Rivera, Gustavo Camps-Valls, and José Moreno. 2012. "Machine Learning Regression Algorithms for Biophysical Parameter Retrieval: Opportunities for Sentinel-2 and -3." *Remote Sensing of Environment* 118 (March): 127–39. https://doi.org/10.1016/j.rse.2011.11.002.

Wang, Weile, Jennifer Dungan, Hirofumi Hashimoto, Andrew R. Michaelis, Cristina Milesi, Kazuhito Ichii, and Ramakrishna R. Nemani. 2011. "Diagnosing and Assessing Uncertainties of Terrestrial Ecosystem Models in a Multimodel Ensemble Experiment: 1. Primary Production: ENSEMBLE MODEL UNCERTAINTIES: GPP/NPP." *Global Change Biology* 17 (3): 1350–66. https://doi.org/10.1111/j.1365-2486.2010.02309.x.

Wright, Marvin N., and Andreas Ziegler. 2017. "**Ranger** : A Fast Implementation of Random Forests for High Dimensional Data in *C++* and *r*." *Journal of Statistical Software* 77 (1). https://doi.org/10.18637/jss.v077.i01.

Wu, Chaoyang, Jing M. Chen, and Ni Huang. 2011. "Predicting Gross Primary Production from the Enhanced Vegetation Index and Photosynthetically Active Radiation: Evaluation and Calibration." *Remote Sensing of Environment* 115 (12): 3424–35. https://doi.org/10.1016/j.rse.2011.08.006.

Wu, Genghong, Kaiyu Guan, Chongya Jiang, Bin Peng, Hyungsuk Kimm, Min Chen, Xi Yang, et al. 2020. "Radiance-Based NIR $_{v}$ as a Proxy for GPP of Corn and Soybean." *Environmental Research Letters* 15 (3): 034009. https://doi.org/10.1088/1748-9326/ab65cc.

Xiao, Jingfeng, Frederic Chevallier, Cecile Gomez, Luis Guanter, Jeffrey A. Hicke, Alfredo R. Huete, Kazuhito Ichii, et al. 2019. "Remote Sensing of the Terrestrial Carbon Cycle: A Review of Advances over 50 Years." *Remote Sensing of Environment* 233 (November): 111383. https://doi.org/10.1016/j.rse.2019.111383.

Xie, Xinyao, Ainong Li, Jianbo Tan, Huaan Jin, Xi Nan, Zhengjian Zhang, Jinhu Bian, and Guangbin Lei. 2020. "Assessments of Gross Primary Productivity Estimations with Satellite Data-Driven Models Using Eddy Covariance Observation Sites over the Northern Hemisphere." *Agricultural and Forest Meteorology* 280 (January): 107771. https://doi.org/10.1016/j.agrformet.2019.107771.

Zeng, Yelu, Dalei Hao, Alfredo Huete, Benjamin Dechant, Joe Berry, Jing M. Chen, Joanna

Joiner, et al. 2022. "Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally." *Nature Reviews Earth & Environment* 3 (7): 477–93. https://doi.org/10.1038/s43017-022-00298-5.

Zhang, Zhaoying, Yongguang Zhang, Yao Zhang, Nadine Gobron, Christian Frankenberg, Songhan Wang, and Zhaohui Li. 2020. "The Potential of Satellite FPAR Product for GPP Estimation: An Indirect Evaluation Using Solar-Induced Chlorophyll Fluorescence." *Remote Sensing of Environment* 240 (April): 111686. https://doi.org/10.1016/j.rse.2020.111686.

A. Appendices

A.1. Bitstrings tables

A.1.1. MODIS State 1km

Table A.1.: state_1km bit strings

Bit	State
Cloud State	
00	clear
01	cloudy
10	mixed
11	not set, assumed clear
Cloud Shadow	
1	yes
0	no
Land Water	
000	shallow ocean
001	land
010	ocean coastlines and lake shorelines
011	shallow inland water
100	ephemeral water
101	deep inland water
110	continental/moderate ocean
111	deep ocean

Aerosol Quantity	
00	climatology
01	low
10	average
11	high
Cirrus Detected	
00	none
01	small
10	average
11	high
Cloud Flag	
1	cloud
0	no cloud
Fire Flag	
1	fire
0	no fire
Snow Ice Flag	
1	yes
0	no
Pixel adjacent Cloud	
1	yes
0	no
Salt Pan Cloud	
1	yes
0	no
Snow Mask	
1	yes
0	no

