

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Data-driven understanding of the global carbon and water cycles through ecosystem phenology and water use efficiency

Permalink

<https://escholarship.org/uc/item/3786p2k2>

Author

Lu, Xinchun

Publication Date

2024

Peer reviewed|Thesis/dissertation

Data-driven understanding of the global carbon and water cycles through ecosystem
phenology and water use efficiency

by

Xinchen Lu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Trevor F. Keenan, Chair

Professor Dennis D. Baldocchi

Assistant Professor Manuela Girotto

Spring 2024

Data-driven understanding of the global carbon and water cycles through ecosystem
phenology and water use efficiency

Copyright 2024
by
Xinchen Lu

Abstract

Data-driven understanding of the global carbon and water cycles through ecosystem phenology and water use efficiency

by

Xinchen Lu

Doctor of Philosophy in Environmental Science, Policy and Management

University of California, Berkeley

Associate Professor Trevor F. Keenan, Chair

The carbon and water cycles in global terrestrial ecosystems hold significant importance due to their unique influence on biosphere-climate interactions. In my dissertation, I adopted a data-driven method to deepen our understanding of these global cycles. In the first chapter, I enhanced the AmeriFlux data product, resulting in the creation of a publicly accessible dataset named ONEFlux BETA. This process entailed the downscaling, gap-filling, and partitioning of flux observations from 63 distinct flux tower sites. Subsequent chapters employed this dataset, paired with other global data such as FLUXNET 2015, MODIS observations of phenology, and remote sensing observations of vegetation indices, to examine ecosystem autumn phenology and water use efficiency. Specifically, in the second chapter, I investigated the connection between photosynthesis during the growth season and end-of-season (EOS) senescence across various ecosystems, using data from 40 flux tower sites. This analysis challenged existing theories by revealing no significant negative correlation between growing season photosynthesis and EOS. On the contrary, more productive growth seasons are often associated with a delayed EOS in diverse ecosystems. The third chapter shifted the focus to water use efficiency (WUE) and its responses to water stress, factoring in variables such as vapor pressure deficit (VPD) and soil moisture. The results indicated a more consistent response to VPD across most sites. My dissertation showcases a data-driven approach to understanding ecosystem dynamics, with a particular emphasis on the carbon and water cycles of terrestrial ecosystems, harnessing data from global flux tower sites.

Dedicated to my younger self, who decided to traverse this challenging life experience and the journey across the world to do so.

Contents

Contents	ii
List of Figures	iv
List of Tables	vii
Introduction	1
References	4
1 ONEFlux BETA release for a data driven approach to understand the global carbon and water cycling	6
1.1 Introduction	7
1.2 Methods and Materials	9
Ameriflux BASE	9
ONEFlux pipeline	9
Downscaling the meteorological variables	12
1.3 Results	13
ERA-interim downscaling of the meteorological variables	13
Differences between different partitioning methods	14
Comparison with the FLUXNET 2015 dataset and ReddyProc results	18
Flux tower observation data for each sites	21
1.4 Discussion	22
1.5 Conclusion	25
References	26
1.6 Appendix	35
Transition from chapter 1 to chapter 2	37
2 Reevaluating autumn phenology: the limited role of growing season productivity compared with environmental factors	38
2.1 Introduction	39
2.2 Methods and Materials	41

Flux tower measurements	41
Phenological records from Harvard Forest	42
MODIS global land dynamics product	42
Determination of phenological metrics from flux measurements	42
Modelling of leaf senescence dates	43
2.3 Results	45
No evidence for the carry-over effects of growing season photosynthesis on EOS	45
Control of environmental factors over EOS in different ecosystems	47
The introduction of growing season photosynthesis did not improve EOS pre- dictions	49
2.4 Discussion	53
The carry-over effects of “growing season photosynthesis” on EOS	53
The environmental controls over EOS in different ecosystems	55
Matches and mismatches between flux estimated and remotely sensed pheno- logical metrics	55
2.5 Conclusion	56
2.6 Acknowledgement	57
References	58
2.7 Appendix	62
Transition from chapter 1 to chapter 2	67
3 Disentangling the influence of soil moisture and atmospheric dryness on the coupled ecosystem carbon and water cycling using machine learning	68
3.1 Introduction	69
3.2 Methods and Materials	71
Flux tower observation	71
Analytical methods	72
3.3 Results	73
Modeling of WUE in different ecosystems	73
GPP, ET and WUE Responses to Water Stress: ALE Analysis Insights . . .	75
Variations in WUE Across Aridity Gradients	77
3.4 Discussion	80
3.5 Conclusion	82
3.6 Acknowledgement	82
References	83
3.7 Appendix	86
Conclusion	91

List of Figures

1.1	The diagram of the ONEFlux pipeline. Adapted from the figure in Pastorello et al. 2020	11
1.2	Comparative analysis of NEE partitioning methods at US-ARc and CA-TP2. . .	15
1.3	Boxplot comparison of RMSE values in $\text{gC}/\text{m}^2/\text{day}$ for ecosystem respiration and gross primary productivity across various IGBP ecosystem types, demonstrating the close agreement between daytime and nighttime NEE partitioning methods.	17
1.4	Synthesized seasonal and diurnal cycles of GPP and RECO from nighttime and daytime partitioning methods across sites	18
1.5	Linear regression analysis of ecosystem flux variables: comparing FLUXNET 2015 with ONEFlux BETA Release from 20 sites.	19
1.6	Comparative analysis of the averaged seasonal cycles of NEE, RECO, GPP, LE, and H estimates from OneFlux Beta and FLUXNET2015.	29
1.7	Comparative analysis of the averaged seasonal cycles of NEE, RECO, GPP, LE, and H estimates from OneFlux Beta and FLUXNET2015.	30
1.8	Comparative analysis of the seasonal cycles of NEE, RECO, GPP, LE, and H estimates from OneFlux BETA and Reddyproc.3.4 ONEFlux BETA product for each site.	31
1.9	The location of the sites in this release, with the IGBP type of each site denoted.	32
1.10	Intended use of the downloads of the ONEFlux BETA release.	33
1.11	Word cloud from the detailed intention of use from the user downloads.	34
2.1	The relationship between growing season gross primary productivity (GPP) and phenological end of season (EOS) metrics from eddy-covariance (EC) observations or MODIS observations at deciduous broadleaf forest (DBF), evergreen needleleaf forest (ENF), mixed forest (MF), grassland (GRA), and all sites (ALL). All the time series have been detrended and the differences introduced by different sites were set as random effects.	46

2.2	Partial correlations between environmental variables and end of growing season (EOS) in Harvard forest using ground phenological records (a) and the MCD12Q2 product (b). The environmental variables included here were growing season photosynthesis (Photosynthesis), summer global radiation (R_g), autumn R_g , summer air temperature (T_{air}), autumn T_{air} , summer vapor pressure deficit (VPD), and autumn VPD. All the time series have been detrended. Here, summer is defined as June, July and August while fall is defined as September, October and November.	48
2.3	Partial correlations between detrended environmental variables and flux-derived EOS in different ecosystems. The environmental variables tested include growing season photosynthesis, summer global radiation (R_g), autumn R_g , summer air temperature (T_{air}), autumn T_{air} , summer vapor pressure deficit (VPD), and autumn VPD. All the time series have been detrended. The overall linear regression model performance (correlation) using all predictors was labeled.	50
2.4	Analysis of the flux-inferred end of growing season in deciduous broadleaf forest (DBF) sites with three different structural equation models (see the method section). The β coefficients determined for each predictor in the structural equation models are denoted, with significant coefficients ($p < 0.05$) denoted with a *. The end of growing season (EOS) and predictors at each site have been detrended. (a) predicts EOS using the direct influence of environmental variables; (b) predicts EOS using the direct influence of environmental variables along with the indirect influence through predicted photosynthesis, and (c) predicts EOS using the direct influence of both environmental variables and eddy-covariance inferred estimates of growing season photosynthesis.	51
2.5	Projections of future delays in autumn senescence dates with increasing air temperature, as estimated by three different models. The error bars indicate the standard deviations of the predictions within the 95th percentile of model performance from the MCMC process.	53
A2.1	The time series of gross primary productivity (GPP) at the site of US-Vcp in the year of 2007, indicating a double-peaked growing season. I excluded such double-peaked sites.	63
A2.2	Similar to Fig.1., but using the daytime partitioning method.	64
A2.3	Similar to Fig.1., but using the daytime partitioning method.	65
A2.4	The relationship between growing season photosynthesis (GPP) and ground observed leaf senescence or remotely sensed end of season (EOS).	66
3.1	WUEs of different ecosystems during the growing season from the flux tower observations. The blue bars denote the mean WUEs of sites in each IGBP group, and the error bars present the standard deviation. The number of sites used is also presented in (d).	74

3.2	Normalized ALE of GPP, ET and WUEs to changes in the different percentiles of VPD and SWC across different sites. The central line indicates the mean normalized ALE across sites, and the dashed area indicates the standard deviation across sites.	76
3.3	Normalized sensitivities of GPP, ET and WUEs to VPD and SWC in both arid and mesic sites based on aridity index. The bar denotes the mean sensitivities of sites in that IGBP group, and the error bar presents the standard deviation. The number of sites used were also presented.	78
3.4	Normalized sensitivities of GPP, ET and WUEs to VPD and SWC in different ecosystems. The bar denotes the mean sensitivities of sites in that IGBP group, and the error bar presents the standard deviation. The number of sites used were also presented.	79
A3.1	Similar to Figure 3.2, with the model trained using XGBOOST.	88
A3.2	The overall R^2 of the trained random forest models to predict fluxes in different sites.	89
A3.3	Feature importance of SWC and VPD in models for GPP, ET and WUE.	90

List of Tables

1.1	The RMSE and correlation between ECMWF ERA interim data and flux tower observations. The RMSE and correlation were noted for each environmental variables.	13
1.2	Regression Analysis of ONEFlux Beta Release Versus FLUXNET 2015 Data Across Key Flux Measurements, Showing Mean Coefficients of Determination (R^2), Slopes, and Intercepts with Their Respective Standard Deviations across sites	20
1.3	Summary of pipeline failure causes at AmeriFlux BASE sites	22
A1.1	Description of the sites published in this release. “n/a” indicates that information for that site was not provided by the principal investigators.	35
2.1	The model performance in predicting leaf senescence date at Harvard Forest. Here, the CDD, SIAM and PIA model represent the cold degree days model, spring phenology impacted autumn phenology model and the photosynthesis impacted phenology model, respectively; and the RMSE denotes the root mean square error of each model.	52
2.2	Comparison of predictability (R) of flux tower inferred and remotely sensed end of growing season EOS in deciduous broadleaf forest (DBF), evergreen needleleaf forest (ENF), mixed forest (MF), grassland (GRA) and ALL sites. Here, I used the International Geosphere-Biosphere Program (IGBP) scheme to identify the biome types of different sites, and the values in the table indicate the correlation between a multiple linear model prediction and observations in both models. . .	56
A2.1	Description for the sites used in this study. Lat, Lon indicated the latitude and longitude of the site respectively; and FY and LY presented the first and last year of observations available. Here, I used the International Geosphere-Biosphere Program (IGBP) scheme to identify the biome types of different sites: evergreen needleleaf forest (ENF), grassland (GRA), deciduous broadleaf forest (DBF), open shrubland (OSH), mixed forest (MF), woody savanna (SVA), wetland (WET) and evergreen broadleaf forest (EBF) sites.	62

A3.1 The detailed information of the sites used in the study. Here, I used the International Geosphere-Biosphere Program (IGBP) scheme to identify the biome types of different sites: evergreen needleleaf forest (ENF), grassland (GRA), deciduous broadleaf forest (DBF), open shrubland (OSH), mixed forest (MF), woody savanna (SVA), wetland (WET) and evergreen broadleaf forest (EBF) sites. The AI_type indicated the aridity index based classification of the sites.	86
---	----

Acknowledgments

Writing this part of the dissertation stirs a whirlwind of emotions, symbolizing the impending culmination of my transformative journey as a Ph.D. student at UC Berkeley. I wish to commence by extending my heartfelt gratitude to my esteemed advisor, Prof. Trevor F. Keenan. His unwavering support, both academically and in navigating my personal career trajectories, has been a cornerstone of my development. Prof. Keenan's commitment to fostering an environment of curiosity and independence allowed me the liberty to explore diverse academic topics, always underpinned by his insightful guidance.

Recognizing his dedication to rigorous, well-designed studies has imparted a lasting influence on my approach to research, irrespective of the future avenues I pursue. Being his first graduate student here, I deeply value his endorsement of my aspiration to join the industry upon graduation. Embarking on this journey as an international student necessitated traversing geographical and cultural frontiers. The realization of my academic endeavor would have been unattainable without the steadfast support of my parents, Mrs. Shaoping Lyu and Ms. Li Chen. Their unwavering faith and multifaceted support have been my anchor, fueling my pursuit of dreams even across oceans. The restrictions imposed by the COVID-19 pandemic have constrained my ability to visit them, accentuating the significance of family ties. I earnestly vow to prioritize reconnecting with my roots and spending more cherished moments with them in the years to come.

Further acknowledgment is due to Prof. Dennis Baldocchi and Prof. Manuela Giritto, esteemed members of my Ph.D. committee. The reservoir of knowledge and perspectives gleaned from them, coupled with collaborative learning within their research groups, has been instrumental. Prof. Baldocchi's course on flux tower observations and Prof. Giritto's insights into the intricate interplay between hydrological and global carbon cycles have enriched my academic repertoire, significantly shaping my dissertation.

Navigating the academic landscape of UC Berkeley has afforded me the privilege of collaborating with a myriad of brilliant minds. The collective wisdom and diverse experiences of Dr. Yao Zhang, Dr. Xiangzhong Luo, Dr. Chi Chen, Dr. Cecilia Chavana-Bryant, Dr. Sha Zhou, Dr. Julia Green, Dr. Hyungsuk Kim, Dr. Julia Yang and others have been pivotal in molding my academic inclinations and fostering my research independence. The camaraderie and shared intellectual pursuits with fellow graduate students Sophie Ruher, Huiqi Wang, Ngoc Nguyen, and Jackie Rue have been a source of inspiration and support. Witnessing their growth and anticipating their future contributions to science is truly exhilarating.

Reflecting on this journey, I marvel at the confluence of exploration, dedication, and self-realization that has characterized my time here. My interest in climate studies started as I was a master student in Beijing Normal University. And I would like to appreciate all the scientists who have inspired me and supported me along the way. Especially for my advisor

at Beijing, Prof. Xianglan Li and Prof. Xiao Cheng, who encouraged me to pursue what I am most passionate about and have supported me through every career decisions that I made.

My Twitter profile succinctly encapsulates my journey: “PhD candidate, love science & love cats, and hope this will eventually lead to a job!” This journey, imbued with challenges and triumphs, has been an extraordinary chapter of my life, for which I am profoundly grateful. With eager anticipation and unwavering resolve, I look forward to embracing the myriad possibilities and adventures that the next chapter holds.

Introduction

Global terrestrial ecosystems play a pivotal role in the Earth's climate system, with the carbon cycle acting as a fundamental component in balancing atmospheric constituents (Keenan and Willam 2018). Concurrently, the terrestrial biosphere absorbs an average of 15% of atmospheric carbon annually through primary productivity, yet returns nearly all of it via ecosystem respiration (Friedlingstein et al. 2022). The equilibrium of this cycle has been challenged by human activities, notably fossil fuel combustion and land-use changes since the Industrial Revolution, leading to elevated atmospheric CO_2 concentrations. Nevertheless, terrestrial ecosystems have demonstrated resilience by functioning as natural land carbon sinks, absorbing approximately one-third of human-induced CO_2 emissions and storing carbon in plant biomass and soils (Ballantyne et al. 2012).

The global water cycle is also integral to climate regulation, with evapotranspiration (ET), the transfer of water from the land to the atmosphere through both evaporation from soil and surfaces and transpiration from plants which act as its second-largest component, following precipitation (Pr) (Oki and Kanae 2006; Green et al. 2019). Global warming is expected to intensify the global hydrological cycle, leading to increases in both ET and precipitation; however, the precise magnitude and spatial distribution of these impacts remain uncertain (Konapala et al. 2020). Measuring ET accurately at large scales is challenging due to observational and measurement constraints, necessitating reliance on various methods including models, remote sensing by proxy, and upscaling of point measurements, each of which is sensitive to the specific input data and applied algorithms or models (Mueller et al. 2011; Fisher et al. 2020). Therefore, accurately constraining land ET is essential for predicting alterations in surface climate, such as heatwaves and droughts, assessing the repercussions on ecosystems and water resources, and crafting effective adaptation strategies, highlighting the pivotal role of the global terrestrial ecosystem water cycle in addressing climate change (Konapala et al. 2020).

The interconnectedness of carbon and water cycles is foundational to ecological functioning, forming a symbiotic relationship that governs the balance of life on Earth. Within this interconnected framework, Water-Use Efficiency (WUE) emerges as a pivotal trait, quantifying the adeptness of plants in utilizing water to achieve carbon gain (Keenan et al. 2013; Beer et al. 2009). It serves as the nexus between global carbon, water, and energy budgets, pro-

viding essential insights into plant responses to water stress. At the ecosystem level, WUE is calculated as the ratio of carbon uptake through photosynthesis to water lost through evapotranspiration, revealing variations across different plant species due to factors like stomatal conductance, photosynthetic rate, and plant water status (Lavergne et al. 2019). This metric is instrumental for understanding how different species, particularly those with varying levels of drought resistance, regulate water loss and adapt to water stress. Ultimately, a nuanced understanding of WUE is vital for predicting plant adaptations to climate change and enhancing the resilience of ecosystems to drought scenarios (Yi et al. 2019).

To study the terrestrial ecosystem carbon and water cycles at global scales, several observational methods have been developed. For instance, gas chambers, inventories and surveys, flux tower observations, satellite observations, isotope analysis, and ecosystem models have been devised to monitor the ecosystem carbon cycles (Ryu et al. 2019). Similarly, for observing global water cycles, the development of flux towers and satellite models has been instrumental (Fisher et al. 2020). Among these, flux tower observations stand out as a prominent method. Flux towers are ground-based observation stations that continuously monitor exchanges of carbon dioxide, water vapor, and energy between the ecosystem and the atmosphere, providing high temporal resolution data (Baldocchi 2008). This enables a deeper understanding of the diurnal and seasonal variations in carbon fluxes across various ecosystems. With this network, scientists can develop a data-driven approach to assess the flows of carbon, water, and energy to and from the terrestrial biosphere, spanning the diverse time and spatial scales over which the biosphere operates.

In my dissertation, I leverage data-driven approaches to investigate the global water and carbon cycles. Chapter 1 addresses the enhancement of the AmeriFlux network dataset, essential for monitoring land-atmosphere interactions. The AmeriFlux BASE dataset, with data from 514 operational sites in 2021, lacked critical partitioned data, rendering it unsuitable for studies focusing on ecosystem productivity, phenology, and water use efficiency (Chu et al. 2023; Novick et al. 2018). To remedy this, I improved the data quality and reliability for 63 sites by utilizing the ONEFlux processing pipeline and integrating additional data from the ECMWF ERA-Interim dataset. This improvement, detailed in Chapter 1, is crucial, as the partitioned data it provides forms the basis for my detailed examination of ecosystem productivity, phenology, and water use efficiency in Chapters 2 and 3, offering a comprehensive and cohesive dataset for these specific studies.

In Chapter 2, the examination shifts focus to the dynamics of the ecosystem phenology, emphasizing its pivotal role in regulating carbon, water, and energy fluxes in ecosystems. By analyzing data from 40 eddy-covariance flux tower sites, along with additional observational tools, a novel discovery emerged. It was revealed that heightened productivity during the growing season is not associated with its ecosystem senescence dates, which runs counter to several prevailing assumptions in the field.

Chapter 3 investigates the hypothesis that water use efficiency (WUE) in various ecosystems responds distinctly to water stress, influenced by factors like plant physiology and the extent of atmospheric and soil aridity. Utilizing flux tower observations from 64 sites across 753 site-years, I assess the response of three WUE metrics—WUE(ET), WUE(T), and iWUE—to dry conditions. The analysis reveals that while ecosystem WUE is impacted by both decreased soil water content (SWC) and increased atmospheric water demand (vapor pressure deficit, VPD), a more consistent response is observed to VPD, especially in arid regions. This study aims to clarify the complex relationship between WUE and water stress, exploring how adjusting WUE for VPD with iWUE demonstrates less variation across different environmental conditions, thereby muting the effects of atmospheric water stress.

To summarize, this dissertation serves as a testament to the pivotal role of methodical, data-driven approaches in unraveling the multifaceted dynamics governing global water and carbon cycles, offering a foundation for further scholarly inquiries in the domain.

References

- Baldocchi, D. (2008). “‘Breathing’ of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems”. In: *Australian Journal of Botany* 56.1, pp. 1–26. ISSN: 1444-9862.
- Ballantyne, A. et al. (2012). “Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years”. In: *Nature* 488.7409, pp. 70–72. ISSN: 0028-0836.
- Beer, C. et al. (2009). “Temporal and among-site variability of inherent water use efficiency at the ecosystem level”. In: *Global Biogeochemical Cycles* 23.2. ISSN: 08866236. DOI: 10.1029/2008GB003233.
- Chu, H. et al. (2023). “AmeriFlux BASE data pipeline to support network growth and data sharing”. In: *Scientific Data* 10.1, p. 614. ISSN: 2052-4463.
- Fisher, J. B. et al. (2020). “ECOSTRESS: NASA’s Next Generation Mission to Measure Evapotranspiration From the International Space Station”. In: *Water Resources Research* 56.4. ISSN: 0043-1397 1944-7973. DOI: 10.1029/2019wr026058.
- Friedlingstein, P. et al. (2022). “Global carbon budget 2022”. In: *Earth System Science Data Discussions* 2022, pp. 1–159. ISSN: 1866-3591.
- Green, J. K. et al. (Jan. 2019). “Large influence of soil moisture on long-term terrestrial carbon uptake”. en. In: *Nature* 565.7740, pp. 476–479. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-018-0848-x.
- Keenan, T. and C. Willam (2018). “The terrestrial carbon sink”. In: *Annual Review of Environment and Resources* 43, pp. 219–243. ISSN: 1543-5938.
- Keenan, T. et al. (2013). “Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise”. In: *Nature* 499.7458, pp. 324–327. ISSN: 1476-4687. DOI: 10.1038/nature12291.
- Konapala, G. et al. (2020). “Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation”. In: *Nature communications* 11.1, p. 3044. ISSN: 2041-1723.
- Lavergne, A. et al. (2019). “Observed and modelled historical trends in the water-use efficiency of plants and ecosystems”. In: *Global change biology* 25.7, pp. 2242–2257. ISSN: 1354-1013.
- Mueller, B. et al. (2011). “Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations”. In: *Geophysical Research Letters* 38.6. ISSN: 0094-8276.
- Novick, K. et al. (2018). “The AmeriFlux network: A coalition of the willing”. In: *Agricultural and Forest Meteorology* 249, pp. 444–456. ISSN: 0168-1923.
- Oki, T. and S. Kanae (2006). “Global hydrological cycles and world water resources”. In: *science* 313.5790, pp. 1068–1072. ISSN: 0036-8075.
- Ryu, Y., J. A. Berry, and D. D. Baldocchi (2019). “What is global photosynthesis? History, uncertainties and opportunities”. In: *Remote sensing of environment* 223, pp. 95–114. ISSN: 0034-4257.

- Yi, K. et al. (2019). “Linking variation in intrinsic water-use efficiency to isohydricity: a comparison at multiple spatiotemporal scales”. In: *New Phytologist* 221.1, pp. 195–208. ISSN: 0028-646X.

Chapter 1

ONEFlux BETA release for a data driven approach to understand the global carbon and water cycling

Xinchen Lu^{1,2}, Trevor F. Keenan^{1,2}

1, Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

2, Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA

This chapter details the tools and techniques I used to develop the AmeriFlux ONEFlux Beta dataset, which is now published at <https://ameriflux.lbl.gov/data/download-data-oneflux-beta/>.

Abstract

Flux towers are advanced ground-based instruments that continuously monitor the exchange of substances like carbon dioxide, water vapor, and energy between land and atmosphere. They provide valuable and continuous data necessary for thorough ecosystem studies. From the beginning, AmeriFlux researchers recognized the importance of a coordinated, network-based approach to their work. This innovative strategy led to new insights into the interaction between land and atmosphere. By 2020, the AmeriFlux network had expanded significantly, with 514 active sites and a total of 2537 years worth of data. Impressively, many of these sites have been collecting data for over a decade. However, there was a limitation in using this data for some research purposes, especially studies requiring gap-filling, and studies that focused on ecosystem respiration and/or photosynthesis. This was mainly because the data had gaps and was not presented in a way that is comparable to existing data products. In this chapter, I describe efforts to improve data quality for 63 AmeriFlux sites using the

Open Network-Enabled Flux (ONEFlux) processing pipeline. This is an upgrade from a 2019 preliminary version that covered only 17 sites. To address the data gaps, I used the ECMWF ERA-Interim dataset, and then incorporated this improved data into the ONEFlux system. The result indicated that overall, this data set showed a consistency with formerly available data set of FLUXNET 2015, with the average R^2 of the daily sum of net ecosystem exchange, ecosystem respiration, gross primary production, latent heat and sensible heat at 0.95, 0.96, 0.97, 0.99 and 0.99 respectively for the sites that have overlapping observations. While there are some discrepancies with different flux tower observation processing approaches such as results from ReddyProc, their products tended to show similar seasonal cycles. As a result, this data set has been published as an AmeriFlux ONEFlux beta product, and has been downloaded more than 168 times since its release.

1.1 Introduction

To understand the intricate interactions between the biosphere and climate, it is imperative to understand the carbon exchanges between ecosystems and the atmosphere (Booth et al. 2012; Fan et al. 1998). Concurrently, gaining insight into the associated water and energy fluxes is equally essential (Jung et al. 2010). Eddy covariance flux towers have proven to be invaluable tools in this regard (Aubinet et al. 2012; Baldocchi 2008). These towers continuously measure and estimate exchanges of carbon dioxide, water vapor, and other gas between the atmosphere and underlying ecosystems (Baldocchi 2003). By capturing turbulent eddies, or air currents, and the gas they carry, they provide uninterrupted observations. Such continuous monitoring offers profound insights into ecosystem functions, carbon and water balances, and the effects of land-use changes on carbon, water and energy fluxes (Goulden et al. 1996).

Different ecosystems may respond in varied ways to climate change (Gian-Reto et al. 2002). To capture this diversity, network-based approaches have been introduced to offer a comprehensive view of biosphere-climate interactions across different scenarios (Jones et al. 2010). These methods not only pave the way for multi-scale studies but also furnish a data-driven foundation for understanding broad ecological interactions (Rundel et al. 2009). A prime example of this strategy is the National Ecological Observatory Network (NEON) program (Kampe et al. 2010). This initiative was designed to augment our capacity to extrapolate research findings to wider spatial realms by instituting a standardized set of measurements or indicators. Furthermore, it fostered collaboration among researchers both internally and across disparate monitoring networks. For scientists probing the global dynamics of water, carbon, and energy cycles, a robust technique involves amalgamating and synthesizing flux observations from multiple eddy covariance towers (Baldocchi 2008). By adopting this approach, we have synthesized global observations that have enabled nuanced research into myriad topics. For instance, with the aid of these globally dispersed flux tower observations,

we can now discern the contrasting responses of isohydric and anisohydric species to arid conditions (Yi et al. 2019).