A.1.2. MODIS QC Scan

Table A.2.: qc_scan bit strings

Bits	State
Modland	
00	ideal quality - all bands
01	less than ideal quality - some or all bands
10	product not produced due to cloud effects
11	product not produced for other reasons
Band quality (apply for all bands)	
0000	highest_quality
0111	noisy detector
1000	dead detector, data interpolated in L1B
1001	solar zenith >= 86 degrees
1010	solar zenith >= 85 and < 86 degrees
1011	missing input
1100	internal constant used
1101	correction out of bounds
1110	L1B data faulty
1111	not processed due to deep ocean or clouds
Atmospheric Correction	
0	no
1	yes
Adjacency Correction	
0	no
1	yes

A.2. Complete LM metrics

A.2.1. Monthly LM metrics

Table A.3.: Summary of Linear models for GPP estimation using the vegetation indices on a monthly basis (per site).

	adj.r.squared	statistic	AIC	BIC	p.value
Michigan					
All	0.918	48.285	59.460	64.802	1.120×10^{-7}
CCI	0.864	109.210	66.171	68.842	1.483×10^{-8}
NIRv	0.856	102.166	67.214	69.885	2.366×10^{-8}
EVI	0.850	97.219	67.983	70.654	3.341×10^{-8}
NDVI	0.699	40.436	80.515	83.186	9.481×10^{-6}
Bartlett					
All	0.889	34.995	60.577	65.920	7.627×10^{-7}
CCI	0.887	133.808	58.692	61.363	3.495×10^{-9}
EVI	0.880	125.716	59.691	62.363	5.466×10^{-9}
NIRv	0.878	123.297	60.001	62.672	6.280×10^{-9}
NDVI	0.784	62.655	70.289	72.960	6.363×10^{-7}
Borden					
CCI	0.782	133.796	177.703	182.616	1.107×10^{-13}
All	0.778	33.452	181.071	190.897	3.485×10^{-11}
NDVI	0.755	114.882	182.191	187.104	9.428×10^{-13}
NIRv	0.746	109.643	183.534	188.447	1.790×10^{-12}
EVI	0.738	105.340	184.673	189.586	3.086×10^{-12}
All					
All	0.771	62.572	327.544	341.368	3.117×10^{-22}
CCI	0.732	200.153	336.514	343.426	1.766×10^{-22}
NDVI	0.638	129.875	358.619	365.531	8.818×10^{-18}
NIRv	0.583	103.147	369.129	376.041	1.528×10^{-15}
EVI	0.566	96.172	372.136	379.048	6.694×10^{-15}

A.2.2. Weekly LM metrics

Table A.4.: Summary of Linear models for GPP estimation using the vegetation indices on a weekly basis (per site).

	adj.r.squared	statistic	AIC	BIC	p.value
Bartlett					
All	0.809	77.188	271.583	285.326	1.513×10^{-24}
EVI	0.787	267.599	276.529	283.400	8.712×10^{-26}
CCI	0.783	260.798	278.010	284.881	1.795×10^{-25}
NIRv	0.770	241.637	282.352	289.224	1.495×10^{-24}
NDVI	0.633	125.324	316.317	323.189	2.444×10^{-17}
Michigan					
All	0.799	64.530	263.238	276.284	4.684×10^{-21}
NIRv	0.782	229.980	265.761	272.284	1.060×10^{-22}
EVI	0.780	228.274	266.140	272.663	1.275×10^{-22}
CCI	0.701	151.227	286.110	292.633	2.138×10^{-18}
NDVI	0.609	100.592	303.638	310.161	1.116×10^{-14}
Borden					
All	0.600	53.920	761.297	779.032	3.178×10^{-27}
CCI	0.577	193.515	766.307	775.174	3.589×10^{-28}
EVI	0.572	189.314	768.107	776.974	8.756×10^{-28}
NIRv	0.563	182.981	770.864	779.732	3.434×10^{-27}
NDVI	0.499	141.470	790.399	799.267	5.537×10^{-23}
All					
CCI	0.560	356.273	1,445.058	1,455.963	1.022×10^{-51}
All	0.558	89.013	1,449.447	1,471.255	1.971×10^{-48}
NDVI	0.461	239.572	1,501.991	1,512.895	2.115×10^{-39}
NIRv	0.459	237.796	1,502.954	1,513.858	3.417×10^{-39}
EVI	0.452	230.689	1,506.838	1,517.743	2.370×10^{-38}

A.2.3. Daily LM metrics

Table A.5.: Summary of Linear models for GPP estimation using the vegetation indices on a daily basis (per site).

	adj.r.squared	statistic	AIC	BIC	p.value
Bartlett					
All	0.866	485.890	1,178.271	1,200.533	6.162×10^{-129}
CCI	0.832	1,489.205	1,243.242	1,254.374	2.354×10^{-118}
EVI	0.813	1,310.145	1,275.087	1,286.219	1.762×10^{-111}
NIRv	0.800	1,201.359	1,296.213	1,307.345	6.406×10^{-107}
NDVI	0.657	576.639	1,458.698	1,469.830	7.912×10^{-72}
Michigan					
All	0.713	91.565	669.582	687.525	2.499×10^{-38}
EVI	0.702	345.680	671.830	680.802	3.248×10^{-40}
NIRv	0.664	289.199	689.807	698.778	2.366×10^{-36}
CCI	0.610	229.492	711.553	720.524	1.121×10^{-31}
NDVI	0.521	159.932	741.759	750.730	3.569×10^{-25}
All					
All	0.606	320.497	4,522.636	4,550.986	1.640×10^{-166}
CCI	0.591	1,201.855	4,550.671	4,564.846	1.355×10^{-163}
EVI	0.494	812.661	4,727.696	4,741.871	3.301×10^{-125}
NIRv	0.475	755.095	4,757.393	4,771.569	9.120×10^{-119}
NDVI	0.447	672.798	4,801.777	4,815.952	3.873×10^{-109}
Borden					
All	0.521	104.996	2,181.484	2,205.188	4.203×10^{-60}
CCI	0.441	303.652	2,237.254	2,249.106	1.838×10^{-50}
EVI	0.428	287.124	2,246.624	2,258.475	1.973×10^{-48}
NIRv	0.380	235.703	2,277.328	2,289.180	8.977×10^{-42}
NDVI	0.273	144.920	2,338.369	2,350.221	1.617×10^{-28}

A.3. Complete GAM metrics

A.3.1. Monthly GAM metrics

Table A.6.: Summary of the GAM models output for monthly GPP estimation using the every VI as an individual non-linear term

site	index	edf	f	p_value	AIC
Michigan	evi_mean	3.817030	87.35043	0.000000e+00	46.78342
Michigan	nirv_mean	4.044975	75.37774	0.000000e+00	48.41677
Bartlett	cci_mean	2.239036	66.85647	0.000000e+00	54.60874
Michigan	cci_mean	2.602984	57.48746	0.000000e+00	59.36252
Bartlett	evi_mean	1.000019	125.71069	0.000000e+00	59.69151
Bartlett	nirv_mean	1.000030	123.28934	0.000000e+00	60.00142
Michigan	ndvi_mean	2.806127	31.20277	5.690552e-06	68.06940
Bartlett	ndvi_mean	1.000069	62.64577	-3.515872e-07	70.28904
Borden	cci_mean	1.190879	96.46348	0.000000e+00	177.64856
Borden	ndvi_mean	1.809503	53.64331	0.000000e+00	181.16487
Borden	nirv_mean	1.000106	109.62110	0.000000e+00	183.53397
Borden	evi_mean	1.000118	105.31638	0.000000e+00	184.67361
All	cci_mean	2.404895	72.32912	0.000000e+00	333.05822
All	ndvi_mean	2.954201	51.51928	0.000000e+00	341.40009
All	nirv_mean	3.770256	31.19524	0.000000e+00	355.55683
All	evi_mean	3.789326	29.98655	0.000000e+00	357.45536

Table A.7.: Summary of the GAM models output for monthly GPP estimation using the all VIs per site category as non-linear terms covariates

site	index	term	edf	f	p_value	AIC
Borden	All	s(evi_mean)	1.000007	1.561538	0.2202315969	181.0712
Borden	All	s(nirv_mean)	1.000006	1.832159	0.1850688800	181.0712
Borden	All	s(cci_mean)	1.000100	2.667602	0.1118925771	181.0712
All	All	s(evi_mean)	5.154191	4.079875	0.0004261942	4411.7779
All	All	s(nirv_mean)	4.545057	2.844737	0.0125690226	4411.7779
All	All	s(cci_mean)	6.218002	13.246895	0.0000000000	4411.7779