Numerous flux tower networks have been established around the world. Notable examples include FLUXNET (Baldocchi et al. 2001) and EuroFlux (Aubinet et al. 1999; Berbigier et al. 2001). Other regional networks like Oz-Flux and Asia-Flux are also emerging as significant contributors to this research (Beringer et al. 2022; Isaac et al. 2017; Yu and Hirano 2021). In the Americas, the AmeriFlux network stands out as the leading continental-level flux tower network. Founded over two decades ago by a trailblazing group of scientists, AmeriFlux started as individual monitoring efforts at separate sites and site-clusters (Novick et al. 2018). These towers facilitate the continuous observation of variables like the net ecosystem exchange of CO₂ (NEE), evapotranspiration (ET), and other crucial land-atmosphere interactions.

The roots of the FLUXNET 2015 dataset can be traced back to pioneering efforts like the Marconi datasets of 2000 and the influential 2007 La Thuile dataset. These earlier datasets laid the groundwork for advanced environmental data collection and analysis. Building upon this legacy, the FLUXNET 2015 dataset, sourced from flux tower networks, represents the culmination of these efforts and provided data for more than 212 locations (Falge et al. 2001; Pastorello et al. 2020). Notably, 63 of these sites offer over a decade of data. Its extensive scope and reliability have established FLUXNET 2015 as both a benchmark and a vital tool, frequently employed in various ecological research efforts. However, a significant limitation of the FLUXNET 2015 dataset is that it has not been updated with data after 2014. The data encompasses observations only up to 2014, excluding any subsequent records, and scientists need to rely on other data sources such as satellite observations to study more recent events (Schimel et al. 2019).

The AmeriFlux network has undergone notable growth over the past decade, encompassing 514 registered sites and amassing 2537 site-years of data, with a significant portion documenting observations after 2014. However, a sizable number of these sites do not offer gap-filled and partitioned data, curbing their applicability in various research endeavors. Additionally, in the AmeriFlux BASE dataset, principal investigators of each site were allowed to choose the variables of interest to upload to the network, making it difficult to compare with other datasets, such as FLUXNET 2015. In 2019, the Ameriflux data team published a BETA version of its gap-filled and partitioned data to address this issue. However, its coverage had not been extended to more sites and therefore limited its applications.

The AmeriFlux data team used the Open Network-Enabled Flux (ONFlux) data pipeline to generate the BETA version of the gap-filled and partitioned data (Pastorello et al., 2020). This code pipeline combines code developed by both the community and the team, guaranteeing a uniform and superior quality of data output. Generally, this pipeline consists of four major components, which handle quality control, meteorological variables gap-filling, CO₂

flux processing, LE and H processing respectively. However, a major challenge is that the meteorological variable gap-filling is not an integrative part of the pipeline, and therefore gap-filled meteorological data needs to be provided in order for ONEFlux to run.

In this chapter, I detail my efforts using the ONEFlux codes to equip AmeriFlux with a gap-filled and partitioned data product, which has since been published in the form of the ONEFlux Beta release. To do so, I developed an approach to use the ERA interim data to gap-fill the meteorological variables, and applied these outputs to generate an update for the ONEFlux data release. This effort aimed to provide a more up-to-date and easy to use dataset for the global flux data user community.

1.2 Methods and Materials

Ameriflux BASE

By 2020, the most prominent and widely used AmeriFlux data is the Ameriflux BASE product. In the AmeriFlux BASE product, the data team undertakes a rigorous quality control process. This process involves merging quality-checked flux-met data, updating the BASE DOI, and making the amalgamated BASE-BADM accessible to the AmeriFlux community. However, many of the sites did not provide a gap-filled and partitioned product. In this chapter, we used the BASE flux tower data product along with ONEFlux codes to generate a new data set of uniformly gap-filled and partitioned flux data.

As noted previously, by 2020, AmeriFlux had registered over 514 sites (Chu et al. 2023). However, not all of these sites could be processed due to various reasons. Our study specifically targeted sites that had both registered and made their data publicly available. Furthermore, the Ameriflux BASE datasets contained data of different processing levels, and we only considered sites with a processing level exceeding level 2. A key challenge was that some sites did not possess crucial variables required for the ONEFlux pipeline. For example, numerous sites lacked measurements like canopy height or USTAR, preventing the successful execution of the processing pipeline. Another point of divergence within the BASE dataset was the data presentation style. For instance, the Harvard Forest site (US-Ha1) did not offer latent heat measurements, rendering it unprocessable in its entirety. After trying to run this pipeline at all sites, we managed to process data from 63 sites successfully.

ONEFlux pipeline

In this chapter, I utilized the ONEFlux code pipeline, developed for the FLUXNET2015 Dataset for the data generation. This pipeline is a result of a collaboration between the European Ecosystem Fluxes Database, the integrated carbon observation system (ICOS), and AmeriFlux. It seamlessly incorporates both community and team-developed code, ensuring

a consistent and high-quality data output (refer to Figure 1.1). Starting with half-hourly data from FLUXNET sites, the pipeline places a strong emphasis on harmonization and quality control and would generate flux products for each site.

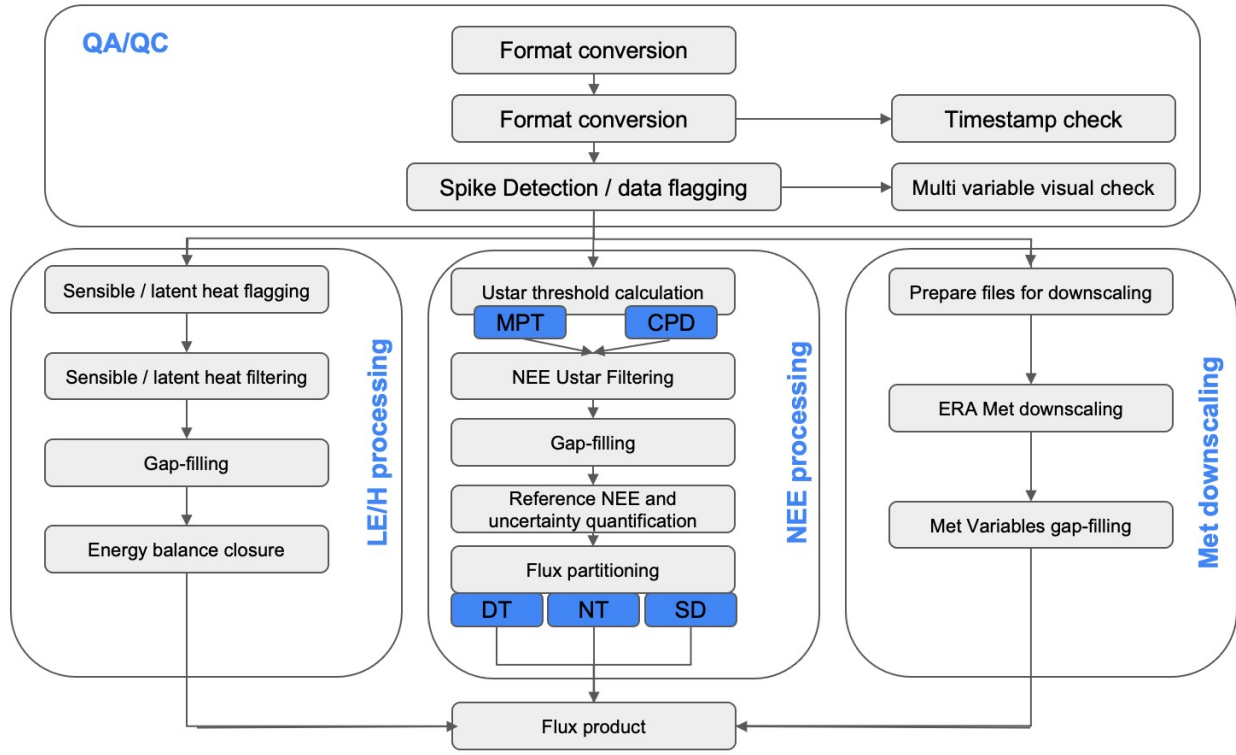


Figure 1.1: The diagram of the ONEFlux pipeline. Adapted from the figure in Pastorello et al. 2020

Broadly, the pipeline encompasses four major components. It initiates with the QA/QC process. The QA/QC process starts with format conversion, ensuring data is in a uniform structure for analysis. Next, plots for visual checks on variables are generated to identify anomalies. This is followed by spike detection, pinpointing abrupt, unusual data changes. Finally, data flagging based on the QA/QC process is marked for potential issues for further examination.

Following the initial QA/QC process, the pipeline divides into three distinct segments. The first segment focuses on sensible and latent heat. In this segment, the data undergoes several steps: it is first flagged for any noticeable issues, then filtered to remove irrelevant or erroneous data points. After filtering, the data is gap-filled using the Marginal Distribution Sampling (MDS) method (Aubinet et al. 2012), addressing any missing values to ensure a continuous dataset. Lastly, this segment includes an energy balance closure check.

Next, the net ecosystem exchange (NEE) processing takes place. Here, the USTAR threshold

is computed using two methodologies. The Moving Point Test (MPT) and Change-Point Detection (CPD) are methods used for USTAR threshold filtering in flux data analysis. MPT involves progressively adjusting the USTAR value, observing changes in statistical properties of the data at each point to find the optimal threshold (Gu et al. 2005). CPD, on the other hand, employs statistical algorithms to detect shifts in data properties like mean or variance, related to USTAR hereby identifying the appropriate threshold (Barr et al. 2013). Both methods are crucial for ensuring data reliability, especially in complex environments where turbulence significantly impacts gas exchange and were used in this pipeline. Then, the pipeline gap-filled NEE and partitioned it into gross primary production (GPP) and ecosystem respiration (RECO) using three different methods. The nighttime (NT), daytime (DT), and sundown (SD) methods each uniquely partition NEE into RECO and GPP. The nighttime method uses night data to model RECO, applying 14-day windows to account for dynamic environmental factors, with GPP derived as the residual of NEE and RECO (Reichstein et al. 2005). The daytime method combines day and night data, employing light-response curves for GPP and similar temperature relations for RECO, optimized over 8-day periods to factor in slower changes (Lasslop et al. 2010). In contrast, the sundown method, tailored for forest sites, relies on immediate post-sundown data to avoid advection effects, not requiring low turbulence filtering for NEE (Gorsel et al. 2009). Note that the sundown partitioning method would only apply for forest sites where the CO₂ concentration profile for storage fluxes was available.

Another segment of the pipeline concerns meteorological variable processing, where gaps in meteorological variables are filled using the ECMWF Re-Analysis (ERA) interim product downscaled from each site from the European Centre for Medium-Range Weather Forecasts (ECMWF).

In addition to the ONEFlux pipeline, we also run the ReddyProc package (Wutzler et al. 2018). This R-based package also offers a suite of tools for the post-processing of eddy covariance flux data, facilitating tasks such as data filtering, gap-filling, and flux partitioning. This contrast highlights ONEFlux’s emphasis on initial data integrity and ReddyProc’s adaptability for further analytical processing.

Downscaling the meteorological variables

In this chapter, I used the ERA-Interim dataset to downscale and fill gaps in the meteorological variables for each site. This dataset stems from a global atmospheric reanalysis undertaken by the European Centre for Medium-Range Weather Forecasts (ECMWF), as articulated by Dee et al. 2011. As our methodology shares similarities with those applied to the FLUXNET 2015 datasets, our specific focus was leveraging the ERA-Interim data for targeted tasks. Our primary use of this dataset centered on downsizing meteorological readings and addressing any missing data within the site observations. Key variables under

consideration were air temperature, global radiation, wind speed, vapor pressure deficit, and longwave incoming radiation. This refinement is pivotal for ensuring a robust dataset with uniform readings, making it compatible with the ONEFlux pipeline, which notably lacks an intrinsic capability for meteorological variable downscaling and gap-filling. It is pertinent to highlight that the ECMWF has since introduced an updated version of ERA-5 (Hersbach et al. 2020). However, we opted against its use to maintain alignment with the FLUXNET2015 processing standards, aiming for a dataset with heightened comparability.

When addressing gaps in the meteorological data, we identified two distinct categories. The first consists of short gaps, defined as those spanning less than six time stamps (6 half-hours). For these, we employed linear interpolation directly, without resorting to ERA-Interim data. The second category encompasses larger gaps. To fill these, we utilized ERA-Interim data, but only after establishing a linear regression relationship between the ERA-Interim data and our site-specific meteorological observations, and then filled the gap using the established relationships.

1.3 Results

ERA-interim downscaling of the meteorological variables

As it is crucial for us to establish a reliable relationship between the ERA-Interim meteorological data and the flux tower observations, we recorded the mean correlation and RMSE from the regression from each site. A summary can be found in table 1.1.

Meteorological Variables	Correlation	Correlation std	RMSE	RMSE std
Air Temperature	0.94	0.03	2.38°C	0.98°C
Atmospheric Pressure	0.91	0.08	0.38kPa	0.07kPa
VPD	0.64	0.17	0.65kPa	1.18kPa
Wind Speed	0.45	0.18	0.35m/s	0.26m/s
Precipitation	0.91	0.08	0.03m	0.07m
SW _{IN}	0.91	0.05	67W/m ⁻²	16.11W/m ⁻²
LW _{IN}	0.85	0.09	21W/m ⁻²	4.9W/m ⁻²

Table 1.1: The RMSE and correlation between ECMWF ERA interim data and flux tower observations. The RMSE and correlation were noted for each environmental variables.

The analysis of ERA-Interim data against flux tower meteorological observations, incorporating mean correlations and RMSE along with their standard deviations across sites, reveals varied levels of data consistency. Air Temperature exhibits a strong average correlation (0.94) with minimal variation (std 0.03) and a moderate RMSE (2.38°C), albeit with higher

variability (std 0.98°C). Atmospheric Pressure shows a high mean correlation (0.91) with slightly more variability (std 0.08) and a low RMSE (0.38 kPa) with minimal variation (std 0.07 kPa).

Vapor Pressure Deficit (VPD) displays a moderate correlation (0.64) but with considerable variability (std 0.17), and a relatively high RMSE (0.65 kPa) with significant variation across sites (std 1.18 kPa). Wind Speed has the lowest average correlation (0.45) with notable variability (std 0.18), and a RMSE of 0.35 m/s, again with substantial variation (std 0.26 m/s).

Precipitation maintains a consistently high correlation (0.91) with moderate variation (std 0.08) and a very low RMSE (0.03 m), although the variation is larger (std 0.07 m). Radiation parameters, Shortwave Incoming Radiation (SW_IN) and Longwave Incoming Radiation (LW_IN), show strong correlations (0.91 and 0.85 respectively) with relatively low variability, and RMSEs of 67 W/m² (std 16.11 W/m²) and 21 W/m² (std 4.9 W/m²) respectively, indicating a generally consistent replication with some site-specific differences.

Differences between different partitioning methods

First, as an example, I examined the differences between the two partitioning methods (daytime and nighttime partitioning methods) at two sites, namely US-ARc and CA-TP2 (Figure 1.2). At the US-ARC site, a grassland ecosystem, both partitioning methods demonstrate a commendable consistency in their estimations, albeit with minor mismatches. This suggests that for grassland ecosystems, the chosen partitioning approaches are robust and yield reliable data for both GPP and RECO.

Conversely, at the CA-TP2 site, an evergreen needleleaf forest, the partitioning methods to some degree diverged, especially noticeable in the year 2005, where the daytime method estimates significantly lower values for GPP and RECO than the nighttime method.

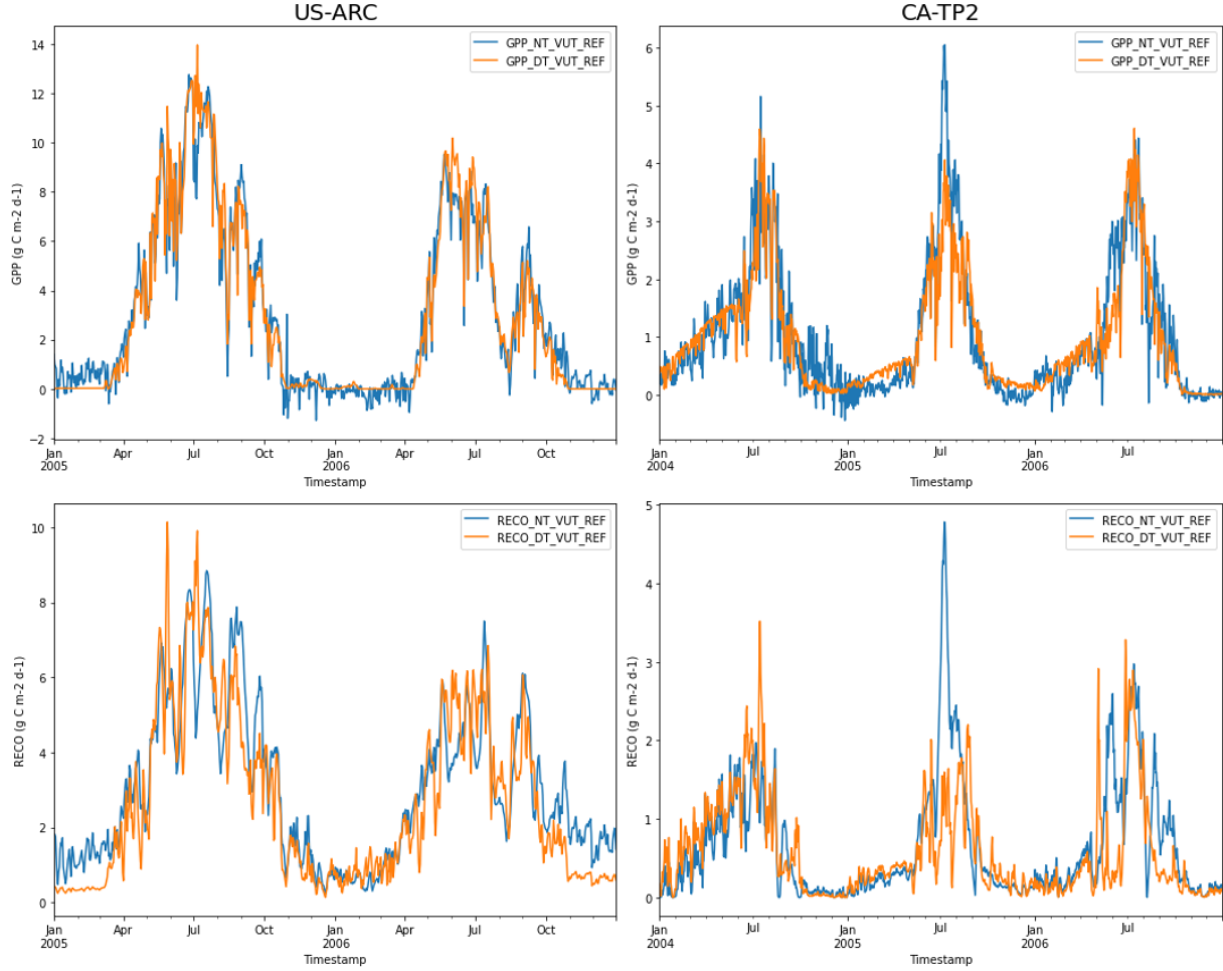


Figure 1.2: Comparative analysis of NEE partitioning methods at US-ARC and CA-TP2.

Also, I analyzed the RMSE values for daytime and nighttime NEE partitioning methods across various IGBP ecosystem types to discern their comparability in estimating GPP and Reco (Figure 1.3). Overall, the analysis revealed that median RMSE values for both methods across most ecosystems, such as Evergreen Needleleaf Forests (ENF) and Mixed Forests (MF), exhibit a close alignment, typically within a narrow range around $1.0 \text{ gC/m}^2/\text{day}$ for Reco and slightly above $0.5 \text{ gC/m}^2/\text{day}$ for GPP.

This comparability of daytime and nighttime estimates suggests a consistent level of uncertainty in the estimation of ecosystem respiration and photosynthesis, irrespective of the partitioning method applied. The findings highlight that the divergence in RMSE values between the two methods is generally minimal, supporting the notion that either method

can be selected based on factors other than accuracy, such as data availability or the specific research context, without a significant compromise on the reliability of the results.

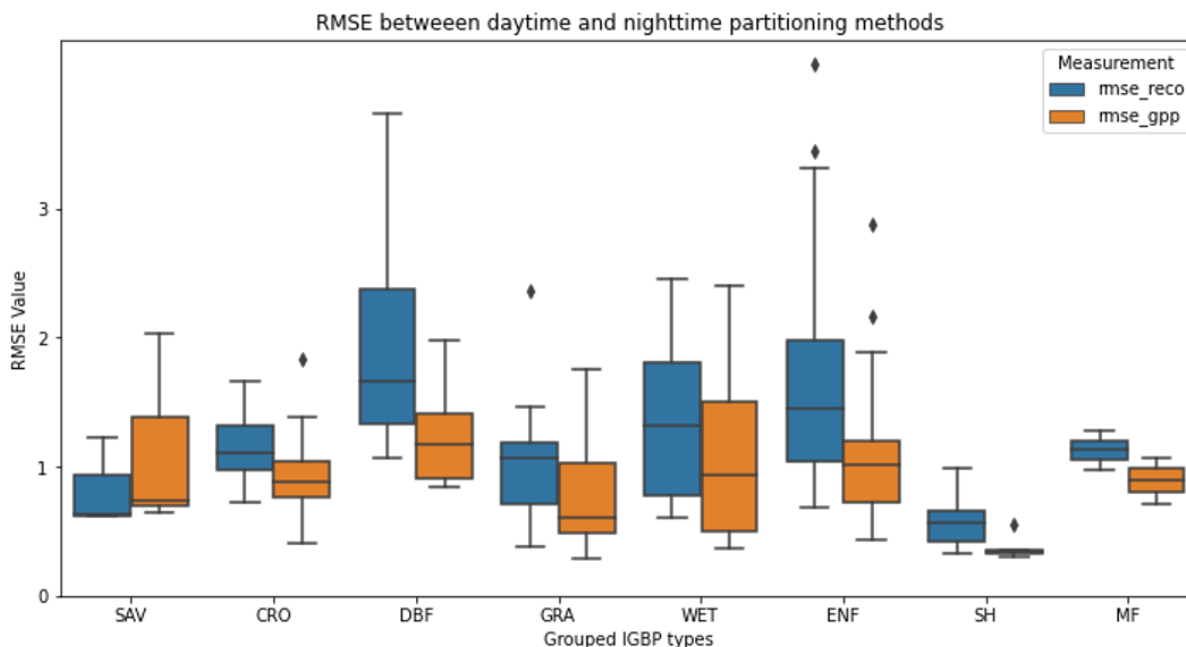


Figure 1.3: Boxplot comparison of RMSE values in $\text{gC}/\text{m}^2/\text{day}$ for ecosystem respiration and gross primary productivity across various IGBP ecosystem types, demonstrating the close agreement between daytime and nighttime NEE partitioning methods.

I also analyzed the seasonal and diurnal cycles of the two partitioning methods by averaging them across all sites (Figure 1.4). The synthesized data revealed a high degree of concordance between the two methods, as illustrated by the largely overlapping curves. Nonetheless, it was observed that the daytime partitioning method consistently resulted in lower peak values than the nighttime partitioning method, across both the seasonal and diurnal cycles. This distinction suggests that the daytime partitioning method may provide a more conservative estimate of peak GPP and RECO. The insights garnered from this comparison are invaluable, offering a retrospective reflection on the impact of methodological choices on the interpretation of ecosystem productivity and respiration dynamics.

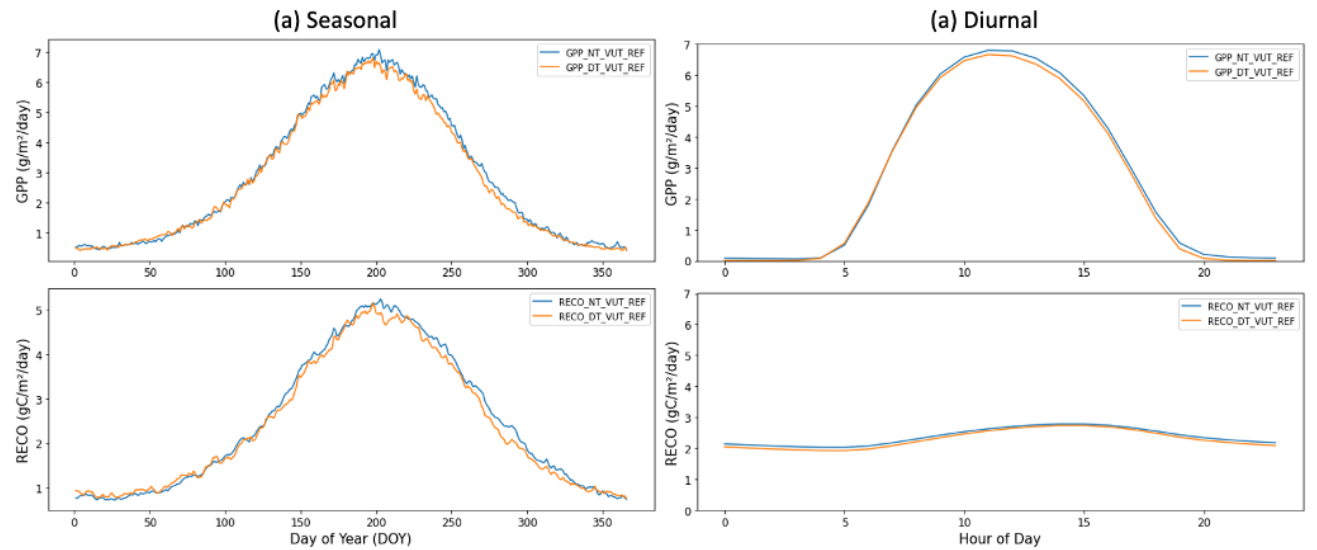


Figure 1.4: Synthesized seasonal and diurnal cycles of GPP and RECO from nighttime and daytime partitioning methods across sites

Comparison with the FLUXNET 2015 dataset and ReddyProc results

I also compared the daily output of fluxes (NEE, GPP, RECO, LE and H) with the existing FLUXNET 2015 data set. As some of the flux towers in this release were not included in the FLUXNET 2015 data set, or they may not have an overlapping time window of observation, we only used the 20 sites that have overlapping observations to conduct this comparison. The results are shown in Figure 1.5.

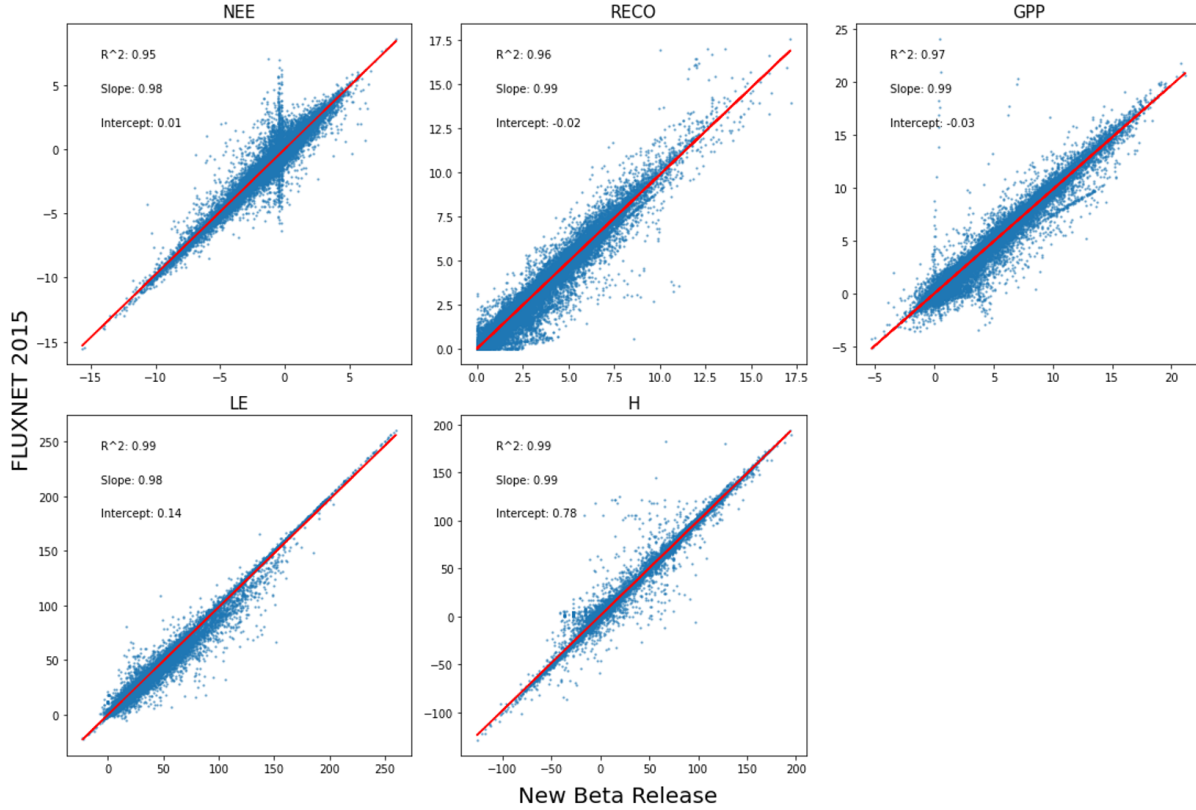


Figure 1.5: Linear regression analysis of ecosystem flux variables: comparing FLUXNET 2015 with ONEFlux BETA Release from 20 sites.