A.3.2. Weekly GAM metrics

Table A.8.: Summary of the GAM models output for weekly GPP estimation using the every VI as an individual non-linear term

site	index	edf	f	p_value	AIC
Michigan	evi_mean	3.663726	75.39276	0	247.2844
Michigan	nirv_mean	3.209376	71.49144	0	256.7204
Bartlett	evi_mean	3.336909	71.86998	0	272.0930
Bartlett	cci_mean	2.583077	86.30588	0	275.5742
Michigan	cci_mean	3.520166	43.76791	0	277.5737
Bartlett	nirv_mean	3.272579	65.98375	0	278.0024
Michigan	ndvi_mean	2.675203	39.90565	0	293.9057
Bartlett	ndvi_mean	1.000087	125.30174	0	316.3173
Borden	evi_mean	3.560463	46.50232	0	762.7106
Borden	cci_mean	1.000274	193.41184	0	766.3073
Borden	nirv_mean	2.853504	54.04690	0	767.5482
Borden	ndvi_mean	1.406799	80.37055	0	790.2338
All	cci_mean	4.538741	69.58626	0	1433.6512
All	evi_mean	4.748375	53.58396	0	1468.9599
All	nirv_mean	4.334480	55.98519	0	1473.8814
All	ndvi_mean	3.223579	71.73012	0	1480.0666

Table A.9.: Summary of the GAM models output for weekly GPP estimation using the all VIs per site category as non-linear terms covariates

site	index	term	edf	f	p_value	AIC
Michigan	All	s(evi_mean)	4.488458	5.8177458	0.0001315731	240.5091
Michigan	All	s(nirv_mean)	1.000195	9.7009559	0.0028789998	240.5091
Michigan	All	s(cci_mean)	1.000043	0.0175387	0.8952502747	240.5091
Bartlett	All	s(evi_mean)	1.000051	3.1454363	0.0806776572	269.6656
Bartlett	All	s(nirv_mean)	1.000129	0.5997498	0.4413933495	269.6656
Bartlett	All	s(cci_mean)	2.128825	4.2538207	0.0095109230	269.6656
Borden	All	s(evi_mean)	1.001795	3.0976955	0.0804966509	760.4175
Borden	All	s(nirv_mean)	1.976616	0.6214367	0.5421077236	760.4175
Borden	All	s(cci_mean)	1.000123	2.6980022	0.1027452161	760.4175
All	All	s(evi_mean)	5.154191	4.0798752	0.0004261942	4411.7779
All	All	s(nirv_mean)	4.545057	2.8447368	0.0125690226	4411.7779
All	All	s(cci_mean)	6.218002	13.2468950	0.0000000000	4411.7779

A.3.3. Daily GAM metrics

Table A.10.: Summary of the GAM models output for daily GPP estimation using the every VI as an individual non-linear term

site	index	edf	f	p_value	AIC
Michigan	evi_mean	4.457179	77.72707	0	652.5681
Michigan	nirv_mean	2.807163	86.23596	0	687.8352
Michigan	cci_mean	4.218361	48.29421	0	704.7992
Michigan	ndvi_mean	2.980300	55.52978	0	722.8092
Bartlett	cci_mean	5.975808	236.42420	0	1215.1703
Bartlett	evi_mean	5.365946	250.56443	0	1222.4305
Bartlett	nirv_mean	5.461424	226.11564	0	1244.9206
Bartlett	ndvi_mean	5.198974	137.44546	0	1374.8503
Borden	evi_mean	5.569955	55.49361	0	2202.2988
Borden	nirv_mean	5.804760	46.01083	0	2230.5977
Borden	cci_mean	3.191225	78.29535	0	2231.2904
Borden	ndvi_mean	4.912790	45.99191	0	2235.6639
All	cci_mean	6.266321	173.70000	0	4515.9138
All	evi_mean	6.642583	148.32392	0	4574.2505
All	ndvi_mean	5.308406	155.72087	0	4615.5740
All	nirv_mean	6.317347	140.36124	0	4618.3302