In figure 1.5, the regression analyses offer a validation of the new beta release of flux measurements when compared to the well-curated FLUXNET 2015 dataset. The evaluated flux variables all exhibit a remarkable degree of linear correlation with R^2 values spanning from 0.95 to 0.99. Notably, the slopes of these regression analyses are uniformly proximal to unity, ranging from 0.98 to 0.99, which underscores the proportionality and thus the direct comparability between the two datasets.

Furthermore, the negligible y-intercepts, hovering close to zero, are indicative of an absence of systematic error, bolstering confidence in the BETA release. Such strong alignment is not only statistically significant but also of great ecological importance, as it affirms the reliability of the beta release for quantifying ecosystem dynamics. We also validated this data set at each site, and calculated the site-specific metrics for each flux variable (Table 1.2).

Meteorological Variables	Correlation	Correlation std	RMSE	RMSE std
Air Temperature	0.94	0.03	2.38°C	0.98°C
Atmospheric Pressure	0.91	0.08	0.38kPa	0.07kPa
VPD	0.64	0.17	0.65kPa	1.18kPa
Wind Speed	0.45	0.18	0.35m/s	0.26m/s
Precipitation	0.91	0.08	0.03m	0.07m
SW _{IN}	0.91	0.05	67W/m ⁻²	16.11W/m ⁻²
LW _{IN}	0.85	0.09	21W/m ⁻²	4.9W/m ⁻²

Table 1.2: Regression Analysis of ONEFlux Beta Release Versus FLUXNET 2015 Data Across Key Flux Measurements, Showing Mean Coefficients of Determination (R^2), Slopes, and Intercepts with Their Respective Standard Deviations across sites

The mean coefficients of determination (R^2_{mean}) for Net Ecosystem Exchange (NEE), Gross Primary Production (GPP), Ecosystem Respiration (RECO), Latent Heat (LE), and Sensible Heat (H) are impressively high, ranging from 0.94 for NEE to 0.99 for both LE and H. This indicates a strong linear relationship across all flux variables. The mean slopes ($\text{Slope}_{\text{mean}}$) for these variables are similarly close to unity, varying from 0.97 for GPP and RECO to 0.99 for LE and H, suggesting that the ONEFlux beta release measurements are nearly proportional to those of FLUXNET 2015. The standard deviations for these slopes ($\text{Slope}_{\text{std}}$) are low, indicating consistency in this relationship across different datasets. The mean intercepts ($\text{Intercept}_{\text{mean}}$) are notably low for NEE, GPP, and RECO (ranging from 0.004 to 0.02), but slightly higher for LE and H (both at 0.1). However, the standard deviation for the intercept of H is notably high (1.05), suggesting more variability in this parameter compared to others.

Overall, the high R^2 values combined with slopes close to one and low intercepts for most flux measurements confirm the reliability of the ONEFlux beta release in replicating the FLUXNET 2015 data. The consistency across different flux variables underscores the potential of the beta release for accurate ecological modeling and analysis.

I also analyzed the seasonal cycles of the fluxes by both ONEFlux BETA release and the FLUXNET 2015 data set (Figure 1.6). The time series plots for each flux metric exhibit a high degree of similarity, indicating that both datasets yield closely corresponding estimations of ecosystem exchanges. This correspondence underscores the reliability of these independent flux measurement networks and supports their use in understanding terrestrial ecosystem dynamics. The consistency across the datasets reinforces the confidence in the flux data as a basis for further ecological and biogeochemical research. The congruency between the OneFlux Beta and FLUXNET2015 datasets thus reinforces the robustness of observational networks in capturing the dynamism of ecosystem processes, forming a dependable basis for advancing our scientific understanding of terrestrial carbon and energy fluxes.

At the same time, we also compared the ONEFlux BETA release with the results from ReddyProc, the results are shown in Figure 1.7. The comparative analysis between the ONEFlux beta release and the ReddyProc results highlights a significant linear relationship, as evidenced by R^2 values ranging from 0.82 to 0.97. This strong correlation demonstrates a substantive agreement between the datasets, suggesting that both processing methodologies can yield consistent flux estimations. Nevertheless, the presence of some scatter and the slight variance in slope and intercept values point towards discrepancies likely stemming from the distinct processing frameworks and quality assurance/quality control (QA/QC) protocols employed by each system.

While the slopes are near unity, indicating a close proportional relationship, the deviations, however small, may be attributable to methodological divergences in handling raw data and algorithmic differences in data interpretation. Moreover, these mismatches, although minor, underscore the critical need for harmonization between different processing systems to ensure data interoperability and to facilitate a comprehensive synthesis of ecosystem observations.

I also compared the seasonal cycles from the ONEFlux BETA release and the results from the ReddyProc processing results (Figure 1.8). The seasonal cycles of NEE, LE, and H between the two processing approaches exhibit a strong alignment, indicating a reliable estimation of these fluxes despite the methodological differences. In contrast, the GPP and RECO estimates show slight divergence between OneFlux and ReddyProc, which may be attributed to variances within the respective processing pipelines. Such discrepancies, particularly in the peak values of GPP and the mid-season values of RECO, could reflect differences in algorithmic interpretations of photosynthetic activity and respiratory processes.

Flux tower observation data for each sites

By the time we generated this data set, we had 271 sites from AmeriFlux BASE that had more than one year of observations. However, we were unable to generate a data product for all sites. More specifically, there were 208 sites where the pipeline failed. We summarized the reason that the pipeline failed at each of this site (Table 1.3). The table categorizes the specific reasons for the pipeline failures at these sites. The most common issue, encountered in 80 sites, was a processing level of 1 or lower, indicating insufficient data processing. Possible issues included the wrong VPD units. Additionally, 30 sites lacked USTAR values, which are critical for the USTAR filtering process of the flux data. Another significant hindrance was the absence of measurement height in 60 sites. Lastly, 38 sites were excluded due to the lack of other essential data, essential for comprehensive and accurate analysis. This detailed breakdown helps in identifying the key areas for improvement in data collection and processing at AmeriFlux BASE sites, guiding future efforts to enhance data quality and pipeline efficiency.

Reason the pipeline not running properly	Number of sites
Processing level ≤ 1	80
No USTAR values	30
No measurement height	60
Lack of other essential data	38

Table 1.3: Summary of pipeline failure causes at AmeriFlux BASE sites

For each site, we generated a gap-filled and partitioned data product aligning with the FLUXNET 2015 dataset. We provided two versions for each site: the FULLSET and the SUBSET, both sourced from the ONEFlux pipeline. The FULLSET encompasses all processing variables, including every data product variant, quality indicators, uncertainty measures, and intermediary variables for a comprehensive understanding of each step. Conversely, the SUBSET offers a curated selection tailored to the anticipated needs of most users, highlighting pivotal quality and uncertainty variables. Additionally, we supplied the AUXMETEO and ERA products. AUXMETEO presents results from downscaling micrometeorological variables via ERA-Interim reanalysis data, shedding light on linear regression and variable-specific error/correlation. AUXNEE focuses on NEE processing, detailing USTAR filtering results, applied thresholds, and model efficacy. ERA showcases a comprehensive record (1989-2017) of downscaled micrometeorological variables rooted in the ERA-Interim reanalysis data. In summary, we generated flux tower data from 63 sites in total. The earliest flux tower observation dates back to 1984, while the most recent one is from 2018 (Figure 1.9). On average, each site has approximately 7.78 years of observations. The majority of these sites are situated in Canada and the United States. Collectively, these sites span 11 distinct vegetation types based on the International Geosphere-Biosphere Programme (IGBP) classification. Specifically, the sites encompass Evergreen Needleleaf Forest (ENF, 20 sites), Croplands (CRO, 9 sites), Grasslands (GRA, 9 sites), Deciduous Broadleaf Forest (DBF, 7 sites), Permanent Wetlands (WET, 6 sites), Open Shrublands (OSH, 3 sites), and two each for Mixed Forest (MF), Woody Savannas (WSA), Closed Shrublands (CSH), and Water Bodies (WAT), with a single site for Savannas (SAV). Detailed information of each site could be found in the Appendix table 1.1.

1.4 Discussion

In this chapter, we presented the effort to produce an updated version of the ONEFlux BETA release, in which we produced the gap-filled and partitioned flux tower observations for 63 sites, which significantly increased the flux data availability for the Ameriflux data set. Among the 63 sites, there were 46 sites that were not included in the earlier version of the Ameriflux BETA release; at the same time, there were 22 sites that were not included in the FLUXNET 2015 data sets. Since the publication of this version, it has been downloaded

more than 168 times, and its applications in the flux tower community include the use for multi-site synthesis, remote sensing studies and teaching.

The comparison of ERA-Interim data with flux tower observations reveals varied reliability across meteorological parameters. Air Temperature and Atmospheric Pressure exhibit high correlations, indicating strong consistency, though Air Temperature shows some site-specific variations in RMSE. Vapor Pressure Deficit (VPD) and Wind Speed, with moderate correlations and higher RMSEs, are more influenced by local environmental conditions. Precipitation and Radiation parameters display high correlations, but their RMSE variability points to localized differences. These results highlight the strengths and limitations of ERA-Interim data in replicating diverse atmospheric conditions.

With regard to the NEE partitioning, the analysis at US-ARC and CA-TP2 sites elucidates the relative robustness of partitioning methods across ecosystems. At US-ARC, a congruence between daytime and nighttime approaches affirms their reliability for grassland ecosystems. However, the CA-TP2 site displayed notable methodological disparities, particularly in the year 2005, reflecting the complexity of forest ecosystems and the importance of method selection in accurate carbon flux interpretation (Figure 4). Additionally, RMSE evaluations across various IGBP types suggest that both partitioning methods maintain a comparable level of uncertainty, permitting flexibility in method choice based on research design rather than accuracy alone (Figure 5). This underscores the broader applicability of these methods in ecological carbon cycle research.

In consolidating our findings, the linear regression analysis presented in Figure 4 substantiates the ONEFlux beta release data by comparing it to the FLUXNET 2015 dataset. The analysis reveals high coefficients of determination (R^2) for all flux variables, indicating robust linear correlations and minimal systematic error. The close-to-unity slopes affirm the proportional relationship between the datasets, corroborating the beta release's validity in capturing ecosystem flux dynamics (Table 2). Concurrently, the comparison with ReddyProc results, shown in Figure 5, exhibits a significant linear relationship across datasets, reinforcing the potential of varied processing methodologies to yield consistent estimations. Nonetheless, the slight variances in slope and intercept values signal the impact of differing processing and QA/QC protocols. Such findings highlight the need for methodological harmonization to enhance data compatibility and ecological synthesis.

In comparison to other flux datasets, the ONEFlux BETA offers one of the most current and comprehensive products, primarily emphasizing sites in Canada and the United States. For context, the Marconi dataset from 2000 contains data from 38 sites over 97 site-years, gathered during a workshop and exclusive to its participants. LaThuile 2007 is a broader compilation with data from 252 sites across 965 site-years, establishing a consistent code base for data comparability and broadening user access with a three-tiered restriction. FLUXNET2015,

on the other hand, presents data from 212 sites amassing 1532 site-years, with the advantage of a jointly established standardized code base and wider data accessibility featuring a two-tiered restriction. The ICOS and Drought-2018 teams have also released eddy covariance flux observational data from 52 stations spanning 1989-2018, some beyond the ICOS network, all processed through the ONEFlux pipeline and aligning with standards set by ICOS L2, AmeriFlux, and FLUXNET2015.

Since the release of this dataset, the Ameriflux data team advanced by launching the Ameriflux Fluxnet dataset. Specifically, FLUXNET data offerings, enriched and gap-filled products that include variables such as partitioned gross primary productivity, have played a pivotal role in global datasets, serving model validation and benchmarking for years. In this context, the Ameriflux data team collaborated with international partners like ICOS to formulate the ONEFlux (Open Network-Enabled Flux) codes, subsequently contributing to the birth of the FLUXNET2015 data product. Moreover, the Ameriflux data team is capitalizing on the premium, standardized BASE data product as a foundation for the ONEFlux codes, aiming to devise the next-generation FLUXNET data product tailored for AmeriFlux sites. The protocols and procedures refined for the BASE data-processing pipeline are now being adapted to curate the FLUXNET product. By 2022, the Ameriflux data team had unveiled the AmeriFlux FLUXNET data product for 79 sites, and they remain committed to regularly updating and releasing the AmeriFlux FLUXNET datasets in harmony with other flux network partners.

As of December 2023, this ONEFlux Beta data product has been downloaded for 168 times, with the main areas of intended use summarized in Figure 1.10. 'Research on fluxes synthesis' emerges as the most frequent intended use, with over 61 instances, indicating a strong focus within the user community on synthesizing flux data for research purposes. The next prominent categories, 'Research - Remote Sensing' and 'Research - Land modeling/model', show approximately 37 and 29 instances, respectively, illustrating significant interest in applying the dataset to these research areas. Uses categorized as 'Other' and 'Research - Other' have a modest representation, each with around 10 mentions. Notably, the dataset's role in educational activities, as represented by 'Education (Teacher or Student)', is the least reported use, with under 10 instances.

Further analyzing the word cloud of the detailed intention of use by the users, I found that the central terms such as "model," "analysis," and "ecosystem," suggest a thoughtful engagement with ecological data, while the significant presence of "GPP" and "NEE" reflects a considerate emphasis on understanding carbon dynamics within ecosystems (Figure 1.11). The term "validation" echoes the conscientious approach users take towards ensuring data integrity, and "remote sensing" indicates a reliance on advanced techniques for ecological observation.

Additionally, the inclusion of terms like "RUBISCO" and "drought" points to an exploration of plant physiology and environmental stress factors, demonstrating the product's utility in delicately probing the complexities of ecological systems. This visualization shares the narrative of the product's use and also paints a broader picture of the diligent and methodical efforts by researchers to elucidate the interactions within ecosystems under varying environmental conditions.

1.5 Conclusion

In this chapter, we have detailed the rigorous efforts to refine the AmeriFlux datasets via the ONEFlux pipeline, aligning it with the FLUXNET 2015 dataset. The meticulous process has resulted in a robust dataset that demonstrates high coefficients of determination, with R^2 values of 0.94 for NEE, 0.95 for GPP, and 0.99 for both LE and H when compared against the FLUXNET 2015 dataset, indicating strong linear correlations. This dataset has proven reliable across diverse ecosystems, a cornerstone for accurately capturing the dynamics of carbon and water exchanges. The ONEFlux beta release, through its rigorous validation process and high-quality data, has become an important tool for researchers to explore and understand the intricacies of biosphere-atmosphere interactions.

References

- Aubinet, M., T. Vesala, and D. Papale (2012). *Eddy covariance: a practical guide to measurement and data analysis*. Springer Science & Business Media. ISBN: 94-007-2350-4.
- Aubinet, M. et al. (1999). “Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology”. In: *Advances in ecological research*. Vol. 30. Elsevier, pp. 113–175. ISBN: 0065-2504.
- Baldocchi, D. (2008). “‘Breathing’ of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems”. In: *Australian Journal of Botany* 56.1, pp. 1–26. ISSN: 1444-9862.
- Baldocchi, D. et al. (2001). “FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities”. In: *Bulletin of the American Meteorological Society* 82.11, pp. 2415–2434. ISSN: 0003-0007.
- Baldocchi, D. D. (2003). “Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future”. In: *Global change biology* 9.4, pp. 479–492.
- Barr, A. G. et al. (2013). “Use of change-point detection for friction–velocity threshold evaluation in eddy-covariance studies”. In: *Agricultural and Forest Meteorology* 171, pp. 31–45.
- Berbigier, P., J.-M. Bonnefond, and P. Mellmann (2001). “CO₂ and water vapour fluxes for 2 years above Euroflux forest site”. In: *Agricultural and Forest Meteorology* 108.3, pp. 183–197. ISSN: 0168-1923.
- Beringer, J. et al. (2022). “Bridge to the future: Important lessons from 20 years of ecosystem observations made by the OzFlux network”. In: *Global Change Biology* 28.11, pp. 3489–3514. ISSN: 1354-1013.
- Booth, B. B. et al. (2012). “High sensitivity of future global warming to land carbon cycle processes”. In: *Environmental Research Letters* 7.2, p. 024002. ISSN: 1748-9326.
- Chu, H. et al. (2023). “AmeriFlux BASE data pipeline to support network growth and data sharing”. In: *Scientific Data* 10.1, p. 614. ISSN: 2052-4463.
- Dee, D. P. et al. (2011). “The ERA-Interim reanalysis: Configuration and performance of the data assimilation system”. In: *Quarterly Journal of the royal meteorological society* 137.656, pp. 553–597. ISSN: 0035-9009.
- Falge, E. et al. (2001). “Gap filling strategies for long term energy flux data sets”. In: *Agricultural and Forest Meteorology* 107.1, pp. 71–77.
- Fan, S. et al. (1998). “A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models”. In: *Science* 282.5388, pp. 442–446. ISSN: 1095-9203.
- Gian-Reto, W. et al. (2002). “Ecological responses to recent climate change”. In: *Nature* 416.6879, pp. 389–395.

- Gorsel, E. van et al. (2009). “Estimating nocturnal ecosystem respiration from the vertical turbulent flux and change in storage of CO₂”. In: *Agricultural and forest meteorology* 149.11, pp. 1919–1930. ISSN: 0168-1923.
- Goulden, M. L. et al. (1996). “Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy”. In: *Global change biology* 2.3, pp. 169–182. ISSN: 1354-1013.
- Gu, L. et al. (2005). “Objective threshold determination for nighttime eddy flux filtering”. In: *Agricultural and Forest Meteorology* 128.3-4, pp. 179–197.
- Hersbach, H. et al. (2020). “The ERA5 global reanalysis”. In: *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049. ISSN: 0035-9009.
- Isaac, P. et al. (2017). “OzFlux data: network integration from collection to curation”. In: *Biogeosciences* 14.12, pp. 2903–2928. ISSN: 1726-4170.
- Jones, K. B. et al. (2010). “Design and importance of multi-tiered ecological monitoring networks”. In: *Long-Term Ecological Research: Between Theory and Application*, pp. 355–374. ISSN: 9048187818.
- Jung, M. et al. (2010). “Recent decline in the global land evapotranspiration trend due to limited moisture supply”. In: *Nature* 467.7318, pp. 951–954. ISSN: 0028-0836.
- Kampe, T. U. et al. (2010). “NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure”. In: *Journal of Applied Remote Sensing* 4.1, p. 043510. ISSN: 1931-3195.
- Lasslop, G. et al. (2010). “Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation”. In: *Global Change Biology* 16.1, pp. 187–208. ISSN: 1354-1013.
- Novick, K. et al. (2018). “The AmeriFlux network: A coalition of the willing”. In: *Agricultural and Forest Meteorology* 249, pp. 444–456. ISSN: 0168-1923.
- Pastorello, G. et al. (2020). “The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data”. In: *Scientific data* 7.1, pp. 1–27. ISSN: 2052-4463.
- Reichstein, M. et al. (2005). “On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm”. In: *Global change biology* 11.9, pp. 1424–1439.
- Rundel, P. W. et al. (2009). “Environmental sensor networks in ecological research”. In: *New Phytologist* 182.3, pp. 589–607. ISSN: 0028-646X.
- Schimel, D., F. D. Schneider, and JPL Carbon and Ecosystem Participants (2019). “Flux towers in the sky: global ecology from space”. In: *New Phytologist* 224.2, pp. 570–584. ISSN: 0028-646X.
- Wutzler, T. et al. (2018). “Basic and extensible post-processing of eddy covariance flux data with REddyProc”. In: *Biogeosciences* 15.16, pp. 5015–5030. ISSN: 1726-4170.
- Yi, K. et al. (2019). “Linking variation in intrinsic water-use efficiency to isohydricity: a comparison at multiple spatiotemporal scales”. In: *New Phytologist* 221.1, pp. 195–208. ISSN: 0028-646X.

Yu, G. and T. Hirano (2021). “Review and future perspective of AsiaFlux”. In: *Journal of Agricultural Meteorology* 77.1, pp. 1–1.

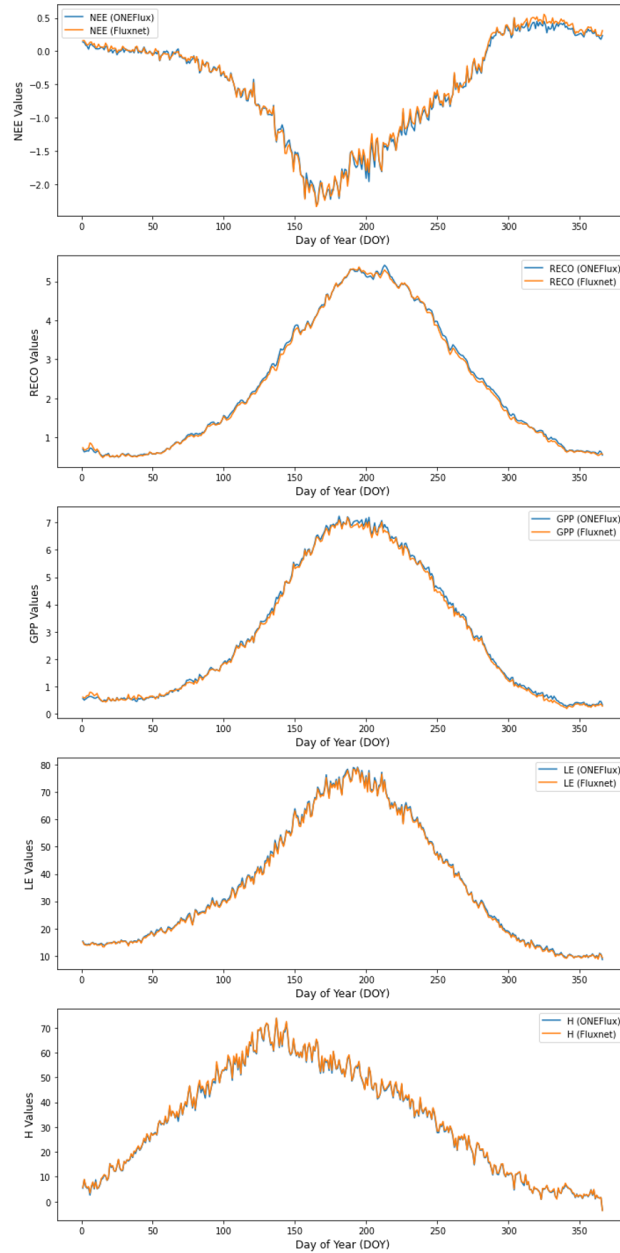


Figure 1.6: Comparative analysis of the averaged seasonal cycles of NEE, RECO, GPP, LE, and H estimates from OneFlux Beta and FLUXNET2015.

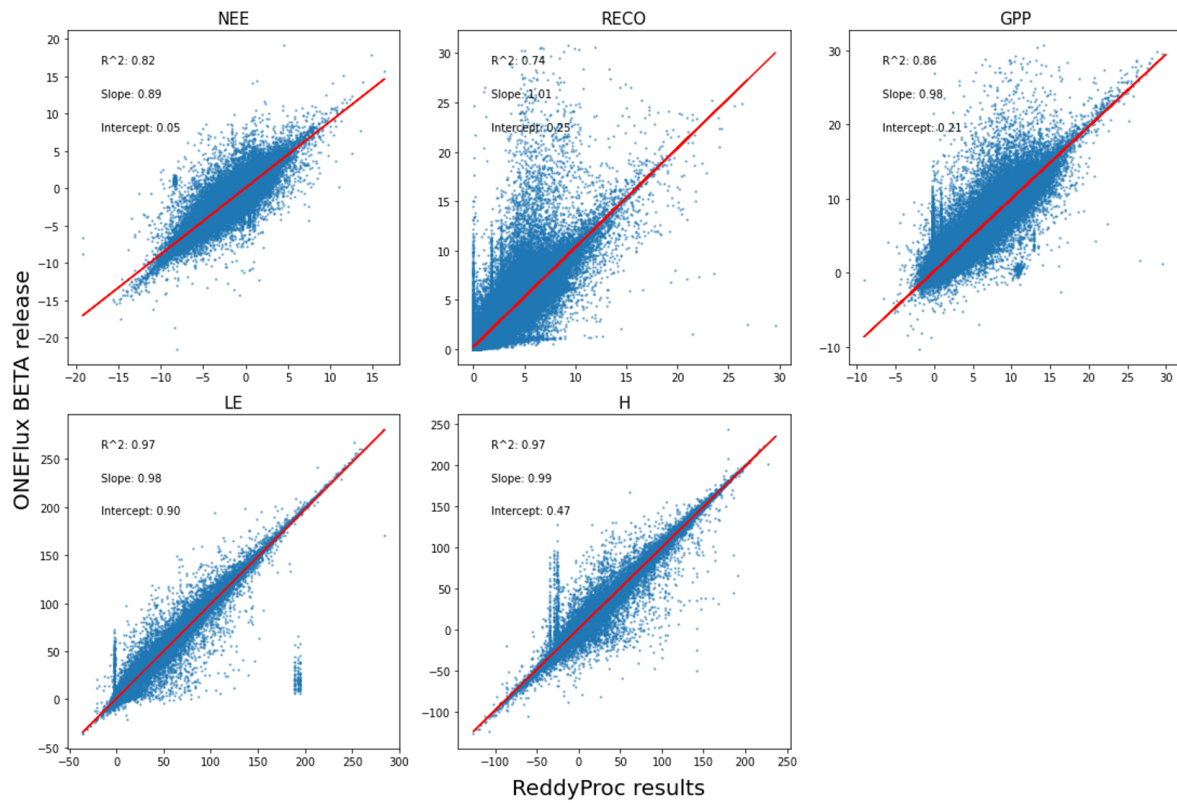


Figure 1.7: Comparative analysis of the averaged seasonal cycles of NEE, RECO, GPP, LE, and H estimates from OneFlux Beta and FLUXNET2015.