Table A.11.: Summary of the GAM models output for daily GPP estimation using the all VIs per site category as non-linear terms covariates

site	index	term	edf	f	p_value	AIC
Michigan	All	s(evi_mean)	3.012297	15.313284	0.000000e+00	630.891
Michigan	All	s(nirv_mean)	1.953289	11.098293	1.465150e-05	630.891
Michigan	All	s(cci_mean)	3.316744	2.334852	5.228136e-02	630.891
Bartlett	All	s(evi_mean)	5.075549	6.261981	3.136046e-06	1139.659
Bartlett	All	s(nirv_mean)	3.933390	2.321316	5.093414e-02	1139.659
Bartlett	All	s(cci_mean)	5.523614	9.113037	0.000000e+00	1139.659
Borden	All	s(evi_mean)	1.001559	22.542139	3.092503e-06	2147.143
Borden	All	s(nirv_mean)	4.720920	4.428707	2.739846e-04	2147.143
Borden	All	s(cci_mean)	3.652172	2.062896	9.147704e-02	2147.143
All	All	s(evi_mean)	5.154191	4.079875	4.261942e-04	4411.778
All	All	s(nirv_mean)	4.545057	2.844737	1.256902e-02	4411.778
All	All	s(cci_mean)	6.218002	13.246895	0.000000e+00	4411.778

A.4. Residuals distributions

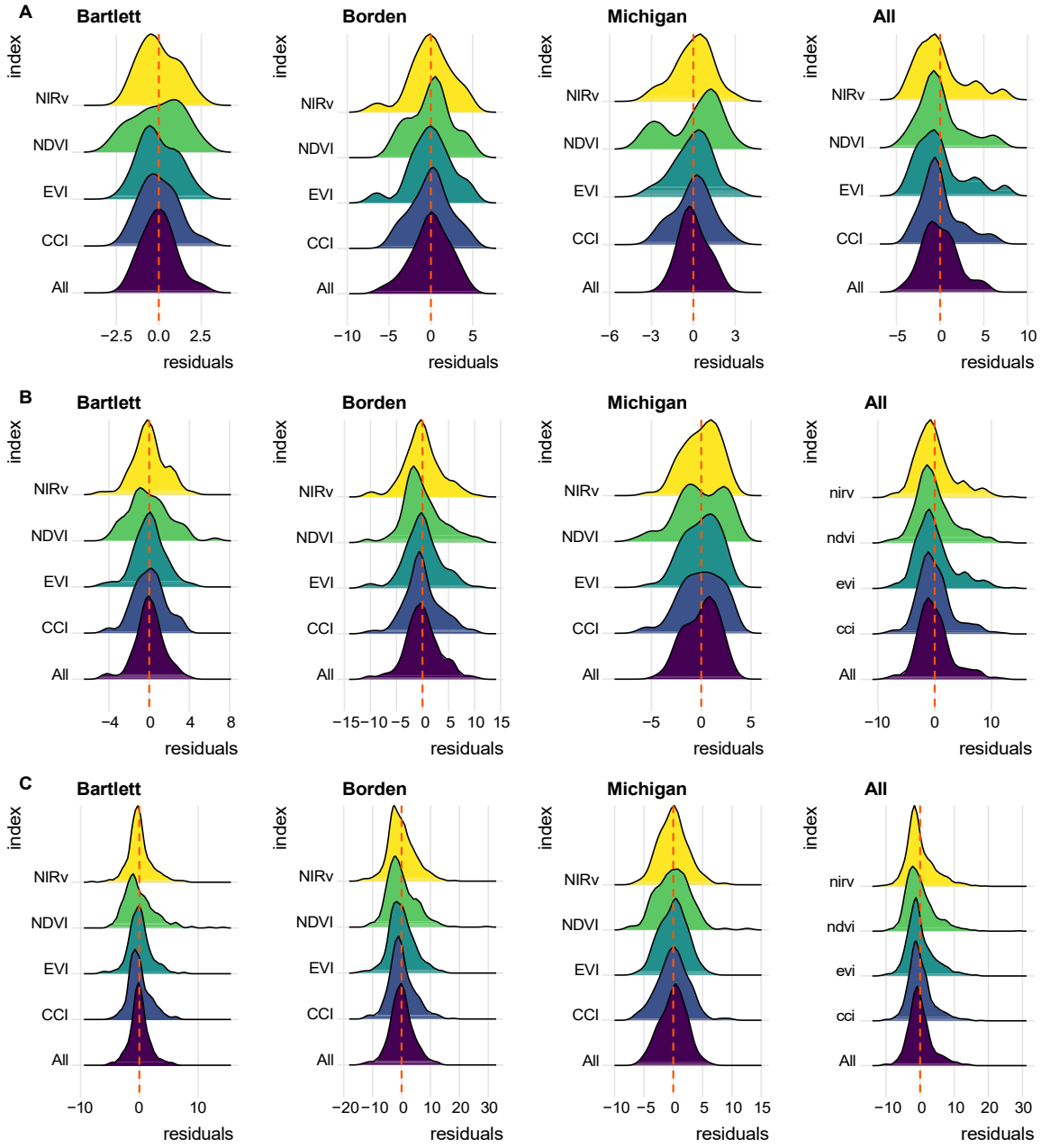

Figure A.1.: Residuals distributions for each of the LMs for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis.