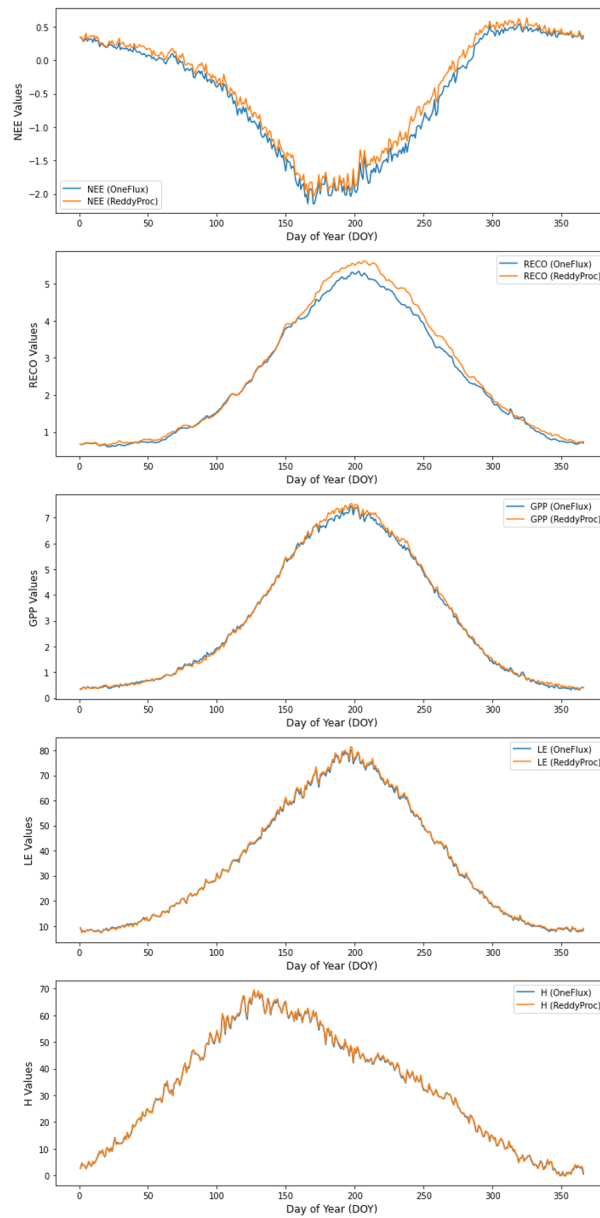


Figure 1.8: Comparative analysis of the seasonal cycles of NEE, RECO, GPP, LE, and H estimates from OneFlux BETA and Reddyproc.3.4 ONEFlux BETA product for each site.

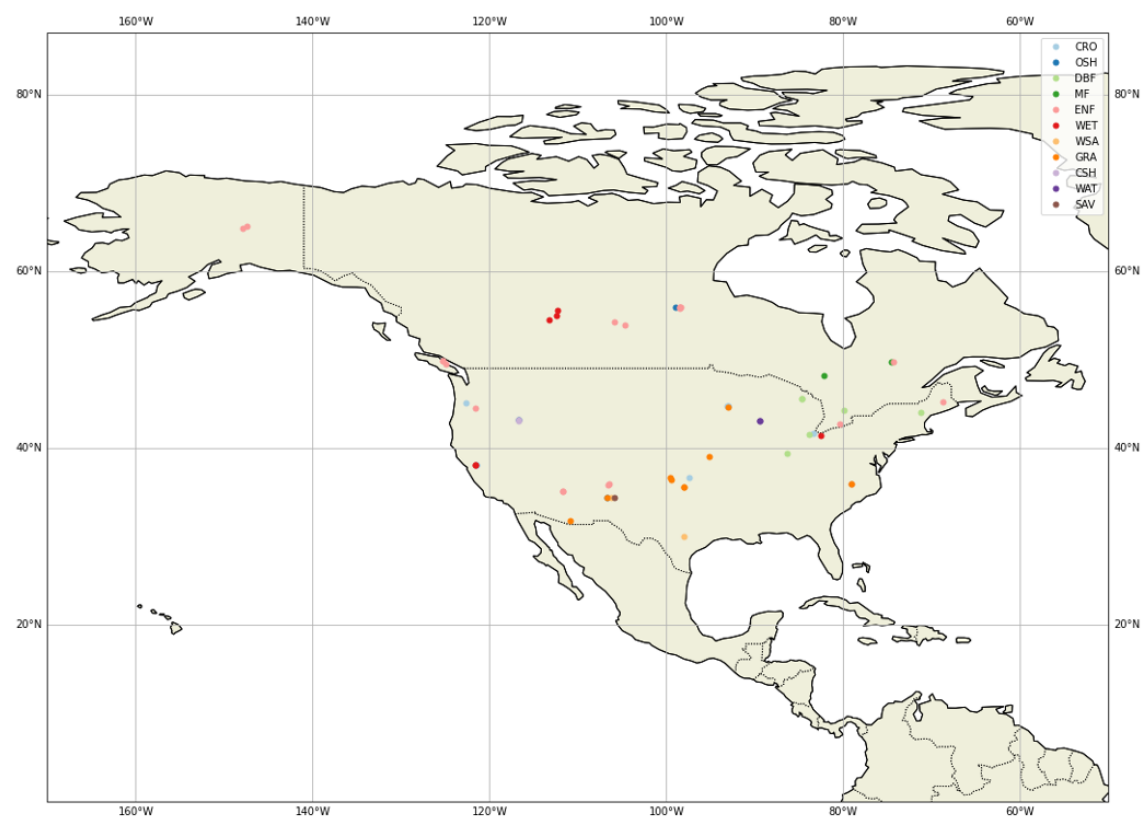


Figure 1.9: The location of the sites in this release, with the IGBP type of each site denoted.

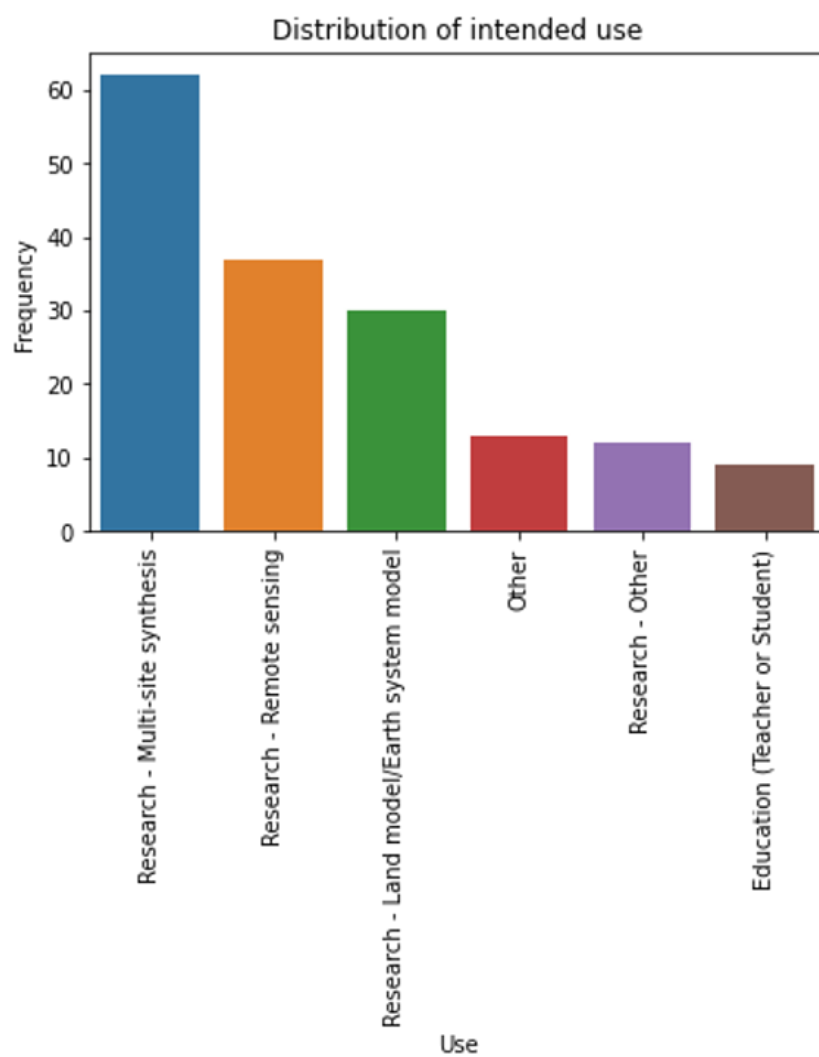


Figure 1.10: Intended use of the downloads of the ONEFlux BETA release.

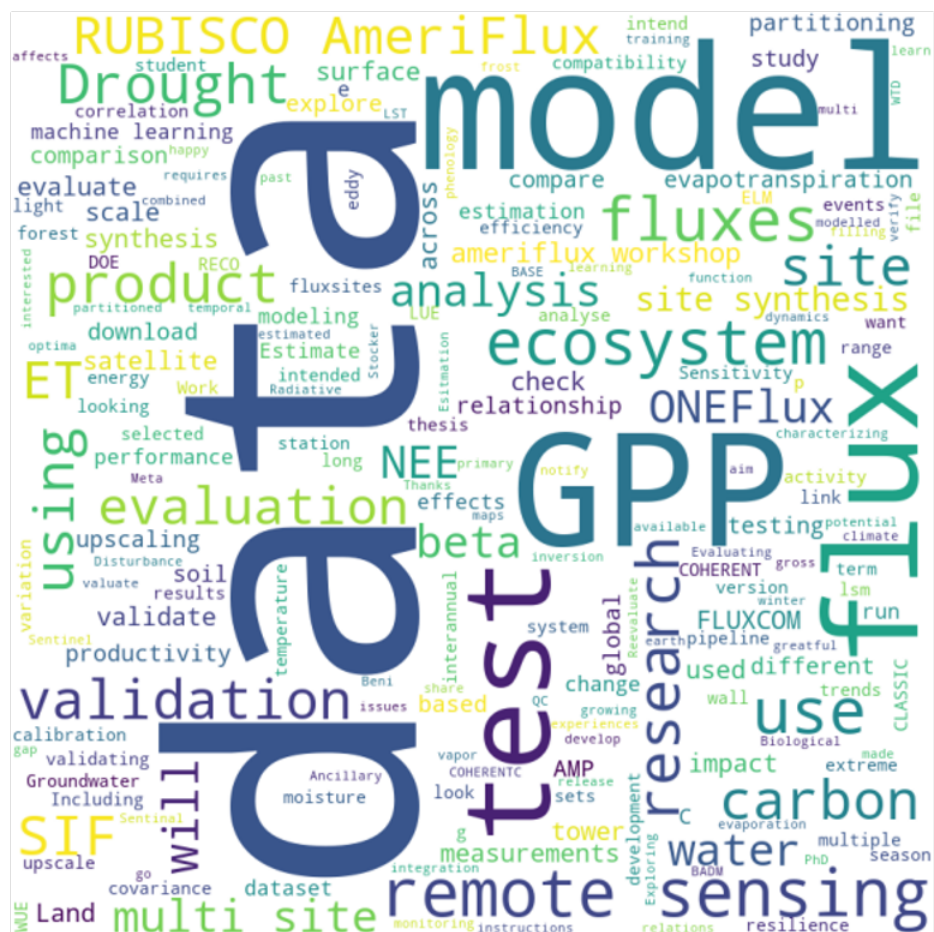


Figure 1.11: Word cloud from the detailed intention of use from the user downloads.

1.6 Appendix

Table A1.1: Description of the sites published in this release. “n/a” indicates that information for that site was not provided by the principal investigators.

Site	Start Year	End Year	Latitude	Longitude	Elevation
US-ARM	2003	2018	36.6058	-97.4888	314
US-Ses	2007	2017	34.3349	-106.7442	1604
US-Ro6	2017	2017	44.6946	-93.0578	282
US-UMB	2007	2017	45.5598	-84.7138	234
CA-Qc2	2007	2010	49.7598	-74.5711	n/a
CA-SF2	2001	2006	54.2539	-105.8775	520
CA-WP1	2003	2009	54.9538	-112.467	540
US-FR2	2005	2008	29.9495	-97.9962	271.9
US-Bar	2005	2017	44.0646	-71.2881	272
US-Ro1	2004	2016	44.7143	-93.0898	290
US-Ro5	2017	2017	44.691	-93.0576	283
US-ARb	2005	2006	35.5497	-98.0402	424
US-ARc	2005	2006	35.5465	-98.04	424
US-OWC	2015	2016	41.3795	-82.5125	174
US-Bi1	2016	2017	38.0992	-121.4993	-2.7
US-UMd	2007	2017	45.5625	-84.6975	239
US-Seg	2007	2017	34.3623	-106.702	1596
US-Dk1	2001	2008	35.9712	-79.0934	168
CA-WP3	2004	2006	54.47	-113.32	n/a
US-Rls	2014	2017	43.1439	-116.7356	1608
CA-WP2	2004	2006	55.5375	-112.3343	n/a
US-AR1	2009	2012	36.4267	-99.42	611
US-Vcm	2007	2017	35.8884	-106.5321	3030
US-Vcp	2007	2017	35.8642	-106.5967	2500
US-Men	2014	2017	43.0772	-89.403	260
US-Tw4	2013	2017	38.1027	-121.6413	-5
US-Ho1	1996	2018	45.2041	-68.7402	60
CA-Cbo	1994	2014	44.3167	-79.9333	120
US-Dk2	2001	2008	35.9736	-79.1004	168
US-Fuf	2005	2010	35.089	-111.762	2180
CA-Gro	2003	2014	48.2167	-82.1556	340
US-Dk3	2001	2008	35.9782	-79.0942	163
US-AR2	2009	2012	36.6358	-99.5975	646

US-Bi2	2017	2017	38.1091	-121.5351	-5
US-CRT	2011	2013	41.6285	-83.3471	180
CA-NS3	2001	2005	55.9117	-98.3822	260
US-Mpj	2008	2017	34.4385	-106.2377	2196
US-Ro4	2014	2017	44.6781	-93.0723	274
US-Wgr	2015	2015	45.1129	-122.656	52
CA-Ca1	1996	2010	49.8673	-125.3336	300
US-Oho	2004	2013	41.5545	-83.8438	230
US-Fmf	2005	2010	35.1426	-111.7273	2160
CA-NS2	2001	2005	55.9058	-98.5247	260
US-MMS	1999	2017	39.3232	-86.4131	275
US-Me2	2002	2017	44.4526	-121.5589	1253
CA-Ca2	1999	2010	49.8705	-125.2909	300
CA-NS1	2001	2005	55.8792	-98.4839	260
CA-TP4	2002	2017	42.7102	-80.3574	184
US-Uaf	2003	2017	64.8663	-147.8555	155
US-SRG	2008	2017	31.7894	-110.8277	1291
US-Tw1	2011	2017	38.1074	-121.6469	-5
US-Prr	2010	2016	65.1237	-147.4876	210
US-Pnp	2016	2017	43.0896	-89.4158	260
US-Rws	2014	2017	43.1675	-116.7132	1425
CA-NS5	2001	2005	55.8631	-98.485	260
US-Tw2	2012	2013	38.0969	-121.6365	-5
CA-Qfo	2003	2010	49.6925	-74.3421	382
US-Rms	2014	2017	43.0645	-116.7486	2111
US-KFS	2007	2017	39.0561	-95.1907	310
CA-NS6	2001	2005	55.9167	-98.9644	244
CA-Ca3	2001	2010	49.5346	-124.9004	n/a
US-Wjs	2007	2017	34.4255	-105.8615	1931
CA-SJ2	2002	2010	53.945	-104.649	580

Transition from chapter 1 to chapter 2

In Chapter 1, I detailed the efforts made to generate a gap-filled and partitioned dataset for the Ameriflux ONEFlux BETA release. This updated version expanded the dataset from 17 sites to over 63 sites, offering standardized and synthesized flux products for each location up to 2018. These sites represent a diverse range of ecosystem types, encompassing 11 distinct vegetation categories based on the International Geosphere-Biosphere Programme (IGBP) classification. Specifically, the sites include Evergreen Needleleaf Forest (ENF, 20 sites), Croplands (CRO, 9 sites), Grasslands (GRA, 9 sites), Deciduous Broadleaf Forest (DBF, 7 sites), Permanent Wetlands (WET, 6 sites), Open Shrublands (OSH, 3 sites). There are also two sites each for Mixed Forest (MF), Woody Savannas (WSA), Closed Shrublands (CSH), and Water Bodies (WAT), and a single site for Savannas (SAV).

The provision of these datasets has paved the way for us to study global carbon and water cycles from a data-driven perspective. This means we are now in a position to address many unresolved questions in this domain using flux tower observations from around the world. In the subsequent chapter 2, I flux tower observations from Ameriflux and other global scales to examine the response of ecosystem autumn phenology to climate change.

The duration of the growing season profoundly affects the carbon, water, and energy exchanges within global terrestrial ecosystems. While numerous studies have indicated a trend toward an earlier onset of the growing season, primarily attributed to warmer spring temperatures, the factors determining the conclusion of the growing season (EOS) are not as thoroughly understood in most ecosystems. A recent hypothesis posits that the productivity during the growing season can directly influence the timing of its end, suggesting a more robust growing season might culminate earlier. This proposed relationship, however, has not been sufficiently explored using in-situ data from a range of ecosystems. In Chapter 2, I delved into the ONEFlux BETA release, among other tools, to examine this supposition.

Chapter 2

Reevaluating autumn phenology: the limited role of growing season productivity compared with environmental factors

Xinchen Lu^{1,2}, Trevor F. Keenan^{1,2}

1, Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

2, Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA

*An edited version of this chapter is published in Global Change Biology:
<https://doi.org/10.1111/gcb.16104>*

Abstract

The length of the growing season has a large influence on the carbon, water and energy fluxes of global terrestrial ecosystems. While there has been mounting evidence of an advanced start of the growing season mostly due to elevated spring air temperatures, the mechanisms that control the end of the growing season (EOS) in most ecosystems remain relatively less well understood. Recently, a strong lagged control of EOS by growing season photosynthesis has been proposed, suggesting that more productive growing seasons lead to an earlier EOS. However, this relationship has not been extensively tested with in-situ observations across a variety of ecosystems. In this chapter, I use observations from 40 eddy-covariance flux tower sites in temperate and boreal ecosystems in the northern hemisphere with more than 10 years of observations (594 site-years), ground observations of phenology, satellite

observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) and three leaf senescence models to test the extent of a relationship between growing season photosynthesis and end of season senescence. The results suggest that there is no significant negative relationship between growing season photosynthesis and observed leaf senescence, flux-inferred EOS estimates, or remotely sensed phenological metrics, in most ecosystems. On the contrary, while I found negative effects of summer air temperatures and autumn vapor pressure deficit on EOS, which indicated the environmental factors control over autumn phenology, more productive growing seasons were typically related to a later, not earlier, EOS. Compared with environmental factors, incorporating the growing season photosynthesis in empirical and process-based models did not improve their performances. Our results challenge recent reports of negative effects of photosynthesis on EOS timing, and suggest those results may not hold over a large range of ecosystems.

2.1 Introduction

Global terrestrial ecosystems take up around one third of human emission of CO₂ (Friedlingstein et al. 2020). With recent climate change, the seasonality of terrestrial ecosystems, a sensitive indicator of biosphere-climate interactions, has been reported to be changing (Piao et al. 2007; Seddon et al. 2016). Longer growing seasons have been associated with increased carbon uptake by many ecosystems (Keenan et al. 2014; Baldocchi et al. 2005; Churkina et al. 2005),. However, the dominant controls of the growing season length remain relatively poorly understood.

While rising air temperatures are a major driver in the advancing Start of Growing Season (SOS) in temperate and boreal ecosystems (Chmielewski and Rötzer 2002; Song et al. 2010; Yu et al. 2013; Chuine and Régnière 2017; Nizinski and Saugier 1988), other factors such as changes in photoperiod, soil moisture, and genetic variations among species also play significant roles. Meanwhile, modelling of the timing of the end of growing season (EOS) has been challenging, as the mechanisms that control the EOS have not been thoroughly investigated or well understood (Lang et al. 2019; Zhang et al. 2020; Tang et al. 2014), and many studies found a much smaller scale of EOS changes compared with the SOS changes (Jeong et al. 2011; Park et al. 2016). There are many factors that may contribute to the changes of EOS, such as the limiting factors of plant productivity of air temperature, radiation, soil moisture and VPD (Archetti et al. 2013; Körner and Basler 2010). For instance, dry conditions during the growing season could have significant impacts on the autumn phenology. Specifically, Zhang et al. 2020 elucidate that water availability, shaped by soil moisture and mean annual temperature, is increasingly recognized as a critical limiting factor for late-season photosynthesis, suggesting that dry conditions during the growing season can precipitously advance the EOS. Complementing this, Wolf et al. 2016 demonstrate that warming-induced early vegetation activity, while initially compensating for reduced carbon

uptake during summer droughts, leads to earlier soil moisture depletion, thereby potentially advancing autumn phenological events due to increased water stress. Therefore, it is evident that dry conditions experienced in earlier seasons can significantly influence the timing and nature of autumn phenology. At the same time, other factors such as nutrient supply, sink limitation and photoperiod, as well as a potential influence of soil moisture limitation on EOS globally (Buermann et al. 2018; Liu et al. 2016; Lian et al. 2021) may also play a role in EOS changes.

Recently, it has been suggested that growing season photosynthesis may have a significant negative impact on EOS in some plants. That is to say, if the plants had a very productive growing season, they are more likely to senesce their leaves early (Zani et al. 2020). *ibid.* used observations from the European Phenology Network and controlled experiments to test the effects of growing season photosynthesis on the timing of leaf senescence and found that more productive growing seasons led to earlier leaf senescence in central European forests. They hypothesized a strong role of sink limitation in these species. However, they modelled the gross primary productivity (GPP) following the Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS) model, which may introduce some uncertainties due to the model parameterization, and the study was limited to specific species in central Europe. Also, their results contradict the findings of most free-air CO₂ enrichment (FACE) experiments, in which the leaf senescence dates are typically either delayed or remained unchanged (Norby 2021). An earlier senescence under higher GPP would have large implications for expected EOS changes under climate change, but further testing with more diverse and distributed phenology and photosynthesis observations is needed.

Typically, phenological metrics of different species are recorded by visual observations, or digital cameras at specific sites (Richardson et al. 2018). The most successful efforts include the European phenology network and the USA National Phenology Network that provide more than 60 years of observations (Vliet et al. 2003). In recent years, the seasonal cycles of remotely sensed observations such as vegetation indices (VIs), sun-induced fluorescence and vegetation optical depth, and flux tower observations have also been widely used to estimate key phenological metrics at landscape and larger scales (D’Odorico et al. 2015; Gonsamo et al. 2012; Joiner et al. 2014; Wu et al. 2017; Yang and Noormets 2021; Zhang et al. 2019). These metrics provide insight into the changes of seasonal cycles of greenness and/or carbon fluxes in a diverse array of ecosystems, with significant potentials for understanding phenology-climate interactions (Tang et al. 2016). These global and regional observations also provide a unique opportunity to test phenological hypotheses in different ecosystems, particularly when co-located eddy-covariance (Novick et al. 2018; Pastorello et al. 2020), ground and remote sensing observations can be combined.

In this chapter, I used flux observations from 40 eddy-covariance sites that have measurement records of more than 10 years to test the relationship between growing season photosynthesis and EOS across a variety of ecosystems. These sites represent a large subsection of

main biomes in North America and Europe (8 biome types according to the International Geosphere-Biosphere Program (IGBP)) from 1992 to 2017 and have a total of 589 site-years of observations. In addition, I used long-term ground phenological records co-located with eddy-covariance measurements at a site in the northeastern US (Harvard Forest), along with phenological metrics from the MODIS global land cover dynamics product at all sites studied, and three leaf senescence models. Doing so, I aim to address the following questions:

1. Are there significant carry-over effects of growing-season photosynthesis on EOS in most sites and ecosystems?
2. What are the most important factors that contribute to EOS changes?
3. Does the introduction of growing-season photosynthesis improve model predictions of EOS?

2.2 Methods and Materials

Flux tower measurements

I used 40 sites from the FLUXNET 2015 dataset (<http://fluxnet.fluxdata.org/>, (Baldocchi 2003; Pastorello et al. 2020)) and AmeriFlux (<https://ameriflux.lbl.gov/>, (Novick et al. 2019; Keenan and Willam 2018)) in this chapter (Table A1). I focused on ecosystems that had a strong seasonal cycle of photosynthesis (e.g., no tropical sites were selected). Also, I removed sites that had a double growing season (i.e., double peaks in the growing season GPP most likely due to site management and/or water stress), as indicated by the GPP time series (Figure A2.1). For sites in both the FLUXNET 2015 and AmeriFlux databases, I chose the source with a longer record of observations.

For 12 out of 15 sites within AmeriFlux that do not provide gap-filled and partitioned data, I used the R package REddyProc to gap-fill and partition net ecosystem change (NEE) into GPP and ecosystem respiration (R_{Eco}) using the nighttime partitioning model (Reichstein et al. 2005, Wutzler et al. 2018). Similar results were found when using the daytime partitioning method (Figure A2.2, Lasslop et al. 2010).

To determine growing season photosynthesis, I summed up the hourly or half-hourly GPP within the period. I determined the start of the growing season as outlined in section 2.4 and used the first day that has daylength of fewer than 11.2 hours as the end date for the calculation of end of growing season photosynthesis (in order to avoid introducing spurious correlations between EOS and growing season photosynthesis, following Zani et al. 2020). Also, following the method used in Fu et al. 2017, I calculated the environmental conditions (air temperature (T_{air}), vapor pressure deficit (VPD) and global radiation (R_g)) in different

seasons in the Northern Hemisphere (Summer: June, July and August; Fall: September, October and November).

Phenological records from Harvard Forest

I used observations from the Phenology of Woody Species at Harvard Forest dataset from 1991 to 2017, obtained from the Harvard Forest Data Archive (hf003-08, O’Keefe 2000; Richardson et al. 2006). Phenological observations (spring and fall) have been recorded for more than 33 woody species around the Harvard Forest starting from 1991 (for spring phenology) and 1992 (for autumn phenology). Leaf phenology dates were recorded by looking at the percentage of the leaves that have emerged (spring) and dropped (fall). I used the dates within the dataset when 50% of leaves have emerged or dropped each year. The mean fall dates at species-level provided with the data set were used and I focused on the two dominant species within the eddy-covariance tower footprint, i.e., Red Oak and Red Maple, in the forest Finzi et al. 2020.

MODIS global land dynamics product

I used the MODIS land cover dynamics product (MCD12Q2, (Zhang et al. 2003)) downloaded from Oak Ridge National Laboratory’s Distributed Active Archive Center (with a spatial resolution of 500 m, combined from Terra and Aqua, collection 6). This dataset provides a global estimate of several phenological metrics since 2001. Technically, it identifies phenological metrics derived from the seasonal time series of MODIS-observed enhanced vegetation index (EVI). The original EVI time series was gap-filled and smoothed, and then fit to a logistic model. Detailed information of the process can be found in *ibid.*, Zhang et al. 2006 and Ganguly et al. 2010. I used the date when EVI last crossed 15% (the MCD12Q2 Dormancy date) of the segment EVI amplitude, EOS, as the leaf senescence metric. I only used the pixel on which the flux tower centered from the MCD12Q2 data set.

Determination of phenological metrics from flux measurements

To determine the phenological metrics (i.e., SOS and EOS) from flux-estimated GPP, I applied a widely used double-logistic curve fitting method (Gonsamo et al. 2012). The method has proven to work well in most sites and ecosystems (D’Odorico et al. 2015; Lu et al. 2018). To retrieve the key phenological dates, I used the half-hourly GPP estimates from the FLUXNET or AmeriFlux data sets. First, I summed up the gap-filled half-hourly GPP at the daily scale; then, I filtered out years when the data coverage was not sufficient (fewer than 100 days of GPP estimates); finally, I fit the daily GPP time series to the following double-logistic model:

$$GPP_t = a_1 + \frac{a_2}{1 + \exp(-d_1(t - b_1))} + \frac{a_3}{1 + \exp(-d_2(t - b_2))} \quad (2.1)$$

The seven free parameters (a_1 , a_2 , a_3 , b_1 , b_2 , d_1 and d_2) were determined using non-linear curve fitting. $a_2 - a_1$ and $a_3 - a_1$ represent the difference between the winter background value and the amplitude of the summer peak growing season values. d_1 and d_2 are the transition curvature parameters, while b_1 and b_2 are the midpoints in DOYs of these transitions for green-up and senescence/abscission, respectively.

This method identifies the SOS as the start of the slope of the ascending curve and EOS is identified as the end of the descending curve (the inflection point). I followed the method used in Gonsamo et al. 2012 to identify the thresholds from the logistic model (1), which estimates the SOS and EOS as:

$$SOS = b_1 - \frac{4.5622}{d_1} \quad (2.2)$$

$$EOS = b_2 + \frac{4.5622}{d_2} \quad (2.3)$$

I compared the EOS estimates from both flux tower estimated GPP and remotely sensed EVI (Figure A2.3). The EOS estimated by the two methods were comparable, but mismatches can be evident in some ecosystems. Overall, I found that the flux-inferred and remotely sensed EOS metrics were significantly correlated ($p < 0.01$), and especially for deciduous broadleaf forest (DBF) and mixed forest (MF) sites, though GPP EOS had a higher dynamic range in all ecosystems. For ecosystems where the seasonality of canopy greenness is more difficult to detect (e.g., grassland, GRA), the two EOS estimates varied substantially.

In order to compare the relationship between growing season photosynthesis, leaf senescence metrics and environmental conditions, I linearly detrended all time series in order to reduce the likelihood of extraneous correlations.

Modelling of leaf senescence dates

First, to analyze the relationship between environmental factors and the timing of senescence, I developed structural equation models using the Structural Equation Models Optimization in Python (semopy) package (Igolkina and Meshcheryakov 2020). This package provides a concise way to test the structural relationships between different variables and to build structural equation models. I used structural equation models to compare three different ways of empirically predicting the EOS using environmental conditions. In model a, I used all environmental variables (summer and autumn T_{air} , summer and autumn R_g , summer and autumn VPD) to predict the EOS directly; in model b, I used summer environmental variables (summer T_{air} , summer R_g and summer VPD) to predict the growing season photosynthesis and used predicted growing season photosynthesis with autumn environmental variables (autumn T_{air} , autumn R_g and autumn VPD) to predict the EOS; in model c, I used growing season photosynthesis from flux observations and the autumn environmental

variables (autumn T_{air} , autumn R_g and autumn VPD) to predict the EOS.

In addition, I compared three different process-oriented leaf senescence models by examining their performance for predicting autumn senescence dates at the Harvard forest site. The first model I assessed is the cold-degree day model that combines the use of both air temperature and photoperiod (Delpierre et al., 2009). When thresholds of air temperature (T_b) and photoperiod (P_{start}) are reached, the control of air temperature and photoperiod over autumn leaf senescence is modelled via the cumulative cold-degree days as follows:

$$CDD(d) = T(b) - T(d) \times \frac{P_d}{P_{\text{start}}} \quad \text{if } T(b) > T(d) \quad (2.4)$$

$$CDD(d) = 0 \quad \text{if } T(b) \leq T(d) \quad (2.5)$$

Where the CDD(d) is the cold-degree day at day(d) with the air temperature of $T(d)$ and photoperiod $P(d)$. Then the CDD(d) of each day is accumulated. The modeled leaf senescence date (Y_{mod}) was set as the first day that the accumulated CDD (aCDD) reached a critical threshold as follows:

$$Y_{\text{mod}} = d \quad \text{if } a \times CDD > Y_{\text{crit}} \quad (2.6)$$

Here, Y_{crit} is a threshold to be determined. In this manuscript, I refer to this model as the CDD model.

As summer dry conditions may also impact the autumn leaf senescence, a growing-season mean temperature (TPM) model was proposed (Liu et al., 2019). In the TPM model, the summer growing season temperature was introduced and used, the threshold of Y_{crit} is set as:

$$Y_{\text{crit}} = a + b \times Temp_{\text{summer}} \quad (2.7)$$

Here, $Temp_{\text{summer}}$ is the average summer temperature of the associated years. It has also been reported that spring phenology is associated with autumn phenology, and a revised version of the CDD model has been proposed to accommodate these effects (Keenan and Richardson, 2015). In this spring phenology influenced autumn phenology model (SIAM), the threshold of Y_{crit} is set as:

$$Y_{\text{crit}} = a + b \times S_a \quad (2.8)$$

Here, S_a is the spring phenology anomaly of the associated years. Also, as Zani et al. 2020 suggests, growing season photosynthesis might have an impact on the end of the growing season. I therefore also modified the model as follows to represent this hypothesis:

$$Y_{\text{crit}} = a + b \times GPP \quad (2.9)$$

Here, the GPP represents the growing season canopy photosynthesis. In this manuscript, I referred to this model as the photosynthesis-influenced autumn phenology (PIA) model. To estimate the free parameters in the three different models, I used the Pymcmcstat Markov-chain Monte Carlo python package (Miles 2019). I estimated the parameters used in the three models 100 times, and sampled parameters resulting in the 95th percentile of model performance.

To examine differences in projections of future autumn phenology change from the 3 different autumn phenology models, I estimated the delays of autumn senescence date across 5 °C of projected future temperature changes. For the SIAM model, I predicted the sensitivities of spring phenology dates to air temperature to be 3 d⁻¹ °C⁻¹ according to the model ensemble assessed in Migliavacca et al. 2012. For the PIA model, I predicted a 6 gC yr⁻¹ m⁻² °C⁻¹ increase of growing season photosynthesis with the increase in air temperature as this is the overall slope at Harvard Forest according to our results.

2.3 Results

No evidence for the carry-over effects of growing season photosynthesis on EOS

At the site level, I focused on the observations from Harvard Forest, which provide a long-term record of eddy-covariance measurements and phenological metrics of Red Oak (the most dominant species in the forest, Figure A2.4). I found positive but non-significant relationships between growing season photosynthesis and EOS dates either when using ground phenological records or MODIS estimates ($p > 0.05$).

I then examined the relationship between growing season photosynthesis and flux-estimated EOS across different ecosystems in the eddy-covariance flux network (Figure 2.1). I found a significant and positive relationship between growing season photosynthesis and EOS in evergreen needleleaf forest sites (ENF) (slope = 0.24, $p = 0.02$). In other ecosystems, however, I did not find significant relationships between them. In DBF, GRA, and MF sites, I found a positive yet non-significant growing season GPP – EOS relationship. When all sites were pooled together, I found a significant and positive, though weak, relationship between growing season photosynthesis and flux-estimated EOS (slope = 0.1, $p = 0.01$).

I also tested the relationships between growing season photosynthesis and MODIS EOS in different sites and ecosystems. Except for MF sites, where I found a negative yet non-significant relationship between growing season photosynthesis and remotely sensed EOS (slope close to 0, $p = 0.97$), the growing season photosynthesis-EOS relationship tends to be positive yet non-significant. Overall, I found no significant relationship between growing season photosynthesis and remotely sensed EOS (slope = 0.03, $p = 0.27$). I also tested the

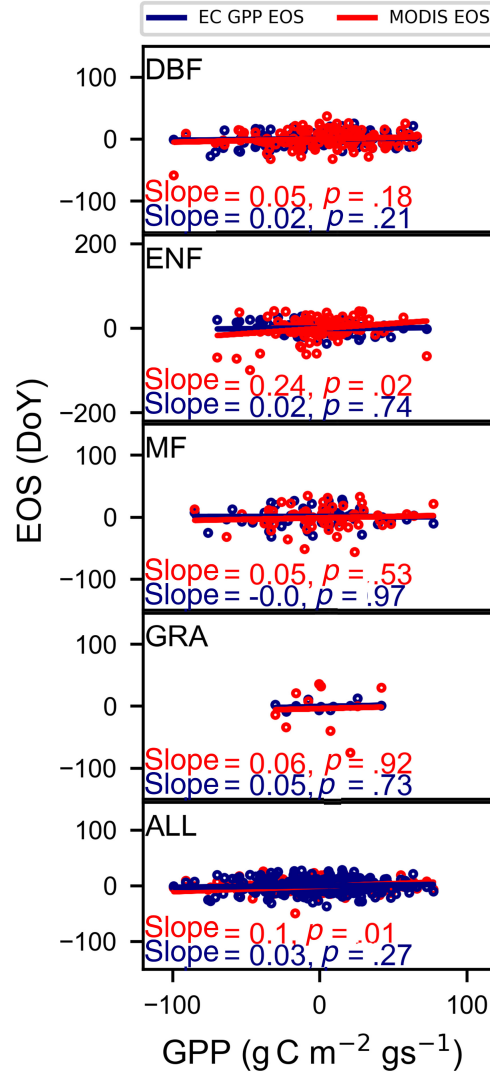


Figure 2.1: The relationship between growing season gross primary productivity (GPP) and phenological end of season (EOS) metrics from eddy-covariance (EC) observations or MODIS observations at deciduous broadleaf forest (DBF), evergreen needleleaf forest (ENF), mixed forest (MF), grassland (GRA), and all sites (ALL). All the time series have been detrended and the differences introduced by different sites were set as random effects.

relationships between growing season photosynthesis and MODIS EOS in different sites and ecosystems. Except for MF sites, where we found a negative yet non-significant relationship between growing season photosynthesis and remotely sensed EOS (slope close to 0, $p = .97$), the growing season photosynthesis–EOS relationship tends to be positive yet non-significant. Overall, we found no significant relationship between growing season photosynthesis and remotely sensed EOS (slope = 0.03, $p = .27$).

Control of environmental factors over EOS in different ecosystems

I examined the multi-variable relationships between environmental factors and EOS metrics in different ecosystems. First, I performed a partial correlation analysis between different environmental variables and EOS estimates at Harvard Forest (Figure 2.2). For ground observations of Red Oak, I found a positive partial correlation between growing season photosynthesis and leaf senescence metrics, while I found weaker positive relationships with MODIS remotely sensed metrics. I also found contrasting effects of global radiation (R_g), air temperature and VPD in the spring and in the autumn. Both data sets indicated that higher summer R_g , lower summer temperature and lower summer VPD were associated with a later EOS; while more autumn R_g , higher autumn air temperature and higher autumn VPD led to an earlier EOS.

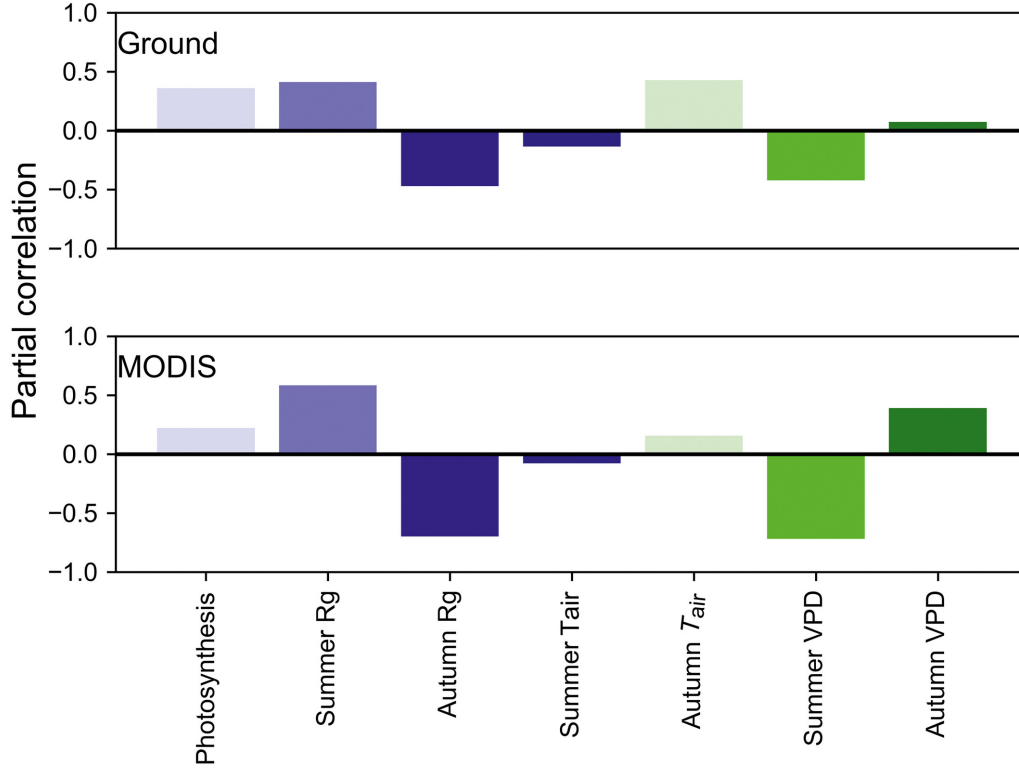


Figure 2.2: Partial correlations between environmental variables and end of growing season (EOS) in Harvard forest using ground phenological records (a) and the MCD12Q2 product (b). The environmental variables included here were growing season photosynthesis (Photosynthesis), summer global radiation (R_g), autumn R_g , summer air temperature (T_{air}), autumn T_{air} , summer vapor pressure deficit (VPD), and autumn VPD. All the time series have been detrended. Here, summer is defined as June, July and August while fall is defined as September, October and November.

At the cross-site scale, I found more complicated relationships between flux-estimated EOS and different environmental variables (Figure 2.3). First, I did not find growing season photosynthesis to be the most important factor when it came to predicting EOS in any of the ecosystems. In all ecosystems, I found a positive contribution (partial correlation) between growing season photosynthesis and EOS ($r = 0.12, 0.11, 0.03$ and 0.32 for DBF, ENF, MF and GRA sites respectively). Overall, I found a positive partial correlation between growing season photosynthesis and EOS ($r = 0.12$). The controls of environmental factors were different in different ecosystems; and the same environmental factors exhibited different impacts in different seasons. For global radiation, for example, the autumn R_g contributed most positively to the variation of EOS except for GRA; while the summer R_g effects were mixed in different ecosystems (marginal effects found in ENF, while positive effects were found in DBF sites and negative effects in GRA and MF sites). For air temperature, interestingly, I found that summer T_{air} contributed mostly negatively to EOS (although sometimes marginally) except for GRA sites; while the autumn T_{air} contributions were mixed in different ecosystems I tested. Finally, for VPD, I also found that summer VPD effects were mostly positive except for GRA sites, while the autumn VPD effects were all negative. Overall, when all sites were used, I found a marginal contribution of summer R_g ; positive contributions of autumn R_g , autumn T_{air} and summer VPD; negative contributions of summer T_{air} and autumn VPD.

The introduction of growing season photosynthesis did not improve EOS predictions

Using a structural equation model analysis, I found that the introduction of growing season photosynthesis did not improve the overall model performance when using all DBF sites (Figure 2.4). I tested three different structural equation models. In the first model, I used a simple multiple linear regression model to predict EOS using only meteorological variables ($R = 0.43$). In the second model, I used summer environmental factors to predict growing season photosynthesis, and then used the growing season photosynthesis and autumn environmental factors to predict EOS ($R = 0.34$). Finally, I used flux-estimated photosynthesis along with autumn environmental factors to predict EOS ($R = 0.39$). I found positive coefficients of photosynthesis in the latter two models ($\beta = 0.06$ for model b and $\beta = 0.06$ for model c).

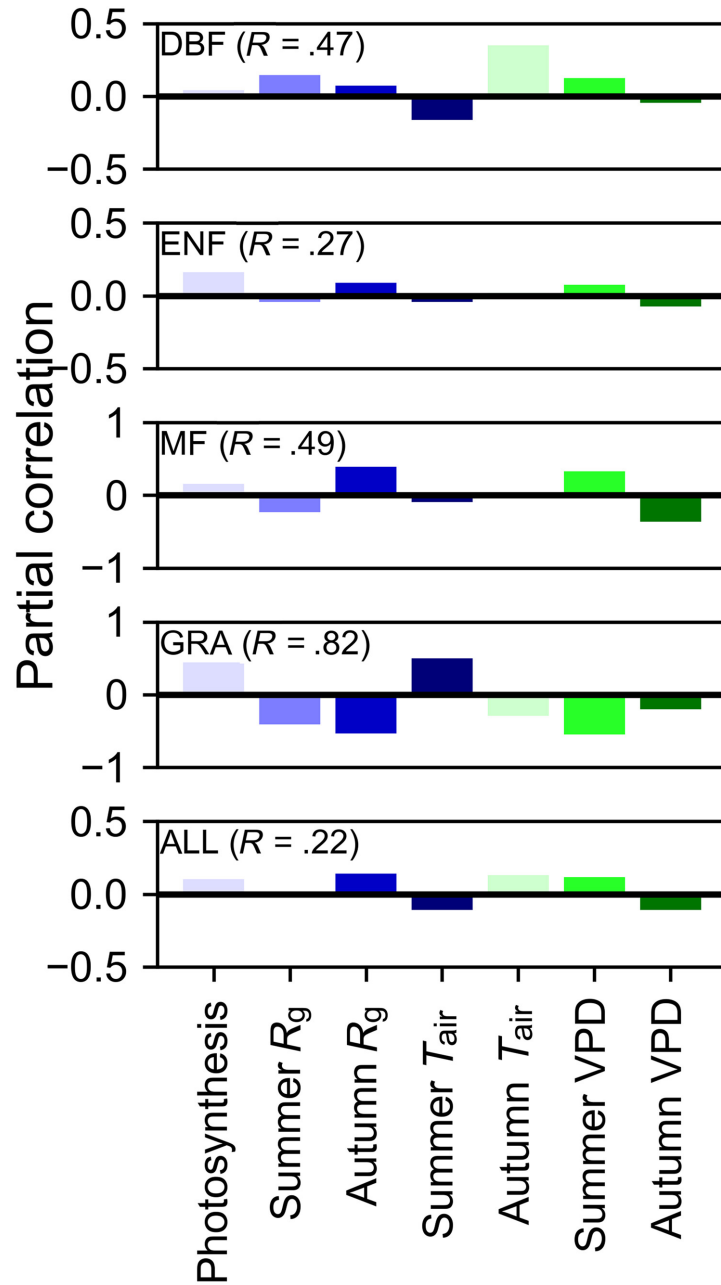


Figure 2.3: Partial correlations between detrended environmental variables and flux-derived EOS in different ecosystems. The environmental variables tested include growing season photosynthesis, summer global radiation (R_g), autumn R_g , summer air temperature (T_{air}), autumn T_{air} , summer vapor pressure deficit (VPD), and autumn VPD. All the time series have been detrended. The overall linear regression model performance (correlation) using all predictors was labeled.

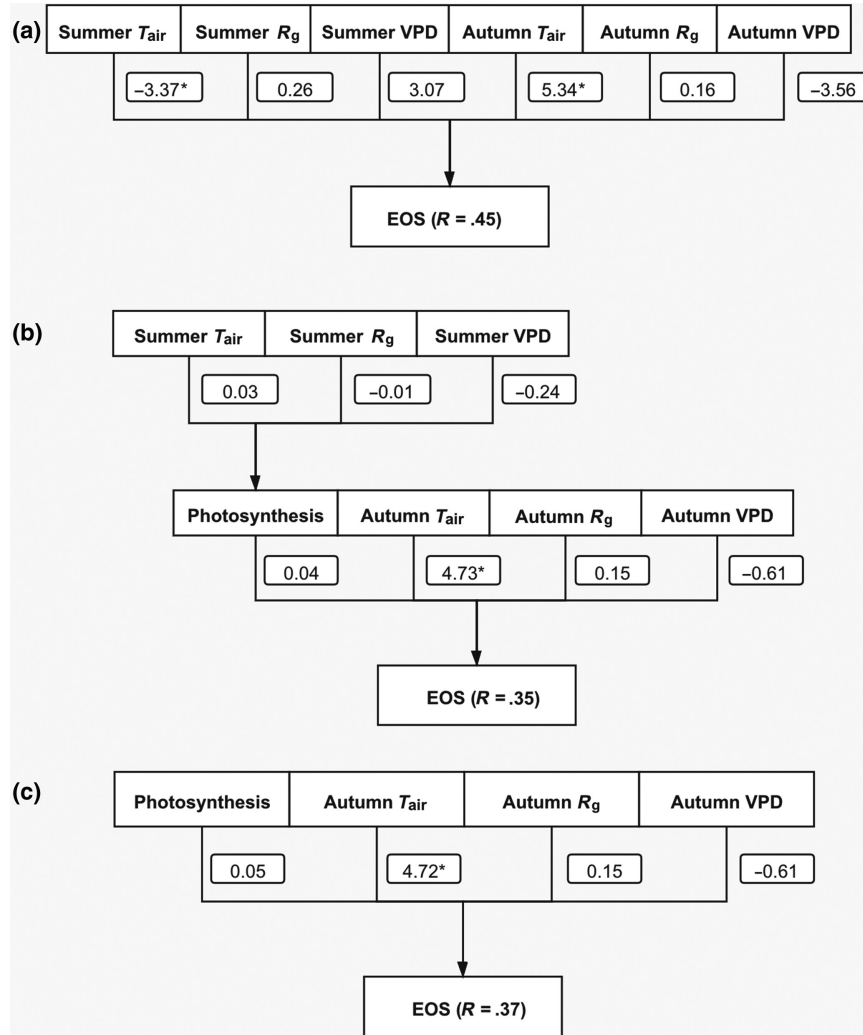


Figure 2.4: Analysis of the flux-inferred end of growing season in deciduous broadleaf forest (DBF) sites with three different structural equation models (see the method section). The β coefficients determined for each predictor in the structural equation models are denoted, with significant coefficients ($p < 0.05$) denoted with a *. The end of growing season (EOS) and predictors at each site have been detrended. (a) predicts EOS using the direct influence of environmental variables; (b) predicts EOS using the direct influence of environmental variables along with the indirect influence through predicted photosynthesis, and (c) predicts EOS using the direct influence of both environmental variables and eddy-covariance inferred estimates of growing season photosynthesis.

At Harvard Forest, I also tested the performance of three different end of growing season models to predict the leaf senescence dates (Table 2.1). In the resulting comparison it is important to note that the number of free parameters is different in different models: the CDD model requires three free parameters while the SIAM and PIA models require four free parameters. The results showed that the performance of the three models was generally comparable. Overall, however, I found that the PIM model, which introduced the growing season photosynthesis predictor, underperformed the CDD and the SIAM model. For Red Oak, the most dominant species in the forest, all three models did a relatively worse job when predicting the leaf senescence date, though within a comparable range to previous studies for other species at the forest.

Table 2.1: The model performance in predicting leaf senescence date at Harvard Forest. Here, the CDD, SIAM and PIA model represent the cold degree days model, spring phenology impacted autumn phenology model and the photosynthesis impacted phenology model, respectively; and the RMSE denotes the root mean square error of each model.

Model	Red Oak		Red Maple	
	Correlation	RMSE	Correlation	RMSE
CDD	0.48	4.87	0.65	2.63
SIAM	0.48	4.87	0.7	2.48
PIA	0.44	4.93	0.63	2.67
TPM	0.50	4.85	0.71	2.47

While the three models (CDD, SIAM and PIA) showed similar performance for estimating the autumn senescence date, their predictions of future autumn senescence delays diverged (Figure 2.5). Using the parameterized models, I found a similar response of autumn senescence delays in both the CDD and PIA models. This was mainly due to the fact that the coefficient relating growing season photosynthesis and the threshold of autumn senescence was estimated to be small in the PIA model. In both species, I also found that the SIAM model tended to predict a smaller delay in autumn senescence dates, especially for Red Maple where the SIAM model performed the best. The divergent predictions by different models despite similar performance for long-term observations of natural variability points to the need for future studies, especially controlled experiments, to fully tease apart the controls of autumn senescence to rising temperature.

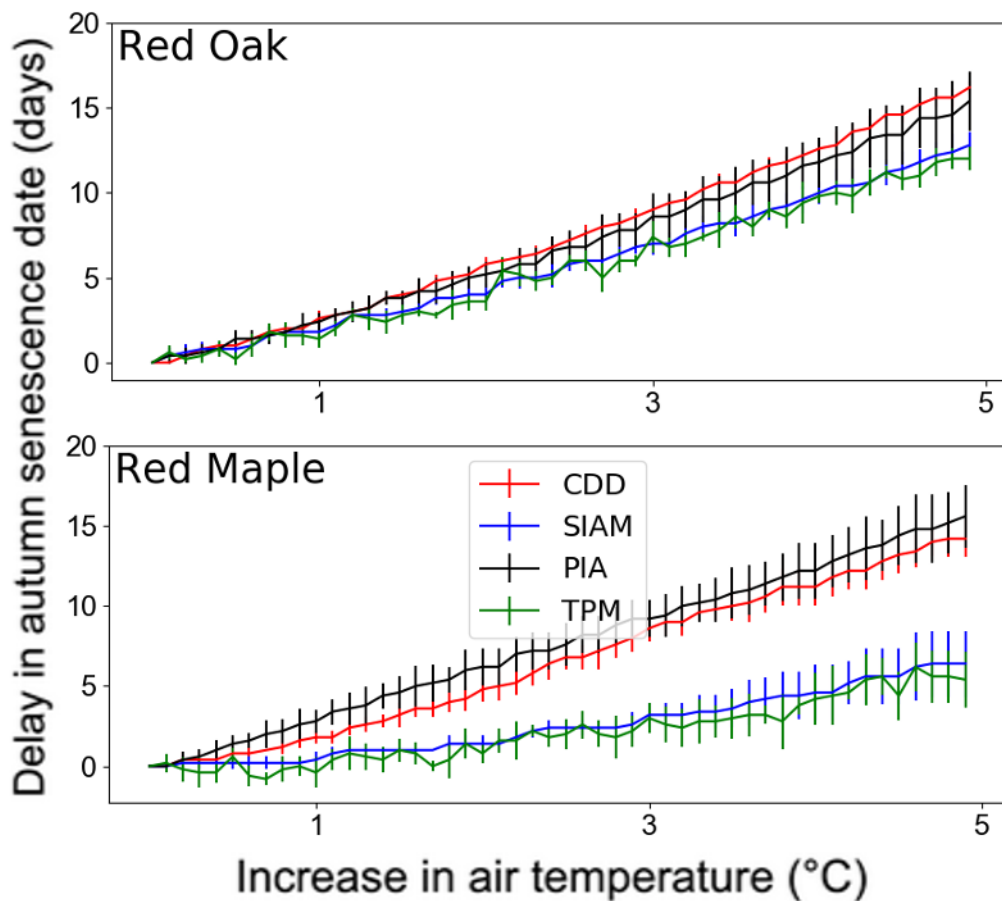


Figure 2.5: Projections of future delays in autumn senescence dates with increasing air temperature, as estimated by three different models. The error bars indicate the standard deviations of the predictions within the 95th percentile of model performance from the MCMC process.

2.4 Discussion

The carry-over effects of “growing season photosynthesis” on EOS

In the present study, our primary objective was to use flux tower observations from 40 sites across the temperate and boreal ecosystems in the Northern Hemisphere to test the carry-over effect hypothesis, which proposed significant negative effects of growing season photosynthesis on the date of leaf senescence and the end of the growing season (Zani et al. 2020).

I did not find significant carry-over effects in the ground and remote sensing observations for Harvard Forest, nor in most of the eddy-covariance ecosystems examined. For instance, I found positive but insignificant relationships in all three ground-observed phenological records at Harvard Forest. On the contrary, I found that in most sites, a more productive growing season either led to a delayed EOS or contributed little to the changes of EOS.

In addition, I also tested the possible improvement of model performance by introducing the use of growing season photosynthesis. I found that the growing season photosynthesis influenced models did not outperform the previous empirical models (structural equation model) or process-based models (CDD and SIAM models). And with future warming, our results show that model choice can lead to considerable differences in the future phenological shifts. This is true even in the absence of considering a CO₂ effect on photosynthesis, which would likely lead to more divergent predictions between the CDD and PIA models. That said, our results indicate small, and in most cases positive, coefficients between growing season GPP and Y_{crit}, which suggests that incorporating a CO₂ effect on photosynthesis is unlikely to greatly influence PIA model projections for our studied sites.

Our findings contrast with recent results from Zani et al. 2020, which report a strong influence of growing season photosynthesis on autumn senescence. The differences may result from several factors, including differences in the methods used, ecosystems focused on, and geographical and environmental patterns. First, the methods used in the two studies have notable differences. In *ibid.*, the results were based on a model of growing season photosynthesis; in the present study, however, I used flux tower estimated GPP based on direct eddy covariance observations. Also, *ibid.* used phenology records from central Europe, while I use phenology estimates across a broad geographical range of boreal and temperate North America and Europe, made from multiple approaches (the flux-estimated, remotely sensed and ground observations at Harvard Forest).

In addition, the ecosystems of interest are different. In the present study, I focused on 8 different types of biomes across North America and Europe, while the *ibid.* study mostly focused on central Europe deciduous forests. The carry-over effects, especially the effects as predicted from the sink limitation hypothesis proposed as an explanation of their observations by *ibid.*, are most likely to happen in ecosystems in which nutrients, water and/or sink capacity are more likely to be limiting factors (Fatichi et al. 2014). I did find, however, that growing season photosynthesis contributed marginally negatively to remotely sensed EOS in evergreen needleleaf forests when using the nighttime partitioning method (most of the ENF sites are in the boreal ecosystems where there might be sink limitations, see Figure A2.2).

Finally, the geographical and environmental patterns for the sites tested in the two studies are also different. For instance, the soil nutrient availability of the Central Europe forests may be very different that of the other sites I tested (Ackerman et al. 2019).

The environmental controls over EOS in different ecosystems

I examined how the environmental variables in the autumn and in the previous seasons contributed to the changes of EOS in different ecosystems. Air temperature in summer and in autumn were one of the most important factors that determined the EOS in different ecosystems (Figure 2.3). This indicated that air temperature was contributing largely to the EOS in different ecosystems, similar to results found in Zhang et al. 2020. Also, I found that autumn VPD contributed negatively to the EOS timing, while the effects of summer VPD were mostly positive in different ecosystems. These results indicated that the water limitation possibly contributed to an earlier EOS.

At the same time, I found that the environmental controls over EOS could be different in different sites. For instance, in most ecosystems, the results indicated that the summer air temperature was negatively correlated with EOS in most ecosystems while I found a positive correlation in GRA sites. One possible explanation can be that the two GRA sites used in the present study (CA-Let, annual mean air temperature at 5.4 °C and IT-Mbo, annual mean air temperature at 5.1 °C) are both in cold regions and the warmer air temperature was usually associate with productive and lengthier growing season. To sum up, the environmental controls of EOS can be complex and different at different sites and in different ecosystems. Overall, I failed to find a negative carry-over effect of growing season photosynthesis on EOS.

Due to the complex controls on EOS in different ecosystems, future studies based on controlled experiments would be exceptionally beneficial. For instance, Zani et al. 2020 tested the hypothesis of sink limitation with three groups with different treatments and found that growing season photosynthesis had a negative relationship with the leaf senescence date. However, Norby et al., (2021) pointed out that in most FACE experiments, some of which also have a warming component, the leaf senescence dates in most experiments were either delayed or unchanged. Networks of phenological measurements such as PhenoCam (Richardson et al. 2018) combined with flux measurement and or other experiments could also be of significance, especially for testing the sink limitation hypothesis.

Matches and mismatches between flux estimated and remotely sensed phenological metrics

Remote sensing of terrestrial ecosystem phenology is challenging, especially in ecosystems with low seasonality (e.g., ENF) or with significant snow presence (e.g., arctic tundra ecosystems). For instance, White et al., 2009 found that even when using the same input data sets, the derived remotely sensed phenological metrics are largely method dependent in Northern America. Also, Wu et al. 2017 found that the NDVI-estimated phenological metrics matched poorly with flux-tower estimated photosynthesis phenology metrics. Recently, Lu et al. 2018

found that satellite-observed SIF outperformed conventional VIs when estimating phenological metrics even with much coarser spatial resolutions at relatively homogeneous sites (50 km). In the present study, I used the MODIS land dynamics product (MCD12Q2) as it is one of the few publicly available datasets that have phenological metrics from satellite observations globally, and previous studies have shown that it effectively captures phenological dynamics (D’Odorico et al., 2015).

It is also worth noting that methods such as the double logistic curve fitting method to determine EOS used in this chapter assume that there is an inflection point within the time series of GPP. In the autumn, however, the senescence of photosynthesis can be more subtle, and the inflection points can be more difficult to detect. For instance, I found that the MODIS phenological metrics performed better in DBF sites where the inflection points were more significant. Overall, when comparing all sites, I found a significant relationship between flux-estimated and remotely sensed phenological metrics (Figure. A3).

I also compared the predictability of the two EOS metrics by using a linear regression model that used all predictors (Table 2.2). The results indicated that overall, in DBF, MF and GRA sites, the flux tower inferred EOS are more predictable than the MODIS phenology products; while in ENF, the remotely sensed metrics from MODIS are easier to predict. This may result from the subtle and slow changes of GPP in the fall in ENF sites. In summary, I argue that flux estimated and remotely sensed phenological metrics both contain meaningful information despite the mismatches in some cases.

Table 2.2: Comparison of predictability (R) of flux tower inferred and remotely sensed end of growing season EOS in deciduous broadleaf forest (DBF), evergreen needleleaf forest (ENF), mixed forest (MF), grassland (GRA) and ALL sites. Here, I used the International Geosphere-Biosphere Program (IGBP) scheme to identify the biome types of different sites, and the values in the table indicate the correlation between a multiple linear model prediction and observations in both models.

IGBP	DBF	ENF	MF	GRA
Tower EOS	0.47	0.27	0.49	0.82
MCD12Q2	0.49	0.33	0.25	0.81

2.5 Conclusion

I tested the relationships between growing season photosynthesis and estimates of leaf senescence and EOS in different ecosystems. I did not find negative carry-over effects of growing season photosynthesis on leaf senescence and EOS in most sites and most ecosystems, in contrast to recent reports. I also found that the controls of EOS in different ecosystems

were different and, in most cases, a more productive growing season was related to a later, not earlier, leaf senescence. Our results challenge the notion that EOS is negatively affected by growing season photosynthesis, and highlight the need for controlled experiments to distinguish competing controls on the timing of fall senescence.

2.6 Acknowledgement

XL and TFK acknowledge support from a NASA Terrestrial Ecology Program IDS Award NNH17AE86I. XL was supported in part by funding provided to the AmeriFlux Management Project by the U.S. Department of Energy's Office of Science under Contract No. DE-AC02-05CH11231. XL was also supported by the grant of Future Investigators in NASA Earth and Space Science and Technology (No. 80NSSC21K1602). TFK was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 as part of their Regional and Global Climate Modeling program through the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation Scientific Focus Area (RUBISCO SFA) project. Funding for AmeriFlux data resources was provided by the U.S. Department of Energy's Office of Science, and can be downloaded from <https://ameriflux.lbl.gov/data/>. The study used the FLUXNET 2015 data set which can be accessed at <https://fluxnet.org/data/fluxnet2015-dataset/>.

References

- Ackerman, D., D. B. Millet, and X. Chen (2019). “Global estimates of inorganic nitrogen deposition across four decades”. In: *Global Biogeochemical Cycles* 33.1, pp. 100–107. ISSN: 0886-6236.
- Archetti, M. et al. (2013). “Predicting climate change impacts on the amount and duration of autumn colors in a New England forest”. In: *PLoS One* 8.3, e57373. ISSN: 1932-6203.
- Baldocchi, D. D. et al. (2005). “Predicting the onset of net carbon uptake by deciduous forests with soil temperature and climate data: a synthesis of FLUXNET data”. In: *International Journal of Biometeorology* 49, pp. 377–387. ISSN: 0020-7128.
- Baldocchi, D. D. (2003). “Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future”. In: *Global change biology* 9.4, pp. 479–492.
- Buermann, W. et al. (2018). “Widespread seasonal compensation effects of spring warming on northern plant productivity”. In: *Nature* 562.7725, pp. 110–114. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0555-7.
- Chmielewski, F.-M. and T. Rötzer (2002). “Annual and spatial variability of the beginning of growing season in Europe in relation to air temperature changes”. In: *Climate research* 19.3, pp. 257–264. ISSN: 0936-577X.
- Chuine, I. and J. Régnière (2017). “Process-based models of phenology for plants and animals”. In: *Annual review of ecology, evolution, and systematics* 48, pp. 159–182. ISSN: 1543-592X.
- Churkina, G. et al. (2005). “Spatial analysis of growing season length control over net ecosystem exchange”. In: *Global Change Biology* 11.10, pp. 1777–1787. ISSN: 1354-1013.
- D’Odorico, P. et al. (2015). “The match and mismatch between photosynthesis and land surface phenology of deciduous forests”. In: *Agricultural and Forest Meteorology* 214-215, pp. 25–38. ISSN: 01681923. DOI: 10.1016/j.agrformet.2015.07.005.
- Faticchi, S., S. Leuzinger, and C. Körner (2014). “Moving beyond photosynthesis: from carbon source to sink-driven vegetation modeling”. In: *New Phytologist* 201.4, pp. 1086–1095. ISSN: 0028-646X.
- Finzi, A. C. et al. (2020). “Carbon budget of the Harvard Forest Long-Term Ecological Research site: Pattern, process, and response to global change”. In: *Ecological Monographs* 90.4, e01423. ISSN: 0012-9615.
- Friedlingstein, P. et al. (2020). “Global carbon budget 2020”. In: *Earth System Science Data Discussions* 2020, pp. 1–3. ISSN: 1866-3591.
- Fu, Z. et al. (2017). “Climate controls over the net carbon uptake period and amplitude of net ecosystem production in temperate and boreal ecosystems”. In: *Agricultural and Forest Meteorology* 243, pp. 9–18. ISSN: 01681923. DOI: 10.1016/j.agrformet.2017.05.009.
- Ganguly, S. et al. (2010). “Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product”. In: *Remote Sensing of Environment* 114.8, pp. 1805–1816. ISSN: 0034-4257.

- Gonsamo, A. et al. (2012). “Land surface phenology from optical satellite measurement and CO₂ eddy covariance technique”. In: *Journal of Geophysical Research: Biogeosciences* 117.G3. ISSN: 01480227. DOI: 10.1029/2012jg002070.
- Igolkina, A. A. and G. Meshcheryakov (2020). “semopy: A Python Package for Structural Equation Modeling”. In: *Structural Equation Modeling: A Multidisciplinary Journal* 27.6, pp. 952–963. ISSN: 1070-5511 1532-8007. DOI: 10.1080/10705511.2019.1704289.
- Jeong, S.-J. et al. (2011). “Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008”. In: *Global change biology* 17.7, pp. 2385–2399. ISSN: 1354-1013.
- Joiner, J. et al. (2014). “The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange”. In: *Remote Sensing of Environment* 152, pp. 375–391. ISSN: 00344257. DOI: 10.1016/j.rse.2014.06.022.
- Keenan, T. and C. Willam (2018). “The terrestrial carbon sink”. In: *Annual Review of Environment and Resources* 43, pp. 219–243. ISSN: 1543-5938.
- Keenan, T. F. et al. (2014). “Net carbon uptake has increased through warming-induced changes in temperate forest phenology”. In: *Nature Climate Change* 4.7, p. 598. ISSN: 1758-6798.
- Körner, C. and D. Basler (2010). “Phenology under global warming”. In: *Science* 327.5972, pp. 1461–1462. ISSN: 0036-8075.
- Lang, W. et al. (2019). “A new process-based model for predicting autumn phenology: how is leaf senescence controlled by photoperiod and temperature coupling?” In: *Agricultural and Forest Meteorology* 268, pp. 124–135. ISSN: 0168-1923.
- Lasslop, G. et al. (2010). “Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation”. In: *Global Change Biology* 16.1, pp. 187–208. ISSN: 1354-1013.
- Lian, X. et al. (2021). “Seasonal biological carryover dominates northern vegetation growth”. In: *Nature communications* 12.1, pp. 1–10. ISSN: 2041-1723.
- Liu, Q. et al. (2016). “Delayed autumn phenology in the Northern Hemisphere is related to change in both climate and spring phenology”. In: *Global change biology* 22.11, pp. 3702–3711. ISSN: 1354-1013.
- Lu, X. et al. (2018). “Seasonal patterns of canopy photosynthesis captured by remotely sensed sun-induced fluorescence and vegetation indexes in mid-to-high latitude forests: A cross-platform comparison”. In: *Science of The Total Environment* 644, pp. 439–451. ISSN: 0048-9697.
- Migliavacca, M. et al. (2012). “On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model”. In: *Biogeosciences* 9.6, pp. 2063–2083.
- Miles, P. R. (2019). “pymcmcstat: A python package for bayesian inference using delayed rejection adaptive metropolis”. In: *Journal of Open Source Software* 4.38, p. 1417. ISSN: 2475-9066.

- Nizinski, J. J. and B. Saugier (1988). “A model of leaf budding and development for a mature *Quercus* forest”. In: *Journal of Applied Ecology*, pp. 643–652. ISSN: 0021-8901.
- Norby, R. J. (2021). “Comment on “Increased growing-season productivity drives earlier autumn leaf senescence in temperate trees””. In: *Science* 371.6533.
- Novick, K. et al. (2018). “The AmeriFlux network: A coalition of the willing”. In: *Agricultural and Forest Meteorology* 249, pp. 444–456. ISSN: 0168-1923.
- Novick, K. A., A. G. Konings, and P. Gentine (2019). “Beyond soil water potential: An expanded view on isohydricity including land–atmosphere interactions and phenology”. In: *Plant, cell & environment* 42.6, pp. 1802–1815. ISSN: 0140-7791.
- O’Keefe, J. (2000). “Phenology of Woody Species at Harvard Forest since 1990, Long Term Ecological Research Network”. In.
- Park, T. et al. (2016). “Changes in growing season duration and productivity of northern vegetation inferred from long-term remote sensing data”. In: *Environmental Research Letters* 11.8, p. 084001. ISSN: 1748-9326.
- Pastorello, G. et al. (2020). “The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data”. In: *Scientific data* 7.1, pp. 1–27. ISSN: 2052-4463.
- Piao, S. et al. (2007). “Growing season extension and its impact on terrestrial carbon cycle in the Northern Hemisphere over the past 2 decades”. In: *Global Biogeochemical Cycles* 21.3. ISSN: 1944-9224.
- Reichstein, M. et al. (2005). “On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm”. In: *Global change biology* 11.9, pp. 1424–1439.
- Richardson, A. D. et al. (Mar. 2018). “Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery”. In: *Sci Data* 5, p. 180028. ISSN: 2052-4463 (Electronic) 2052-4463 (Linking). DOI: 10.1038/sdata.2018.28.
- Richardson, A. D. et al. (2006). “Phenology of a northern hardwood forest canopy”. In: *Global Change Biology* 12.7, pp. 1174–1188. ISSN: 1354-1013.
- Seddon, A. W. et al. (2016). “Sensitivity of global terrestrial ecosystems to climate variability”. In: *Nature* 531.7593, pp. 229–232. ISSN: 1476-4687.
- Song, Y. et al. (2010). “Trends of the thermal growing season in China, 1951–2007”. In: *International Journal of Climatology: A Journal of the Royal Meteorological Society* 30.1, pp. 33–43.
- Tang, J. et al. (2016). “Emerging opportunities and challenges in phenology: a review”. In: *Ecosphere* 7.8, e01436. ISSN: 21508925. DOI: 10.1002/ecs2.1436.
- Tang, X. et al. (Dec. 2014). “How is water-use efficiency of terrestrial ecosystems distributed and changing on Earth?” en. In: *Scientific Reports* 4.1, p. 7483. ISSN: 2045-2322. DOI: 10.1038/srep07483.
- Vliet, A. J. van et al. (2003). “The European phenology network”. In: *International Journal of Biometeorology* 47.4, pp. 202–212. ISSN: 1432-1254.

- Wolf, S. et al. (2016). “Warm spring reduced carbon cycle impact of the 2012 US summer drought”. In: *Proceedings of the National Academy of Sciences* 113.21, pp. 5880–5885. ISSN: 0027-8424.
- Wu, C. et al. (2017). “Land surface phenology derived from normalized difference vegetation index (NDVI) at global FLUXNET sites”. In: *Agricultural and Forest Meteorology* 233, pp. 171–182. ISSN: 01681923. DOI: 10.1016/j.agrformet.2016.11.193.
- Wutzler, T. et al. (2018). “Basic and extensible post-processing of eddy covariance flux data with REddyProc”. In: *Biogeosciences* 15.16, pp. 5015–5030. ISSN: 1726-4170.
- Yang, L. and A. Noormets (2021). “Standardized flux seasonality metrics: a companion dataset for FLUXNET annual product”. In: *Earth System Science Data* 13.4, pp. 1461–1475. ISSN: 1866-3508.
- Yu, Z. et al. (2013). “Effects of seasonal snow on the growing season of temperate vegetation in China”. In: *Global change biology* 19.7, pp. 2182–2195.
- Zani, D. et al. (2020). “Increased growing-season productivity drives earlier autumn leaf senescence in temperate trees”. In: *Science* 370.6520, pp. 1066–1071. ISSN: 0036-8075.
- Zhang, X., M. A. Friedl, and C. B. Schaaf (2006). “Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements”. In: *Journal of Geophysical Research: Biogeosciences* 111.G4. ISSN: 0148-0227.
- Zhang, X. et al. (2003). “Monitoring vegetation phenology using MODIS”. In: *Remote sensing of environment* 84.3, pp. 471–475. ISSN: 0034-4257.
- Zhang, Y. et al. (2019). “Can vegetation optical depth reflect changes in leaf water potential during soil moisture dry-down events?” In: *Remote Sensing of Environment* 234. ISSN: 00344257. DOI: 10.1016/j.rse.2019.111451.
- Zhang, Y. et al. (2020). “Large and projected strengthening moisture limitation on end-of-season photosynthesis”. In: *Proceedings of the National Academy of Sciences* 117.17, pp. 9216–9222. ISSN: 0027-8424.

2.7 Appendix

Table A2.1: Description for the sites used in this study. Lat, Lon indicated the latitude and longitude of the site respectively; and FY and LY presented the first and last year of observations available. Here, I used the International Geosphere-Biosphere Program (IGBP) scheme to identify the biome types of different sites: evergreen needleleaf forest (ENF), grassland (GRA), deciduous broadleaf forest (DBF), open shrubland (OSH), mixed forest (MF), woody savanna (SVA), wetland (WET) and evergreen broadleaf forest (EBF) sites.

Site	IGBP	Lat	Lon	FY	LY	DOI
FI-Hyy	ENF	61.85	61.85	1996	2014	10.18140/FLX/1440158
CA-Obs	ENF	53.99	53.99	1997	2010	10.18140/FLX/1440044
CA-Let	GRA	49.71	49.71	1998	2008	10.17190/AMF/1436318
US-MMS	DBF	39.32	39.32	1999	2014	10.18140/FLX/1440083
US-NC2	ENF	35.8	35.8	2005	2017	10.17190/AMF/1246083
CA-Ojp	ENF	53.92	53.92	1996	2010	10.17190/AMF/1375199
IT-MBo	GRA	46.01	46.01	2003	2013	10.18140/FLX/1440170
US-EML	OSH	63.88	63.88	2008	2017	10.17190/AMF/1418678
BE-Vie	MF	50.3	50.3	1996	2014	10.18140/FLX/1440130
CA-Oas	DBF	53.63	53.63	1996	2010	10.18140/FLX/1440043
US-Syv	MF	46.24	46.24	2001	2017	10.18140/FLX/1440091
US-Vcm	ENF	35.89	35.89	2007	2017	10.17190/AMF/1246121
NL-Loo	ENF	52.17	52.17	1996	2014	10.18140/FLX/1440178
DE-Hai	DBF	51.08	51.08	2000	2012	10.18140/FLX/1440148
CA-Cbo	DBF	44.32	44.32	1994	2014	10.17190/AMF/1498755
CA-Ca1	ENF	49.87	49.87	1996	2010	10.17190/AMF/1480300
CA-Gro	MF	48.22	48.22	2003	2014	10.18140/FLX/1440034
US-UMB	DBF	45.56	45.56	2007	2017	10.18140/FLX/1440093
CA-Man	ENF	55.88	55.88	1994	2008	10.18140/FLX/1440035
US-Ha1	DBF	42.54	42.54	1991	2012	10.18140/FLX/1440071
CA-Ca3	ENF	49.53	49.53	2000	2010	10.17190/AMF/1480302
US-UMd	DBF	45.56	45.56	2007	2017	10.18140/FLX/1440101
IT-Col	DBF	41.85	41.85	1996	2014	10.18140/FLX/1440167
US-NR1	ENF	40.03	40.03	1998	2017	10.18140/FLX/1440087
BE-Bra	MF	51.31	51.31	1996	2014	10.18140/FLX/1440128
DE-Tha	ENF	50.96	50.96	1996	2014	10.18140/FLX/1440152
US-Blo	ENF	38.9	38.9	1997	2007	10.18140/FLX/1440068
US-Wjs	SAV	34.43	34.43	2007	2017	10.17190/AMF/1246120
US-SRG	GRA	31.79	31.79	2008	2017	10.18140/FLX/1440114

US-Wkg	GRA	31.74	31.74	2004	2017	10.18140/FLX/1440096
DE-Lnf	DBF	51.33	51.33	2002	2012	10.18140/FLX/1440150
US-Me2	ENF	44.45	44.45	2002	2017	10.18140/FLX/1440079
US-GLE	ENF	41.37	41.37	1999	2017	10.18140/FLX/1440069
US-Los	WET	46.08	46.08	2000	2017	10.18140/FLX/1440076
US-Ho1	ENF	45.2	45.2	1996	2017	10.17190/AMF/1246061
DK-Sor	DBF	55.49	55.49	1996	2014	10.18140/FLX/1440155
IT-Cpz	EBF	41.71	41.71	1997	2009	10.18140/FLX/1440168
CH-Lae	MF	47.48	47.48	2004	2014	10.18140/FLX/1440134
US-WCr	DBF	45.81	45.81	1999	2014	10.18140/FLX/1440095

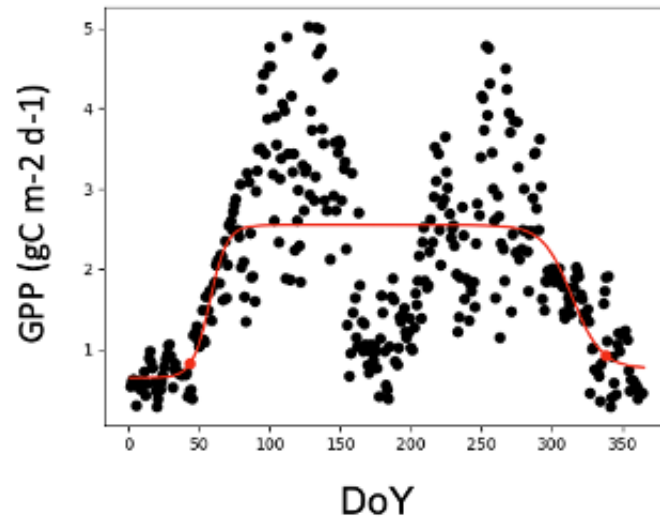


Figure A2.1: The time series of gross primary productivity (GPP) at the site of US-Vcp in the year of 2007, indicating a double-peaked growing season. I excluded such double-peaked sites.

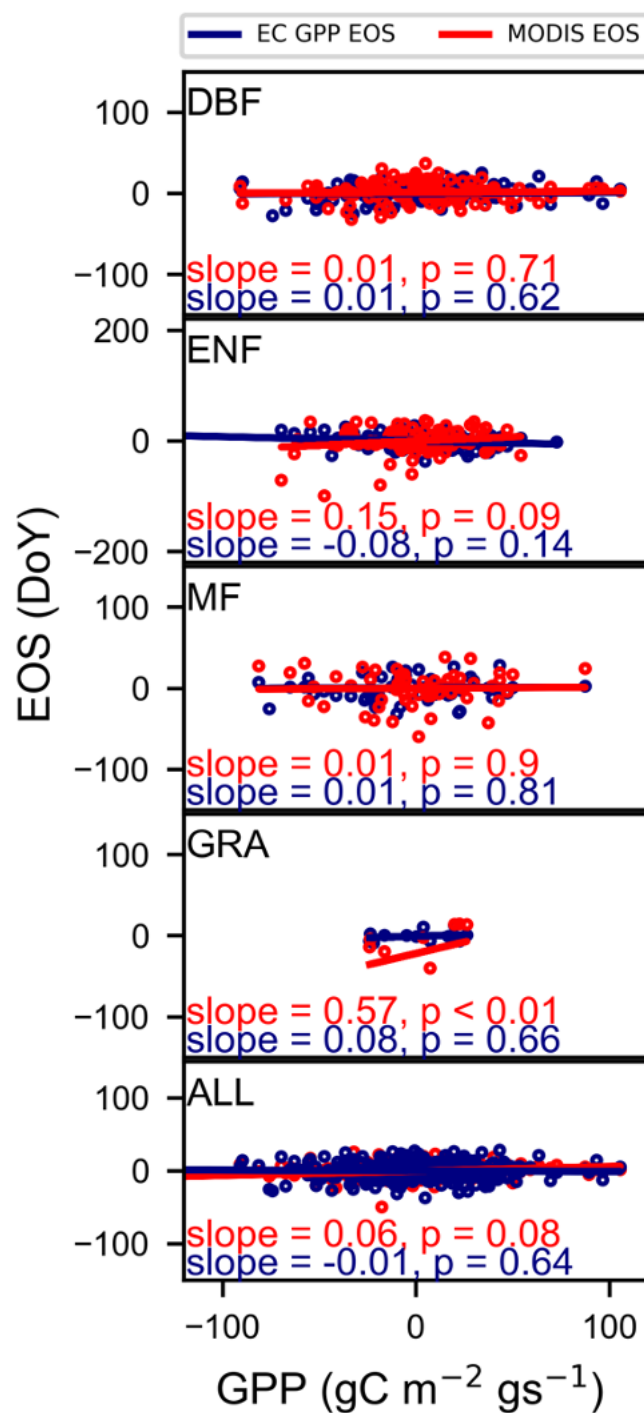


Figure A2.2: Similar to Fig.1., but using the daytime partitioning method.

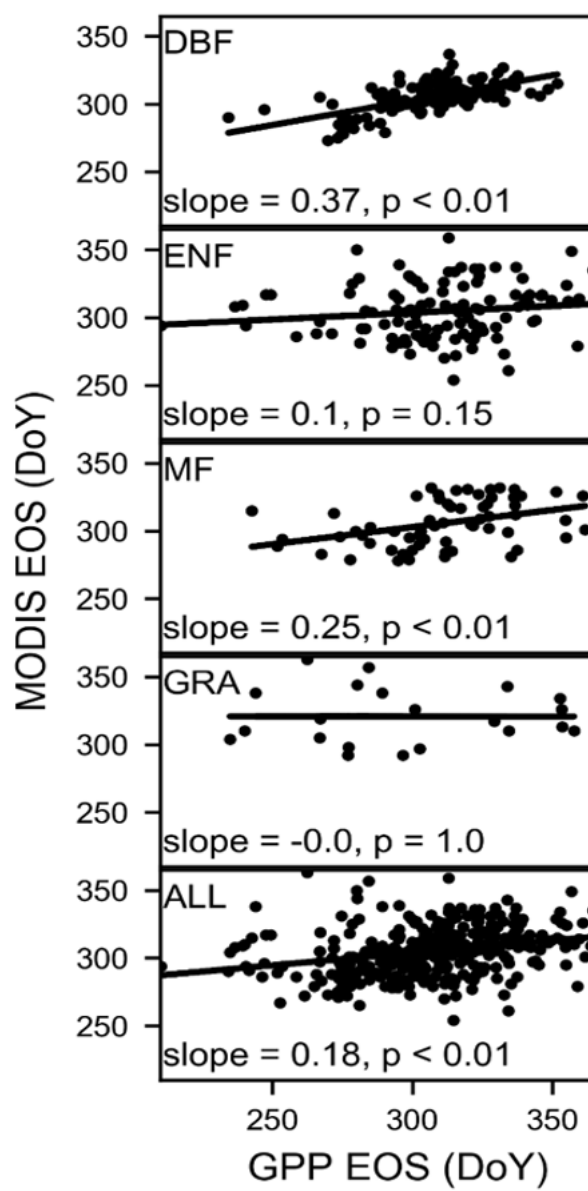


Figure A2.3: Similar to Fig.1., but using the daytime partitioning method.

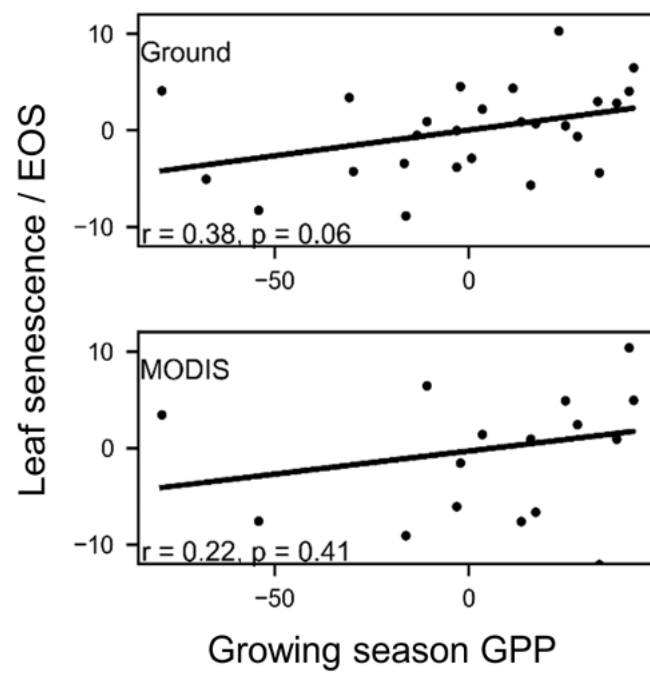


Figure A2.4: The relationship between growing season photosynthesis (GPP) and ground observed leaf senescence or remotely sensed end of season (EOS).

Transition from chapter 2 to chapter 3

In Chapter 2, I explored the interplay between growing season photosynthesis and indices of leaf aging and the end of the growing season (EOS) across varied ecosystems. Contrary to recent studies, our findings did not identify a negative impact of heightened growing season photosynthesis on leaf aging and EOS across most sites and ecosystems. Interestingly, the factors influencing EOS varied by ecosystem. In the majority of cases, heightened productivity during the growing season correlated with a delayed, rather than expedited, leaf aging process. This challenges the prevailing idea that increased growing season photosynthesis negatively influences EOS, underscoring the necessity for more controlled studies to pinpoint the primary determinants of autumn leaf aging.

Building on the ONEFlux BETA flux tower data presented in Chapter 1, we also delved into other pertinent aspects of global water and carbon cycles. Chapter 3 zeroes in on ecosystem water use efficiency (WUE), characterized as the ratio between carbon accumulation and water expenditure.

Plant adaptation to water scarcity often involves adjustments in stomatal behavior, pivotal in modulating both water and carbon exchanges in plants. The water use efficiency (WUE) is instrumental in shedding light on plant reactions under drought conditions. However, the way WUE adapts to such stress varies across ecosystems and hinges on factors like time scales, plant biological processes, and the severity of atmospheric and soil dryness. Current academic discussions introduce a plethora of WUE metrics, adding layers of complexity to the comprehension of WUE's relationship with water stress. Within this chapter, we harness data from flux tower observations spanning 64 sites and 3 WUE metrics.

Chapter 3

Disentangling the influence of soil moisture and atmospheric dryness on the coupled ecosystem carbon and water cycling using machine learning

Xinchen Lu^{1,2}, Trevor F. Keenan^{1,2}

1, Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

2, Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA

Abstract

Defined as the ratio of carbon gain and water loss, water use efficiency (WUE) is an important key ecosystem property that links the global carbon, water and energy budgets. As plants respond to water stress by regulating the openness of stomata, which control the exchanges of water and carbon for plants, WUE also provides important information regarding the response of plants to dry conditions. However, how WUE responds to water stress in different ecosystems depends on the temporal scale, plant physiology and the extent of atmospheric and soil aridity. In addition, in the current literature, there are multiple widely used WUE metrics, which further complicates understanding of WUE-water stress relationships. In this chapter, we used flux tower observations from 64 sites that have more than 5 years of observations (753 site-years) to assess the response to dry conditions of 3 WUE metrics. More Specifically, we used WUE(ET) (defined as the ratio of gross primary productivity (GPP) over evapotranspiration (ET)) and WUE(T) (defined as the ratio of GPP over transpiration (T)); Besides, we also used iWUE which is defined as the ratio of GPP over ET

adjusted by vapor pressure deficit (VPD) to correct for the influence of ambient humidity on plant transpiration rates. Our results indicate that while both decreased soil water content (SWC) and increased atmospheric water demand (vapor pressure deficit, VPD) could impact ecosystem WUE, most sites showed a more consistent response to VPD as both WUE(ET) and WUE(T) decrease with a higher VPD, with sites in more arid regions showing lowest sensitivities.

3.1 Introduction

Water-use efficiency (WUE) is a key trait that quantifies the efficiency of plants in using water for carbon gain, linking the global carbon, water, and energy budgets, and providing important information regarding the response of plants to water stress (Cowan and GD 1977; Keenan et al. 2013; Lavergne et al. 2019; Prentice et al. 2014; Tanner and Sinclair 1983; Yi et al. 2019). At ecosystem scale, WUE is defined as the ratio of carbon uptake (photosynthesis) to the amount of water lost through evapotranspiration (ET) or transpiration (T) (Bréda et al. 2006; Hatfield and Dold 2019). WUE can vary greatly between plant species and is influenced by factors such as stomatal conductance, photosynthetic rate, and plant water status. Plants respond to water stress by regulating the aperture of stomata, which control the release of water from leaves (Green et al. 2019). Different species employ different strategies in response to water stress. For example, drought-resistant species have much higher WUE than less resilient species and can better regulate water loss during drought (Anderegg et al. 2018; Konings and Gentine 2017; Ponton et al. 2002; Tang et al. 2014). Understanding WUE is thus crucial for predicting plant responses to climate change and for developing strategies to enhance the resilience of ecosystems to drought (Feng et al. 2019).

Understanding how different ecosystems respond to water stress by regulating their WUE can be challenging as many factors may contribute to differences in responses. Most tree ring based studies found that in response to dry years, WUE increased as trees became more efficient at using water, suggesting that they may be able to cope with future droughts (Andreu-Hayles et al. 2011). Similarly, Ponce-Campos et al., 2013 compared the response of above-ground net primary production to hydroclimatic conditions in the late twentieth century and early twenty-first century, and found a common increase in ecosystem water-use efficiency across biomes that indicates sensitivity to water availability, with biome-scale resilience to interannual variability associated with drought in the early twenty-first century. However, at a shorter time scale, WUE may respond to dry conditions differently, as WUE may decrease with higher VPD (Roby et al. 2020). In a recent study, Baldocchi et al. 2021 found no significant change in water use efficiency over time, as multiple factors, including CO₂ effects on stomata, soil moisture, vapor pressure deficit, temperature impacts on vapor pressure, and groundwater access, muted out the impact of each other.

At the same time, water stress is characterized by the effects of both decreased soil water

content (SWC) and increased atmospheric water demand, measured by vapor pressure deficit (VPD) (Ciais et al. 2005; Fu et al. 2022; Stocker et al. 2019). In order to optimize carbon gains while minimizing water loss in response to high VPD, plants regulate stomatal conductance (Katul et al. 2010). At the same time, the hydraulic transfer from soils to leaves is hindered and stomatal closure is reinforced when the SWC drops below a critical stress level, thereby exacerbating the negative impact on plant growth. To our knowledge, there have been no studies that specifically look at how different ecosystem-level WUE respond to water stress as indicated by VPD and SWC.

The existence of several different definitions of WUE adds to the complexity of understanding the WUE-drought relationship. WUE(ET) can be defined as the ratio of carbon gain (i.e. gross primary productivity (GPP)) to water used (i.e. evapotranspiration (ET)). Similarly, transpiration use efficiency (WUE(T)) is generally defined as the ratio of carbon gained to water lost through transpiration, but there are variations in the definition, such as the inclusion of soil evaporation. Additionally, intrinsic water use efficiency (iWUE), which is calculated as the ratio of carbon gain to stomatal conductance, is also used as a measure of plant water use efficiency (Green et al. 2019). The discrepancies in the definitions and methodologies for calculating and reporting WUE and TUE would help to make it challenging for a better understanding of plant water use efficiency.

Finally, the concept of WUE, rooted in the ratio of GPP to ET (or T), is influenced by both these components. Recent studies examining the effects of VPD and SWC on GPP have produced conflicting results, making it challenging to deduce their respective roles and the intertwined dynamics of the carbon and water cycles (Kimm et al. 2020; Sulman et al. 2016). Furthermore, while the influences of VPD and SWC on GPP have garnered attention, it is equally paramount to dissect how ET behaves across various ecosystems when subjected to shifts in VPD and SWC. In light of these complexities, this manuscript dedicates a considerable section to examining the responses of both GPP and ET to water stress. Through this aspect, we want to provide a more enriched, layered understanding of the dynamics that dictate how WUE evolves and adapts under varying water stress conditions.

A promising strategy to address this issue is the application of the Accumulated Local Effects (ALE) method. ALE is a statistical tool designed to disentangle and quantify the independent influences of covarying variables, making it a promising approach to analyze their individual impacts on WUE (see methods, (Apley and Zhu 2020)). This is particularly important given the inherent covariation between VPD and SWC, which makes understanding their distinct contributions to different WUE metrics a formidable challenge. With this method, we can isolate and quantify the distinct effects of VPD and SWC, thereby minimizing the potential analytical distortions that might arise due to their inherent covariation. By employing the ALE analysis, we can gain nuanced insights into the specific roles and interactions of VPD and SWC with respect to WUE, laying the groundwork for improved predictive modeling and enhanced management recommendations.

In the current study, we utilized flux tower observations from 64 sites (753 site-years) distributed globally to investigate a fundamental question: how is water use efficiency (WUE) influenced during dry conditions (i.e. high VPD or low SWC) in various ecosystems at a daily scale? To address this question, we concentrated on three primary concerns:

1. How does WUE respond to changes in aridity across diverse ecosystems?
2. What are the key factors contributing to these differences, such as vapor pressure deficit (VPD) and/or soil water content (SWC)?
3. How do varying definitions of WUE influence derived WUE-drought response relationships?

3.2 Methods and Materials

Flux tower observation

We used flux tower observations from 64 sites (732 site-years of measurements) that have more than 5 years of observations from Fluxnet 2015 (Piao et al. 2020) and the AmeriFlux ONEFlux BETA release (<https://ameriflux.lbl.gov/data/download-data-oneflux-beta/>). These sites represent a large range of the ecosystems globally (9 IGBP types) that have observations from 1981 to 2017 (Table A3.1). On average, these sites have an average of 12 years of observations, and they also represent a generally large gradient of aridity index (AI, which was defined as the ratio of precipitation over potential evapotranspiration) ranging from 0.30 to 3.46 (average AI at 1.35). We group the sites based on the IGBP types: 1) evergreen needle leaf (ENF) sites; 2) grassland (GRA) sites; 3) deciduous broadleaf forest and evergreen broadleaf forest sites as broadleaf forest (BF) sites; 4) mixed forest sites; 5) open shrubland and closed shrubland sites and shrubland (SH) sites; 5) woody savanna and savanna sites as savanna (SV) sites. Cropland sites were excluded from the analysis as some of the sites may be impacted by irrigation or other water management (e.g. crop rotation).

In the present study, the gap-filled daily aggregates (or average) from the site observations were used. We considered the main environmental variables of air temperature (T_{air}), vapor pressure deficit (VPD), global radiation (R_g) and soil water content (SWC); and the carbon and water fluxes of gross primary production (GPP) and evapotranspiration (ET) converted from latent heat. As both the FLUXNET 2015 and ONEFlux BETA data sets do not provide the partitioned data of ET (i.e. transpiration (T) from plants), we used a recent study that compared 3 different ways of ET partitioning and used the average daily aggregates of the 3 different methods (Tramontana et al., 2020).

As we want to focus on the growing season where the ecosystems were likely to be impacted by water stress, similar to (Anderegg et al. 2018); (Fu et al. 2022), we applied filters to

identify these days: 1) daytime average air temperature $> 15^{\circ}\text{C}$; 2) solar radiation was found to be significant as photosynthetic photon flux density over a 24-hour period exceeded $500 \mu\text{mol m}^{-2} \text{s}^{-1}$; 3) water movement was established through the presence of evaporative demand with VPD exceeding 0.5 kPa. We calculated three different and commonly used WUE metrics as follows:

$$WUE(ET) = \frac{GPP}{ET} \quad (3.1)$$

$$WUE(T) = \frac{GPP}{T} \quad (3.2)$$

$$iWUE = \frac{GPP \times \sqrt{VPD}}{ET} \quad (3.3)$$

Here, $WUE(ET)$ was defined as the ratio of carbon gain (GPP) to water loss (ET) in order to quantify and represent the water-carbon trade-off for plants (Cowan and GD 1977 eq. 3.1). An alternative approach is to look at the WUE as transpiration efficiency ($WUE(T)$) instead, as soil evaporation and plant transpiration respond to water stress differently and $WUE(T)$ may better capture the plants' responses (Condon et al. 1993, eq. 3.2). As a way to quantify the ecosystems' response to environmental conditions, several studies use intrinsic WUE, by using the mean daylight VPD to in the equation (Beer et al. 2009; Zhou et al. 2015). In the present study, we quantified $iWUE$ using eq. 3.3.

Analytical methods

In the present study, we built random forest models to predict the responses of GPP, ET, WUE and TUE (Breiman, 2001; Cutler et al., 2007) at each site. More specifically, we trained random forest models at each site for each measurement at daily scale (namely, WUEs, GPP and ET) by using air temperature ($^{\circ}\text{C}$), global radiation (W m^{-2}), VPD (kPa), SWC (%), leaf area index (m^2/m^2), and day of year. As VPD was used in the calculation of $iWUE$, we excluded it from this part of the analysis. For each site, we only used the days where the ecosystems are likely to be impacted by water stress and all measurements were averaged or aggregated at daily scales, 30% of the daily measurements were used as a test set. We confirmed that the machine learning method used did not influence our results by examining alternative approaches (e.g. XGBoost) (Figure A3.1, Chen and Guestrin 2016).

As we want to disentangle the effects of each predictor, especially the VPD and SWC impacts, we implemented an Accumulated Local Effects (ALE) analysis based on the trained random forest models (Apley and Zhu 2020). ALE is a machine learning model-agnostic interpretation method that effectively captures the impacts of each feature while accounting for potential interactions with other features within the model. By aggregating the local influences of SWC and VPD, we aim to obtain a comprehensive understanding of the global

impact these two variables have on the model's predictions. This approach enabled us to overcome the limitations of partial dependence plots (PDPs), which may produce biased results when feature interactions are present. Consequently, the application of ALE provided a more reliable means for assessing the individual contributions of SWC and VPD in our machine learning model, thus promoting confidence in our analysis. In the context of our study, the application of the ALE method would help to answer our overarching research questions by isolating the specific influences of SWC and VPD on WUE.

For each site, we trained a random forest model to predict the fluxes, then applied the ALE method to analyze how each environmental factor contributed to the overall effects. We also evaluated the level of water stress by splitting the data set of each site according to their corresponding VPD and SWC percentile with regard to all sites. Then, we calculated the average ALE of each percentile bucket of each site, then analyzed the overall normalized ALE by averaging across sites. ALE allows for more specific comparisons between controls on WUE across ecosystems, showing how changes in environmental factors translate to model predictions. The ALE analysis was carried out using the Alibi python package for machine learning model inspection and interpretation (Klaire et al. 2021).

3.3 Results

Modeling of WUE in different ecosystems

At the ecosystem level, we observed a considerable variation in water use efficiencies (WUEs) across distinct ecosystems (Figure 3.1). Employing three different WUE metrics in our analysis, we discerned analogous patterns in the diversification of WUEs among various ecosystems. Generally, our results show that forest ecosystems (BF, MF, and NF) exhibited higher WUEs, with no statistically significant differences between them on average. In contrast, grassland (GRA) and shrubland (SH) ecosystems demonstrated lower WUEs, particularly for shrublands, which displayed approximately half the efficiency compared to forest ecosystems. Moreover, our results indicated that our Random Forest (RF) model was largely successful in capturing the variations in GPP, ET and WUEs among different sites, achieving an average R^2 value of at least 0.7 for the analyzed fluxes across all sites (Figure A3.2). Overall, it was more challenging to model WUEs compared with GPP and ET.

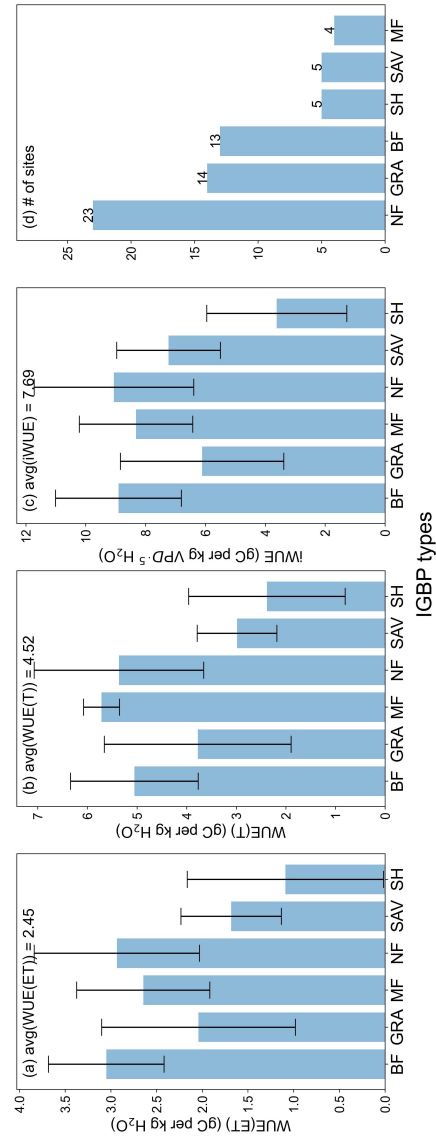


Figure 3.1: WUEs of different ecosystems during the growing season from the flux tower observations. The blue bars denote the mean WUEs of sites in each IGBP group, and the error bars present the standard deviation. The number of sites used is also presented in (d).

GPP, ET and WUE Responses to Water Stress: ALE Analysis Insights

Applying the ALE analysis, we investigated the differential responses of GPP and ET to water stress (Figure 3.2). We found that both SWC and VPD significantly contributed to the variation of GPP across the aridity gradient. The normalized GPP, on average, increased from approximately -0.15 to 0.1 as SWC increased. However, upon reaching a generally higher SWC, the normalized GPP ceased to increase further. In the case of VPD, we similarly observed that the normalized GPP substantially declined with increasing atmospheric demand but did not continue to decrease when attaining a generally higher VPD value.

In contrast, our analysis revealed that the normalized ET, on average, was more closely associated with SWC. As SWC increased, ET followed the pattern and then plateaued. However, ET remained generally stable, albeit with variations from site to site, across different gradients of VPD.

Regarding WUE(ET), we observed a generally consistent decline as VPD increased. This pattern was largely consistent across sites, with more than 80% of the sites demonstrating a significant overall ALE of VPD. In contrast, the relationships between SWC and WUE were less well-defined across sites. On average, SWC's contribution to WUE remained relatively constant, albeit with considerable variation. Similar findings were observed when employing WUE(T), although the uncertainty was notably higher, which is likely attributable to the differences among the three distinct partitioning methods.

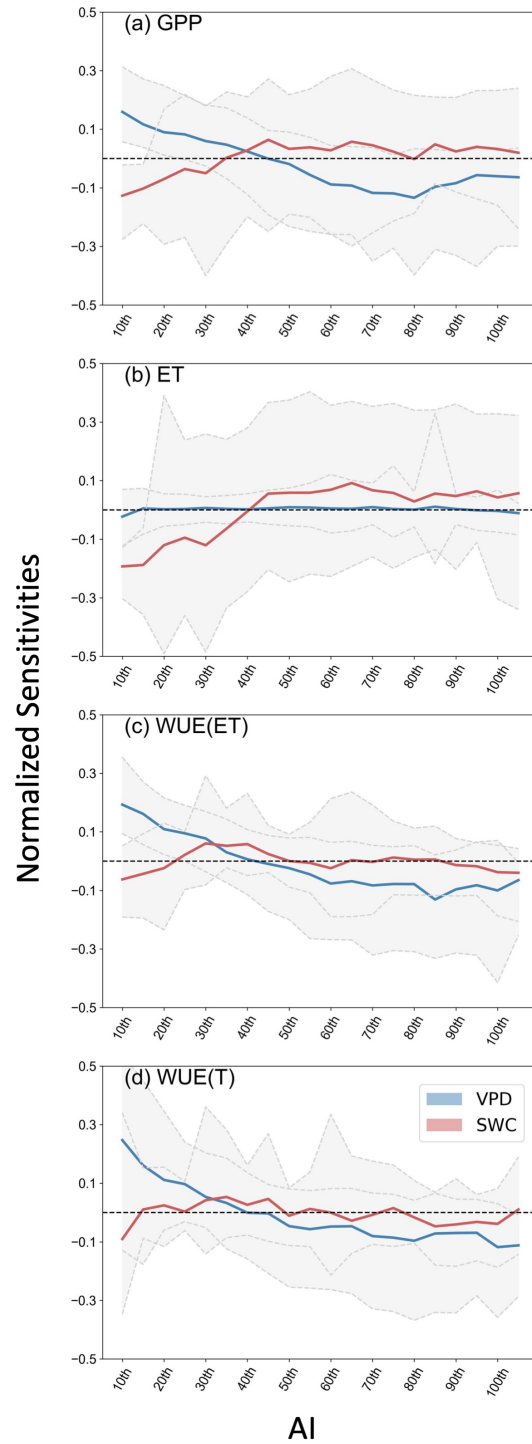


Figure 3.2: Normalized ALE of GPP, ET and WUEs to changes in the different percentiles of VPD and SWC across different sites. The central line indicates the mean normalized ALE across sites, and the dashed area indicates the standard deviation across sites.

Variations in WUE Across Aridity Gradients

We also examined the normalized sensitivities of GPP, ET, and WUEs to both VPD and SWC in generally arid (with aridity index < 0.65) and mesic (with aridity index > 0.65) sites (Figure 3.3). For GPP, we determined that arid sites were generally more sensitive to SWC than VPD, whereas mesic sites exhibited greater sensitivity to VPD, with SWC having a comparatively minimal impact on average. In the case of ET, SWC had a more substantial contribution compared to VPD for both arid and mesic sites, as the majority of sites demonstrated minimal or no VPD sensitivities. For both WUE(ET) and WUE(T), we observed higher sensitivities to VPD in both metrics, while SWC displayed much lower sensitivities.

Similarly, we looked at the normalized sensitivities of GPP, ET, and WUEs to both VPD and SWC across different ecosystems (Figure 3.4). In terms of GPP responses, we observed that most ecosystems exhibited sensitivities to both VPD and SWC. For NF sites, GPP displayed a higher sensitivity to VPD than to SWC, whereas for SAV sites, GPP was more sensitive to SWC. Regarding ET, we discovered that most ecosystems demonstrated more sensitive responses to SWC compared to VPD, while BF, MG, and NF sites exhibited, on average, lower sensitivities. For both WUE(ET) and WUE(T), the majority of sites showed a more consistent response to VPD compared to SWC, with more arid ecosystems such as GRA and SAV displaying relatively lower sensitivities.

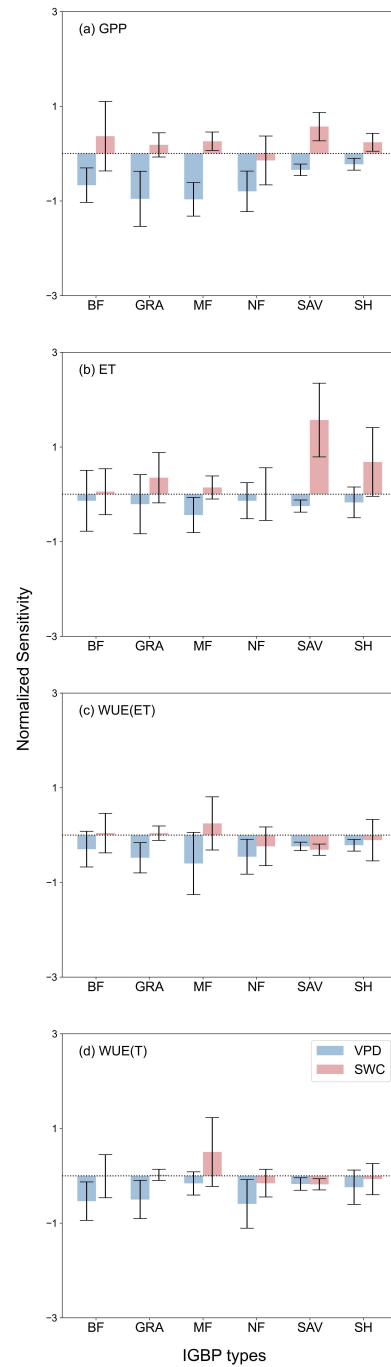


Figure 3.3: Normalized sensitivities of GPP, ET and WUEs to VPD and SWC in both arid and mesic sites based on aridity index. The bar denotes the mean sensitivities of sites in that IGBP group, and the error bar presents the standard deviation. The number of sites used were also presented.

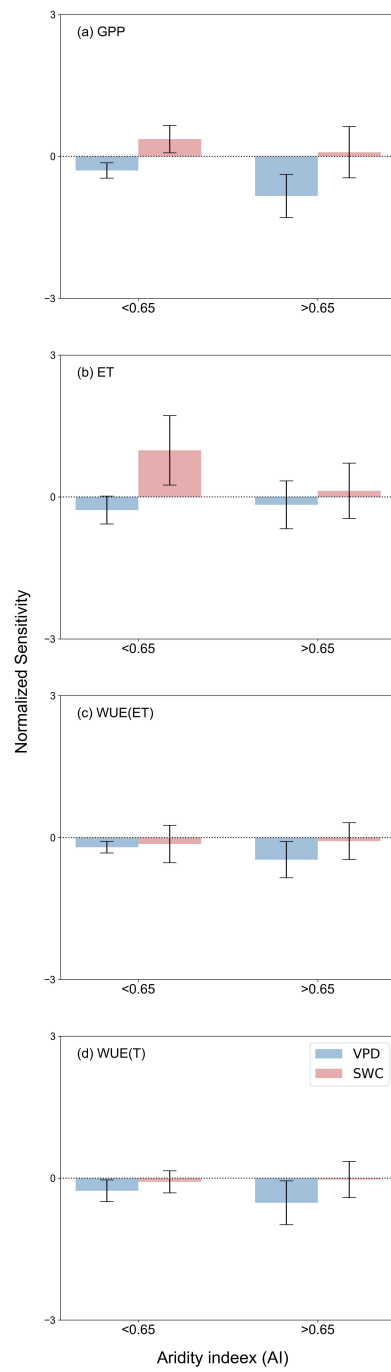


Figure 3.4: Normalized sensitivities of GPP, ET and WUEs to VPD and SWC in different ecosystems. The bar denotes the mean sensitivities of sites in that IGBP group, and the error bar presents the standard deviation. The number of sites used were also presented.

3.4 Discussion

The present study utilized an ALE analysis to explore the regulations of ecosystem-level WUEs, and its components (namely, GPP and ET), in response to high VPD and low SWC across 64 flux tower sites. Firstly, we looked at how the GPP responded to dry conditions. The findings demonstrated that GPP was sensitive to both SWC and VPD at most of these sites. In particular, a significant impact on GPP was observed when SWC was low, causing a marked decline in GPP, as illustrated in Fig. 2. These findings align with the results reported by Fu et al. 2022 and Stocker et al. 2019. Furthermore, the study noted that GPP did not decline further when VPD reached exceedingly high levels, suggesting that stomatal regulation may play a role during these conditions (Novick et al. 2016). The ongoing debate on whether GPP is more sensitive to SWC or VPD could be affected by the degree of aridity at the site, as well as the interactions between SWC and VPD, which may manifest as either covarying or uncoupled effects. Future research should continue to investigate the complex interplay between SWC and VPD in regulating GPP, accounting for factors such as plant functional types, soil properties, and other environmental variables to refine our understanding of ecosystem-level WUE and its responses to changing climatic conditions (Keenan et al., 2013).

Our findings also indicated that, across most sites, ET exhibited greater sensitivity to SWC than to VPD. This observation can be attributed to two primary factors. First, given that ET incorporates soil evaporation, which is more strongly associated with soil moisture, it is reasonable to expect a heightened sensitivity to SWC (Condon et al. 1993). Second, the relatively modest responses of ecosystems to ET on average might suggest the presence of plant-level regulatory mechanisms and a lack of uniformity in site-specific responses to atmospheric aridity (Katul et al. 2010).

In our investigation of daily-scale WUE regulation across various ecosystems, we observed that the majority of sites exhibited a more consistent relationship with VPD. Most sites experienced a decrease in both WUE(ET) and WUE(T) when VPD was high. This finding can be ascribed to the heightened sensitivity of GPP to increasing atmospheric demand, while ET displayed a lesser sensitivity to fluctuations in VPD. Consequently, this resulted in an overall decline in WUE(ET) and WUE(T) across most sites. Although we also discovered that SWC contributed significantly to our random forest models (Figure A3.3), the relationships between SWC and WUE were generally more inconsistent across sites. Interestingly, our analysis revealed that the sensitivities to both VPD and SWC were considerably lower in more arid sites compared to mesic sites. This observation may suggest an adaptive and mitigative strategy employed by ecosystems with more frequent exposure to dry conditions.

In theory, there might be an increase in WUE at extremely low VPD conditions. However, this pattern was not distinctly observable across sites, potentially due to the site-specific nature of low VPD values and the consequent variability in plant responses. This indicates

that while there may be a subtle relationship between WUE and low VPD, detecting and quantifying this relationship is challenging, likely due to the inherent variability in environmental conditions across different ecosystems.

In the current study, we were unable to explore the responses of iWUE to VPD as VPD was used in its calculation (Beer et al., 2009). It is also important to note that there are several distinct definitions of iWUE across different research communities; in this chapter. Additionally, there are also studies that used the slope of GPP over ET as a metric of WUE Baldocchi et al. 2021, In the current study, as I was conducting an analysis at daily scale, I was unable to estimate the daily WUE using this method. Given the existence of multiple definitions of WUE, particularly within the flux tower research community, these varied interpretations have contributed valuable insights from diverse perspectives. However, this variety can also complicate the interpretation and comparison of results across different studies. If the definition of iWUE, as proposed in (Zhou et al. 2015), primarily aims to establish a consistent and interchangeable relationship between GPP and ET at the ecosystem level (i.e., to predict GPP based on ET or vice versa), the application of such an indicator may prove more effective than WUE(ET) and WUE(T).

Lastly, we would like to acknowledge several limitations of the present study. First, although the ALE method employed in this chapter attempted to differentiate the effects of VPD and SWC, it may not be perfect, particularly when the interactions between these two predictors are strong. Nonetheless, ALE is considered more suitable for studies like this in comparison to PDP and offers an analytical approach for addressing these issues in our research.

Another limitation of the current study is the emphasis on aerodynamic (r_a) and stomatal resistances (r_s) in the assessment of WUE, potentially leading to an oversimplified interpretation of ecosystem responses to water stress. While the study considers SWC, it's important to note that SWC may not be as direct a forcing as water potential on plant physiological processes. A more nuanced approach would involve examining the direct impacts of water potential on WUE, which could offer a clearer understanding of the mechanisms underpinning plant responses, particularly under varying soil moisture conditions. Furthermore, the study's analysis does not fully address the dynamic interactions between VPD and the planetary boundary layer (PBL) feedbacks, which are crucial in shaping ecosystem WUE. Specifically, VPD can respond to transpiration (E) during drought through PBL feedback mechanisms, an aspect that warrants deeper exploration. The coupling between VPD and soil moisture, critical for understanding plant physiological responses, is also not thoroughly examined. A more in-depth investigation into these complex feedback loops and interactions would provide valuable insights into ecosystem adaptations to concurrent atmospheric and edaphic changes, thereby enhancing our understanding and predictive capabilities regarding ecosystem behavior under environmental stress.

3.5 Conclusion

In conclusion, water use efficiency (WUE) represents a crucial ecosystem property that interconnects the global carbon, water, and energy budgets. In this chapter, we analyzed flux tower observations from 64 sites, encompassing over five years of data (753 site-years), to investigate the response of WUE metrics—WUE(ET) and WUE(T) to dry conditions on a daily scale, utilizing the ALE method in analytical models. Our results indicate that both decreased soil water content (SWC) and increased atmospheric water demand (VPD) can influence ecosystem WUE. However, the majority of sites exhibited a more consistent response to VPD, with both WUE(ET) and WUE(T) declining as VPD increased. This outcome can be ascribed to the greater sensitivity of GPP compared to ET in relation to VPD. Furthermore, our findings reveal that more arid sites tend to exhibit a reduced response to dry conditions, potentially indicating an adaptive strategy.

3.6 Acknowledgement

This work was supported by the grant of Future Investigators in NASA Earth and Space Science and Technology (No. 80NSSC21K1602) to XC and TFK. TFK was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 as part of their Regional and Global Climate Modeling program through the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation Scientific Focus Area (RUBISCO SFA) project. Funding for AmeriFlux data resources was provided by the U.S. Department of Energy's Office of Science, and can be downloaded from <https://ameriflux.lbl.gov/data/>. The study used the FLUXNET 2015 dataset which can be accessed at <https://fluxnet.org/data/fluxnet2015-dataset/>.

References

- Anderegg, W. R. et al. (2018). “Hydraulic diversity of forests regulates ecosystem resilience during drought”. In: *Nature* 561.7724, pp. 538–541. ISSN: 1476-4687.
- Andreu-Hayles, L. et al. (June 2011). “Long tree-ring chronologies reveal 20th century increases in water-use efficiency but no enhancement of tree growth at five Iberian pine forests: IBERIAN PINE PHYSIOLOGICAL AND GROWTH CHANGES”. en. In: *Global Change Biology* 17.6, pp. 2095–2112. ISSN: 13541013. DOI: 10.1111/j.1365-2486.2010.02373.x.
- Apley, D. W. and J. Zhu (2020). “Visualizing the effects of predictor variables in black box supervised learning models”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.4, pp. 1059–1086. ISSN: 1369-7412.
- Baldocchi, D., S. Ma, and J. Verfaillie (2021). “On the inter-and intra-annual variability of ecosystem evapotranspiration and water use efficiency of an oak savanna and annual grassland subjected to booms and busts in rainfall”. In: *Global Change Biology* 27.2, pp. 359–375. ISSN: 1354-1013.
- Beer, C. et al. (2009). “Temporal and among-site variability of inherent water use efficiency at the ecosystem level”. In: *Global Biogeochemical Cycles* 23.2. ISSN: 08866236. DOI: 10.1029/2008GB003233.
- Bréda, N. et al. (2006). “Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences”. In: *Annals of Forest Science* 63.6, pp. 625–644. ISSN: 1286-4560.
- Chen, T. and C. Guestrin (2016). “Xgboost: A scalable tree boosting system”. In: pp. 785–794.
- Ciais, P. et al. (2005). “Europe-wide reduction in primary productivity caused by the heat and drought in 2003”. In: *Nature* 437.7058, pp. 529–533. ISSN: 0028-0836.
- Condon, A., R. Richards, and G. Farquhar (1993). “Relationships between carbon isotope discrimination, water use efficiency and transpiration efficiency for dryland wheat”. In: *Australian Journal of Agricultural Research* 44.8, pp. 1693–1711. ISSN: 1444-9838.
- Cowan, I. R. and F. GD (1977). “Stomatal function in relation to leaf metabolism and environment.” In.
- Feng, X. et al. (Apr. 2019). “Beyond isohydricity: The role of environmental variability in determining plant drought responses”. en. In: *Plant, Cell & Environment* 42.4, pp. 1104–1111. ISSN: 0140-7791, 1365-3040. DOI: 10.1111/pce.13486.
- Fu, Z. et al. (Feb. 2022). “Atmospheric dryness reduces photosynthesis along a large range of soil water deficits”. en. In: *Nature Communications* 13.1, p. 989. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28652-7.
- Green, J. K. et al. (Jan. 2019). “Large influence of soil moisture on long-term terrestrial carbon uptake”. en. In: *Nature* 565.7740, pp. 476–479. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-018-0848-x.

- Hatfield, J. L. and C. Dold (2019). “Water-use efficiency: advances and challenges in a changing climate”. In: *Frontiers in plant science* 10, p. 103.
- Katul, G. et al. (Mar. 2010). “A stomatal optimization theory to describe the effects of atmospheric CO₂ on leaf photosynthesis and transpiration”. en. In: *Annals of Botany* 105.3, pp. 431–442. ISSN: 1095-8290, 0305-7364. DOI: 10.1093/aob/mcp292.
- Keenan, T. et al. (2013). “Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise”. In: *Nature* 499.7458, pp. 324–327. ISSN: 1476-4687. DOI: 10.1038/nature12291.
- Kimm, H. et al. (June 2020). “Redefining droughts for the U.S. Corn Belt: The dominant role of atmospheric vapor pressure deficit over soil moisture in regulating stomatal behavior of Maize and Soybean”. en. In: *Agricultural and Forest Meteorology* 287, p. 107930. ISSN: 01681923. DOI: 10.1016/j.agrformet.2020.107930.
- Klaise, J. et al. (2021). “Alibi explain: Algorithms for explaining machine learning models”. In: *The Journal of Machine Learning Research* 22.1, pp. 8194–8200. ISSN: 1532-4435.
- Konings, A. G. and P. Gentine (Feb. 2017). “Global variations in ecosystem-scale isohydricity”. In: *Glob Chang Biol* 23.2, pp. 891–905. ISSN: 1365-2486 (Electronic) 1354-1013 (Linking). DOI: 10.1111/gcb.13389.
- Lavergne, A. et al. (2019). “Observed and modelled historical trends in the water-use efficiency of plants and ecosystems”. In: *Global change biology* 25.7, pp. 2242–2257. ISSN: 1354-1013.
- Novick, K. A. et al. (2016). “The increasing importance of atmospheric demand for ecosystem water and carbon fluxes”. In: *Nature climate change* 6.11, pp. 1023–1027. ISSN: 1758-6798.
- Piao, S. et al. (2020). “Characteristics, drivers and feedbacks of global greening”. In: *Nature Reviews Earth & Environment* 1.1, pp. 14–27. ISSN: 2662-138X.
- Ponton, S. et al. (Apr. 2002). “Comparison of water-use efficiency of seedlings from two sympatric oak species: genotype x environment interactions”. en. In: *Tree Physiology* 22.6, pp. 413–422. ISSN: 0829-318X, 1758-4469. DOI: 10.1093/treephys/22.6.413.
- Prentice, I. C. et al. (Jan. 2014). “Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology”. en. In: *Ecology Letters* 17.1. Ed. by J. Penuelas, pp. 82–91. ISSN: 1461023X. DOI: 10.1111/ele.12211.
- Roby, M. C., R. L. Scott, and D. J. P. Moore (Oct. 2020). “High Vapor Pressure Deficit Decreases the Productivity and Water Use Efficiency of Rain-Induced Pulses in Semiarid Ecosystems”. en. In: *Journal of Geophysical Research: Biogeosciences* 125.10. ISSN: 2169-8953, 2169-8961. DOI: 10.1029/2020JG005665.
- Stocker, B. D. et al. (2019). “Drought impacts on terrestrial primary production underestimated by satellite monitoring”. In: *Nature Geoscience* 12.4, pp. 264–270. ISSN: 1752-0894.
- Sulman, B. N. et al. (2016). “High atmospheric demand for water can limit forest carbon uptake and transpiration as severely as dry soil”. In: *Geophysical Research Letters* 43.18, pp. 9686–9695. ISSN: 00948276. DOI: 10.1002/2016gl069416.

- Tang, X. et al. (Dec. 2014). “How is water-use efficiency of terrestrial ecosystems distributed and changing on Earth?” en. In: *Scientific Reports* 4.1, p. 7483. ISSN: 2045-2322. DOI: 10.1038/srep07483.
- Tanner, C. B. and T. R. Sinclair (1983). “Efficient water use in crop production: Research or re-search?” In: *Limitations to efficient water use in crop production*, pp. 1–27.
- Yi, K. et al. (2019). “Linking variation in intrinsic water-use efficiency to isohydricity: a comparison at multiple spatiotemporal scales”. In: *New Phytologist* 221.1, pp. 195–208. ISSN: 0028-646X.
- Zhou, S. et al. (May 2015). “Daily underlying water use efficiency for AmeriFlux sites”. en. In: *Journal of Geophysical Research: Biogeosciences* 120.5, pp. 887–902. ISSN: 2169-8953, 2169-8961. DOI: 10.1002/2015JG002947.

3.7 Appendix

Table A3.1: The detailed information of the sites used in the study. Here, I used the International Geosphere-Biosphere Program (IGBP) scheme to identify the biome types of different sites: evergreen needleleaf forest (ENF), grassland (GRA), deciduous broadleaf forest (DBF), open shrubland (OSH), mixed forest (MF), woody savanna (SVA), wetland (WET) and evergreen broadleaf forest (EBF) sites. The AI_type indicated the aridity index based classification of the sites.

SiteID	First Year	Last Year	Lat	Lon	IGBP	AI_type
US-IB2	2004	2011	41.8406	-88.241	GRA	>0.65
ES-LJu	2004	2013	36.9266	-2.7521	OSH	>0.65
IT-Lav	2003	2014	45.9562	11.2813	ENF	>0.65
US-MMS	1999	2014	39.3232	-86.4131	DBF	>0.65
CH-Dav	1997	2014	46.8153	9.8559	ENF	>0.65
AU-Dry	2009	2014	-15.2588	132.3706	SAV	>0.65
US-SRG	2008	2017	31.7894	-110.8277	GRA	<0.65
CH-Cha	2005	2014	47.2102	8.4104	GRA	>0.65
IT-SRo	1999	2012	43.7279	10.2844	ENF	>0.65
IT-Ren	1998	2013	46.5869	11.4337	ENF	>0.65
US-Whs	2007	2014	31.7438	-110.0522	OSH	<0.65
IT-MBo	2003	2013	46.0147	11.0458	GRA	>0.65
CH-Fru	2005	2014	47.1158	8.5378	GRA	>0.65
CZ-BK1	2004	2014	49.5021	18.5369	ENF	>0.65
US-Me2	2002	2017	44.4523	-121.5574	ENF	<0.65
US-SRM	2004	2014	31.8214	-110.8661	WSA	<0.65
FI-Hyy	1996	2014	61.8474	24.2948	ENF	>0.65
CA-TP3	2002	2014	42.7068	-80.3483	ENF	>0.65
DE-Gri	2004	2014	50.95	13.5126	GRA	>0.65
CH-Oe1	2002	2008	47.2858	7.7319	GRA	>0.65
CA-Obs	1997	2010	53.9872	-105.1178	ENF	>0.65
US-NR1	1998	2014	40.0329	-105.5464	ENF	>0.65
US-Oho	2004	2013	41.5545	-83.8438	DBF	>0.65
MY-PSO	2003	2009	2.973	102.3062	EBF	>0.65
US-Ton	2001	2014	38.4309	-120.966	WSA	<0.65
US-Me3	2004	2009	44.3154	-121.6078	ENF	<0.65
IT-Ro1	2000	2008	42.4081	11.93	DBF	>0.65
DK-Sor	1996	2014	55.4859	11.6446	DBF	>0.65
US-WCr	1999	2014	45.8059	-90.0799	DBF	>0.65

BE-Vie	1996	2014	50.3049	5.9981	MF	>0.65
US-GLE	2004	2014	41.3665	-106.2399	ENF	>0.65
CA-TP2	2002	2007	42.7744	-80.4588	ENF	>0.65
CZ-BK2	2004	2012	49.4944	18.5429	GRA	>0.65
ZM-Mon	2000	2009	-15.4391	23.2525	DBF	>0.65
DE-Tha	1996	2014	50.9626	13.5651	ENF	>0.65
AU-DaS	2008	2014	-14.1593	131.3881	SAV	>0.65
CH-Lae	2004	2014	47.4783	8.3644	MF	>0.65
CA-SF3	2001	2006	54.0916	-106.0053	OSH	>0.65
IT-Tor	2008	2014	45.8444	7.5781	GRA	>0.65
AT-Neu	2002	2012	47.1167	11.3175	GRA	>0.65
IT-Col	1996	2014	41.8494	13.5881	DBF	>0.65
IT-Ro2	2002	2012	42.3903	11.9209	DBF	>0.65
US-Blo	1997	2007	38.8953	-120.6328	ENF	>0.65
CA-Qfo	2003	2010	49.6925	-74.3421	ENF	>0.65
DE-Lnf	2002	2012	51.3282	10.3678	DBF	>0.65
FI-Sod	2001	2014	67.3624	26.6386	ENF	>0.65
FR-LBr	1996	2008	44.7171	-0.7693	ENF	>0.65
IT-Cpz	1997	2009	41.7052	12.3761	EBF	>0.65
DE-Obe	2008	2014	50.7867	13.7213	ENF	>0.65
AU-Stp	2008	2014	-17.1507	133.3502	GRA	<0.65
RU-Fyo	1998	2014	56.4615	32.9221	ENF	>0.65
IT-Noe	2004	2014	40.6062	8.1517	CSH	<0.65
CA-Oas	1996	2010	53.6289	-106.1978	DBF	>0.65
US-Syv	2001	2014	46.242	-89.3477	MF	>0.65
US-Wkg	2004	2014	31.7365	-109.9419	GRA	<0.65
US-Var	2000	2014	38.4133	-120.9508	GRA	<0.65
CA-SF2	2001	2006	54.2539	-105.8775	ENF	>0.65
CA-Gro	2003	2014	48.2167	-82.1556	MF	>0.65
AU-DaP	2007	2013	-14.0633	131.3181	GRA	>0.65
NL-Loo	1996	2014	52.1666	5.7436	ENF	>0.65
DE-Hai	2000	2012	51.0792	10.4522	DBF	>0.65
AU-How	2001	2014	-12.4943	131.1523	WSA	>0.65
US-SRC	2008	2014	31.9083	-110.8395	OSH	<0.65
CA-TP1	2002	2014	42.6609	-80.5595	ENF	>0.65

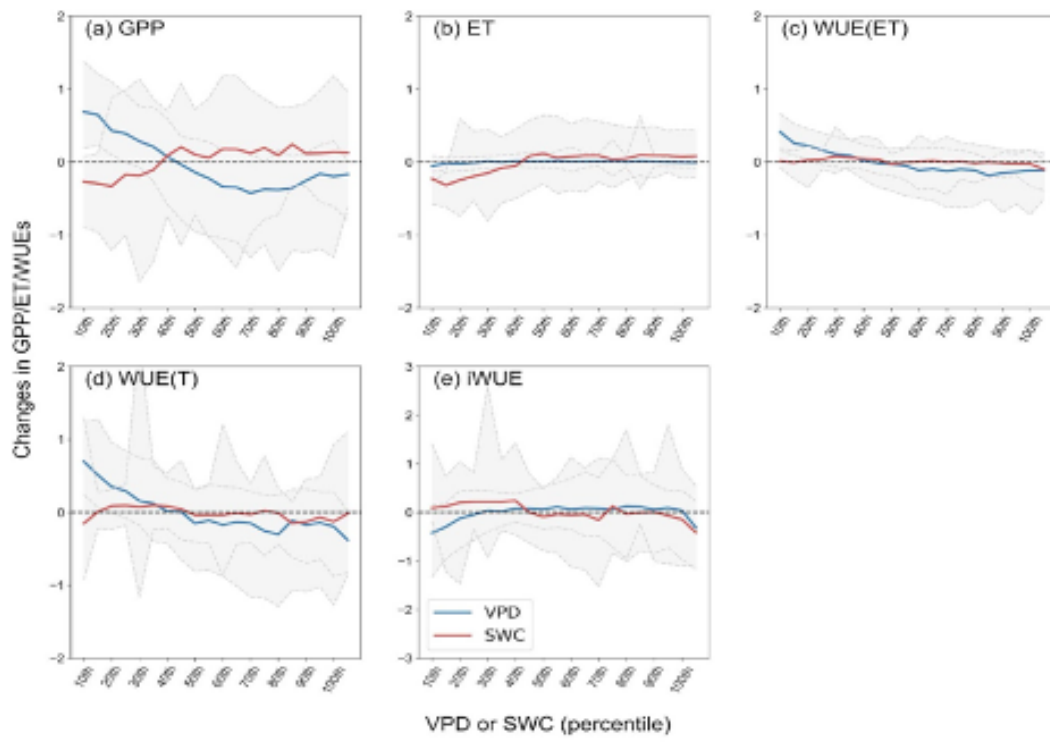


Figure A3.1: Similar to Figure 3.2, with the model trained using XGBOOST.

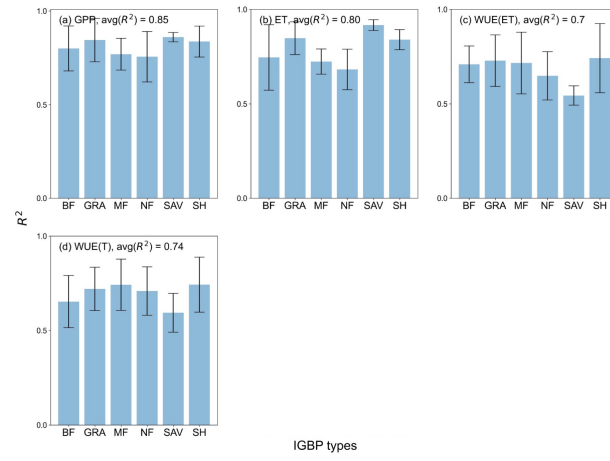


Figure A3.2: The overall R^2 of the trained random forest models to predict fluxes in different sites.

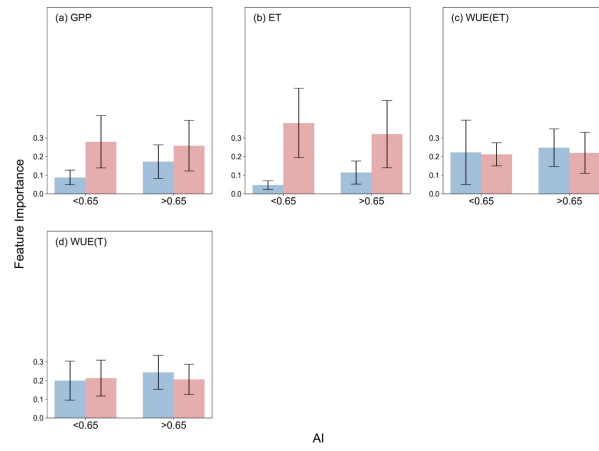


Figure A3.3: Feature importance of SWC and VPD in models for GPP, ET and WUE.

Conclusion

My journey into flux tower observations began in 2016 while working on my master's thesis, which was centered on the remote sensing of phenology. Over the subsequent seven years, my interest and dedication to this area of study deepened. In this dissertation, I progressed from processing flux tower data to providing an updated version for the research community, and then moved to more intricate analyses. This work has, in my view, significantly expanded upon and enriched my prior experiences.

Understanding the carbon and water cycles of global terrestrial ecosystems is crucial, especially given the rising concerns about climate change and environmental sustainability. This dissertation aimed to provide insights into these relationships through a data-driven approach. Leveraging globally available datasets, I have furthered my knowledge in this domain by focusing on three main aspects: refining the AmeriFlux data product, analyzing ecosystem autumn phenology, and evaluating water use efficiency across diverse ecosystems.

The first chapter highlights the enhancement of the AmeriFlux datasets through the ONE-Flux BETA release, which significantly expanded flux data availability by including observations for 63 sites, with many not previously featured in earlier releases or the FLUXNET 2015 datasets. This update has been widely adopted, evidenced by over 168 downloads, for applications ranging from multi-site synthesis to remote sensing studies. The evaluation of ERA-Interim meteorological data against flux tower observations revealed varied reliability, with strong correlations in some parameters but moderate in others, indicating the impact of local environmental conditions. The analysis of Net Ecosystem Exchange (NEE) partitioning methods across different ecosystems highlighted the importance of method selection for accurate carbon flux interpretation, underscoring the dataset's utility in ecological research. The comparison with other datasets like FLUXNET2015 and the incorporation of feedback from a wide range of uses, including model validation and educational purposes, demonstrate the dataset's critical role in advancing understanding of biosphere-atmosphere interactions.

The second chapter challenges the recently proposed hypothesis that a productive growing season leads to earlier leaf senescence. My research indicates the opposite: a more productive growing season actually are associated with delayed end-of-season senescence across different ecosystems. This finding contradicts the existing hypothesis, shedding light on the complex

dynamics between growing season productivity and the timing of autumn senescence. It emphasizes the importance of further research and targeted experiments to unravel the specific mechanisms.

My third chapter focused on the pivotal role of water use efficiency (WUE) within global ecosystems. Analyzing data from 64 flux tower sites, I discovered that while soil water content and atmospheric water demand influence WUE, there was a more consistent response to atmospheric demand across most sites. The introduction of an adjusted metric, iWUE, showed less variation and proved effective in capturing the effects of atmospheric water stress. My findings further suggest that ecosystems in more arid regions might exhibit adaptive strategies to cope with drier conditions.

This research contributes to a nuanced understanding of the interrelated roles of global carbon and water cycles in shaping terrestrial ecosystems. Through a data-driven examination, it has begun to unravel pivotal insights concerning these cycles. The enhancement of the AmeriFlux data product, discussed in the first chapter, represents a step forward in refining the tools available for future investigations, underscoring the critical need for precise and comprehensive datasets to elucidate land-atmosphere interactions. The subsequent chapters, by questioning prevailing views on growing season productivity and illuminating the variations in water use efficiency, underscore the intricate complexities of these global cycles and their significant impacts on ecosystems.