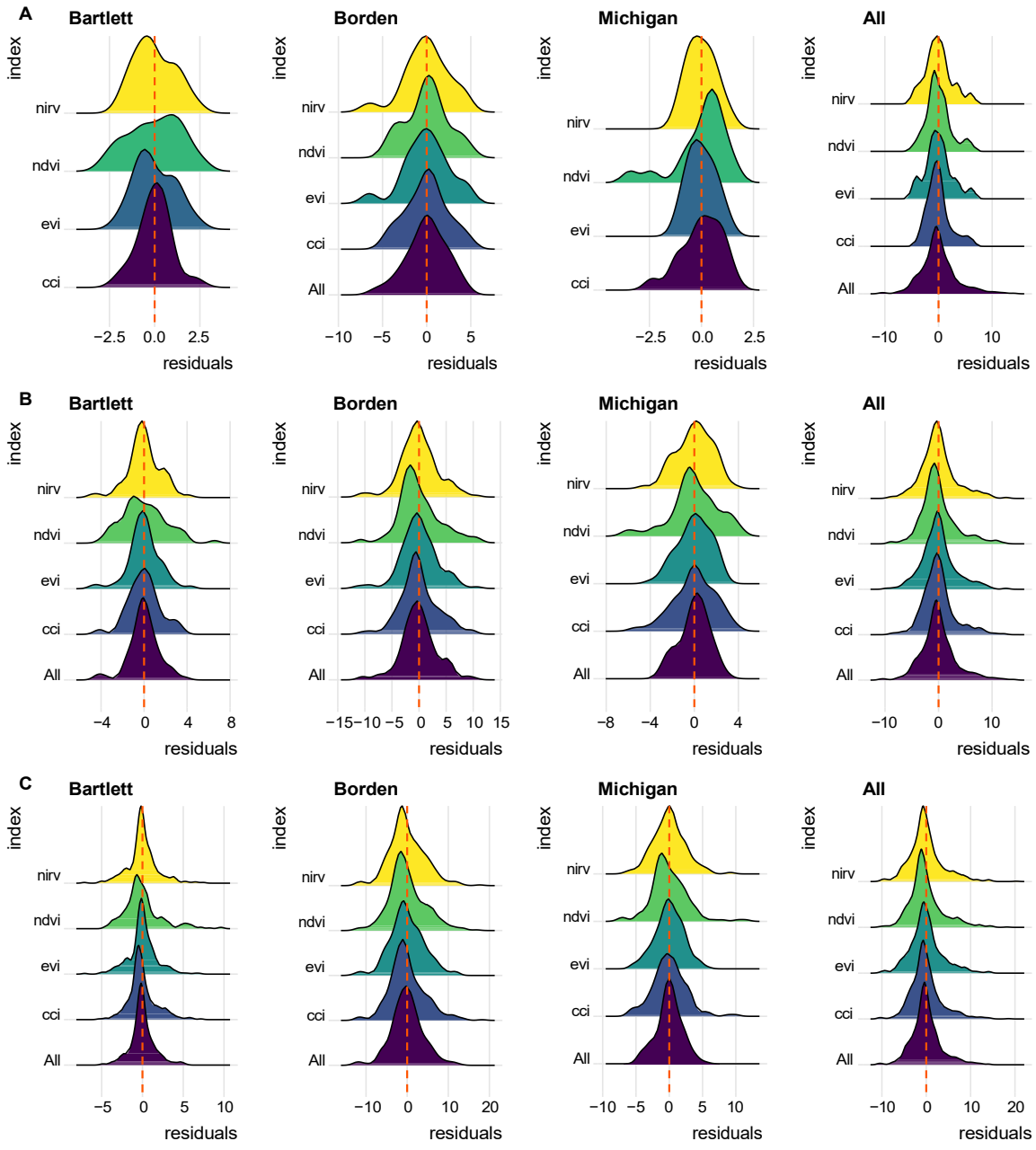

Figure A.2.: Residuals distributions for each of the GAMs for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis.