

**Combining remotely sensed estimates of structure and function to improve quantification
of forest productivity**

Paige Tatum Williams

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial
fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Forestry

Randolph H. Wynne, Co-Chair
Valerie A. Thomas, Co-Chair
David J. Harding
John R. Seiler

August 22, 2024
Blacksburg, Virginia

Keywords: forest, remote sensing, productivity, structure, function

Copyright

Combining remotely sensed estimates of structure and function to improve quantification of forest productivity

Paige Tatum Williams

ABSTRACT

Gross primary productivity (GPP) describes the total photosynthesis (i.e., carbon fixation) for an ecosystem and is an important component of the global land carbon budget. Accurate measurements of forest carbon sequestration are crucial, as climate, soils, and management practices influence carbon uptake. Disturbances such as wildfires, diseases, insect infestations, and extreme weather events significantly impact forest health and functionality. Remote sensing is useful for monitoring global ecosystems and estimating forest vitality. High-resolution images and lidar (Light Detection and Ranging) data offer enhanced insights into forest light utilization. Combining lidar with spectral data provides a comprehensive view of forest ecosystems, integrating both structural and physiological information. This work includes three separate studies under a common objective to establish an improved understanding of forest productivity in different forest compositions using a combination of physiological function from optical imagery and morphological structure from lidar. The first study investigates the photochemical reflectance index (PRI) and how illumination affects the diurnal and vertical distributions of two managed pine stands with varying ages and row orientations, using airborne hyperspectral and lidar data. We developed a novel method to classify canopy illumination into sunlit, shaded, and mixed light areas using the hyperspectral data and a simulated panchromatic band. PRI values varied between sunlit and shaded foliage throughout the day, reflecting differences in foliage efficiency depending on light conditions and forest structure. The second study evaluates how well remotely quantified plant functional traits predict GPP across U.S. forests. Using data from NEON's airborne remote sensing and *in situ* flux tower measurements, the study assessed hyperspectral optical indices as physiological traits and lidar-derived products as morphological and environmental traits. By developing multiple linear regression models with separated and combined trait groups, the best prediction model combined morphological, environmental, and physiological traits with a PRESS R^2 of 0.75. This underscores the importance of integrating various functional traits for accurate forest productivity predictions. The last study detects insect-induced tree mortality with separated and combined models of structure from lidar and physiology from satellite imagery. Although all models tend to overestimate tree mortality, integrating lidar data enhances predictions by 6% offering valuable structural context. A central theme of this work is that lidar-derived structural measurements were crucial in every chapter.

Combining remotely sensed estimates of structure and function to improve quantification of forest productivity

Paige Tatum Williams

GENERAL AUDIENCE ABSTRACT

Plant photosynthesis, converting sunlight and carbon dioxide to chemical energy (carbohydrates) and oxygen, is fundamental to all life on Earth. Measuring how much carbon forests capture is essential to understanding the global carbon cycle. Forests must be healthy to be productive, so forest health monitoring is also vital. Remote sensing is a tool to measure objects without physically touching the object you are measuring. From remote sensing, we can take pictures of forests and record information on how much sunlight is reflected. Remote sensing technologies like lidar use lasers to portray trees and forests in three dimensions (3D). Combining pictures with 3D shapes can provide a more comprehensive understanding of how trees store carbon. This work incorporates three separate studies using a combination of pictures and 3D data. The first study looks at the changes in light utilization of a pine plantation throughout the day and within the canopy. The second study uses a range of forest descriptors with different compositions to predict their productivity. The last study shows how imagery from space and 3D data can locate dead and live trees. The overall thrust of the study shows that incorporating 3D portrayals of the forest enhances our understanding of forest productivity.

Dedication

To Leah Rae

Acknowledgements

Completing a PhD truly takes a village, and I am incredibly fortunate to have been surrounded by a remarkable community of support throughout this journey.

First and foremost, my deepest thanks go to my parents and family, whose unwavering belief in me has been my foundation. Your love, encouragement, and patience kept me going even in the toughest moments. A special shoutout to Uncle Jamie and Aunt Susan, who provided the means to be a successful scholar and wisdom to be a good human.

To my friends, both near and far, thank you for being my escape and for always reminding me to laugh. You've kept me grounded and provided me with so many much-needed distractions when I needed them most.

My grad school friends and IGEP peeps, you were the ones who truly rode this wave with me. We've supported each other through every challenge and celebrated every victory. I couldn't have asked for a better group to share this wild ride with.

To my incredible co-pilot, Daniel Cross, whom I worked with in the field, your persistence, camaraderie, and the fun we had made every challenge worth it. You turned the hardest days into learning experiences and adventures I'll never forget.

To my advisors and committee, thank you for your continued support, encouragement, and guidance. You pushed me to grow as a researcher and thinker, and your belief in me never wavered. To Randy and Val, I hope to continue to make you both proud.

To the NASA SAFE team for their invaluable mentorship. I deeply appreciate the wisdom and guidance you all have shared with me. The impact has been profound, and I am truly grateful for everything you all have done to support me along the way.

A heartfelt thank you to every professor I've had at Virginia Tech. You've shaped my academic journey and influenced who I've become as both a scientist and a person.

Finally, to my dog, Willow, for the countless walks, unconditional love, and always being there when I needed a break. You kept me sane and reminded me of the simple joys in life.

This accomplishment I share with you all. Thank you.

Table of Contents

List of Figures	xi
List of Tables	xvi
Chapter 1 - Introduction and Objectives	1
1.1 Introduction	1
1.1.1 Forest productivity.....	1
1.1.2 Structure and function	1
1.1.3 Remote sensing.....	2
1.2 Focus and Objectives by Chapter.....	3
References	4
Chapter 2 - Diurnal and vertical distributions of illumination and photochemical reflectance index in a loblolly pine plantation	6
2.1 Introduction	7
2.2 Materials and Methods.....	10
2.2.1 Study area	10
2.2.2 Datasets.....	11
2.2.2.1 Canopy lidar and hyperspectral reflectance data	12
2.2.2.2 Contextualization of airborne imagery with eddy covariance flux tower.....	13
2.2.3 Methods	15
2.2.3.1 Delineation of vegetated pixels.....	17
2.2.3.2 Panchromatic band simulation for illumination classification.....	17
2.2.3.3 Photochemical reflectance index	18
2.2.3.4 Canopy structure of diurnal PRI	19
2.2.3.5 Calculation of intersection angle on solar and sensor view vectors	19
2.3 Results	20
2.3.1 Illumination classification of PRI.....	20
2.3.2 Reflectance and PRI as a function of height, stand age, and composition	25
2.3.3 LUE and canopy-scale PRI by illumination class	33
2.3.4 Effects of solar and sensor view angles on PRI.....	34
2.4 Discussion	35
2.5 Conclusion.....	39

References	40
Chapter 3 - Characterizing the influence of varying functional traits on forest productivity using remotely sensed data acquired from different forest compositions across selected United States NEON sites.	46
3.1 Introduction	47
3.2 Materials and Methods	49
3.2.1 Study area	49
3.2.2 Datasets	52
3.2.2.1 NEON terrestrial instrument system (TIS)	53
3.2.2.2 NEON airborne observation platform (AOP)	53
3.2.3 Methods	54
3.2.3.1 Calculation of flux tower parameters from NEON TIS	54
3.2.3.2 Calculation of remote sensing functional traits from NEON AOP	55
3.2.3.3 Regression analysis	57
3.3 Results	58
3.3.1 Variable correlation coefficients	58
3.3.2 Multiple linear regression models	60
3.3.3 Canopy height, age, and productivity	64
3.4 Discussion	66
3.5 Conclusion	70
References	70
Chapter 4 – Detection of tree mortality using Sentinel-2 and airborne laser scanning data	75
4.1 Introduction	76
4.2 Materials and Methods	79
4.2.1 Study area	79
4.2.2 Datasets	81
4.2.2.1 Airborne lidar	81
4.2.2.2 Multispectral imagery – Sentinel-2 MSI	81
4.2.3 Methods	82
4.2.3.1 Forest canopy metrics from lidar	82
4.2.3.2 Multispectral indices from S2	84

4.2.3.3 Tree mortality.....	87
4.2.3.4 Statistical analysis	90
4.2 Results	91
4.3.1 Kolmogorov-Smirnov tests.....	91
4.3.2 Random forest classification	92
4.3.3 Vertical profile of 10 m dead and live examples.....	96
4.4 Discussion	98
4.5 Conclusion.....	101
References	102
Chapter 5 - Conclusion	107
Summary and Conclusions.....	107
Future Directions.....	108
Appendix A – MATLAB Code for Chapter 2 and 3.....	110

List of Figures

Figure 2.1. Study area sites in Parker Tract, NC shown with the 420 m x 420 m bounding boxes in the 20-year old loblolly pine stand centered on the US-NC2 flux tower (plot 1) and a more mature loblolly pine stand (plot 2) overlain on top of (a) high-resolution NAIP imagery from October 2014; (b) G-LiHT's 2 m lidar canopy height model from the October 26, 2013 midday flight (blue to red corresponds to increasing canopy height).....	11
Figure 2.2. Solar zenith angle and solar azimuth angle for the nine G-LiHT flights during October 26-27, 2013. Flight numbers correspond to Tables 2.1 and 2.2.....	13
Figure 2.3 US-NC2 flux tower measurements showing incident PAR (solid line, units: $\mu\text{mol m}^{-2}\text{s}^{-1}$), GPP (dashed line, units: $\mu\text{mol CO}_2 \text{ m}^{-2}\text{s}^{-1}$), and LUE (dotted line, units: $\text{mol CO}_2 \text{ mol}^{-1}$ photons *1000), and air temperature (long dashed line, units: degree Celsius) with corresponding G-LiHT flight acquisition times (red circles) from the FLEX – US Airborne Campaign in Parker Tract, NC.....	14
Figure 2.4 Workflow diagram of the sequential processing done in MATLAB on the canopy hyperspectral reflectance data executed for the flux tower plot that was extended to the other plot.....	16
Figure 2.5. Graph of Plot 1's (US-NC2) 2 m PRI mean values over the two flight days grouped by time of day and the illumination class. Percentages on each bar indicate the percent of pixels within the illumination class.	22
Figure 2.6. Graph of Plot 2's (older pine) 2 m PRI mean values over the two flight days grouped by time of day and the illumination class. Percentages on each bar indicate the percent of pixels within the illumination class.	23

Figure 2.7. Plot 1 (US-NC2 flux tower) image output results at 2 m resolution from the canopy hyperspectral processing shown in order of time across the two days. First row: simulated panchromatic band. Second Row: PRI images scaled to -0.1 - 0.1. Third row: Illumination classification with shadow as purple, mixed as green, and sunlit as yellow. 24

Figure 2.8. Plot 2 (older pine) image output results at 2 m resolution from the canopy hyperspectral processing shown in order of time across the two days. First row: simulated panchromatic band. Second Row: PRI images scaled to -0.1 - 0.1. Third row: Illumination classification with shadow as purple, mixed as green, and sunlit as yellow. 24

Figure 2.9. Plot 1 (US-NC2 flux tower) scatterplot of DSM versus panchromatic reflectance % (first column), shadowed PRI (second column), mixed PRI (third column), and sunlit PRI (fourth column) for the NS flights on October 26th for the morning (first row), midday (second row), and evening (third row).The black plus signs are shown to connect the PRI mean of the understory and overstory to show how the light distribution and PRI responses changes within the canopy throughout the day. 26

Figure 2.10. Plot 2 (older pine stand) scatterplot of DSM versus panchromatic reflectance % (first column), shadowed PRI (second column), mixed PRI (third column), and sunlit PRI (fourth column) for the NS flights on October 26th for the morning (first row), midday (second row), and evening (third row). 28

Figure 2.11. 3D structural results for plot 1 (US-NC2) of the CHM (first column, blue to red indicates increased height), illumination class (middle column, sunlit as yellow, mixed as green, and shadow as purple), and PRI (last column, low values in red to high values in blue) for the morning (first row), midday (second row), and evening (third row) NS flights (black points are

from missing spectral data). Time of flights (EST), solar zenith angle (SZA, in degrees), and the solar azimuth angle (SAA, in degrees) are shown for each row..... 31

Figure 2.12. 3D structural results for plot 2 (older pine) of the CHM (first column, blue to red indicates increased height), illumination class (middle column, sunlit as yellow, mixed as green, and shadow as purple), and PRI (last column, low values in red to high values in blue) for the morning (first row), midday (second row), and evening (third row) NS flights. Time of flights (EST), solar zenith angle (SZA, in degrees), and the solar azimuth angle (SAA, in degrees) are shown for each row..... 32

Figure 2.13. Graphs showing the diurnal relationship of LUE (grey line with squares) to PRI (black line with diamonds) as a function of solar azimuth angle for (a) all PRI values and (b) PRI values separated by illumination class (triangle as shadow, circle as mixed, and box as sunlit). 33

Figure 2.14 Plot 1 results for PRI (y-axis) versus the intersection angle between solar and sensor view vector (in degrees) for the morning, midday, and afternoon NS flights. Color gradient indicates pixel frequency with low in blue and high in red. The 13:30 EST intersection angle range is small due to the small difference in angle of the sun (solar noon) and the NS flight (6 degrees difference).....34

Figure 2.15. Plot 2 results for PRI (y-axis) versus the intersection angle between solar and sensor view vector (in degrees) for the morning, midday, and afternoon NS flights. Color gradient indicates pixel frequency with low in blue and high in red. The 13:30 EST intersection angle range is small due to the small difference in angle of the sun (solar noon) and the NS flight (6 degrees difference). 35

Figure 3.1. Study area map of selected forested NEON sites across the United States overlain on ESRI's National Geographic Style basemap. 50

Figure 3.2. High-resolution orthomosaics (NEON product number DP3.30010.00) of the selected forest NEON flux tower sites across the US. Sites are in alphabetical order (left to right) as shown in Table 3.1.....	51
Figure 3.3. Correlation matrix for all RS mean variables and GPP where positive correlations are shown in dark blue and negative correlations shown in yellow.	59
Figure 3.4. Morphological and environmental traits group’s 2-term model graph showing actual GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the y-axis versus predicted GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the x-axis colored by dominant vegetation types (deciduous as light green, evergreen as dark green, and mixed as mustard yellow) including a grey band around the black fit line showing the 95% confidence region.....	61
Figure 3.5. Physiological traits group’s 3-term model graph showing actual GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the y-axis versus predicted GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the x-axis colored by dominant vegetation types (deciduous as light green, evergreen as dark green, and mixed as mustard yellow) including a grey band around the black fit line showing the 95% confidence region.	62
Figure 3.6. Combination of functional traits group’s 4-term model graph showing actual GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the y-axis versus predicted GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the x-axis colored by dominant vegetation types (deciduous as light green, evergreen as dark green, and mixed as mustard yellow) including a grey band around the black fit line showing the 95% confidence region.....	63
Figure 3.7. Correlation matrix for the NEON RS variables used in the 4-term combination of functional traits group model (CHM mean, DSM standard deviation, PRI standard deviation, and	

WBI mean) where positive correlations are shown in dark blue and negative correlations shown in yellow. These four NEON variables have positive correlation coefficients less than 0.5..... 64

Figure 4.1. Study area map of Mountain Lake Biological station showing the mission area of the 420 m x 420 m bounding box centered on US-xML flux tower overlain on NEON’s high resolution orthorectified camera imagery mosaics (NEON product number DP3.30010.00)..... 80

Figure 4.2. Lidar point cloud showing elevation from the UAS acquisition on July 27th, 2023 at MLBS..... 83

Figure 4.3. S2 image example of the normalized difference vegetation index (NDVI) for the study area. 87

Figure 4.4. Classification image of tree mortality showing pixels with dead material as orange and live material as green. 88

Figure 4.5. Example of dead and live 10 m pixels overlain on NEON’s high resolution orthorectified camera imagery mosaics (NEON product number DP3.30010.00.)..... 89

Figure 4.6. Example of dead and live 10 m pixels with 2D point clouds showing height in meters overlain on NEON’s high resolution orthorectified camera imagery mosaics (NEON product number DP3.30010.00)..... 89

Figure 4.7. Variable importance scores for the lidar metrics only model..... 93

Figure 4.8. Variable importance scores for the S2 multispectral indices only model. 94

Figure 4.8. Variable importance scores for the combined model. 95

Figure 4.10. Example of dead and live 10 m pixels with 3D point clouds showing a gradient of heights in meters in the first column and the Cb50 height classification showing ground as 0 – 2 m in brown, the Cb50 delineation as 2 – 10 m in light green, and above the Cb50 delineation as 10 – 20 m in dark green. 97

List of Tables

Table 2.1. Information for G-LiHT flights over plot 1 (US-NC2 loblolly pine plantation) including flight name, time (EST) and day, solar zenith angle and solar azimuth angle (in degrees), and PRI mean and standard deviations for all pixels and the three illumination classes – shadowed, mixed, and sunlit. Bolded rows indicate flights on the 27th.....	21
Table 2.2. Information for G-LiHT flights over plot 2 (older loblolly pine plantation) including flight name, time (EST) and day, solar zenith angle and solar azimuth angle (in degrees), and PRI mean and standard deviations for all pixels and the three illumination classes – shadowed, mixed, and sunlit. Bolded rows indicate flights on the 27th.....	21
Table 2.3. Complimentary statistics to Figure 2.9 for plot 1 using the DSM elevation to delineate the overstory (13 to 32 meters) and the understory (6 to less than 13 meters) providing statistics for the three illumination classes (shadowed, mixed, and sunlit) with the percent of total pixels (%), percent of panchromatic reflectance’s mean and standard deviation, DSM elevation’s mean and standard deviations (in meters), and PRI mean and standard deviation for each illumination class for morning (10:00 EST), midday (13:30 EST), and afternoon (15:30 EST).....	27
Table 2.4. Complimentary statistics to Figure 2.10 for plot 2 using the DSM elevation to delineate the overstory (13 to 32 meters) and the understory (6 to less than 13 meters) providing statistics for the three illumination classes (shadowed, mixed, and sunlit) with the percent of total pixels (%), percent of panchromatic reflectance’s mean and standard deviation, DSM elevation’s mean and standard deviations (in meters), and PRI mean and standard deviation for each illumination class for morning (10:00 EST), midday (13:30 EST), and afternoon (15:30 EST).29	
Table 3.1. Suitable NEON field sites with descriptions including name, state, latitude, longitude, acquisition date, and dominant vegetation types.	52
Table 3.2. Description of physiological traits, their acronyms, name, and equation from NEON’s AOP hyperspectral products used in this study.	55
Table 3.3. Name and description of morphological and environmental traits from NEON’s AOP lidar products used in this study.....	56
Table 3.4. NEON remote sensing variables grouped into three trait classes – morphological and environmental (lidar data), psychological (hyperspectral data), and a combination of all functional traits (lidar and hyperspectral data).	57
Table 3.5. Multiple regression model results for each model group with number of terms used in the model, NEON’s remote sensing variables, AIC, BIC, R^2 , adjusted R^2 , and PRESS R^2 . Bolded rows indicate the model with the lowest AIC and BIC and highest PRESS R^2 for each model group and correspond to Figures 3.4-3.6.	60
Table 4.1. Forest metrics from lidar calculated in las tools and exported as raster images with 10 m spatial resolution and clipped to 420 m by 420 m bounding box.	83

Table 4.2. Suite of S2 multispectral indices calculated in SNAP’s Thematic Land Processor including the name abbreviation, full name, index description, and the equation, if applicable. NIR: near infrared; R = red; G = green; middle IR= middle infrared; R1 = red edge 1; R2 = red edge 3; R3 = red edge 3.	84
Table 4.3. Modeling groups and remote sensing variables used in each machine learning model group. The lidar group has 14 variables, the S2 group has 28 variables, and the combined has 42 variables.	91
Table 4.4. KS statistics showing which lidar and S2 multispectral indices were insignificant, significant in the positive direction (Prob >D+), or significant in the negative direction (Prob > D-).	92
Table 4.5. Confusion matrix for the lidar metrics only random forest classification model results.	92
Table 4.6. Confusion matrix for the S2 indices only random forest classification model results.	93
Table 4.7. Confusion matrix for the combined lidar and S2 random forest classification model results.	94

Chapter 1 - Introduction and Objectives

1.1 Introduction

1.1.1 Forest productivity

Plant photosynthesis is the main source of energy for all life on Earth (Demmig-Adams and Adams, 2000). Gross primary productivity (GPP) describes ecosystem photosynthesis by measuring the rate at which vegetation converts light into chemical energy (Gilabert et al., 2015). Since GPP plays an important, fundamental role in the global land carbon budget, accurate measurements of GPP in forests are critically needed to inform climate change mitigation strategies (Bellassen and Luyssaert, 2014; Moomaw et al., 2020). Because carbon uptake by forests can be reduced or increased by climate change, fires, drought, pathogens, pests, and management practices, it is crucial to analyze the health and productivity of forested ecosystems over time and space to reduce uncertainties in their role in the global carbon cycle.

Disturbances such as wildfires, diseases, insect infestations, and extreme weather events are significant forces that alter the health and functionality of forest ecosystems. Specifically, insect infestations cause significant damage and physiological stress to trees which can lead to reductions in photosynthesis, weakening the tree, and eventually leading to mortality. Pests and pathogens inducing tree mortality is a significant global issue, affecting forest health and ecosystem dynamics by altering the structure and function of forest environments (Meng et al., 2018).

1.1.2 Structure and function

Ecosystem structure and function provide insight on the dynamics of the the pools and fluxes of the carbon cycle of the ecosystem. Ecosystem structure for forests can be defined as the canopy

height, crown shapes, leaf area and orientation, etc., while the function for forests can be described by the rates of photosynthesis and respiration which are the major CO₂ fluxes of the system. Further analysis of the interactions of forest structure and function will enhance the understanding of forest health and productivity to effectively inform climate models (Muraoka and Koizumi, 2008).

1.1.3 Remote sensing

Lidar (Light Detection and Ranging) captures the vertical and horizontal complexities of forest environments by providing three-dimensional (3D) spatial arrangement of forest canopy height, density, and terrain. Spectral data, obtained from optical sensors, offers valuable insights into vegetation health, composition, and biophysical properties based on the way different wavelengths of light are absorbed and reflected (Campbell et al., 2023). When combined, these datasets offer a more holistic view of complex forested ecosystems by providing morphological and physiological status of vegetation (Dalponte et al, 2008; Fernández-Alonso et al., 2022). Lidar can be used to inform and refine spectral data interpretation by providing the structural context, while spectral data can complement lidar by revealing additional information about vegetation types and conditions that may affect the observed structural characteristics (Huang et al., 2019; Matasci et al., 2018).

Spectroscopy data acquired by satellite remote sensing, which quantifies the amount of reflected solar illumination, is a valuable tool for monitoring global ecosystems and has been used to estimate forest vitality, but the spatial resolution of tens to hundreds of meters is inadequate to understand how foliage intercepts and uses light for photosynthesis (Cheng et al., 2013; Gilabert et al., 2015; Goerner et al., 2011). High-resolution images at meter to sub-meter scales, as well as characterization of vegetation structure using lidar, provide the opportunity to

enhance understanding of forest light utilization in order to improve estimation of GPP.

1.2 Focus and Objectives by Chapter

The focus of this work is to establish an improved understanding of forest productivity in different forest compositions using a combination of physiological function from optical imagery and morphological structure from lidar.

Specific chapter objectives include the following:

Objective 1: The objective is to incorporate high-resolution multitemporal airborne hyperspectral imagery and lidar to assess the diurnal and vertical distributions of the Photochemical Reflectance Index (PRI) for different illumination classes in two managed pine stands with different ages and row orientation.

Objective 2: The objective of this study is to characterize the relationship between functional traits and forest productivity. Specifically, to determine how well GPP is predicted by variations in remotely quantified forest functional traits from different forest compositions across selected United States' NEON sites.

Objective 3: The objective of this study is to use machine learning classification to detect insect-induced tree mortality in a mixed deciduous stand with separated and combined models of lidar metrics from airborne laser scanning data and multispectral indices from Sentinel-2 MSI data.

References

- Bellassen, V., & Luyssaert, S. (2014). Managing forests in uncertain times. *Nature*, 506, 153–155.
- Campbell, J. B., Wynne, R. H., Thomas, V. A. (2023). *Introduction to remote sensing*. United Kingdom: Guilford Publications.
- Cheng, Y.-B., Middleton, E., Zhang, Q., Huemmrich, K., Campbell, P., Corp, L., Cook, B., et al. (2013). Integrating solar induced fluorescence and the photochemical reflectance index for estimating gross primary production in a cornfield. *Remote Sensing*, 5(12), 6857–6879.
- Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1416–1427.
- Demmig-Adams, B. (2003). Linking the xanthophyll cycle with thermal energy dissipation. *Photosynthesis Research*, 76(1–3), 73–80.
- Fernández-Alonso, J. M., Llorens, R., Sobrino, J. A., Ruiz-González, A. D., Alvarez-González, J. G., Vega, J. A., & Fernández, C. (2022). Exploring the Potential of Lidar and Sentinel-2 Data to Model the Post-Fire Structural Characteristics of Gorse Shrublands in NW Spain. *Remote Sensing*, 14(23), 6063.
- Gilabert, M. A., Moreno, A., Maselli, F., Martínez, B., Chiesi, M., Sánchez-Ruiz, S., García-Haro, F. J., et al. (2015). Daily GPP estimates in Mediterranean ecosystems by combining remote sensing and meteorological data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102, 184–197.

- Goerner, A., Reichstein, M., Tomelleri, E., Hanan, N., Rambal, S., Papale, D., Dragoni, D., et al. (2011). Remote sensing of ecosystem light use efficiency with MODIS-based PRI. *Biogeosciences*, 8(1), 189–202.
- Huang, C., Anderegg, W. R. L., & Asner, G. P. (2019). Remote sensing of forest die-off in the Anthropocene: From plant ecophysiology to canopy structure. *Remote Sensing of Environment*, 231, 111233.
- Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Zald, H. S. J. (2018). Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sensing of Environment*, 209, 90–106.
- Meng, R., Dennison, P. E., Zhao, F., Shendryk, I., Rickert, A., Hanavan, R. P., Cook, B. D., et al. (2018). Mapping canopy defoliation by herbivorous insects at the individual tree level using bi-temporal airborne imaging spectroscopy and LiDAR measurements. *Remote Sensing of Environment*, 215, 170–183.
- Moomaw, W. R., Law, B. E., & Goetz, S. J. (2020). Focus on the role of forests and soils in meeting climate change mitigation goals: summary. *Environmental Research Letters: ERL [Web site]*, 15, 045009.
- Muraoka, H., & Koizumi, H. (2009). Satellite Ecology (SATECO)-linking ecology, remote sensing and micrometeorology, from plot to regional scale, for the study of ecosystem structure and function. *Journal of Plant Research*, 122(1), 3–20.

Chapter 2 - Diurnal and vertical distributions of illumination and photochemical reflectance index in a loblolly pine plantation

Abstract

Remotely sensed data coupled with in situ eddy covariance measurements can enhance our understanding of plant physiological responses. The photochemical reflectance index (PRI) index is a narrow-band vegetation function that has been used in remote sensing as a proxy for photosynthetic light use efficiency (LUE), which is called out as the most critical parameter for improving gross primary productivity (GPP) estimates – i.e., forest productivity. This paper explores the effect of illumination on the diurnal and vertical distributions of PRI for two managed pine stands with different ages and row orientation using hyperspectral and lidar data. The hyperspectral imagery and lidar data were acquired from GLiHT (Goddard Lidar Hyperspectral and Thermal imager) during the FLEX Campaign in late October 2013. The flux tower data was acquired from AmeriFlux's US-NC2 eddy covariance flux tower located in a loblolly pine plantation in Parker Tract, NC. We analyzed diurnal and vertical distributions of illumination and PRI in two pine plantation plots with different ages and row orientations. We developed a new method to classify illumination into sunlit, shaded, and mixed light portions of the canopy using the hyperspectral dataset and a simulated panchromatic band. We incorporated an NDVI threshold to focus on the vegetated pixels only to alleviate soil background effects. To decipher the vertical distributions of PRI, we used the provided lidar from the corresponding NS flights. We further investigated the effects of solar and sensor view angles on the PRI responses throughout the day, which showed no significant effect in this dataset. We conclude that PRI responses vary for sunlit and shaded foliage throughout the day producing different foliage efficiencies depending on the light condition and the structure of the forest. The pine crowns show increasing reflectance (more sunlit foliage) with height and an associated decrease in PRI, especially in the morning and afternoon. Overall, the midday flight shows an increase in illumination with a negative PRI shift for understory contributing to the total canopy PRI.

Keywords: forest, hyperspectral, lidar, light use efficiency, photochemical reflectance index

2.1 Introduction

Accurate measurements of carbon sequestration in forests are critically needed to inform climate change mitigation strategies (Bellassen and Luyssaert, 2014; Moomaw et al., 2020). Because carbon uptake by forests is influenced by climate, biotic factors, and management practices, it is crucial to analyze the health and productivity of forested ecosystems over time and space to reduce uncertainties in their role in the global carbon cycle.

Advances in airborne and satellite remote sensing have enabled regional to global estimates of forest gross primary productivity (GPP). However, there is substantial uncertainty in GPP estimations. Li et al. (2018), for example, compared annual GPP estimates from six published satellite-based models against eddy covariance flux tower data, the most accurate measurements of GPP, for deciduous forests in the eastern United States. The comparison of the GPP results were inconsistent and, in some cases, contradictory. The authors hypothesize that the primary reason for the poor performance is that the sometimes large within-biome variation of canopy structure (and thus light environment) and composition (and thus function and its diversity across space and time) is not accurately accounted for in the differing approaches parameterizing key forest attributes. They stated that improved accounting of light use efficiency (LUE), which quantifies the efficiency by which foliage uses light during photosynthesis, is the most important parameter for improving GPP estimates. This finding is consistent with results from the North American Carbon Program model data intercomparison project (Schaefer et al., 2012), in which improved estimates of canopy-scale LUE were shown to be a critical path toward reducing model underestimation of GPP in the summer growing season and overestimation of GPP in the other seasons.

Canopy structure mediates the light environment, by creating sunlit and shadowed foliage components, and thus influences the rates of photosynthesis throughout the forest. Sunlit foliage

during the growing season generally has low photosynthetic efficiency because the amount of available light exceeds the ability to fully use it in photosynthesis. With high temperatures, vapor pressure deficit (VPD), and incoming light, sunlit leaves will activate photoprotective mechanisms. In contrast, shaded leaves can be very efficient in using the available light (Chen et al., 2014, Fauset et al., 2017) but the overall rate of photosynthesis is generally lower. As a result, although sunlit leaves have lower efficiency at certain times of the day, they are the source of a large majority of GPP (Chen et al., 1999). The distribution of light in a forest is mediated not only by canopy structure but also by the orientation of the solar illumination and is therefore not static within a given day or season (Tortini et al., 2017). A better understanding of the influence of canopy structure and solar illumination on the rates of photosynthesis, across time and different forest types, will inform processes controlling terrestrial carbon cycling at regional, continental, and global scales. Monitoring the productivity of forested ecosystems is also necessary to make management decisions geared toward enhancing removal of atmospheric CO₂ and the economic value of timber products (Thomas et al., 2018).

The photochemical reflectance index (PRI) is a narrow-band vegetation function index that has been used in remote sensing as an indicator of photosynthetic light use efficiency in plants (Gamon et al., 1992; Gamon et al., 1997; Gamon et al., 2016). The PRI equation (Eq. 1) uses two-narrow bands within the green wavelength range of the spectrum that are sensitive to the changes in the xanthophyll cycle of a leaf. The 531 nm band is used as the signal wavelength and the 570 nm band as the reference wavelength.

$$PRI = \frac{(531\text{ nm} - 570\text{ nm})}{(531\text{ nm} + 570\text{ nm})} \quad (\text{Eq. 2.1})$$

PRI has been shown to demonstrate a diurnal signal that is complementary to the diurnal signal of photosynthetic LUE (Chen et al., 2020; Möttus et al., 2018, Zhang et al., 2016). PRI

expresses relative down-regulation of photosynthesis primarily induced by high light conditions and water stress activating the photoprotective mechanisms of the leaves (Hall et al., 2008; Peñuelas et al., 1994; Suárez et al., 2008). Furthermore, PRI is sensitive to a range of factors that affect the index response such as: time of day and year, sky conditions, soil background effects, variations in the angle and direction of illumination, and structural changes at leaf and canopy levels (Gamon et al., 1992; Gamon et al., 1997; Gamon et al., 2016; Gamon et al., 2023; Garbulsky et al., 2011; Suarez et al., 2008; Yang, 2022; Zhang et al., 2016). Gamon et al. 2023 details the importance of analyzing diurnal imagery to better understand the PRI-light response, but this data is not yet available at the satellite-scale and is rarely obtained on an airborne basis. A full understanding of the effects of sun and sensor view geometry and stand structure are crucially needed to interpret the LUE-PRI relationship (Cheng et al., 2009; Garbulsky et al., 2011; Hilker et al., 2008).

This study looks at diurnal PRI with variable illumination conditions and structural complexities of forest stands to disentangle the daily functional changes mediated by the forest structure and changing light conditions. The objective is to incorporate high-resolution multitemporal airborne hyperspectral imagery and lidar to assess the diurnal and vertical distributions of PRI for different illumination classes in two managed pine stands with different ages and row orientation. This research enhances our understanding of forest light utilization throughout the day by the use of diurnal airborne remote sensing observations.

2.2 Materials and Methods

2.2.1 Study area

The study area is located in the lower coastal plains of North Carolina in Parker Tract. The forest composition in this area is a mix of deciduous broadleaf forest and managed loblolly pine (*Pinus taeda* L.). Our study focuses on the AmeriFlux tower site US-NC2 in a commercially managed mid-rotation loblolly pine plantation with a North-South (NS) row orientation established in 1992. In 2009, a combination of management practices was implemented on this stand by thinning every 5th row and selective thinning focused on removal of poorly formed or defective trees (reducing basal area more than 50%). The understory present in the thinned rows of this stand contains young red maple (*Acer rubrum* L.), devil's walking stick (*Aralia spinosa* L.), pokeweed (*Phytolacca americana*), beautyberry (*Callicarpa americana*), giant cane (*Arundinaria macrosperma*), and Meadow grass (*Poa spp.*) (Domec et al. 2010).

A neighboring stand without a flux tower was chosen to provide another example of a diurnal signal for a pine plantation stand used to assess structure and composition influences on light use. The other loblolly pine plantation plot is north of the tower plot is older and denser with an East-West (EW) row orientation. Although the flux tower is not in close vicinity to this neighboring plot, it provides a wider distribution of the sensitivity of the diurnal and structural effect on PRI (Figure 2.1).

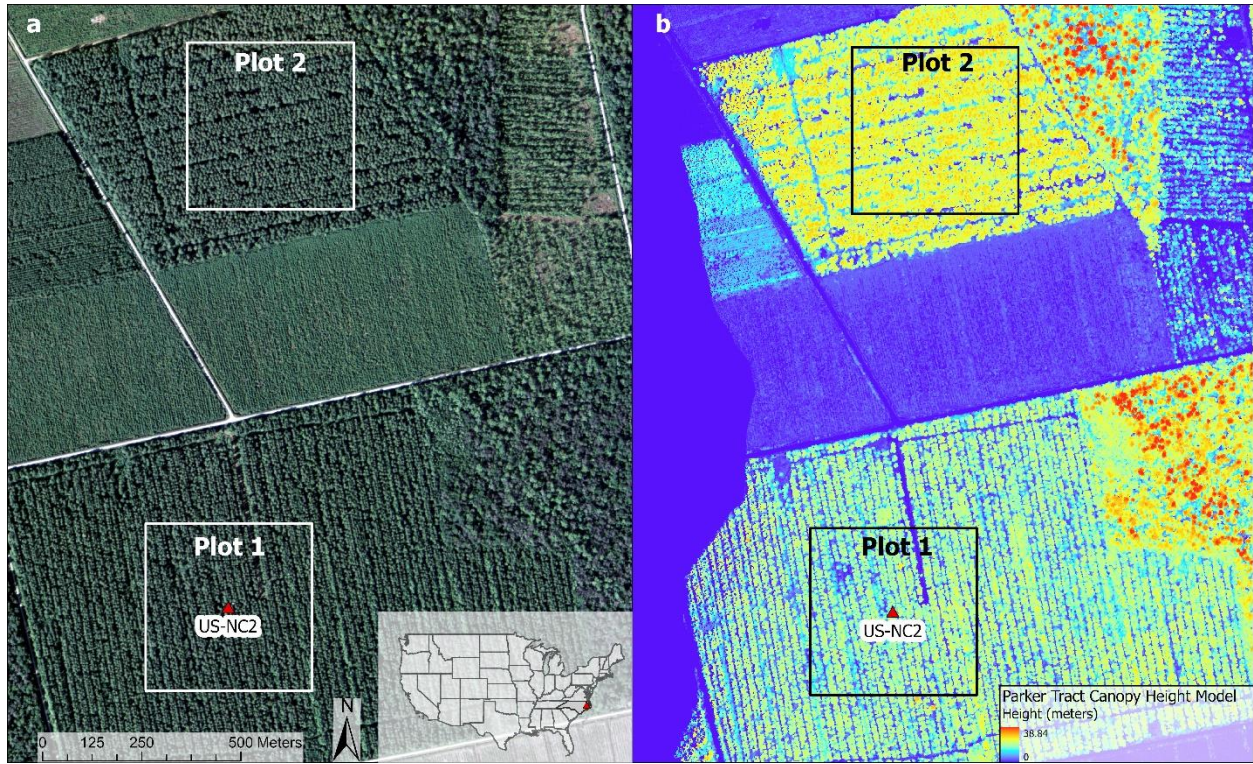


Figure 2.1. Study area sites in Parker Tract, NC shown with the 420 m x 420 m bounding boxes in the 20-year old loblolly pine stand centered on the US-NC2 flux tower (plot 1) and a more mature loblolly pine stand (plot 2) overlain on top of (a) high-resolution NAIP imagery from October 2014; (b) G-LiHT's 2 m lidar canopy height model from the October 26, 2013 midday flight (blue to red corresponds to increasing canopy height).

2.2.2 Datasets

This study incorporates high-resolution multitemporal airborne hyperspectral imagery and lidar from NASA's Goddard Lidar Hyperspectral and Thermal imager (G-LiHT). The hyperspectral imagery and lidar data were acquired during the G-LiHT's FLEX Campaign in late October 2013 over AmeriFlux's US-NC2 eddy covariance flux tower located in Parker Tract, NC. The hyperspectral dataset provides a means to calculate canopy-level PRI and to delineate into illumination classes and the lidar provides the vertical distribution of canopy-level PRI by illumination classes.

2.2.2.1 Canopy LiDAR and hyperspectral reflectance data

The Goddard Lidar, Hyperspectral and Thermal (G-LiHT) sensor suite provides aerial hyperspectral imagery and lidar data and is freely available on <https://glihtdata.gsfc.nasa.gov/> (Cook et al., 2013). The data for this study was acquired during the FLEX – US Airborne Campaign flights on October 26-27, 2013 during morning, midday, and afternoon in Parker Tract, NC (Middleton et al., 2017). Flights were flown NS and Solar Principal Plane (SPP) (diagonally to North South). Due to the orientation of some flight lines, not all flights covered both forest plots.

G-LiHT's hyperspectral imagery data products have been calculated to surface reflectance at 2 m resolution with 114 bands at 5 nm band widths between 420 nm and 950 nm (Cook et al., 2013). The lidar data is collected simultaneously with discrete-return lidar acquisitions with processed products for canopy height and digital surface models at 2 m resolution for structural measurements. G-LiHT also provides ancillary data products along with the hyperspectral and lidar data, which includes day of the year, aircraft elevation (m), aircraft altitude AGL (m), modeled clear-sky irradiance (W/m^2), solar zenith angle (SZA, deg), solar azimuth angle (SAA, deg), observed incoming PAR (W/m^2), clearness index, view angle (VA, deg), view azimuth (VAZ, deg), swatch number, and GTL flag. Figure 2.2 depicts the solar zenith angle and solar azimuth angle for all the flights used in analysis.

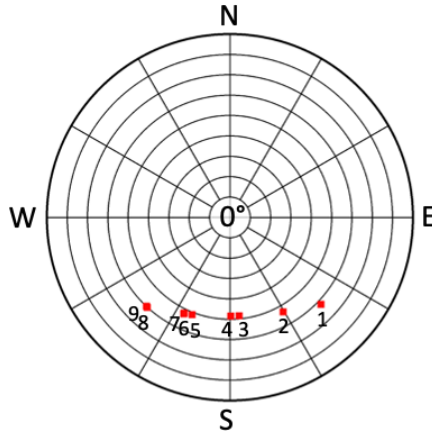


Figure 2.2. Solar zenith angle and solar azimuth angle for the nine G-LiHT flights during October 26-27, 2013. Flight numbers correspond to Tables 2.1 and 2.2.

2.2.2.2 Contextualization of airborne imagery with eddy covariance flux tower.

Eddy covariance techniques were used to measure carbon flux at the US-NC2 flux tower. The 23 m tall tower was established in 2005 in a loblolly pine stand planted at 35.8031°N, -76.66791°W (Noormets et al., 2010). Flux tower measurements reported every 30 minutes are available through a public website maintained by the AmeriFlux and NEON networks (<https://ameriflux.lbl.gov/>). Measurements of Net Ecosystem Production (NEP) from the flux tower were partitioned into GPP and ecosystem respiration using relationships developed between nighttime NPP and air temperature during the period of the flight campaign (Goulden et al., 1996). We computed LUE at half hourly intervals as the ratio of GPP to the incident photosynthetically active radiation (incident PAR) (Middleton et al., 2017). We associated the times of the G-LiHT flights to the corresponding flux measurement timing. G-LiHT's ancillary data product does not provide the exact time of each flight but does include solar zenith angle and solar azimuth in degrees. Time of day was derived from the NOAA sun calculator using the location, time zone, and the solar azimuth values extracted from G-LiHT's

ancillary data products for each flight for the different plots. Since the flux tower measurements are 30-minute averages, we connected the times of flights and flux tower measurements according to the time of flight within the 30-minute averaging window of the flux measurements. Incident PAR, GPP, and LUE were graphed for the two flight days with a red dot delineating the time of the G-LiHT acquisition during the flight campaign (Figure 2.3).

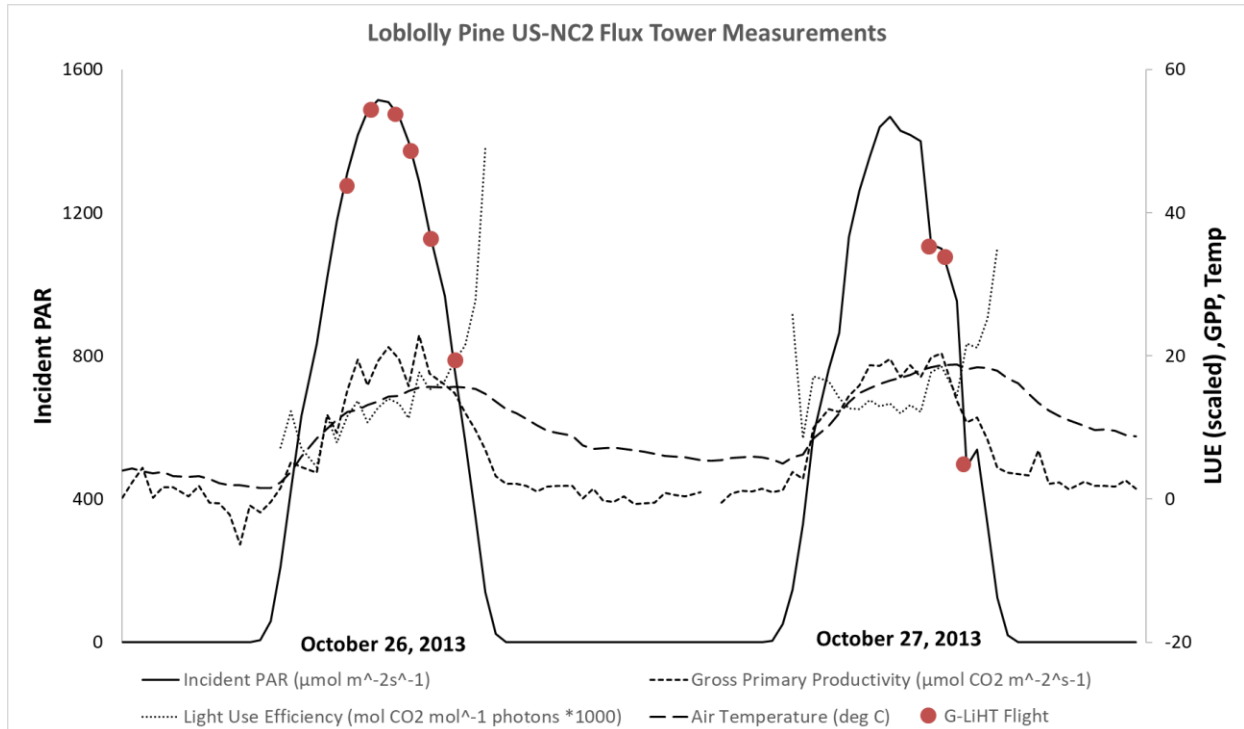


Figure 2.3. US-NC2 flux tower measurements showing incident PAR (solid line, units: $\mu\text{mol m}^{-2}\text{s}^{-1}$), GPP (dashed line, units: $\mu\text{mol CO}_2 \text{ m}^{-2}\text{s}^{-1}$), and LUE (dotted line, units: $\text{mol CO}_2 \text{ mol}^{-1} \text{ photons} * 1000$), and air temperature (long dashed line, units: degree Celsius) with corresponding G-LiHT flight acquisition times (red circles) from the FLEX – US Airborne Campaign in Parker Tract, NC.

2.2.3 Methods

Data processing on the canopy hyperspectral reflectance data was done in MATLAB 2021b (Figure 2.4), while lidar processing, graphs, and image visualization were done in ENVI, ArcGIS Pro 2.8, SAS JMP 16, Fortran 2018, and Microsoft Excel. Images were first clipped to the three plots with a 420 m x 420 m bounding box. The bounding box size was chosen due to the coverage of flux tower measurements (Chu et al., 2021). For the plot without a flux tower, we located the center of the stand and constructed the bounding box around the center location.

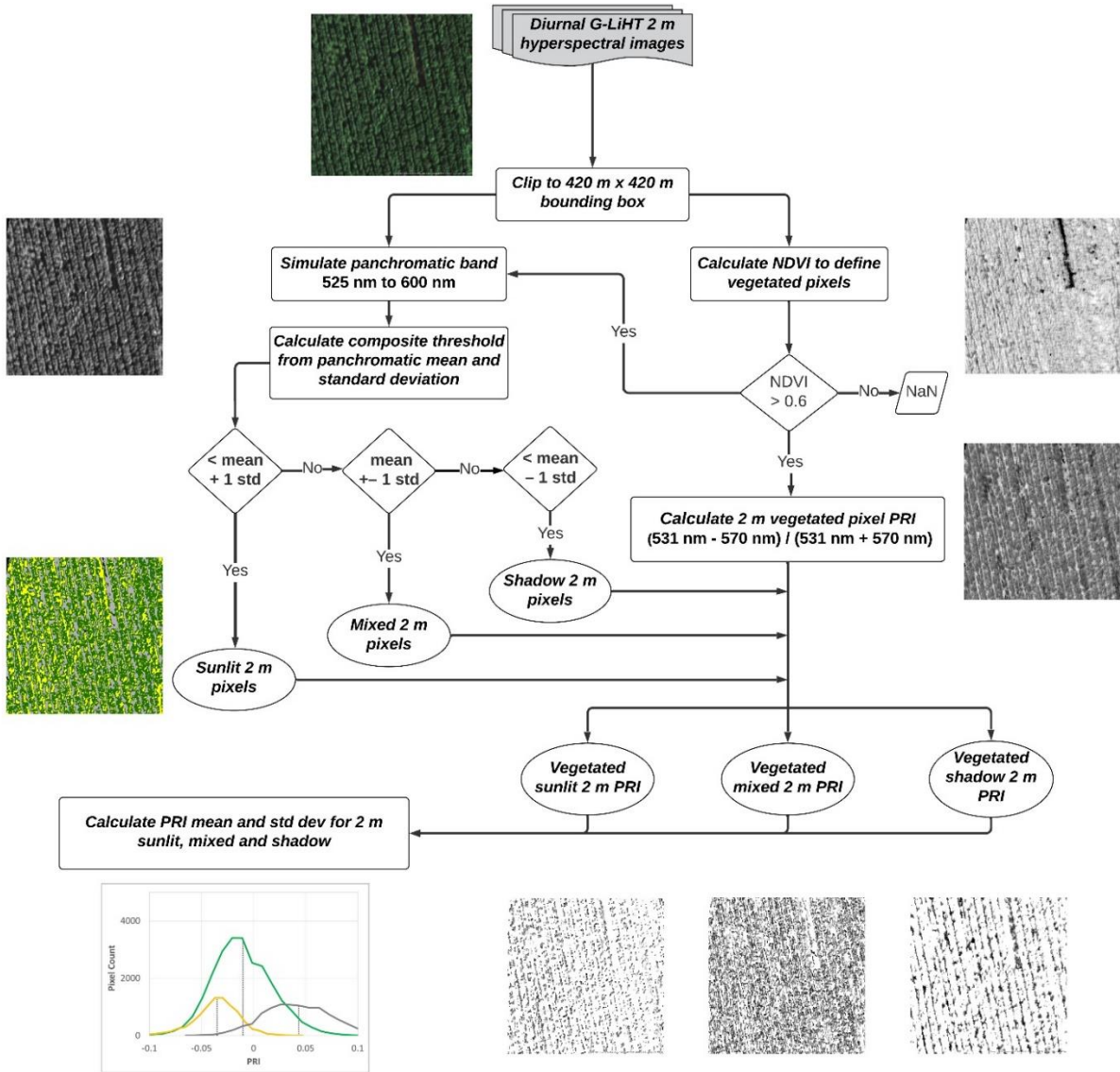


Figure 2.4. Workflow diagram of the sequential processing done in MATLAB on the canopy hyperspectral reflectance data executed for the flux tower plot that was extended to the other plot. Example of AM NS flight image results along with workflow include the following: G-LiHT RGB natural color image; NDVI threshold with the red pixels depicting a value below 0.6; panchromatic image scaled to 0-600; illumination classification with sunlit as yellow, mixed as green, and shadow as purple; vegetated PRI image scaled to -0.1 – 0.1 for each illumination classification (sunlit, mixed, and shadow) scaled to -0.1 – 0.1. Each image is 420 m by 420 m.

2.2.3.1 Delineation of vegetated pixels

A Normalized Difference Vegetation Index (NDVI) was implemented to detect vegetated pixels to alleviate any non-vegetated pixels (ex. ground or water) present in the images. We used the NDVI supplied in G-LiHT's data products which uses the thematic mapper Landsat bands (Tucker, 1979). A threshold of an NDVI value greater than 0.6 was considered vegetated, and values below the threshold set to NaN. This threshold was tested visually across all forest plots with different composition. This visual inspection step was used to improve the location of the threshold in lieu of using a method that corrects for non-photosynthetic scene components such as $fAPAR_{chl}$ (Zhang et al. 2014).

2.2.3.2 Panchromatic band simulation for illumination classification

To dictate the type of light environment of the foliage, we developed a panchromatic band simulation by averaging G-LiHT spectral bands over the range 525 nm to 600 nm at 2 m resolution. The selected range of wavelengths included 18 G-LiHT bands (bands 25-42, 525 nm – 600 nm). The statistical methods for separation of light environments were implemented with quartile distributions using a plot specific composite threshold from all flights. The panchromatic band was classified into sunlit, mixed, and shadow based on the mean and standard deviation of reflectance for each plot. The delineation of each illumination class is specific to each plot's composite threshold from the panchromatic brightness image statistics of all flights combined. Sunlit pixels have reflectance greater than the mean + 1 sigma, mixed between ± 1 sigma, and shadow less than the mean – 1 sigma. At-sensor reflectance mean and standard deviations were calculated from the panchromatic image reflectance.

For this study, the illumination classes are defined as follows: sunlit is directly illuminated foliage by photons coming from the sun; mixed is indirectly illuminated by photons that were reflected inward and downward by sunlit foliage; and shadow is low intensity indirectly illuminated foliage, typically deep within the canopy.

Image results for the panchromatic band and illumination classification outputs were compiled, scaled, and exported for each plot using ArcGIS Pro. Panchromatic image brightnesses were scaled to 0-600. The illumination classification image is separated into three classes with a corresponding color: mixed as green, shadow as purple, and sunlit as yellow.

2.2.3.3 Photochemical reflectance index

PRI was calculated on all flights at 2 m resolution using the 531 nm and 570 nm bands in the green wavelengths range of the spectrum (Eq. 2.1) (Gamon, 1997). For each PRI band, two of G-LiHT's bands were averaged to assure the reflectance bands are closest to the formula requirement for 531 nm (bands 25 and 26) and 570 nm (bands 36 and 37).

The vegetated PRI images were then coupled with a “mask” for each illumination class from the panchromatic band simulation. The mask multiplies the illumination classification image from the panchromatic band simulation times the vegetated PRI image. Minimum, maximum, mean, and standard deviation for vegetated sunlit, mixed, and shadow PRI 2 m image outputs were calculated for each flight for both plots. Image results for the PRI outputs were compiled, scaled, and exported for each plot using ArcGIS Pro. PRI image results were scaled to the range of -.1 to .1 for visualization.

2.2.3.4 Canopy structure of diurnal PRI

To assess how the incoming light over the course of the day is mediated by the structure of the canopy, we coupled the diurnal PRI results from the NS flights with the corresponding lidar acquisitions provided in the G-LiHT product suite (Figures 2.9-2.10). Considering the NS flights cover both plots, these flights were used for visualization. From the lidar data, we have the canopy height model (CHM), digital surface model (DSM) and digital terrain model (DTM) for each plot as image rasters at 2 m resolution.

In addition to the PRI and illumination classification output images, we enhanced to a 3D visualization using lidar point clouds. The lidar point clouds were used to create 3D images of the CHM, the PRI, and illumination classification results from the NS flights for each plot (Figures 2.12-2.13). The CHM classifies the point clouds based on heights above ground. In ArcGIS Pro, the illumination classification was overwritten onto the point clouds by recoding the classes. A similar recoding was done with the PRI image classification on the point clouds but required classifying the PRI values into 5 classes based on the quartile distribution of one flight and applied to the other flights for consistency. An important note is that the point cloud density is approximately 5 points per square meter, and the illumination and PRI classifications images are at a 2 m resolution, so multiple points fall within each 2 m pixel.

2.2.3.5 Calculation of intersection angle on solar and sensor view vectors

To look further into the solar and sensor view angle effects on the diurnal PRI responses, we calculated the intersection angle between the solar vector and sensor view vector. To calculate the intersection angle, we derived the single swath ancillary data products solar zenith angle, solar azimuth angle, view angle, and view azimuth. We used an equation to find the angle

between two points given their azimuth and elevation (Eq. 2.2) (Stack Exchange User 1887118 (nalyd88), 2023).

$$\theta_{u,v} = \cos^{-1}[\cos El_1 \cos El_2 \cos(Az_1 - Az_2) + \sin El_1 \sin El_2] \quad (\text{Eq. 2.2})$$

Where $\theta_{u,v}$ is the intersection angle, El_1 is defined as $90 - \text{solar zenith angle}$; El_2 as $90 - \text{view angle}$; Az_1 is as the solar azimuth; and Az_2 as view azimuth + 180 if the sum is greater than 360, or view azimuth + 180 if the sum is greater than 360. Each ancillary data image was clipped to the plot's boundary and the intersection angle image was calculated in FORTRAN 2018. Images for the resulting intersection angle and PRI were loaded into ENVI 5.7 to create graphs for pixel comparisons. We also compared to NDVI, which does not show a diurnal trend.

2.3 Results

2.3.1 Illumination classification of PRI

Availability of each plot's G-LiHT acquisition with corresponding flight name, time, day, and sun angles were organized into tables for each plot along with PRI mean and standard deviation results for the different illumination classes (Tables 2.1-2.2). Bolded rows indicate a different flight day and numbers correspond with Figure 2.2.

Table 2.1. Information for G-LiHT flights over plot 1 (US-NC2 loblolly pine plantation) including flight name, time (EST) and day, solar zenith angle (zen) and solar azimuth angle (az) (in degrees), and PRI mean and standard deviations for all pixels and the three illumination classes – shadowed, mixed, and sunlit. Bolded rows indicate flights on the 27th.

G-LiHT Name	Time	Day	Sun angles,		PRI							
			deg		All		Shadow		Mixed		Sunlit	
			Zen	Az	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
1. AM NS	10:00	26	61.9	133.6	0.001	0.032	0.035	0.029	-0.005	0.025	-0.024	0.018
2. AM SPP2	11:30	26	53.4	150.5	-0.007	0.026	0.026	0.025	-0.010	0.020	-0.030	0.016
3. Noon SPP1	13:00	26	48.6	174.5	-0.014	0.024	0.014	0.024	-0.016	0.019	-0.042	0.016
4. Noon NS	13:30	26	48.5	179.5	-0.017	0.026	0.018	0.027	-0.017	0.022	-0.035	0.018
5. Noon SPP1	14:00	27	51.3	201.2	-0.008	0.024	0.027	0.020	-0.009	0.019	-0.029	0.016
6. Noon SPP2	14:30	26	52.1	205.4	-0.011	0.025	0.015	0.025	-0.016	0.019	-0.035	0.017
7. Noon NS	14:30	27	52.6	205.6	-0.009	0.027	0.028	0.022	-0.011	0.022	-0.029	0.018
8. PM NS	15:30	26	60.0	223.2	-0.001	0.031	0.034	0.028	-0.005	0.025	-0.023	0.020
9. Noon SPP2	15:30	27	60.0	222.5	-0.011	0.023	0.014	0.020	-0.016	0.018	-0.039	0.016

Table 2.2. Information for G-LiHT flights over plot 2 (older loblolly pine plantation) including flight name, time (EST) and day, solar zenith angle (zen) and solar azimuth angle (az) (in degrees), and PRI mean and standard deviations for all pixels and the three illumination classes – shadowed, mixed, and sunlit. Bolded rows indicate flights on the 27th.

G-LiHT Name	Time	Day	Sun angles,		PRI							
			deg		All		Shadow		Mixed		Sunlit	
			Zen	Az	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
1. AM NS	10:00	26	61.5	134.3	0.002	0.031	0.034	0.029	-0.004	0.024	-0.021	0.015
3. Noon SPP1	13:00	26	48.6	174.4	-0.007	0.022	0.018	0.025	-0.012	0.016	-0.026	0.011
4. Noon NS	13:30	26	48.5	180.3	-0.011	0.024	0.018	0.027	-0.014	0.020	-0.026	0.013
5. Noon SPP1	14:00	27	51.3	201.1	-0.006	0.021	0.026	0.020	-0.008	0.016	-0.024	0.011
6. Noon SPP2	14:30	26	52.1	205.4	-0.008	0.022	0.016	0.025	-0.014	0.016	-0.027	0.011
7. Noon NS	14:30	27	53.0	206.8	-0.008	0.024	0.025	0.022	-0.008	0.021	-0.024	0.014
8. PM NS	15:30	26	60.6	224.1	0.002	0.031	0.035	0.029	-0.003	0.026	-0.019	0.017

Figures 2.5 and 2.6 provide a visualization of the diurnal distributions of the 2 m PRI mean values for the two flight days. The bars display the PRI mean results for all pixels (red) and each illumination class (sunlit as yellow, mixed as green, and shadow as purple). On each bar there is a percentage value for the percent of pixels in each illumination class. The flight days are separated by a black bar. The diurnal effect of PRI mean values is evident on October 26th in each plot but is not as clear on the following day October 27th.

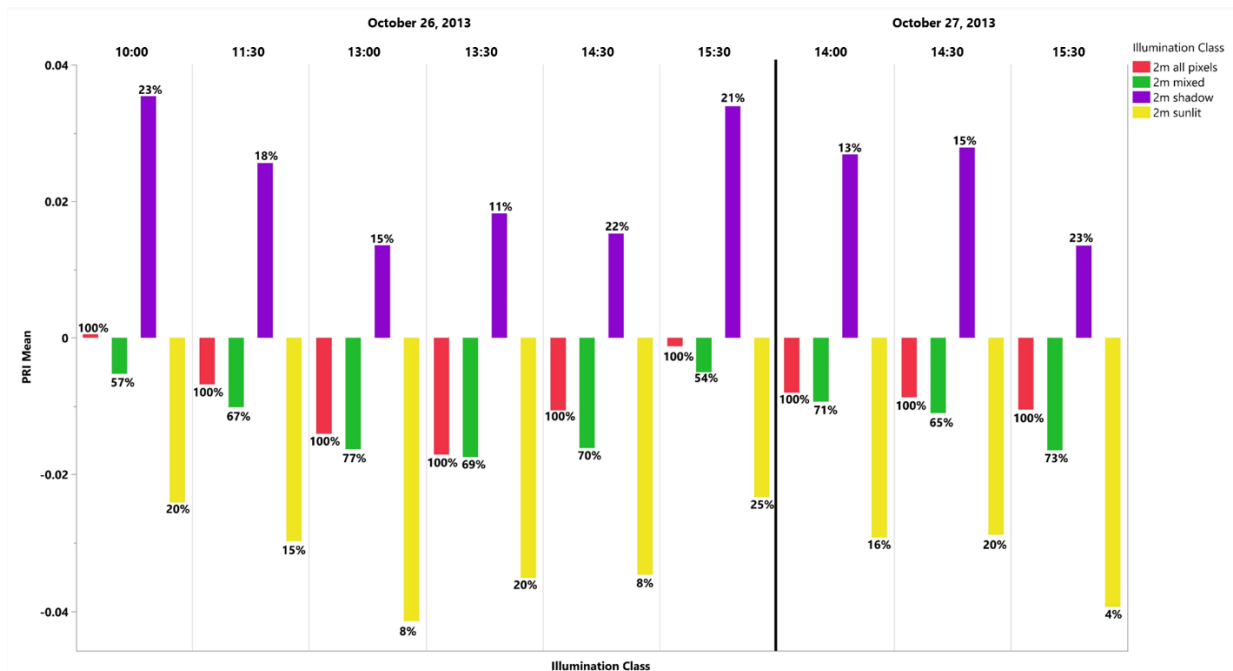


Figure 2.5. Graph of Plot 1's (US-NC2) 2 m PRI mean values over the two flight days grouped by time of day and the illumination class. Percentages on each bar indicate the percent of pixels within the illumination class.

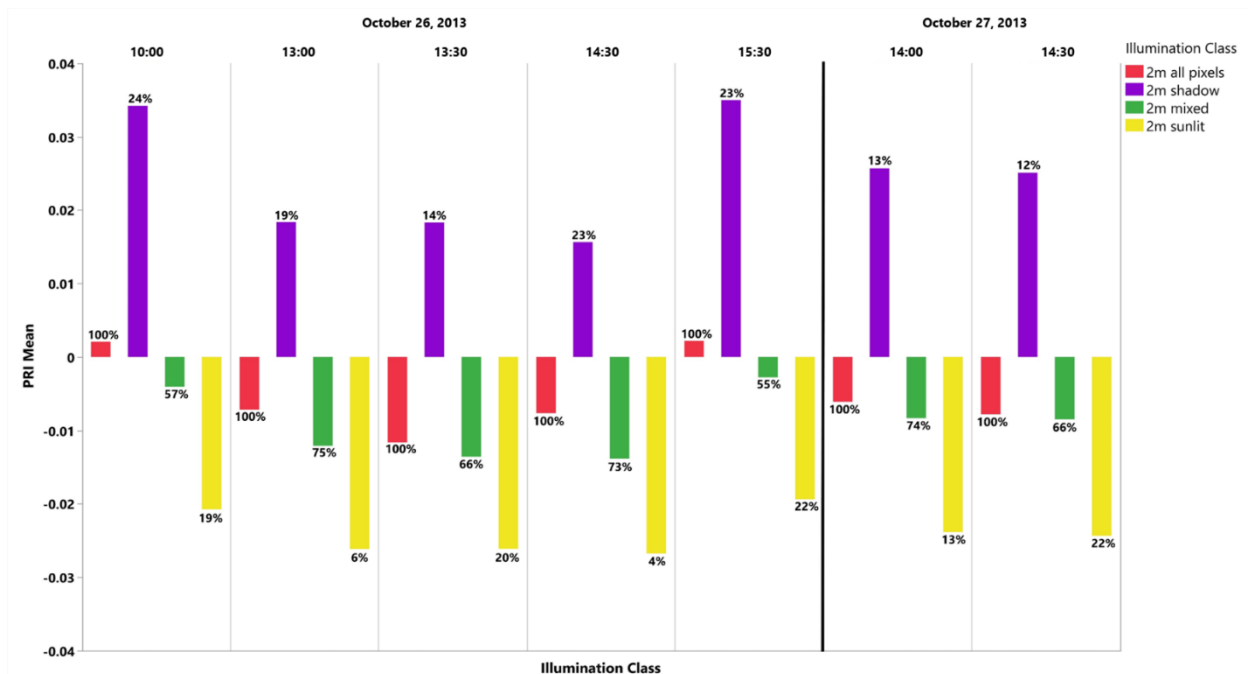


Figure 2.6. Graph of Plot 2's (older pine) 2 m PRI mean values over the two flight days grouped by time of day and the illumination class. Percentages on each bar indicate the percent of pixels within the illumination class.

Figures 2.7 and 2.8 show the output images for each flight and specific plots include the simulated panchromatic band (first row), PRI output for all pixels (second row), and the illumination classification image (third row). The black areas on the left and right sides of the images are missing data values due to the swath of the flights over the plots.

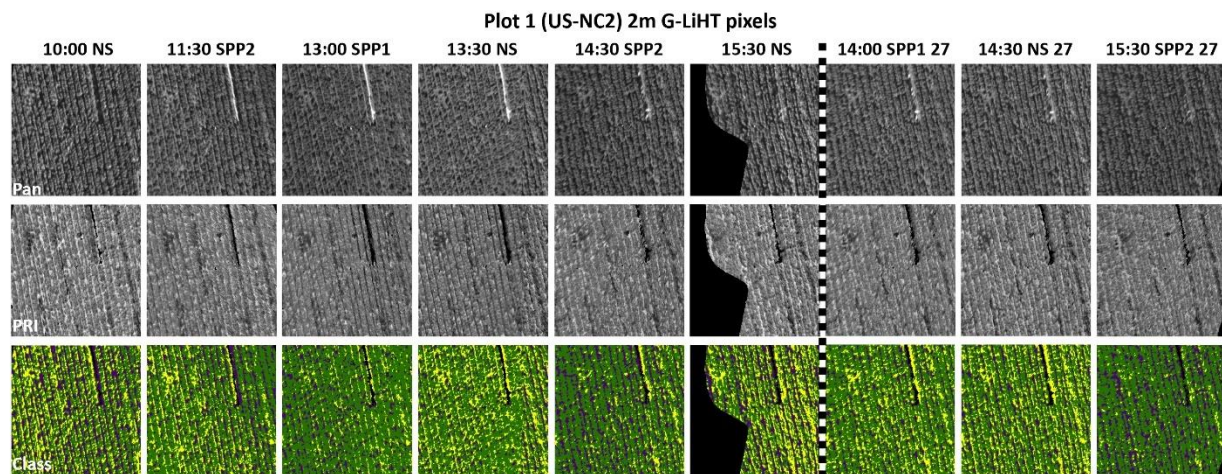


Figure 2.7. Plot 1 (US-NC2 flux tower) image output results at 2 m resolution from the canopy hyperspectral processing shown in order of time across the two days. First row: simulated panchromatic band. Second Row: PRI images scaled to -0.1 - 0.1. Third row: Illumination classification with shadow as purple, mixed as green, and sunlit as yellow.

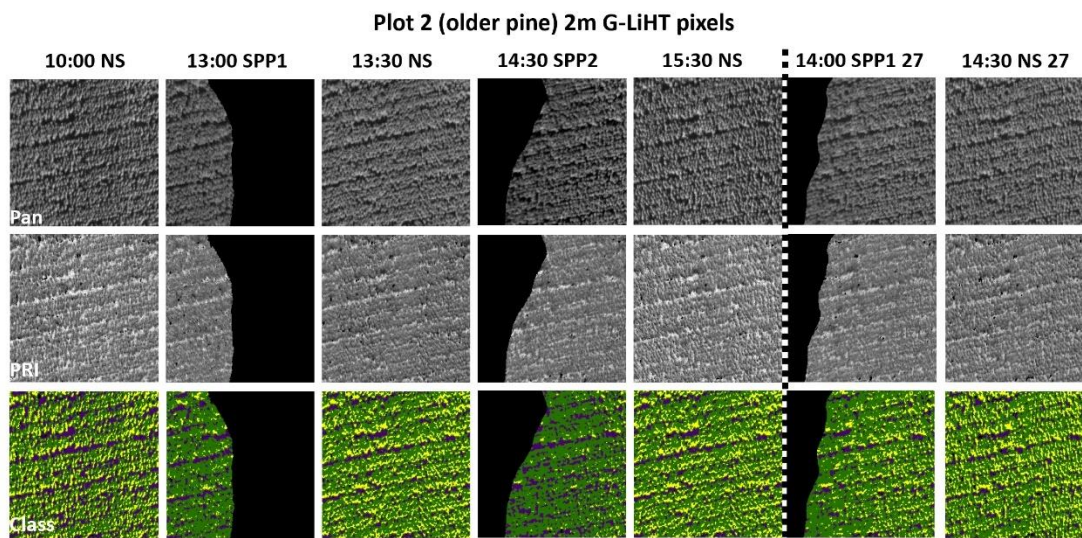


Figure 2.8. Plot 2 (older pine) image output results at 2 m resolution from the canopy hyperspectral processing shown in order of time across the two days. First row: simulated panchromatic band. Second Row: PRI images scaled to -0.1 - 0.1. Third row: Illumination classification with shadow as purple, mixed as green, and sunlit as yellow.

2.3.2 Reflectance and PRI as a function of height, stand age, and composition

Scatterplots of 2 m DSM elevations versus panchromatic reflectance and PRI for the three illumination classes reveal structural controls on the canopy light environment (Figures 2.9 and 2.10). NS flights were used for visualizations because the swath of the data covered both forest plots. Colors indicate the density of the points. Complimentary statistics to Figure 2.9 for plot 1 are shown in Table 2.3 and Figure 2.10 for plot 2 are shown in Table 2.4. The statistics are separated into height classes using the DSM elevation to delineate the overstory (13 to 32 meters) and the understory (6 to less than 13 meters) for the three illumination classes (shadowed, mixed, and sunlit) with the percent of total pixels (%), percent of panchromatic reflectance's mean and standard deviation, DSM elevation's mean and standard deviations (in meters), and PRI mean and standard deviation for each illumination class for morning (10:00 EST), midday (13:30 EST), and afternoon (15:30 EST).

In Figure 2.9 and Table 2.3 for plot 1, the pine crowns show increasing reflectance (more sunlit foliage) with height and an associated decrease in PRI, especially in the morning and afternoon. A deciduous understory filling in thinned rows is mostly in shadow in the morning and afternoon but becomes more sunlit during midday and its PRI decreases by a greater amount than the pine canopy in the overstory. Overall, the midday flight shows an increase in illumination with a negative PRI shift for understory contributing to the total canopy PRI.

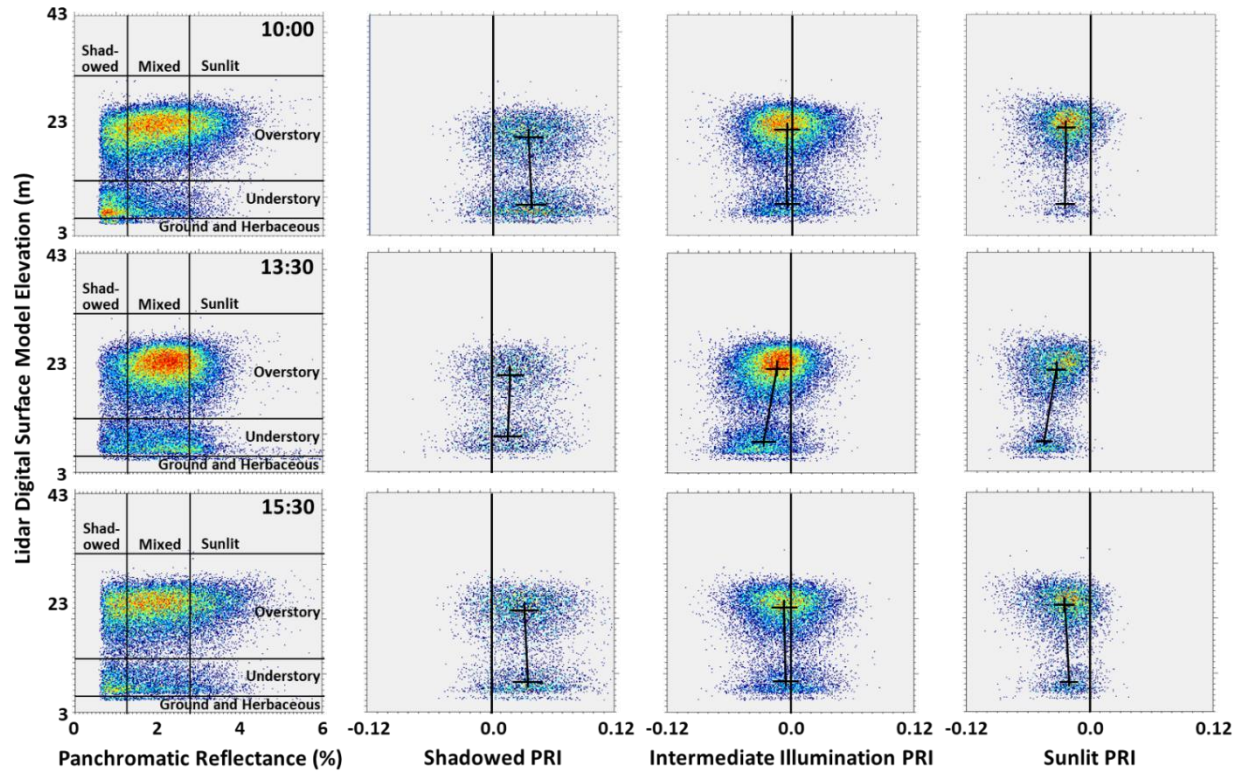


Figure 2.9. Plot 1 (US-NC2 flux tower) scatterplot of DSM versus panchromatic reflectance % (first column), shadowed PRI (second column), mixed PRI (third column), and sunlit PRI (fourth column) for the NS flights on October 26th for the morning (first row), midday (second row), and evening (third row). The black plus signs are shown to connect the PRI mean of the understory and overstory to show how the light distribution and PRI responses changes within the canopy throughout the day.

Table 2.3. Complimentary statistics to Figure 2.9 for plot 1 using the DSM elevation to delineate the overstory (13 to 32 meters) and the understory (6 to less than 13 meters) providing statistics for the three illumination classes (shadowed, mixed, and sunlit) with the percent of total pixels (%), percent of panchromatic reflectance's mean and standard deviation, DSM elevation's mean and standard deviations (in meters), and PRI mean and standard deviation for each illumination class for morning (10:00 EST), midday (13:30 EST), and afternoon (15:30 EST).

DSM class and elevation (m)	Illumination class	Morning (10:00 EST)						
		% of 43,305 pixels	Pan Reflectance (%)		DSM Elevation (m)		PRI	
			Mean	St Dev	Mean	St Dev	Mean	St. Dev.
Overstory 13 to 32	Shadowed	13.7	1.0	0.2	20.8	3.0	0.034	0.028
	Mixed	48.6	2.0	0.4	22.1	2.9	-0.005	0.025
	Sunlit	18.8	3.2	0.4	22.7	2.7	-0.024	0.018
Understory 6 to <13	Shadowed	9.4	0.9	0.2	8.6	1.6	0.038	0.030
	Mixed	8.2	1.8	0.4	8.8	1.7	-0.005	0.025
	Sunlit	1.3	3.1	0.3	8.9	1.9	-0.025	0.019
DSM class and elevation (m)	Illumination class	Midday (13:30 EST)						
		% of 43,218 pixels	Pan Reflectance (%)		DSM Elevation (m)		PRI	
			Mean	St Dev	Mean	St Dev	Mean	St. Dev.
Overstory 13 to 32	Shadowed	7.1	1.0	0.2	20.7	3.2	0.019	0.026
	Mixed	57.2	2.0	0.4	22.0	2.8	-0.015	0.022
	Sunlit	16.6	3.1	0.3	21.9	3.0	-0.033	0.019
Understory 6 to <13	Shadowed	4.0	0.9	0.2	9.1	1.7	0.017	0.028
	Mixed	11.5	2.0	0.4	8.8	1.7	-0.028	0.023
	Sunlit	3.7	3.0	0.3	8.4	1.6	-0.043	0.016
DSM class and elevation (m)	Illumination class	Afternoon (15:30 EST)						
		% of 31,632 pixels	Pan Reflectance (%)		DSM Elevation (m)		PRI	
			Mean	St Dev	Mean	St Dev	Mean	St. Dev.
Overstory 13 to 32	Shadowed	14.6	1.0	0.2	21.5	3.0	0.033	0.028
	Mixed	44.9	2.0	0.3	22.1	2.8	-0.006	0.025
	Sunlit	20.9	3.4	0.5	22.5	2.8	-0.024	0.020
Understory 6 to <13	Shadowed	6.7	0.9	0.2	8.4	1.8	0.036	0.029
	Mixed	9.3	1.9	0.4	8.6	1.8	-0.003	0.025
	Sunlit	3.6	3.3	0.4	8.4	1.7	-0.020	0.018

In Figure 2.10 and Table 2.4, the height of plot 2 is greater, but the depth of the pine crown overstory has decreased. The understory is mostly in shadow due to crown cover, because this is not a thinned stand. Similar to plot 1, we see an increase in illumination with more sunlit and mixed pixels and a shift to lower PRI at midday. Comparing to plot 1, plot 2 is older, taller, and denser with minimal understory due to no management regimes for thinning.

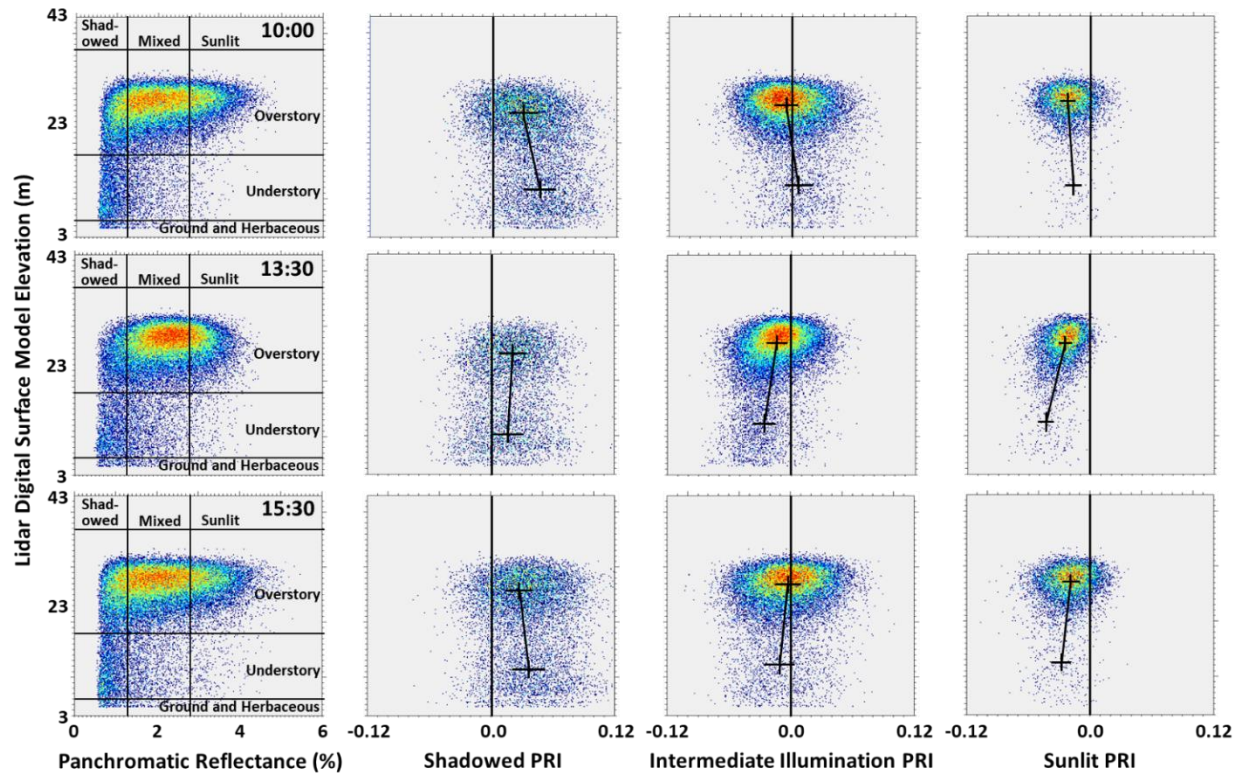


Figure 2.10. Plot 2 (older pine stand) scatterplot of DSM versus panchromatic reflectance % (first column), shadowed PRI (second column), mixed PRI (third column), and sunlit PRI (fourth column) for the NS flights on October 26th for the morning (first row), midday (second row), and evening (third row).

Table 2.4. Complimentary statistics to Figure 2.10 for plot 2 using the DSM elevation to delineate the overstory (13 to 32 meters) and the understory (6 to less than 13 meters) providing statistics for the three illumination classes (shadowed, mixed, and sunlit) with the percent of total pixels (%), percent of panchromatic reflectance's mean and standard deviation, DSM elevation's mean and standard deviations (in meters), and PRI mean and standard deviation for each illumination class for morning (10:00 EST), midday (13:30 EST), and afternoon (15:30 EST).

DSM class and elevation (m)	Illumination class	Morning (10:00 EST)						
		% of 43,305 pixels	Pan Reflectance (%)		DSM Elevation (m)		PRI	
			Mean	St Dev	Mean	St Dev	Mean	St. Dev.
Overstory 13 to 32	Shadowed	17.5	1.0	0.2	25.3	3.1	0.030	0.028
	Mixed	54.3	2.0	0.4	27.0	2.6	-0.005	0.023
	Sunlit	18.6	3.3	0.4	27.7	2.3	-0.021	0.015
Understory 6 to <13	Shadowed	6.4	0.9	0.2	11.6	3.3	0.046	0.031
	Mixed	2.8	1.8	0.4	12.2	3.4	0.007	0.027
	Sunlit	0.3	3.1	0.3	12.2	3.6	-0.017	0.016
DSM class and elevation (m)	Illumination class	Midday (13:30 EST)						
		% of 43,491 pixels	Pan Reflectance (%)		DSM Elevation (m)		PRI	
			Mean	St Dev	Mean	St Dev	Mean	St. Dev.
Overstory 13 to 32	Shadowed	8.6	1.0	0.2	24.8	3.0	0.020	0.026
	Mixed	61.9	2.1	0.4	27.1	2.6	-0.013	0.020
	Sunlit	19.8	3.2	0.3	27.0	2.6	-0.026	0.013
Understory 6 to <13	Shadowed	5.2	0.9	0.2	11.3	3.2	0.017	0.029
	Mixed	3.9	1.9	0.4	12.2	3.4	-0.028	0.022
	Sunlit	0.6	3.2	0.3	12.9	3.5	-0.041	0.015
DSM class and elevation (m)	Illumination class	Afternoon (15:30 EST)						
		% of 43,487 pixels	Pan Reflectance (%)		DSM Elevation (m)		PRI	
			Mean	St Dev	Mean	St Dev	Mean	St. Dev.
Overstory 13 to 32	Shadowed	16.7	1.0	0.2	25.7	3.1	0.033	0.028
	Mixed	52.4	2.0	0.4	27.0	2.7	-0.002	0.025
	Sunlit	21.3	3.4	0.4	27.3	2.4	-0.019	0.017
Understory 6 to <13	Shadowed	5.7	0.9	0.2	11.6	3.3	0.039	0.032
	Mixed	3.2	1.9	0.4	12.1	3.5	-0.010	0.029
	Sunlit	0.7	3.4	0.4	12.3	3.3	-0.029	0.019

In Figure 2.11, the NS evening flight had missing spectral data on the left side, thus the pixels are black. The structure of the illumination classification changes throughout the day, with more shadowed foliage in the morning and evening, while the number of mixed and sunlit pixels increase during midday. Thinning of this stand controlled how the shadows change throughout the day due to differences in the height of the planted loblolly pine trees and deciduous tree undergrowth. This is complemented in the structure of the PRI results where you see a decrease in efficiency is indicated by lower PRI values shown in red in the midday, while higher PRI values in the morning and evening indicate an increase in foliage efficiency. In the midday structure PRI result, the red points seem to occur on the top and left side of the pine trees, which is a function of the angle of the illumination inducing stress on the upper sunlit leaves of the canopy.

In Figure 2.12, the structure of the pine plantation is an east to west orientation with denser trees. There are more prominent shadows in the morning and evening images, while we see an increase in mixed and sunlit pixels during the midday. This is complimented by the PRI results where we see more negative PRI values in the midday when the foliage is getting more sunlit and inducing stress. This pine plantation shows better efficiency in converting sunlight to photosynthetic efficiency in the morning and evening.

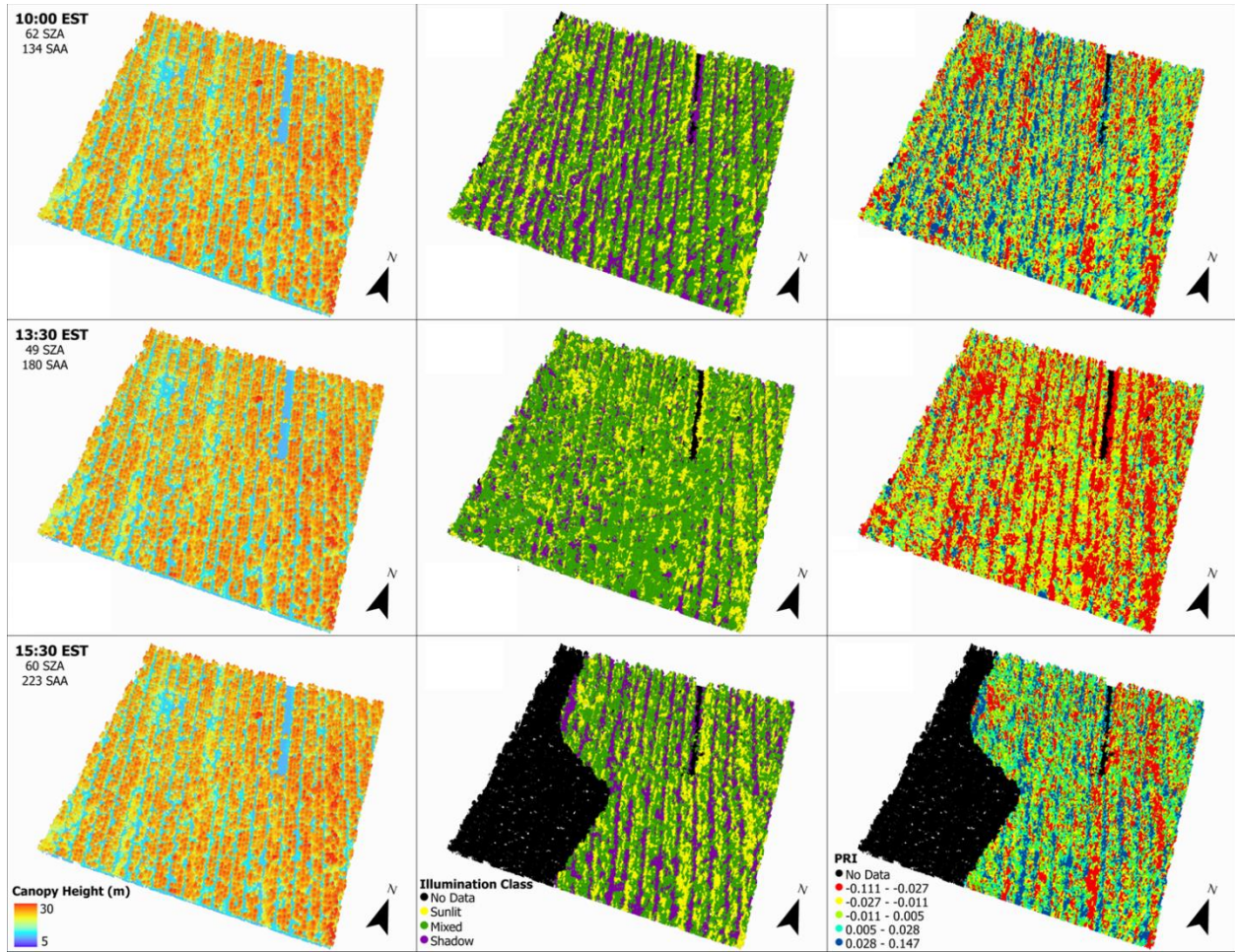


Figure 2.11. 3D structural results for plot 1 (US-NC2) of the CHM (first column, blue to red indicates increased height), illumination class (middle column, sunlit as yellow, mixed as green, and shadow as purple), and PRI (last column, low values in red to high values in blue) for the morning (first row), midday (second row), and evening (third row) NS flights (black points are from missing spectral data). Time of flights (EST), solar zenith angle (SZA, in degrees), and the solar azimuth angle (SAA, in degrees) are shown for each row.

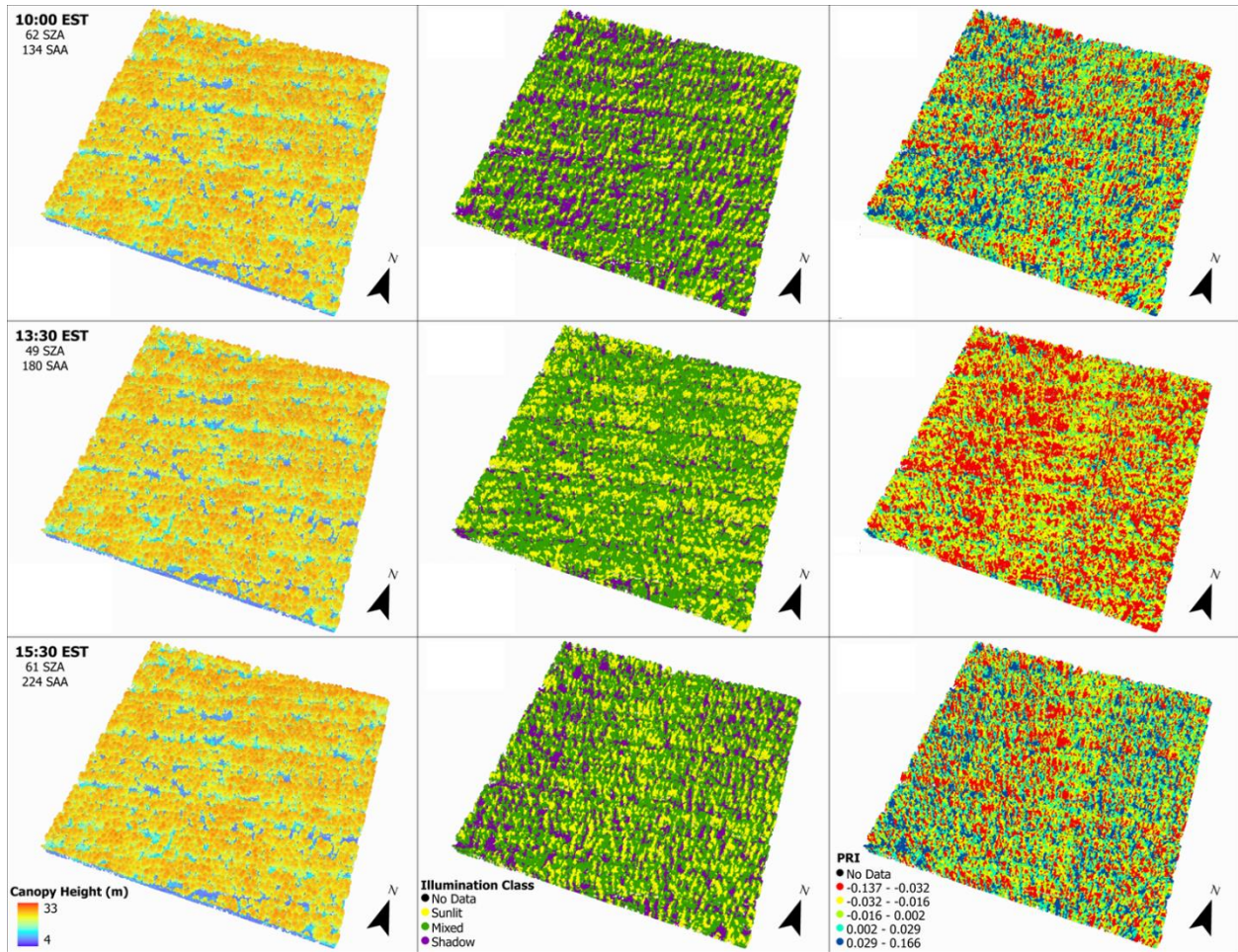


Figure 2.12. 3D structural results for plot 2 (older pine) of the CHM (first column, blue to red indicates increased height), illumination class (middle column, sunlit as yellow, mixed as green, and shadow as purple), and PRI (last column, low values in red to high values in blue) for the morning (first row), midday (second row), and evening (third row) NS flights. Time of flights (EST), solar zenith angle (SZA, in degrees), and the solar azimuth angle (SAA, in degrees) are shown for each row.

2.3.3 LUE and canopy-scale PRI by illumination class

The G-LiHT PRI results for all pixels in plot 1 with the US-NC2 flux tower on October 26th were plotted with flux tower LUE shown in a time series by the solar azimuth angle (Figure 2.13a). LUE has the typical daily J-shape as expected, although the PRI curve is U-shaped. PRI shows higher efficiency in the morning that drops midday then recovers in the afternoon, while LUE shows increasing efficiency throughout the day. We see this similar U-shape for PRI results when separating into illumination classes (Figure 2.13b).

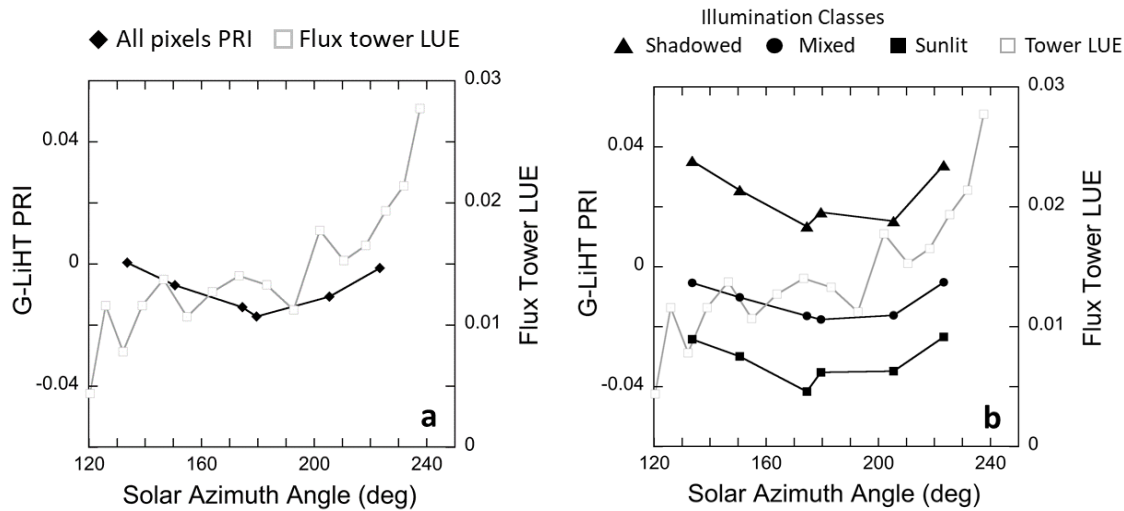


Figure 2.13. Graphs showing the diurnal relationship of LUE (grey line with squares) to PRI (black line with diamonds) as a function of solar azimuth angle for (a) all PRI values and (b) PRI values separated by illumination class (triangle as shadow, circle as mixed, and box as sunlit).

2.3.4 Effects of solar and sensor view angles on PRI

We investigated sun and view angle geometry effects on both PRI and NDVI using the intersection angle calculation. NDVI did not vary diurnally and showed no effect on the intersection angle (not shown). The two forest plots analyzed did not exhibit PRI sensitivity to sun and sensor view geometries (Figures 2.14 and 2.15). We fit a line to the graphs and the slope and R^2 for all times and both plots were insignificant. There is little to no sensitivity with the sun and sensor view geometry in this dataset.

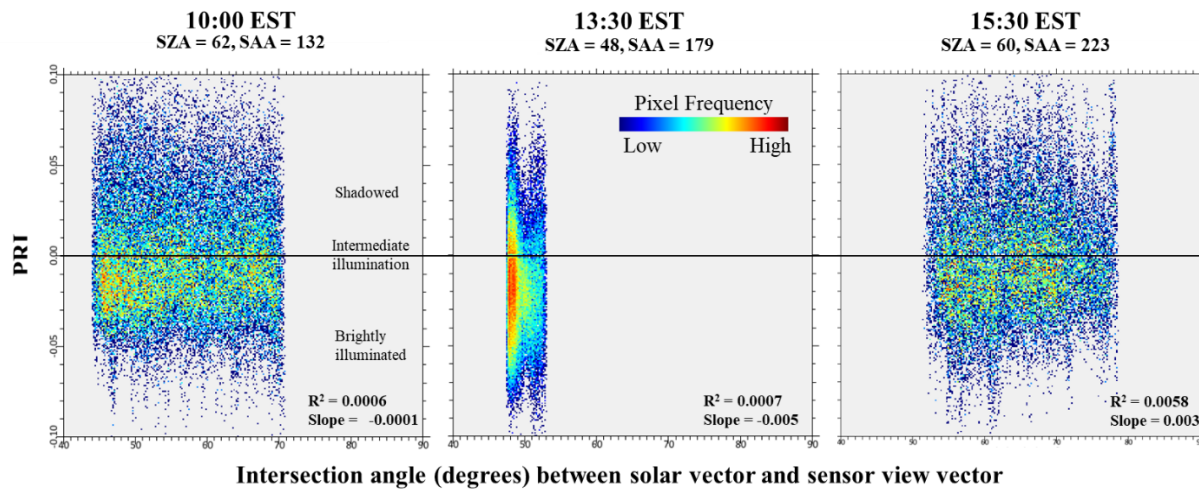


Figure 2.14. Plot 1 results for PRI (y-axis) versus the intersection angle between solar and sensor view vector (in degrees) for the morning, midday, and afternoon NS flights. Color gradient indicates pixel frequency with low in blue and high in red. The 13:30 EST intersection angle range is small due to the small difference in angle of the sun (solar noon) and the NS flight (6 degrees difference).

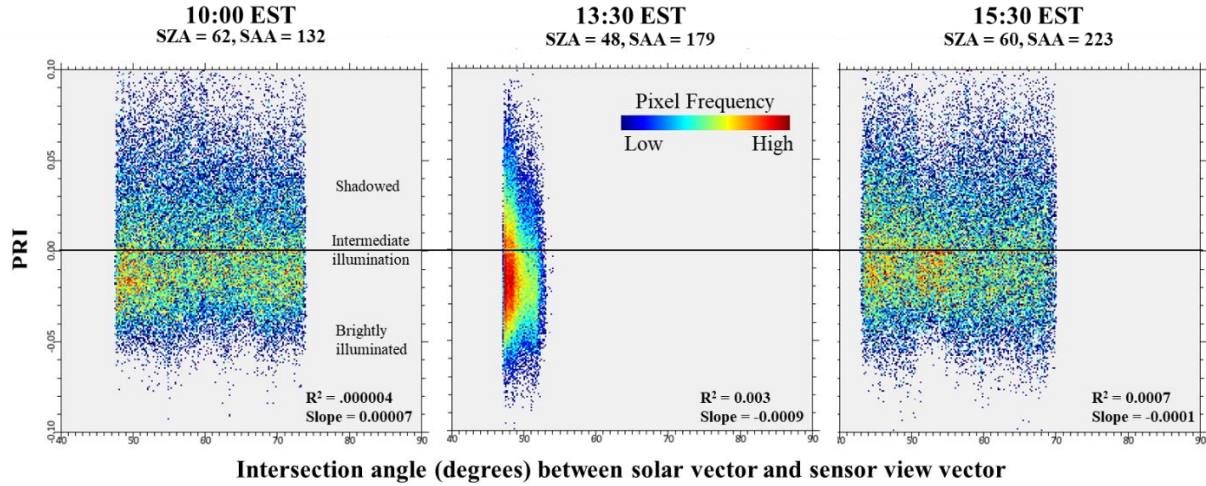


Figure 2.15. Plot 2 results for PRI (y-axis) versus the intersection angle between solar and sensor view vector (in degrees) for the morning, midday, and afternoon NS flights. Color gradient indicates pixel frequency with low in blue and high in red. The 13:30 EST intersection angle range is small due to the small difference in angle of the sun (solar noon) and the NS flight (6 degrees difference).

2.4 Discussion

Our results illustrate the effects of variable illumination calculation of PRI diurnally and vertically to assess forest function and structure from airborne imagery. As expected, areas receiving a greater amount of sunlight have lower PRI than those with a greater amount of shadow over the time range of the G-LiHT flights. LUE from the flux tower data shows a gradual increase through the day, whereas the mean PRI for all three illumination classes decreases at midday (Figure 2.13). The sunlit class PRI is lowest, and the shadow class is highest, suggesting LUE is reduced due to excess solar illumination. PRI is reduced during the mid-day for all three illumination classes, possibly due to the increased amount of direct solar illumination. Middleton et al. 2009 mentions a “bowl-shape” for diurnal PRI responses on sunny days, which are consistent with light-induced accumulation of zeaxanthin (Z) in foliage. On

sunny days, the relative concentration of Z would be expected to increase in the morning and decline throughout the afternoon as the xanthophyll pigment ratios adjust in tandem with illumination conditions.

In Figure 2.3, it is important to note the difference in the flux tower results from the two flight days on 26th versus the 27th. The flights on the second day are situated at the times of the day where the incident PAR was irregular and could affect the results from the second day. This can be further explained with the PRI bar graph results in Figures 2.5 and 2.6. We see a dip in the shadowed PRI and an increase in negative sunlit PRI values at midday on the 26th for all plots, but we do not see this similar trend on the 27th. This could be due to cloud cover that produced lower incident PAR on the 27th which led to more diffuse light and fewer shadows in the images. Middleton et al., 2009 showed that the spectral responses for shaded and sunlit foliage were distinctively different on sunny days with the shaded foliage class maintaining higher PRI values and the sunlit foliage with lower PRI values. On cloudy days with more diffuse light, thus less high light stress on sunlit foliage, the PRI values for the different illumination classes show less separation.

Many studies use two-leaf models with multi-angle observations (Chen et al. 1999; Chen et al., 2020; Cheng et al., 2009; Zhang et al., 2017) and lidar-derived measurements (Gamon et al., 2023; Middleton et al., 2009) to classify illumination classes for sunlit and shaded foliage. This work provides a new method to characterize light levels into different illumination classes using a panchromatic simulation approach. The use of the panchromatic band simulation for the illumination class separation proved optimal with three classes, rather than a binary split into two classes. Separation into illumination classes showed key differences in PRI results for different illumination levels that change throughout the day due to amount and angle of incoming solar

radiation (Tables 2.1-2.2 and Figures 2.5-2.6). Less shadowed pixels are evident in the midday images due to the direct illumination over the plots. This panchromatic simulation approach could be applied to space-borne or airborne imagery to create shadow fractions for each pixel when lidar is not available.

Our work highlights the importance of consideration of canopy structure-related information when using spectral information to monitor canopy physiology. To decipher the vertical distributions of PRI, we used the provided lidar from the corresponding NS flights. In Figure 2.9, the overstory and understory delineation is evident with the vertical distributions of brightness and PRI for each illumination class due to this plot being post-thin. In the first column with panchromatic reflectance, the understory is more illuminated at midday, and this could be attributed to the NS row orientation of the plantation aligning with the sun at midday, thus the understory leaves are more brightly illuminated when the sun is directly above. For the PRI columns, there is a shift in foliage efficiency to lower PRI values for the understory at midday, which is specifically prominent in the mixed illumination class. Also, the spread of the points are denser for the midday row (13:30 EST), thus more mixed pixels, which is complemented with the percentages of mixed pixels at this time of day from Figure 2.5. Figure 2.10 shows how the shadowed, mixed, and sunlit classes shift to more negative PRI values at midday suggesting lower efficiency due to excess illumination. Furthermore, this stand has not been thinned and there is little to no presence of an understory, and the rows are oriented EW. With the vertical distributions of PRI from both plots, we have confidence that the downregulation in efficiency at midday is physiological, and not solely due to the row orientation and understory presence.

Structural mediation shows that the distributions of sunlit and shaded foliage change throughout the course of the day and affect the PRI responses. Shadowed pixels result in positive

PRI values while the sunlit pixels have more negative PRI values. The structural diurnal feature of the illumination class in Figures 2.11 and 2.12 show an increase in sunlit foliage in midday and more shadowed pixels in the morning and evening. PRI shows a decrease in PRI values at midday indicating stress on the foliage, while the morning and evening shows a wider range of values indicating improved foliage efficiency during those times of day.

The diurnal trend of PRI at the leaf scale has shown to have an asymmetric daily cycle with a downregulation around noon (Hall et al., 2008; Mottus et al., 2018). This downregulation at midday can be seen in this study at the canopy scale using airborne data, where LUE and PRI are more closely related at midday and less so in the mornings. Mottus et al., 2018 mentions that there may be only a few days during the peak growing season period that are suitable for daily PRI studies, whereas the data in this study was collected at the end of the growing season. This reiterates the difference of PRI results that we have seen on the different flight days in this study, which led us to focus on the clearer day (normal incident PAR curve) compared to the cloudy day (Figure 2.3).

Furthermore, the airborne data was acquired late in the growing season when productivity is slowing down, thus sources of stress that affect diurnal PRI are not significant (temperature was low, soil moisture was high due to recent rain, and solar radiance was low) compared to what the forest would be experiencing during peak growing season. This late-season negative correlation between LUE and PRI is consistent with the in-situ results of Nakaji et al. (2006) and Nyongesah et al. (2016), which show the expected positive correlation during the growing season but a transition to a negative correlation at the end of the growing season.

Our results conclude that PRI varies diurnally and vertically for different illumination levels depending on the amount of illumination mediated by the canopy structure. This study

shows emphasis on airborne measurements by reevaluating the effect of illumination on canopy scale PRI at different times of the day. We developed a new way for illumination classification using a simulated panchromatic band rather than lidar, then using the lidar to partition the vertical distributions of PRI. The hyperspectral dataset provides a means to calculate PRI and to delineate into illumination classes to separate out shaded, mixed, and sunlit portions of the canopy and the lidar provides the vertical distribution of canopy PRI by illumination classes.

2.5 Conclusion

Our results conclude that PRI varies diurnally and vertically for different illumination levels depending on the amount of illumination mediated by the canopy structure. This study shows emphasis on airborne measurements by reevaluating the effect of illumination on canopy scale PRI at different times of the day. We developed a new way for illumination classification using a simulated panchromatic band rather than lidar, then using the lidar to partition the vertical distributions of PRI. The hyperspectral dataset provides a means to calculate PRI and to delineate into illumination classes to separate out shaded, mixed, and sunlit portions of the canopy and the lidar provides the vertical distribution of canopy PRI by illumination classes.

This study illustrates the effects of variable illumination calculation of PRI from airborne imagery. A diurnal signal for PRI is shown in our results for all illumination classes and shows a difference in PRI response from sunlit and shaded foliage on a clear-sky day. The structure of the canopy (or stand) controls how the light intercepts with the canopy (or stand). Thus, we incorporated lidar in order to measure spectral reflectance (PRI) from different parts of the canopy and combine the effects of structure information to assess forest functioning. These results show the importance of time of day on canopy shadowing (or illumination through the day) and the resulting calculations of PRI. Furthermore, the new method for differentiation of

canopy light levels using a simulated panchromatic band is applicable when lidar data is not available. Documenting diurnal variations in LUE and PRI with in-situ and airborne data could help inform interpretation of satellite observations as well as the design of future observing systems.

References

- Bellassen, V., & Luyssaert, S. (2014). Managing forests in uncertain times. *Nature*, 506, 153–155.
- Breton, C., Blanchet, P., Amor, B., Beauregard, R., & Chang, W.-S. (2018). Assessing the climate change impacts of biogenic carbon in buildings: A critical review of two main dynamic approaches. *Sustainability*, 10(6), 2020.
- Chen, J. M., Liu, J., Cihlar, J., & Goulden, M. L. (1999). Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. *Ecological Modelling*, 124(2–3), 99–119.
- Chen, J., Zhang, Q., Chen, B., Zhang, Y., Ma, L., Li, Z., Zhang, X., et al. (2020). Evaluating multi-angle photochemical reflectance index and solar-induced fluorescence for the estimation of gross primary production in maize. *Remote Sensing*, 12(17), 2812.
- Chen, A., Lichstein, J. W., Osnas, J. L. D., & Pacala, S. W. (2014). Species-independent down-regulation of leaf photosynthesis and respiration in response to shading: evidence from six temperate tree species. *PLOS ONE*, 9(4), e91798.
- Cheng, Y.-B., Middleton, E. M., Hilker, T., Coops, N. C., Black, T. A., & Krishnan, P. (2009). Dynamics of spectral bio-indicators and their correlations with light use efficiency using directional observations at a douglas-fir forest. *Measurement Science and Technology*, 20(9), 095107.

- Cheng, Y.-B., Middleton, E., Zhang, Q., Huemmrich, K., Campbell, P., Corp, L., Cook, B., et al. (2013). Integrating solar induced fluorescence and the photochemical reflectance index for estimating gross primary production in a cornfield. *Remote sensing*, 5(12), 6857–6879.
- Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., et al. (2021). Representativeness of eddy-covariance flux footprints for areas surrounding AmeriFlux sites. *Agricultural and Forest Meteorology*, 301–302, 108350.
- Cook, B., Corp, L., Nelson, R., Middleton, E., Morton, D., Mccorkel, J., Masek, J., et al. (2013). NASA Goddard’s LiDAR, Hyperspectral and Thermal (G-LiHT) airborne imager. *Remote Sensing*, 5(8), 4045–4066.
- Domec, J.-C., King, J. S., Noormets, A., Treasure, E., Gavazzi, M. J., Sun, G., & McNulty, S. G. (2010). Hydraulic redistribution of soil water by roots affects whole-stand evapotranspiration and net ecosystem carbon exchange. *The New Phytologist*, 187(1), 171–183.
- Fauset, S., Gloor, M. U., Aidar, M. P. M., Freitas, H. C., Fyllas, N. M., Marabesi, M. A., Rochelle, A. L. C., et al. (2017). Tropical forest light regimes in a human-modified landscape. *Ecosphere* (Washington, D.C), 8(11), e02002.
- Gamon, J. A., Peñuelas, J., & Field, C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, 41(1), 35–44.
- Gamon, J. A., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, 112(4), 492–501.

- Gamon, J. A., Huemmrich, K. F., Wong, C. Y. S., Ensminger, I., Garrity, S., Hollinger, D. Y., Noormets, A., et al. (2016). A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. *Proceedings of the National Academy of Sciences of the United States of America*, 113(46), 13087–13092.
- Gamon, J. A., Wang, R., & Russo, S. E. (2023). Contrasting photoprotective responses of forest trees revealed using PRI light responses sampled with airborne imaging spectrometry. *The New Phytologist*.
- Garbulsky, M. F., Peñuelas, J., Gamon, J., Inoue, Y., & Filella, I. (2011). The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies: A review and meta-analysis. *Remote Sensing of Environment*, 115(2), 281–297.
- Gilabert, M. A., Moreno, A., Maselli, F., Martínez, B., Chiesi, M., Sánchez-Ruiz, S., García-Haro, F. J., et al. (2015). Daily GPP estimates in Mediterranean ecosystems by combining remote sensing and meteorological data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102, 184–197.
- Goerner, A., Reichstein, M., Tomelleri, E., Hanan, N., Rambal, S., Papale, D., Dragoni, D., et al. (2011). Remote sensing of ecosystem light use efficiency with MODIS-based PRI. *Biogeosciences*, 8(1), 189–202.
- Goulden, M. L., Munger, J. W., Fan, S.-M., Daube, B. C., & Wofsy, S. C. (1996). Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy. *Global Change Biology*, 2(3), 169–182.

- Hilker, T., Coops, N. C., Schwalm, C. R., Jassal, R. P. S., Black, T. A., & Krishnan, P. (2008). Effects of mutual shading of tree crowns on prediction of photosynthetic light-use efficiency in a coastal douglas-fir forest. *Tree Physiology*, 28(6), 825-834.
- Hilker, T., Hall, F. G., Coops, N. C., Lyapustin, A., Wang, Y., Nesic, Z., Grant, N., et al. (2010). Remote sensing of photosynthetic light-use efficiency across two forested biomes: Spatial scaling. *Remote Sensing of Environment*, 114(12), 2863–2874.
- (<https://math.stackexchange.com/users/250230/nalyd88>), N. (n.d.). Finding the angle between two points given their azimuth and elevation angles.
- Li, L., Zhao, Y., Fu, Y., & Xin, Q. (2018). Satellite-based models need improvements to simulating annual gross primary productivity: a comparison of six models for regional modeling of deciduous broadleaf forests. *Remote Sensing*, 10(7), 1008.
- Middleton, E. M., Cheng, Y.-B., Hilker, T., Black, T. A., Krishnan, P., Coops, N. C., & Huemmrich, K. F. (2009). Linking foliage spectral responses to canopy-level ecosystem photosynthetic light-use efficiency at a Douglas-fir forest in Canada. *Canadian Journal of Remote Sensing*, 35(2), 166–188.
- Middleton, E., Rascher, U., Corp, L., Huemmrich, K., Cook, B., Noormets, A., Schickling, A., et al. (2017). The 2013 FLEX—US Airborne Campaign at the Parker Tract loblolly pine plantation in North Carolina, USA. *Remote Sensing*, 9(6), 612.
- Moomaw, W. R., Law, B. E., & Goetz, S. J. (2020). Focus on the role of forests and soils in meeting climate change mitigation goals: summary. *Environmental Research Letters: ERL [Web site]*, 15, 045009.

- Mõttus, M., Hernández-Clemente, R., Perheentupa, V., Markiet, V., Aalto, J., Bäck, J., & Nichol, C. (2018). Measurement of diurnal variation in needle PRI and shoot photosynthesis in a boreal forest. *Remote Sensing*, 10(7), 1019.
- Nakaji, T., Oguma, H., & Fujinuma, Y. (2006). Seasonal changes in the relationship between photochemical reflectance index and photosynthetic light use efficiency of Japanese larch needles. *International Journal of Remote Sensing*, 27(3), 493–509.
- Nyongesah, M. J., Quan, W., & Lu, X. (2016). Remote sensing of assimilating branch light use efficiency using the photochemical reflectance index in *Haloxylon ammodendron* forest. *Journal of Applied Remote Sensing*, 10(2), 026017.
- Peñuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J., & Field, C. B. (1994). Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sensing of Environment*, 48(2), 135–146.
- Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., Ciais, P., et al. (2020). Interannual variation of terrestrial carbon cycle: issues and perspectives. *Global Change Biology*, 26(1), 300–318.
- Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., Davis, K. J., et al. (2012). A model-data comparison of gross primary productivity: results from the North American Carbon Program site synthesis. *Journal of Geophysical Research*, 117(G3), G03010.
- Suárez, L., Zarco-Tejada, P. J., Sepulcre-Cantó, G., Pérez-Priego, O., Miller, J. R., Jiménez-Muñoz, J. C., & Sobrino, J. (2008). Assessing canopy PRI for water stress detection with diurnal airborne imagery. *Remote Sensing of Environment*, 112(2), 560–575.

- Thomas, R. Q., Jersild, A. L., Brooks, E. B., Thomas, V. A., & Wynne, R. H. (2018). A mid-century ecological forecast with partitioned uncertainty predicts increases in loblolly pine forest productivity. *Ecological Applications*, 28(6), 1503–1519.
- Tortini, R., Coops, N. C., Nesic, Z., Christen, A., Lee, S. C., & Hilker, T. (2017). Remote sensing of seasonal light use efficiency in temperate bog ecosystems. *Scientific Reports*, 7(1), 8563.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150.
- Wang, X., Zheng, G., Yun, Z., Xu, Z., Moskal, L. M., & Tian, Q. (2020). Characterizing the spatial variations of forest sunlit and shaded components using discrete aerial lidar. *Remote Sensing*, 12(7), 1071.
- Yang, P. (2022). Exploring the interrelated effects of soil background, canopy structure and sun-observer geometry on canopy photochemical reflectance index. *Remote Sensing of Environment*, 279, 113133.
- Zhang, C., Filella, I., Garbulsky, M., & Peñuelas, J. (2016). Affecting factors and recent improvements of the photochemical reflectance index (PRI) for remotely sensing foliar, canopy and ecosystemic radiation-use efficiencies. *Remote Sensing*, 8(9), 677.
- Zhang, Q., Cheng, Y.-B., Lyapustin, A. I., Wang, Y., Gao, F., Suyker, A., Verma, S., et al. (2014). Estimation of crop gross primary production (GPP): fAPARchl versus MOD15A2 FPAR. *Remote Sensing of Environment*, 153, 1–6.
- Zhang, Q., M. Chen, J., Ju, W., Wang, H., Qiu, F., Yang, F., Fan, W., et al. (2017). Improving the ability of the photochemical reflectance index to track canopy light use efficiency through differentiating sunlit and shaded leaves. *Remote Sensing of Environment*, 194, 1–15.

**Chapter 3 - Characterizing the influence of varying functional traits on forest productivity
using remotely sensed data acquired from different forest compositions across selected
United States NEON sites.**

Abstract

Gross primary productivity (GPP) describes the total photosynthesis (i.e., carbon fixation) for an ecosystem and is an important component of the global land carbon budget. To reduce uncertainties in carbon accounting for different forest ecosystems, it is crucial to analyze the health and productivity of forested ecosystems. Plant functional traits, which are a combination of morphological, physiological, and environmental characteristics, have been shown to be predictive of forest ecosystem carbon dynamics. The objective of this work is to determine how well GPP is predicted by remotely quantified functional traits across forested ecosystems in the United States. Airborne remote sensing observations and in situ flux tower measurements used in this analysis were acquired from selected forested sites from the National Ecological Observatory Network (NEON) data portal. We investigated NEON-provided hyperspectral optical indices and lidar derived products as proxies of remotely sensed plant functional traits. Average midday GPP around the date of flight was calculated by developing a relationship between night respiration and temperature and removing that component from the net surface-atmosphere CO₂ exchange (NSAE) measured by NEON. We used multiple linear regression modeling with a best subset approach for three grouped trait classes: morphological and environmental traits from lidar (2-term model with a PRESS $R^2 = 0.50$ and RMSE of 5.32 $\mu\text{mol CO}_2/\text{m}^2/\text{s}$), physiological traits from hyperspectral (3-term model with a PRESS $R^2 = 0.45$ and RMSE of 5.48 $\mu\text{mol CO}_2/\text{m}^2/\text{s}$), and a combination of all functional traits from lidar and hyperspectral together with a selection criterion for best fit. The best fit GPP model was the 4-term combination of functional traits model using CHM mean, DSM standard deviation, PRI standard deviation, and WBI mean producing a R^2 of 0.87, an adjusted R^2 of 0.84, a PRESS R^2 of 0.75 and RMSE of 3.48 $\mu\text{mol CO}_2/\text{m}^2/\text{s}$. Results show that all three types (i.e., morphological, environmental, and physiological) are important predictors of forest productivity.

Keywords: forest, functional traits, hyperspectral, lidar, NEON, productivity, GPP

3.1 Introduction

Gross primary productivity (GPP, $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) describes the total photosynthesis (i.e., carbon fixation) for an ecosystem, and is an important component of the global land carbon budget. GPP is the total amount of carbon dioxide "fixed" by land plants per unit time through the photosynthetic reduction of CO_2 into organic compounds and is a function of foliage light use efficiency and absorbed sunlight. GPP is regularly modeled from eddy covariance flux tower estimates (Baldocchi et al., 2001; Novick et al., 2018) and is also derived from MODIS EVI (Xiao et al., 2010). However, the flux towers only provide a sample estimate of an area covering roughly 16 hectares (Chu et al., 2021) and the spatial resolution of current space-borne sensors are too coarse for canopy level productivity investigations.

Forest functional traits drive biomass and productivity through multiple mechanisms that influence ecosystem functions (Charturvedi et al., 2011; Hao et al., 2020; Garnier et al., 2015). These mechanisms include three types of traits describing a site: morphological, environmental, physiological. Morphological traits include information about canopy architecture. Environmental traits include information about topography, weather/climate, and soil properties, etc. Physiological traits include indices associated with live green vegetation, vigor, the xanthophyll cycle, light use efficiency, leaf water content, and chlorophyll content. (Garnier et al., 2015; Schneider et al., 2017)

Traits associated with structural (morphological) diversity are an important subset of plant functional traits, the morphological or physiological characteristics that are functionally relevant for growth, reproduction, and survival (Ma et al., 2019) and are at the crossroads between responses to the environment and ecosystem properties (Díaz et al., 2013). Functional traits influence forest ecosystem carbon dynamics. Canopy structure architecture, including the

horizontal and vertical structure of forests and mediation of light availability, affect competitive and complementary light use and ecosystem productivity (Schneider et al., 2017). Metrics that quantify canopy structural diversity are much more strongly associated with forest productivity than traditional biodiversity measures like species richness and phylogenetic diversity (Aponte et al., 2020).

Remotely-sensed functional traits are important predictors of variation in rates of net primary productivity and gross primary productivity (Duran et al., 2019). Remote sensing based GPP estimations can be improved by using footprint size similarity and downscaling techniques, but are still underestimated in certain years and landscapes (Gelybó et al., 2013.) Pierrat et al., 2022 calls for an improved connection between remotely-sensed measurements and GPP, which are often site-specific and vary across biomes. The combination of remote sensing data with flux tower data is a valuable approach for determining site-based GPP, which shows promise for different-scale assessments of carbon sequestration (Chen et al., 2019).

Both physiological and morphological traits are needed for full characterization of functional diversity using remote sensing (Schneider et al., 2017, Ma et al., 2019), necessitating the use of sensors enabling quantification of canopy structure (Aponte et al., 2020) as well as function. Further, high-spatial resolution (or other comparable) data enabling quantification of canopy structure is needed to address multiple scattering and contrasting illumination and to control for varying amounts of vegetation percentage cover in fine spectral resolution measurements (Wang and Gamon, 2019).

The objective of this study is to characterize the relationship between functional traits and forest productivity. Specifically, the goal is to determine how well GPP is predicted by variations

in remotely quantified forest functional traits from different forest compositions across selected United States NEON sites.

3.2 Materials and Methods

3.2.1 Study area

To characterize the influence of functional traits on forest productivity, this study uses a range of forest sites across the United States. Suitable field sites must be forested and have coincident airborne observations with usable in situ flux tower measurements. NEON's sites are broken up into four categories: core terrestrial, relocatable terrestrial, core aquatic, and relocatable aquatic. For this analysis, we chose the core and relocatable terrestrial sites and visually analyzed the sites' land cover in Google Earth Pro to select forested sites. To select the site for analysis, it must also have corresponding flux tower measurements available near the time of the airborne acquisition. NEON flux tower data does have numerous gaps in the datasets due to quality control, thus this is a crucial step, which resulted in only 22 NEON sites with coincident flux tower data and an airborne acquisition (Figure 3.1 and 3.2). In Table 3.1, suitable field sites are listed along with descriptions of each site including name, state, latitude, longitude, acquisition month and year, dominant vegetation, and dominant species.

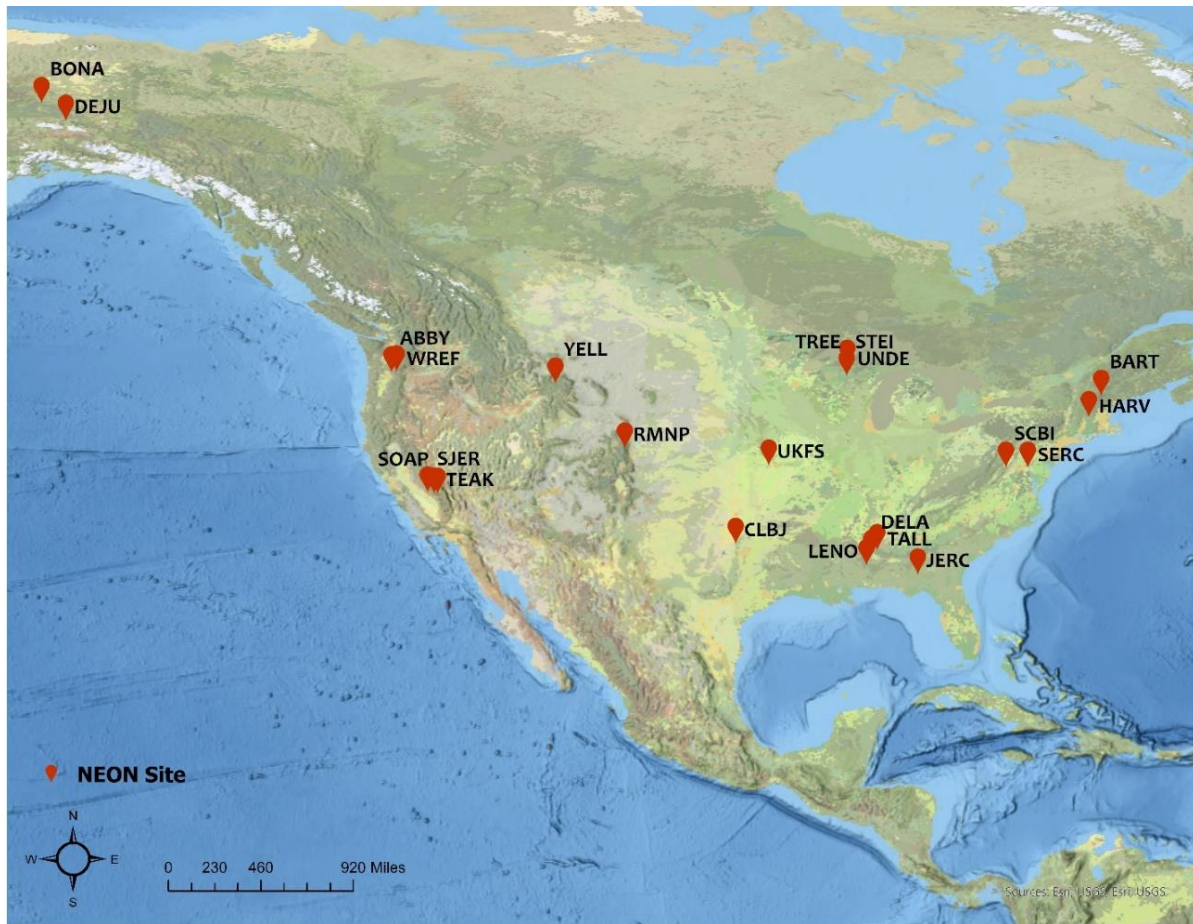


Figure 3.1. Study area map of selected forested NEON sites across the United States overlain on ESRI's National Geographic Style basemap.

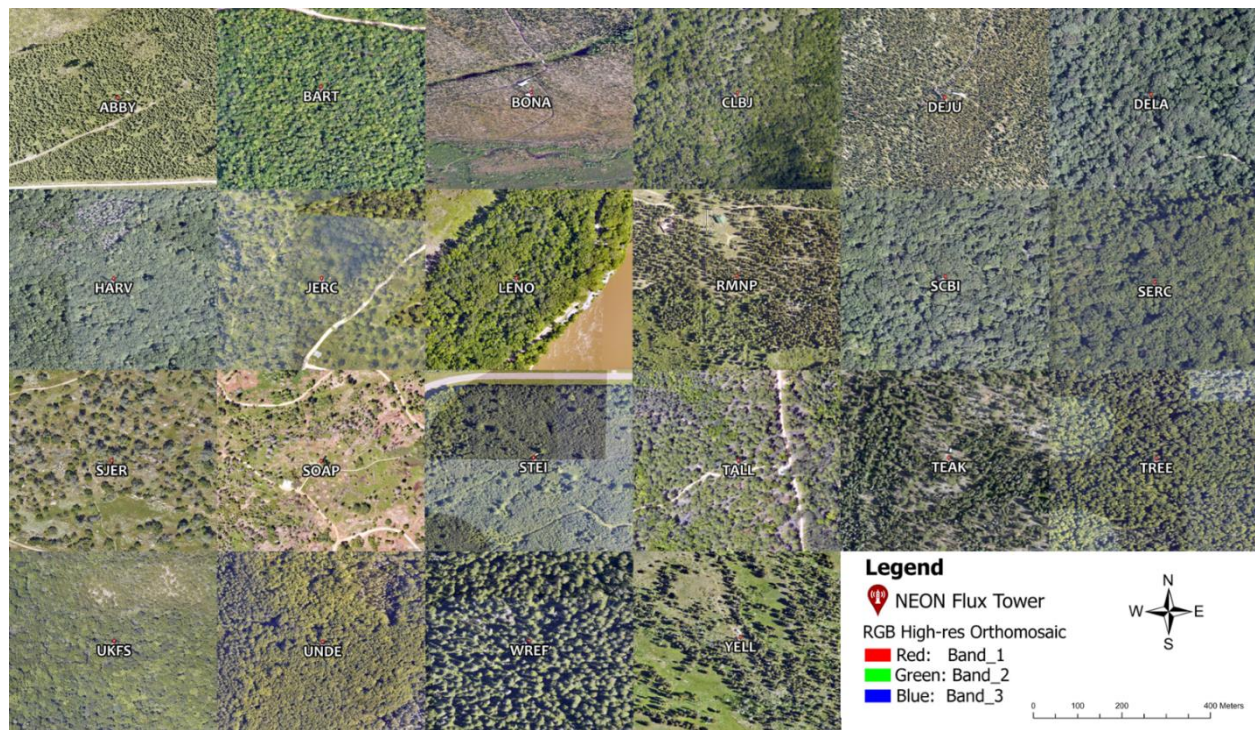


Figure 3.2. High-resolution orthomosaics (NEON product number DP3.30010.00) of the selected forest NEON flux tower sites across the US. Sites are in alphabetical order (left to right) as shown in Table 3.1.

Table 3.1. Suitable NEON field sites with descriptions including name, state, latitude, longitude, acquisition date, and dominant vegetation types.

Field Site	Field Site Full Name	State	Latitude, Longitude	Acquisition Month Year	Dominant Vegetation Types
ABBY	Abby Road	WA	45.762439, -22.33032	Jul 2019	Evergreen Forest Grassland/Herbaceous Shrub/Scrub
BART	Bartlett Experimental Forest	NH	44.063889, -1.287375	Aug 2017	Deciduous Forest Evergreen Forest Mixed Forest
BONA	Caribou-Poker Creeks Research Watershed	AK	65.15401, -47.50258	Aug 2018	Deciduous Forest Evergreen Forest Mixed Forest Woody Wetlands
CLBJ	Lyndon B. Johnson National Grassland	TX	33.40123, -97.57	Apr 2019	Deciduous Forest Grassland/Herbaceous
DEJU	Delta Junction	AK	63.88112, -145.7514	Jul 2018	Evergreen Forest Shrub/Scrub Woody Wetlands
DELA	Dead Lake	AL	32.541727, -87.80388	Apr 2018	Evergreen Forest Woody Wetlands
HARV	Harvard Forest & Quabbin Watershed	MA	42.53691, -72.17265	Aug 2019	Deciduous Forest Evergreen Forest Mixed Forest Woody Wetlands
JERC	The Jones Center At Ichauway	GA	31.194839, -84.46862	Aug 2019	Deciduous Forest Evergreen Forest Mixed Forest Cultivated Crops
LENO	Lenoir Landing	AL	31.853861, -88.16118	May 2021	Deciduous Forest Woody Wetlands
RMNP	Rocky Mountain National Park	CO	40.275903, -105.5460	Jul 2020	Evergreen Forest
SCBI	Smithsonian Conservation Biology Institute	VA	38.892925, -78.13949	May 2019	Deciduous Forest Evergreen Forest Pasture/Hay
SERC	Smithsonian Environmental Research Center	MD	38.890131, -76.56001	Aug 2021	Deciduous Forest Cultivated Crops
SJER	San Joaquin Experimental Range	CA	37.10878, -119.7323	Mar 2021	Evergreen Forest Grassland/Herbaceous Shrub/Scrub
SOAP	Soaproot Saddle	CA	37.03337, -119.2622	Jun 2018	Evergreen Forest Shrub/Scrub
STEI	Steigerwaldt-Chequamegon	WI	45.50894, -89.58637	Aug 2020	Deciduous Forest Mixed Forest Woody Wetlands
TALL	Talladega National Forest	AL	32.95047, -87.39326	Apr 2018	Deciduous Forest Evergreen Forest Mixed Forest
TEAK	Lower Teakettle	CA	37.00583, -119.0060	Jul 2021	Evergreen Forest Shrub/Scrub
TREE	Treehaven	WI	45.49369, -89.58571	Aug 2020	Deciduous Forest Evergreen Forest, Mixed Forest, Woody Wetlands
UKFS	University of Kansas Field Station	KS	39.040431, -95.19215	Jul 2020	Deciduous Forest Pasture/Hay
UNDE	University of Notre Dame Environmental Research Center	MI	46.23391, -89.53725	Aug 2020	Deciduous Forest Mixed Forest Woody Wetlands
WREF	Wind River Experimental Forest	WA	45.82049, -121.9519	Jul 2018	Evergreen Forest
YELL	Yellowstone National Park	WY	44.95348, -110.5391	Jul 2019	Evergreen Forest Grassland/Herbaceous Shrub/Scrub

3.2.2 Datasets

The National Ecological Observatory Network (NEON) provides open data to understand changing ecosystems (<https://data.neonscience.org/>). NEON administers airborne remote sensing acquisitions along with ground flux tower measurements across varying aquatic and terrestrial ecosystems throughout the United States.

3.2.2.1 NEON terrestrial instrument system (TIS)

For NEON's Terrestrial Instrument Systems (TIS), a standard set of measurements are collected to describe ecological processes including 30-minute intervals of eddy covariance measurements of canopy photosynthesis uptake. Meteorological data is collected along with the CO₂ fluxes, including air temperature, precipitation, humidity, frictional velocity, incident photosynthetically active radiation, and net radiation that provide further insight on the site's environmental conditions across time. The provided metrics describing photosynthetic function and environmental conditions can be key indicators of forest productivity and environmental stresses, such as differing temperatures and water availability which influence seasonal growth patterns.

3.2.2.2 NEON airborne observation platform (AOP)

NEON's Airborne Observation Platform (AOP) network regularly (yearly or bi-yearly) acquires airborne remote sensing observations at most sites. The airborne remote sensing observations include a wide range of hyperspectral and lidar derived products. Available hyperspectral products provided as flightlines and 1 km² tiled mosaics at 1 m resolution are as follows: albedo, spectrometer orthorectified surface directional reflectance (10 cm resolution), spectrometer orthorectified at-sensor radiance, high-resolution orthorectified camera imagery, fraction of incident photosynthetically active radiation (fPAR), vegetation indices, canopy water indices. Available lidar products provided discrete return lidar point clouds and derived elevation (DSM and DTM), slope and aspect, and ecosystem structure (CHM).

3.2.3 Methods

This study utilizes a nonparametric and parametric statistical approaches to characterize the influence of varying functional traits on forest productivity with flux tower and remotely sensed data acquired from different forest compositions across selected United States' NEON sites. Methods include three incremental steps of the approach: (1) collection of data, (2) calculation of flux tower parameters from TIS and functional traits from remote sensing data from AOP, and (3) statistical analysis to assess prediction of flux tower derived GPP from remotely sensed functional traits. Collection of data requires selection of coincident airborne observations with usable in situ flux tower measurements. Calculation of flux parameters requires daytime and nighttime measurements to be partitioned into photosynthesis and respiration to derive the midday GPP. The functional traits from remote sensing data includes deriving physiological traits with vegetation and water indices from optical imagery, derivation of morphological and environmental measurements from lidar, and delineating of bounding box for flux tower coverage area. The statistical analysis incorporates multiple linear regression with a best subset approach applied to three modeling groups to provide GPP predictions from a variation of forest functional traits.

3.2.3.1 Calculation of Flux Tower Parameters from NEON TIS

To describe the environmental conditions at the sites, we used net ecosystem exchange (NEE), photosynthetically active radiation (PAR), temperature (T_{air}), ecosystem respiration (Reco), and frictional velocity (U^*) from flux towers. The flux tower data is separated by sensors at different heights on the tower. For this analysis, we chose the top of the tower measurements for each site considering we are looking at top-of-canopy measurements from the airborne

acquisitions. Goulden et al. 2006 provides an eddy covariance mesonet approach to model land-atmosphere exchange from flux tower measurements. Using this approach, we first partitioned NEE into photosynthesis and respiration using daytime and nighttime from a model of Tair and Reco. Gross ecosystem exchange (GEE) was calculated by subtracting Reco from daytime NEE. Gross primary productivity (GPP) was calculated as – GEE. We derived the average midday summer GPP around the days of the airborne flights for each site.

3.2.3.2 Calculation of Remote Sensing Functional Traits from NEON AOP

Functional traits used in this analysis were derived by NEON from hyperspectral data that include a wide range of selected vegetation and water indices at 1 m resolution. Table 3.2 lists NEON's vegetation and water indices used in this study.

Table 3.2. Description of physiological traits, their acronyms, name, and equation from NEON's AOP hyperspectral products used in this study.

Index Group	Index Acronym	Name	Equation
Vegetation	NDVI	Normalized Difference Vegetation Index	$\frac{(NIR - Red)}{(NIR + Red)}$
	SAVI	Soil-adjusted Vegetation Index	$\frac{(NIR - Red)(1 + L)}{(NIR + Red + L)}$ <i>L</i> = soil conditioning index
	EVI	Enhanced Vegetation Index	$2.5 * \frac{(NIR - Red)}{(NIR + 6 * Red - 7.5 * Blue - 1)}$
	ARVI	Atmospherically Resistant Vegetation Index	$\frac{NIR - [Red - \gamma(Blue - Red)]}{NIR + [Red - \gamma(Blue - Red)]}$
	PRI	Photochemical Reflectance Index	$\frac{(\rho_{531} - \rho_{570})}{(\rho_{531} + \rho_{570})}$
	fPAR	Fraction of Incident Photosynthetically Active Radiation	$C[1 - Aexp(-B * LAI)]$ <i>A</i> , <i>B</i> , and <i>C</i> are parameters currently set to 1, 0.4, and 1 respectively
Water	MSI	Moisture Stress Index	$\frac{\rho_{1599}}{\rho_{819}}$
	NDII	Normalized Difference Infrared Index	$\frac{(\rho_{819} - \rho_{1649})}{(\rho_{819} + \rho_{1649})}$
	NDWI	Normalized Difference Water Index	$\frac{(\rho_{847} - \rho_{1241})}{(\rho_{847} + \rho_{1241})}$
	NMDI	Normalized Multi-band Drought Index	$\frac{\rho_{870} - (\rho_{1640} - \rho_{2130})}{\rho_{870} + (\rho_{1640} - \rho_{2130})}$
	WBI	Water Band Index	$\frac{\rho_{970}}{\rho_{900}}$

Functional traits used in analysis derived from lidar are provided at 1 m spatial resolution. Functional traits from NEON's lidar products include a DTM, DSM, CHM, slope, and aspect derived from the DTM (Table 3.3).

Table 3.3. *Name and description of morphological and environmental traits from NEON's AOP lidar products used in this study.*

Name	Description
Digital Terrain Model (DTM)	Bare earth elevation - includes only elevations that relate to the physical terrain surface
Digital Surface Model (DSM)	Actual surface elevation - includes elevations that relate to surface features (e.g. buildings, vegetation)
Canopy Height Model (CHM)	Height of top of canopy above bare earth
Slope	A ratio of rise over run (height over distance) of the bare earth elevation product given in degrees
Aspect	The direction of the steepest slope of the bare earth elevation product (e.g., north, east, south, west) given in degrees clockwise from grid north

A 420 m x 420 m bounding box centered on the flux tower was delineated for each site. Some sites required stitching together multiple 1 km² tiles if the flux tower coverage area was in multiple mosaicked tiles. From the 1 m resolution NEON hyperspectral and lidar products, mean, maximum, minimum, and standard deviations were calculated within the bounding area around the flux towers.

To evaluate the role of a variation of functional traits on forest productivity, we chose to separate the NEON remote sensing variables into three grouped trait classes: morphological and environmental, physiological, and a combination of all functional traits (Table 3.4). The morphological and environmental traits were selected from NEON's AOP lidar products and the physiological traits were selected from NEON's AOP hyperspectral products, thus the combination of all functional traits is a combination of lidar and hyperspectral data.

Table 3.4. NEON remote sensing variables grouped into three trait classes – morphological and environmental (lidar data), physiological (hyperspectral data), and a combination of all functional traits (lidar and hyperspectral data).

Trait Group	NEON Remote Sensing Variables
Morphological Traits	DSM and CHM
Environmental Traits	DTM, Slope, Aspect
Physiological Traits	MSI, NDII, NDWI, NMDI, WBI, fPAR, ARVI, EVI, NDVI, PRI, and SAVI
Combo of Functional Traits	DTM, DSM, Slope, Aspect, CHM, LAI, MSI, NDII, NDWI, NMDI, WBI, fPAR, ARVI, EVI, NDVI, PRI, and SAVI

3.2.3.3 Regression analysis

To represent the relationships among the NEON remote sensing variables used in the model, we calculated and graphed correlation coefficients for all variables. The Multivariate and Correlations tool in JMP Pro 16 was used to explore the correlations among multiple numeric values to evaluate if they will be used in the regression model. Input parameters into the multivariate correlation matrix include means and standard deviations of the functional traits and GPP.

To determine the best fit model for our response variable GPP, we used multiple linear regression with best subsets approach in JMP Pro 16 and STRATA 18 by running a combination of models possible using the regression parameters specified. We ran the best subsets approach for the three trait group classes separately by inputting the variables' mean and standard deviations with the maximum number of terms used in the model set to four. We tested the 1-term, 2-term, 3-term, and 4-term models using a set of criteria for model selection. We used the Bayesian information criterion (BIC) and Akaike information criterion (AIC) to assess model fit (Chakrabarti and Ghosh, 2011). We tested residuals to check normality, correlation coefficients

to check linear correlation between two variables, variance inflation factors (VIF) to check multicollinearity, and P-values to check statistical significance. We also incorporated the PRESS (prediction sum of squares) R^2 statistic as a form of cross validation and to alleviate overfitting (Meyers, 1990). The PRESS metrics give us confidence in the model predictions using the small sample size and highly correlated remote sensing variables. Due to the low number of sites and possible overfitting, machine learning (random forests, CART, etc.) was not the optimal approach to this dataset.

For selection standards for each group, we used the following criteria: lowest BIC and AIC, highest PRESS R^2 , correlation coefficient between -0.55 and 0.55, VIF less than 2, and statistically significant with P-values less than 0.05 for each variable in model. The criteria for the correlation coefficients and variance inflation factors are only applicable to the models with more than one term. We ran the best subsets approach for each trait grouping class to decipher the number of terms to include in the multiple linear regression models and to compare the “best fit” from each group. For the models selected, we assessed model fit by graphs of predicted GPP versus actual GPP for each model trait group.

3.3 Results

3.3.1 Variable correlation coefficients

A multivariate correlation matrix shows the correlation of the target variable (GPP) and predictor variables (NEON remote sensing variables) tested in the models with a color gradient showing negatively correlated in yellow and positively correlated in dark blue (Figure 3.3). The correlation coefficients for EVI, fPAR, and SAVI were high due to similarities in bands used and redundancy of parameters in the equations. EVI is the only index in this group that is directly

calculated from bands. NDVI and ARVI have high correlation because they are similar equations with the addition of the blue band for ARVI, which is considered the adjustment. Furthermore, NDII, NDWI, MSI, and WBI are highly correlated because all are moisture indexes using different bands in the NIR and SWIR regions of the spectrum. Lastly, DSM and DTM are highly correlated (correlation coefficient = .9999) because they include similar information on elevation of the terrain and surface. The correlation coefficients were used to exclude models using highly correlated variables.

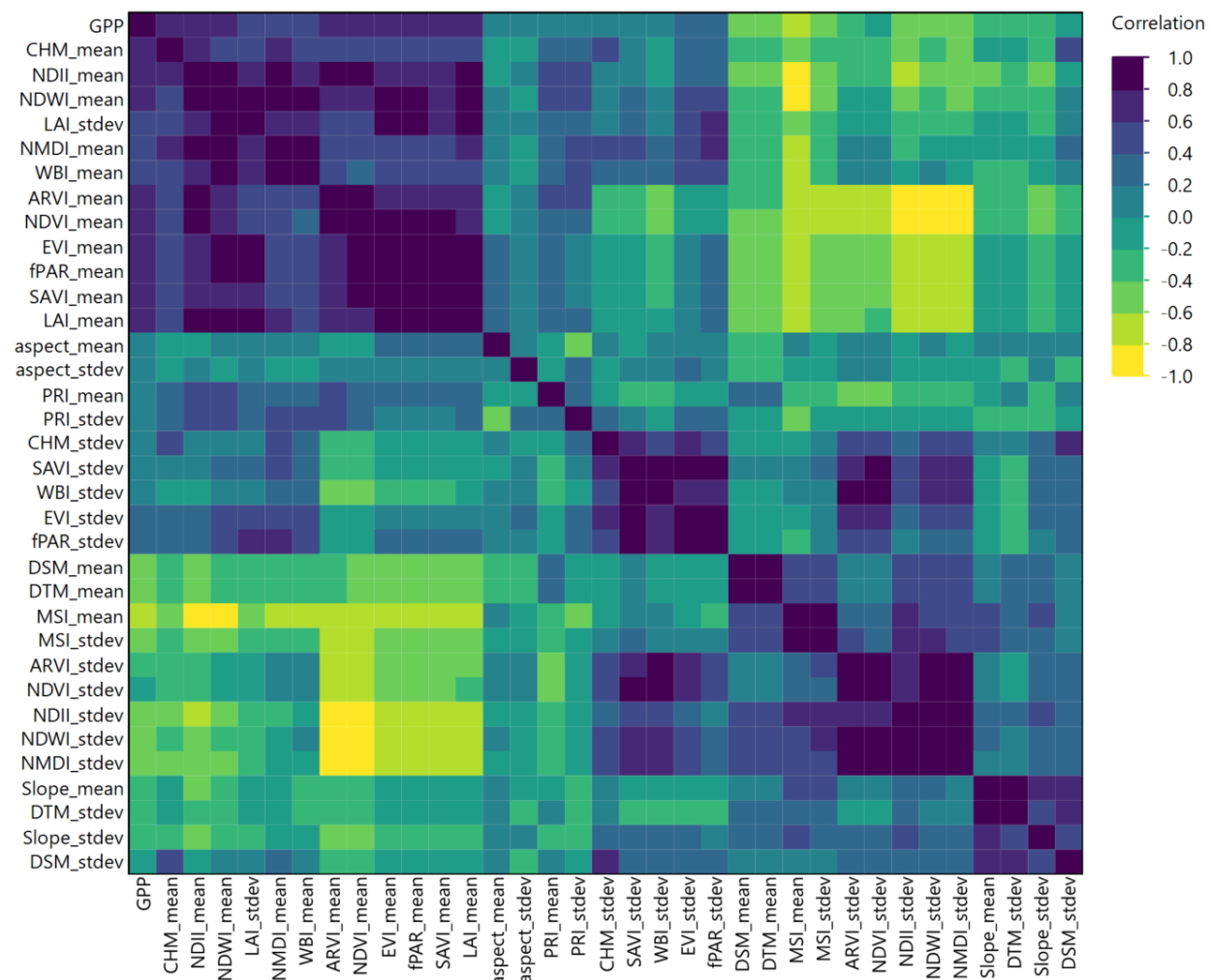


Figure 3.3. Correlation matrix for all RS mean and standard deviation variables and GPP where positive correlations are shown in dark blue and negative correlations shown in yellow.

3.3.2 Multiple linear regression models

The multiple linear regression results for each trait grouping class with added terms are shown in Table 3.5.

Table 3.5. Multiple linear regression model results for each model group with number of terms used in the model, NEON's remote sensing variables, AIC, BIC, R^2 , adjusted R^2 , PRESS R^2 and RMSE (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$). Bolded rows indicate the model with the lowest BIC and highest PRESS R^2 for each model group and correspond to Figures 3.4-3.6.

Model Group	Number of Terms	NEON RS Variables	AIC	BIC	R^2	Adjusted R^2	PRESS R^2	RMSE $\mu\text{mol CO}_2/\text{m}^2/\text{s}$
Morphological & Environmental Traits	1	CHM mean	150.1	152.0	0.47	0.45	0.32	6.51
	2	CHM mean DSM stdev	143.1	145.1	0.67	0.63	0.50	5.32
Physiological Traits	1	fPAR mean	150.8	152.8	0.46	0.43	0.36	6.62
	2	NMDI stdev fPAR stdev	149.6	151.7	0.55	0.50	0.35	6.17
	3	ARVI mean WBI stdev PRI stdev	146.6	148.3	0.66	0.61	0.45	5.48
	4	ARVI mean NMDI mean PRI stdev PRI mean	148.2	149.2	0.70	0.63	0.38	5.36
Combination of Functional Traits	1	CHM mean	150.1	152.0	0.47	0.45	0.32	6.51
	2	CHM mean DSM stdev	143.1	145.1	0.67	0.63	0.50	5.32
	3	CHM mean DSM stdev PRI stdev	137.6	139.3	0.78	0.74	0.60	4.46
	4	CHM mean DSM stdev PRI stdev WBI_mean	129.2	130.1	0.87	0.84	0.75	3.48

For the morphological and environmental trait group regression results, when adding a third term and fourth term to the model, the models fail to meet our selection criteria, thus there are no “good models” to present for the 3-term and 4-term models. Figure 3.4 shows the actual GPP versus predicted GPP for the 2-term model for the morphological/environmental traits group colored by dominant vegetation types.

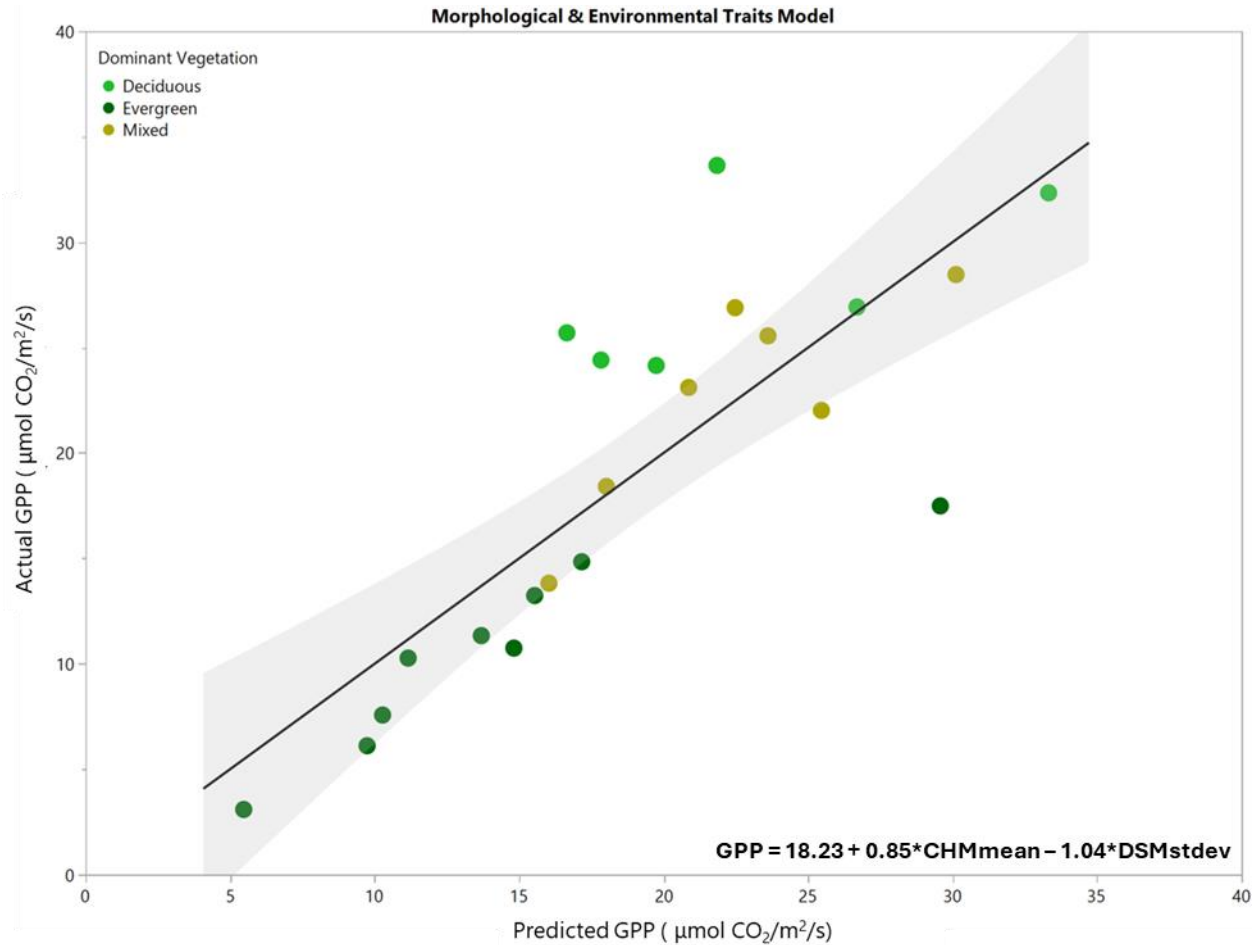


Figure 3.4. Morphological and environmental traits group’s 2-term model graph showing actual GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the y-axis versus predicted GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the x-axis colored by dominant vegetation types (deciduous as light green, evergreen as dark green, and mixed as mustard yellow) including a grey band around the black fit line showing the 95% confidence region.

For the physiological traits group, when adding terms from three to four, we see that the R^2 and adjusted R^2 increase, however the PRESS R^2 decreases, resulting in the 3-term model being the “best fit. The PRESS metrics give us confidence in the model predictions using the small sample size and highly correlated hyperspectral variables. Figure 3.5 shows the actual GPP versus predicted GPP for the 3-term model for the physiological traits group colored by dominant vegetation types.

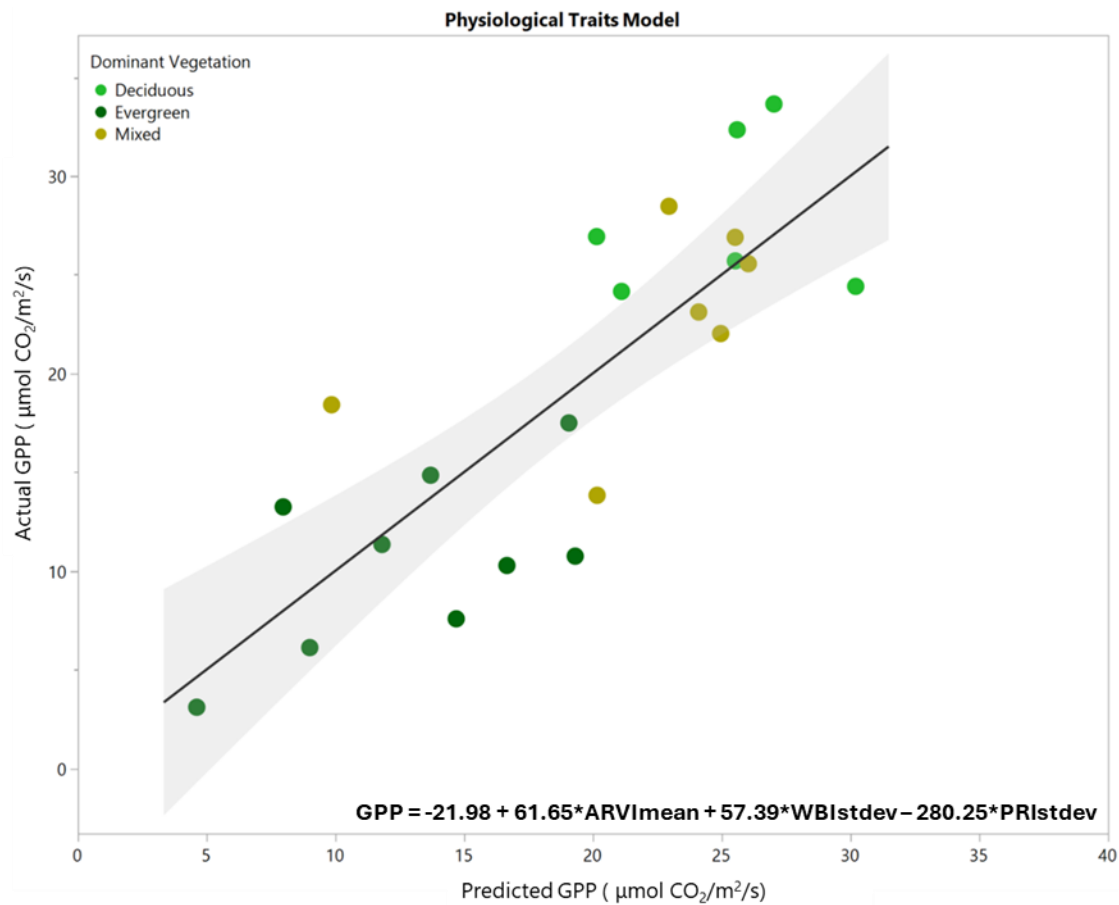


Figure 3.5. Physiological traits group’s 3-term model graph showing actual GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the y-axis versus predicted GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the x-axis colored by dominant vegetation types (deciduous as light green, evergreen as dark green, and mixed as mustard yellow) including a grey band around the black fit line showing the 95% confidence region.

For the combination of functional traits model group, the 4-term model has the lowest BIC and highest PRESS R^2 of all the other models within their respective groups, therefore this model is the most successful predictor of GPP for this dataset. Figure 3.6 shows the actual GPP versus predicted GPP for the 4-term combination of functional traits model group colored by dominant vegetation types.

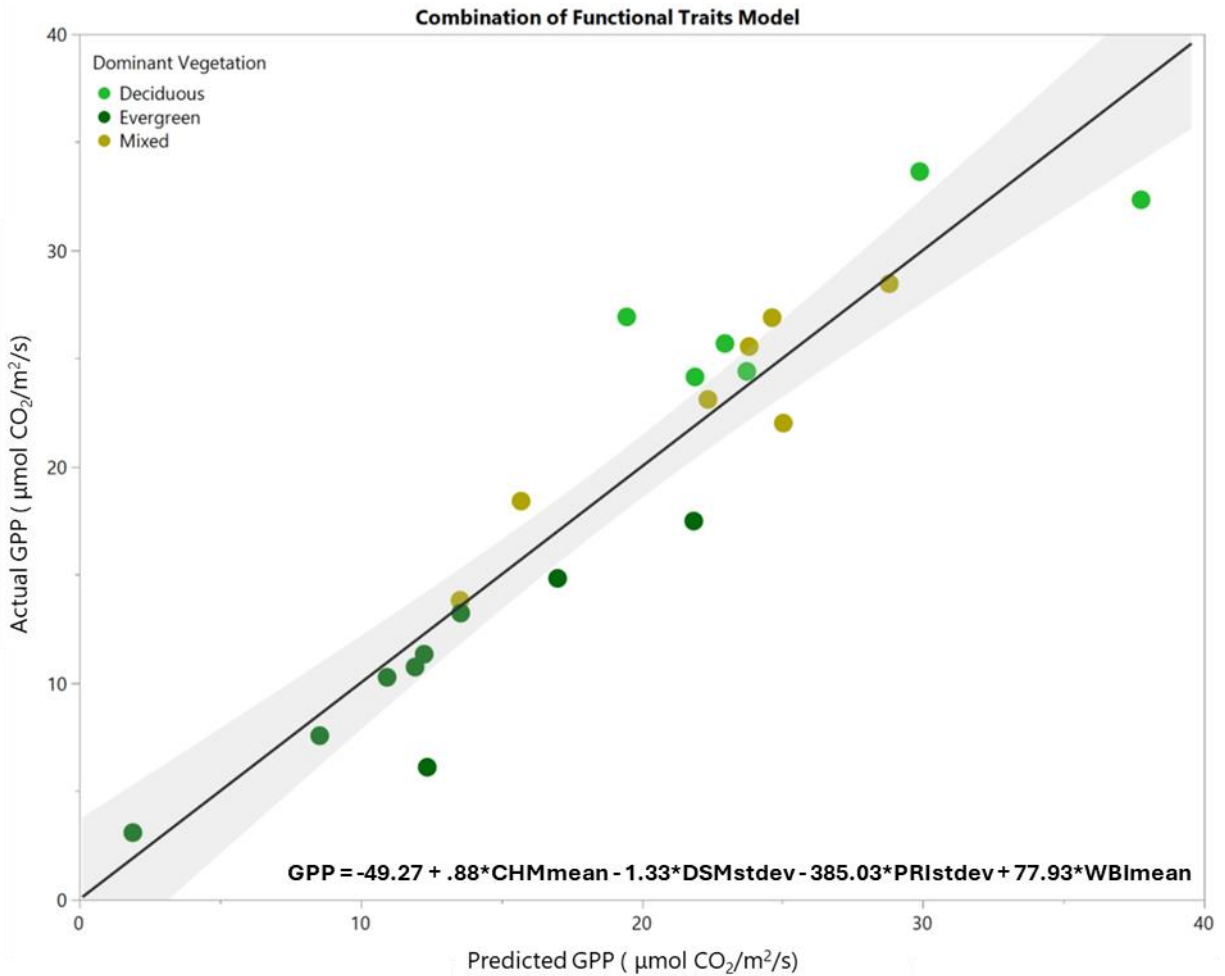


Figure 3.6. Combination of functional traits group's 4-term model graph showing actual GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the y-axis versus predicted GPP (units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) on the x-axis colored by dominant vegetation types (deciduous as light green, evergreen as dark green, and mixed as mustard yellow) including a grey band around the black fit line showing the 95% confidence region.

Figure 3.7 shows the correlation coefficients with a color gradient showing negatively correlated in yellow and positively correlated in dark blue for the input variables used in the 4-term combination of functional traits model group.

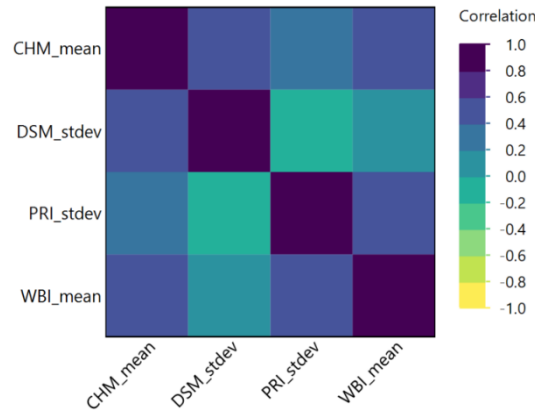


Figure 3.1. Correlation matrix for the NEON RS variables used in the 4-term combination of functional traits group model (CHM mean, DSM standard deviation, PRI standard deviation, and WBI mean) where positive correlations are shown in dark blue and negative correlations shown in yellow. These four NEON variables have correlation coefficients less than 0.5.

3.3.3 Canopy height, age, and productivity

Forest net primary productivity (NPP) varies predictably with stand age, showing a general temporal pattern of rapid increase in young ages, peak growth in middle ages, and slow decline in mature ages (He et al., 2012). In our study, we see a similar trend with GPP and increasing canopy height, where the GPP levels off around a canopy height of 20 m (Figure 3.8), thus suggesting a slowdown in productivity at a given height or age for a forest stand. However, it does seem as though two sites (WREF and UKFS) deter the relationship from a linear trend. UKFS has higher productivity with a moderate canopy height. WREF has very tall trees and low productivity, which could be contributed to the old age of the stand given it is an old-growth forest (NEON, 2018).

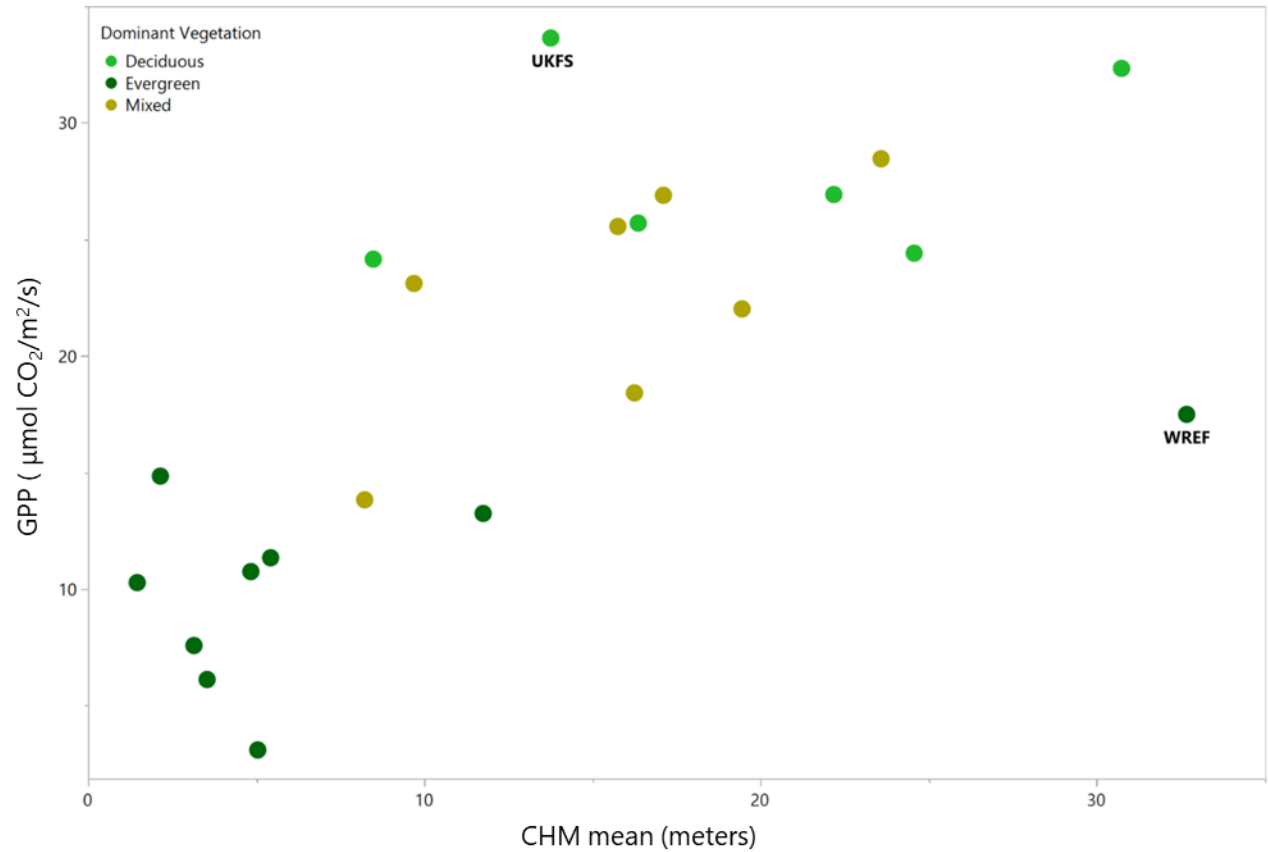


Figure 3.8. GPP (y-axis, units: $\mu\text{mol CO}_2/\text{m}^2/\text{s}$) compared to CHM mean (x-axis, units: meters) colored by dominant vegetation (deciduous as light green, evergreen as dark green, and mixed as brown) show increasing productivity with increasing canopy height. Two sites can be identified as influential sites not following the linear trend of the other sites. UKFS has high GPP and lower mean CHM and WREF has moderate GPP with high mean CHM.

3.4 Discussion

It is important to represent the flux tower data in an effective way to compare to airborne remote sensing observations. A downside to this analysis is the number of sites used because of irregular flux tower measurements or gaps in the data due to quality control, which led to only 22 usable sites out of NEON's total of 47 terrestrial sites.

With the morphological and environmental traits group (lidar only) model producing a PRESS R^2 of 0.50 and the physiological traits group (hyperspectral only) model producing a PRESS R^2 of 0.45, the combination of functional traits model with both lidar and hyperspectral variables exceeds the success of the separated models with a PRESS R^2 of 0.75. While adding terms inherently increases the coefficient of determination, we see the additional terms did not enhance model success (based on our criteria to avoid overfitting: lowest BIC and highest PRESS R^2) for the morphological and environmental traits and physiological traits group because. Our criteria selected models with fewer than the maximum number of variables. For the morphological and environmental group, there are only 10 variables possible for use in the prediction in which most of the variables are highly correlated due to providing similar information on topography. Specifically, DSM and DTM are highly correlated (0.999) and produced high VIF and insignificant P-values in the model results. For the physiological traits group, we see a drop in PRESS R^2 when going from 3-terms to 4-terms. Furthermore, many of the input variables from the hyperspectral indices are correlated (*Section 3.3.1*) – whether using similar bands or looking at similar portions of the electromagnetic spectrum – thus the indices may be providing repeated information causing multicollinearity and overfitting in the models with additional terms of similar indices.

In Figure 3.4, the morphological and environmental traits GPP model shows a better fit for low GPP values, while in Figure 3.5 the physiological traits GPP model shows a better fit for high GPP values. In the combined model in Figure 3.6, the low GPP's and high GPP's are closer to the fit line, showing that using a combination of functional traits enhances model success and forest productivity estimations across a range of GPP values. Thus, using lidar alone or hyperspectral alone in the models were not as strong as they are together. This deepens our understanding of how a range of functional traits (morphological, environmental, and physiological) are important for forest productivity investigations. Furthermore, we have confidence in our representation of the current state of forests in the US by using a dataset with a wide range of forest functional traits. The forest sites used in this study include a wide range of functional traits from different forest ages (young and old growth), species composition (evergreen, deciduous, and mixed forests), spatial range across biomes (as north as Alaska and as south as Alabama), and densities (dispersed and closed canopies).

Figure 3.8 shows that GPP varies linearly with average canopy height (CHM), however there are two sites for which canopy height is not a good predictor of productivity. The two influential points not following the linear trend were the WREF (Wind River Experimental Forest) and UKFS (University of Kansas Field Station) sites. WREF has a max CHM of 73 m with a mean of 32.6 m, which is a high average for CHM compared to other sites. In other words, this site has very tall dispersed evergreen trees but is moderately productive ($GPP = 17.5 \mu\text{mol CO}_2 / \text{m}^2/\text{s}$) in July. Wind River is an old-growth Douglas fir forest plot in the Pacific Northwest, which can experience dry summers and may slow down in productivity during this time of the year and due to its old age (~450 years old) (NEON, 2018). UKFS has a max CHM of 36 m with a mean of 13.7 m. These are reasonable canopy heights for a deciduous plot. However, it is very productive

compared to other sites, with a GPP of $33 \mu\text{mol CO}_2 / \text{m}^2/\text{s}$. The UKFS flux tower is south of an experimental forest in Kansas, which could be contributing to the higher productivity in this forest potentially due to other forest management practices to the north of the plot, or this is simply a highly productive mixed forest site in the Midwest.

Dominant vegetation types describe the range of forest functional traits contributing to productivity. By coloring the points by dominant vegetation type for Figure 3.4-3.6, we can see a similar trend for all models, where the dominantly evergreen sites show lower GPP values, while the dominantly deciduous sites show higher GPP values, with the mixed scattered among with moderate GPP values. Physiologically this makes sense considering that evergreen sites are photosynthesizing all year (thus conserving energy at times), while deciduous sites are only photosynthesizing during the growing season and turn off during the winters. This similar trend is also reflected in Figure 8 where we see lower GPP values for lower CHM mean that are colored as dominantly evergreen, except the WREF site, which is an old-growth forest thus productivity has inevitably slowed down due to age.

All three types of functional traits (morphological, environmental, and physiological) were important predictors of forest productivity. Our results show that a combination of a morphological trait (CHM mean), an environmental trait (DSM stdev), and two physiological traits (PRI stdev and WBI mean) were the strongest predictors of GPP for these forested sites across the US. CHM mean is a measure of the average height of the trees. DSM standard deviation is the variation of the topography and surface. PRI standard deviation is the variation in the green wavelength range of the spectrum used to detect changes in live foliar pigments such as xanthophylls relating to photosynthetic light use efficiency. WBI mean is a measure of canopy water content and used as an indicator for drought stress. Given future projections of the

changing climate, forests may experience lack of water inducing drought stress and increased frequency and severity of forest disturbances inducing structural and morphological changes in canopies (Huang et al., 2019). Therefore, it is imperative to include biophysical measurements of leaf efficiency and moisture content coupled with the structure of the forest canopy.

Our analysis reflects cross-site variability, not temporality or seasonality. Seasonality-wise, these NEON sites are flown during peak greenness. Thus, this model likely won't apply to within-year estimation and seasonal variation. Considering that NEON has been acquiring yearly or bi-yearly airborne acquisitions at these sites since 2017, it would be possible to look at single site variation across time to delineate patterns in productivity across time at a given site.

Forest productivity could be estimated from satellite derived measurements of canopy height, topographic variation, leaf efficiency and moisture content. Canopy height and topographic measurements could be derived from space-borne lidar sensors like ICESat-2 or GEDI. Moisture content and efficiency of vegetation could be derived from multiple space-borne multispectral sensors like Landsat or Sentinel-2 for broad-band indices and hyperspectral sensors like DESIS or PACE for narrow bands indices. However, the spatial resolution of these space-borne sensors hinders the level of specificity of the spectral information derived due to large usually mixed pixels.

The goal of this work was to provide repeatable methods using freely available data that covers a wide range of forest types across the US. We only used NEON's derived products and did not calculate our own suite of indices from the hyperspectral data due to the fact that the data were still being processed for BRDF issues at the time of data collection. Further analysis into other remotely sensed indices (for example chlorophyll content or biomass) were out of the scope of this project but could enhance the model predictions of other productivity metrics.

3.5 Conclusion

In conclusion, this study highlights the critical importance of integrating multiple functional traits—morphological, environmental, and physiological—in models for predicting forest productivity. The combination of lidar and hyperspectral data produced the most accurate predictions, outperforming models alone, with a PRESS R² of 0.75 compared to 0.50 and 0.45 for the morphological-environmental and physiological-only models, respectively. This underscores the value of a comprehensive approach to modeling, which captures a variety of forest characteristics. While our analysis was constrained by the limited number of usable flux tower sites and the inherent challenges of variable correlation and overfitting, it nevertheless offers a robust framework for understanding forest productivity across diverse forest conditions. Future work could further refine these models by incorporating additional spectral indices and exploring temporal variations, leveraging advancements in satellite and airborne remote sensing technologies. This study contributes valuable insights into the complex interplay of forest functional traits and productivity.

References

- Aponte, C., Kasel, S., Nitschke, C. R., Tanase, M. A., Vickers, H., Parker, L., Fedrigo, M., et al. (2020). Structural diversity underpins carbon storage in Australian temperate forests. *Global Ecology and Biogeography*, 29(5), 789–802.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., ... & Wofsy, S. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82(11), 2415-2434.

- Chaturvedi, R., Raghubanshi, A., & Singh, J. (2011). Plant functional traits with particular reference to tropical deciduous forests: A review. *Journal of Biosciences*, 36, 963-981.
- Chen, Y., Shen, W., Gao, S., Zhang, K., Wang, J., & Huang, N. (2019). Estimating deciduous broadleaf forest gross primary productivity by remote sensing data using a random forest regression model. *Journal of Applied Remote Sensing*, 13, 038502 - 038502.
- Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., et al. (2021). Representativeness of eddy-covariance flux footprints for areas surrounding AmeriFlux sites. *Agricultural and Forest Meteorology*, 301–302, 108350.
- Díaz, S., Lavorel, S., de Bello, F., Quétier, F., Grigulis, K., & Robson, T. M. (2007). Incorporating plant functional diversity effects in ecosystem service assessments. *Proceedings of the National Academy of Sciences of the United States of America*, 104(52), 20684–20689.
- Durán, S., Martin, R., Díaz, S., Maitner, B., Malhi, Y., Salinas, N., Shenkin, A., Silman, M., Wiczynski, D., Asner, G., Bentley, L., Savage, V., & Enquist, B. (2019). Informing trait-based ecology by assessing remotely sensed functional diversity across a broad tropical temperature gradient. *Science Advances*, 5.
- Garnier, E., Navas, M.-L., & Grigulis, K. (2015). *Plant functional diversity: Organism traits, community structure, and ecosystem properties*. Oxford University Press.
- Grime, J. P. (1998). Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *The Journal of Ecology*, 86, 902–910.
- Hao, M., Messier, C., Geng, Y., Zhang, C., Zhao, X., & Gadow, K. (2020). Functional traits influence biomass and productivity through multiple mechanisms in a temperate secondary forest. *European Journal of Forest Research*, 1-10.

- He, L., Chen, J., Pan, Y., Birdsey, R., & Kattge, J. (2012). Relationships between net primary productivity and forest stand age in U.S. forests. *Global Biogeochemical Cycles*, 26.
- Huang, C., Anderegg, W. R. L., & Asner, G. P. (2019). Remote sensing of forest die-off in the Anthropocene: From plant ecophysiology to canopy structure. *Remote Sensing of Environment*, 231, 111233.
- Ma, X., Mahecha, M. D., Migliavacca, M., van der Plas, F., Benavides, R., Ratcliffe, S., Kattge, J., et al. (2019). Inferring plant functional diversity from space: the potential of Sentinel-2. *Remote Sensing of Environment*, 233, 111368.
- Madrigal-González, J., Calatayud, J., Ballesteros-Cánovas, J. A., Escudero, A., Cayuela, L., Rueda, M., Ruiz-Benito, P., et al. (2020). Climate reverses directionality in the richness-abundance relationship across the World's main forest biomes. *Nature Communications*, 11(1), 5635.
- Myers, R. H. (1990). Classical and Modern Regression with Applications. United Kingdom: PWS-KENT. pp. 171–178, 370.
- NEON (National Ecological Observatory Network). (2018). Terrestrial Observation System (TOS) Site Characterization Report: Domain 16. NEON.DOC.003899vA
- NEON (National Ecological Observatory Network). Canopy water indices - mosaic (DP3.30019.001). <https://data.neonscience.org> (accessed September 25, 2021)
- NEON (National Ecological Observatory Network). Ecosystem structure (DP3.30015.001). <https://data.neonscience.org> (accessed September 25, 2021)
- NEON (National Ecological Observatory Network). Elevation - LiDAR (DP3.30024.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). fPAR - spectrometer - mosaic (DP3.30014.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). High-resolution orthorectified camera imagery mosaic (DP3.30014.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). LAI - spectrometer - mosaic (DP3.30012.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). Slope and Aspect - LiDAR (DP3.30025.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). Vegetation indices - spectrometer - mosaic (DP3.30026.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). Bundled data products - eddy covariance (DP4.00200.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). Photosynthetically active radiation (PAR) (DP1.00024.001). <https://data.neonscience.org> (accessed September 25, 2021)

NEON (National Ecological Observatory Network). Single aspirated air temperature (DP1.00002.001). <https://data.neonscience.org> (accessed September 25, 2021)

Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J. P., Scott, R. L., & Torn, M. S. (2018). The AmeriFlux network: A coalition of the willing. *Agricultural and Forest Meteorology*, 249, 444–456.

Pierrat, Z., Bortnik, J., Johnson, B., Barr, A., Magney, T., Bowling, D., Parazoo, N., Frankenberg, C., Seibt, U., & Stutz, J. (2022). Forests for forests: combining vegetation indices with solar-induced chlorophyll fluorescence in random forest models improves

- gross primary productivity prediction in the boreal forest. *Environmental Research Letters*, 17.
- Schneider, F. D., Morsdorf, F., Schmid, B., Petchey, O. L., Hueni, A., Schimel, D. S., & Schaepman, M. E. (2017). Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nature Communications*, 8(1), 1441.
- Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., & Siemann, E. (1997). The influence of functional diversity and composition on ecosystem processes. *Science*, 277(5330), 1300–1302.
- Wang, R., & Gamon, J. A. (2019). Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, 231, 111218.
- Xiao, J., Zhuang, Q., Law, B. E., Chen, J., Baldocchi, D. D., Cook, D. R., Oren, R., et al. (2010). A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data. *Remote Sensing of Environment*, 114(3), 576–591.

Chapter 4 – Impact of cankerworm infestation on Sentinel-2 and airborne laser scanning data

Abstract

Disturbances, like insect infestations, affect forest health by altering tree physiology and structure, leading to stress, defoliation, or mortality. These disturbances impact forest ecosystems' structure, composition, and function. Remote sensing technologies, including optical imagery and lidar, offer effective methods for monitoring and mapping tree mortality over large areas, which helps assess carbon emissions and track ecosystem changes. This study explores the benefits of integrating physiological traits from Sentinel-2 MSI (S2) multispectral indices with morphological traits from airborne lidar to detect insect-induced impacts using machine learning methods. The data was collected on July 26, 2023, over a mixed deciduous stand that experienced tree mortality from a fall cankerworm infestation in 2020 at Mountain Lake Biological Station in Southwest Virginia. The airborne laser scanning acquisition was flown at solar noon with an altitude of 90 m $\pm$ 3 m, speed of 12 m/s, and overlap of 55%. With these mission parameters, the lidar dataset produced a point density of 1,200 points/m². A suite of 14 forest metrics was derived from the airborne laser scanning data and 28 vegetation and water multispectral indices from S2. Random forest classification models were generated for three modeling groups: (1) lidar metrics only, (2) S2 multispectral indices only, (3) combination of lidar metrics and S2 multispectral indices. The findings show that combining these datasets improves tree mortality classification accuracy by 6%. While the S2-only model, which relied on NDVI, effectively differentiates between live and dead trees, it does not fully capture post-mortality canopy gaps or vegetation understory emergence. In contrast, lidar metrics like canopy bincentile 50% (Cb50) and canopy standard deviation (Cstd) provide detailed vertical structure information crucial for distinguishing dead from live trees. High LiDAR point density and penetration depth were particularly important for capturing detailed forest features and understanding the impact of tree mortality on canopy structure. Although all models tend to over predict pixels with dead trees, integrating lidar data enhances predictions by offering valuable structural context.

Keywords: forest, tree mortality, lidar, multispectral, UAS

4.1 Introduction

Disturbances like wildfires, disease, insect infestation and extreme weather events lead to changes in plant physiology, chemistry, and structure causing stress, defoliation, or mortality in forests. These disturbances directly impact the structure, composition, and function of forested ecosystems. Insect-induced mortality is a significant forest disturbance affecting global forest health and ecosystem dynamics (Meng et al., 2018).

Remote sensing offers a practical method for mapping effects of tree mortality over large areas, surpassing the limitations of time-consuming field surveys. Furthermore, it enables mapping of the extent and ecological patterns of tree mortality and damage to evaluate potential carbon emissions and track ecosystem changes following tree mortality (Broderick and Asner, 2017; Fernández-Alonso et al., 2022; Huang et al., 2019; Navarro-Cerrillo et al., 2019; Zhao et al., 2019).

Post-disturbance, trees may experience discoloration or defoliation, therefore changing leaf color or exposing branches which affects the components of spectral reflectance (Huang et al., 2019; Navarro-Cerrillo et al., 2019; Wulder et al., 2006). Shifts of vegetation canopy spectra from healthy green leaf to a brown leaf show increases in reflectance within the visible range, decreases in the near-infrared (NIR), and increases in the short-wave infrared (SWIR) portion of the spectrum (Huang et al., 2019). The relatively higher canopy reflectance in the visible portion of the spectrum is due to reductions in chlorophyll and carotenoid concentrations decreasing spectral absorbance, commonly known as the blue shift (Rock et al., 1988). Reflectance of the visible region is shown to increase gradually after tree mortality (White et al., 2007). With vegetation greenness decreasing, there is a reduction in the typically high reflectance features in the NIR regions (especially the NIR shoulder) due to internal structure of the leaf. The typically

high absorption feature in the SWIR region of the spectrum is indicative of water content, thus the increases in this region would indicate lower moisture content in the leaves which could be induced by drought stress or damages from disturbances (Huang et al., 2019; White et al., 2007)

Spectral information is effective in estimating tree mortality, with inclusion of multispectral indices improving predictions, specifically NDVI (Normalized Difference Vegetation Index), which capitalizes on vegetation greenness and health (Bergmüller and Vanderwel, 2022; Bishop et al., 2014; Goldbergs and Upenieks, 2024; Li et al., 2023). However, spectral measurements have been shown to overpredict tree mortality, thus further advancements are needed to achieve more precise predictions at the individual tree level (Bergmüller and Vanderwel, 2022). Satellite optical imagery (like Sentinel-2 or Landsat) only provide a two-dimensional (2D) assessment of the top-of-canopy reflectance at a moderate resolution, excluding information within the canopy.

LiDAR (Light Detection and Ranging) captures the vertical and horizontal complexities of forest environments by providing three-dimensional (3D) spatial arrangement of forest canopy height, density, and terrain (Campbell et al., 2022). Airborne lidar offers high resolution data products that can be used to assess structural characteristics of forests. Applications of lidar for post-disturbance analyses have been shown to provide valuable information on tree position and height in complex landscapes (Wing et al., 2010). Multitemporal lidar studies following mortality of dominant trees reveal an increase in gap fraction and openness of the canopy and a decrease in tree heights, which become more pronounced over time (Hardiman et al., 2013). Lidar can provide detailed information on canopy gap dynamics that are induced by tree mortality post-disturbance (Dalagnol et al., 2021).

Most insect-induced forest disturbance studies using Sentinel-2 (S2) focus on the bark beetle affecting spruce trees (Bárta et al., 2021; Bozzini et al., 2023; Cessna et al., 2021; Goldbergs and

Upenieks, 2024), as it is prevalent in boreal forests, although it does not always lead to mortality. Studies have indicated that using a time-series of S2 multispectral indices, rather than individual bands, has greater potential for the early detection or prediction of insect infestations (Bárta et al., 2021; Bergmüller and Vanderwel, 2022; Li et al., 2023). However, this approach coupled with lidar data has not been thoroughly examined following extensive fall cankerworm outbreaks, which cause significant tree mortality that is patchier and more episodic in complex ecosystems.

The fusion of lidar and optical imagery offers a more holistic view of complex forested ecosystems by providing morphological and physiological status of vegetation (Dalponte et al., 2008; Fernández-Alonso et al., 2022). Lidar can be used to inform and refine spectral data interpretation by providing the structural context, while spectral data can complement lidar by revealing additional information about vegetation types and conditions that may affect the observed structural characteristics (Huang et al., 2019; Matasci et al., 2018). This integration approach enables better monitoring of forest dynamics, more detailed forest assessments, improved biomass and productivity estimates, and better monitoring of ecological changes, ultimately enhancing forest management and conservation efforts (Huang et al., 2019; Li et al., 2023; Navarro-Cerrillo et al., 2019; Xu et al., 2015; Zhao et al., 2019).

When combined, these datasets have been shown to improve characterization of key forest ecosystem attributes (Bishop et al., 2014; Fernández-Alonso et al., 2022; Meng et al., 2018; Romero Rameriz et al., 2018; Xu et al., 2015; Zhao et al., 2019). However, the biophysical effects that underpin the spectral and structural signatures of insect-induced tree mortality are not well understood, and comparisons of different remote sensing metrics have been limited (Meng et al., 2018; Wulder et al., 2006). Advanced data pattern recognition techniques, such as machine

learning, may enhance the unraveling of the complex relationship between the ecophysiology of tree mortality and remote sensing (Ferreira et al., 2018).

The objective of this study is to use machine learning classification to look at the impact of insect-induced tree mortality in a mixed deciduous stand with separated and combined models of lidar metrics from airborne laser scanning data and multispectral indices from Sentinel-2 MSI data.

4.2 Materials and Methods

4.2.1 Study area

Mountain Lake Biological Station (MLBS) is located in a mixed deciduous forest within the Jefferson National Forest of southwest Virginia (latitude and longitude: 37.378314, -80.524847) (Figure 4.1). This station is owned by the Forest Service and hosted by the University of Virginia and is also a terrestrial field site for NEON (National Ecological Observatory Network) that has been collecting data since 2017 (<https://www.neonscience.org/field-sites/mlbs>). The forest is a second-growth forest dominated by red maple (*Acer rubrum*) and white oak (*Quercus alba*) with pockets of eastern white pine (*Pinus strobus*) and an understory predominantly composed of cinnamon fern (*Osmundastrum cinnamomeum*), witch-hazel (*Hamamelis virginiana*) and shadbush (*Amelanchier spp.*). It has infertile, acidic soils that are sandy loams. The site located at 1,184 m elevation experiences cold winters and moderate summers, with an average annual temperature of 9° C and an average annual precipitation of 122 cm (Wilbur et al., 2021). N. red oak, stripped maple in understory with pockets of hemlock and heavy cinnamon fern in the understory

The forest at this field station has experienced an infestation of the fall cankerworm (*Alsophila pometaria*) that began in the early summers of 2018 and 2019 (Wilbur et al., 2021). This North American native pest causes patchy and episodic outbreaks that can defoliate the canopy, typically affecting hardwood species (Darr and Coyle, 2021). The VDOF Forest Health Staff reported spotty defoliation from the fall cankerworm that mainly caused mortality on the older trees already weakened by other factors, while others recovered in subsequent years (Chamberlin and Mooneyham, 2020).

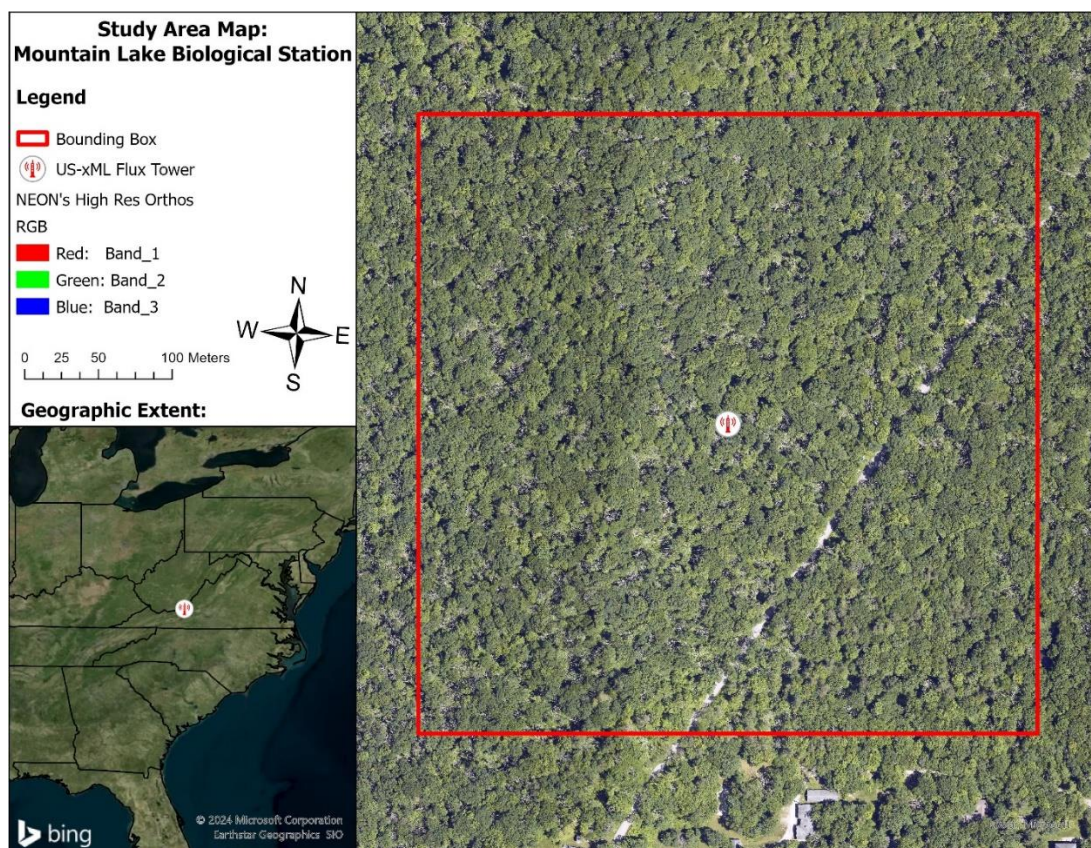


Figure 4.2. Study area map of Mountain Lake Biological station showing the mission area of the 420 m x 420 m bounding box centered on US-xML flux tower overlain on NEON's high resolution orthorectified camera imagery mosaics (NEON product number DP3.30010.00) including the geographic extent of the location in the Eastern US with imagery from Bing ESRI basemaps.

4.2.2 Datasets

4.2.2.1 Airborne lidar

The UAS (Uncrewed Aerial System) used in this study included the SkyFish M6 hexacopter mounted with a Headwall's sensor package with a Velodyne LiDAR puck-LITE. The Velodyne LiDAR puck-LITE is a dual-return system with 16 scan lines at wavelength 903 nm. The LiDAR has a range of 100 m and generates up to ~600,000 points/second, across a 360° horizontal field of view and a 30° vertical field of view.

On July 26, 2023, we employed an UAS acquisition in the mission area at MLBS under clear sky conditions at 13:20 EST. Our study area is a 420 m x 420 m mission area centered on the AmeriFlux tower US-xML (Figure 4.1). This mission area size was chosen to match flux tower coverage and the capabilities of the UAS given altitude and battery power. The mission was flown at solar noon with an altitude of 90 m +/- 3 m, speed of 12 m/s, and overlap of 55%. With these mission parameters, the LiDAR dataset produced a point density of 1,200 points/m². Software used in data processing included mission planning in UgCS, pre-flight calibrations in Headwall Hyperspec III, high-performance GPS/IMU post-processing in POSPac UAV 8.9, and LiDAR post-processing in Headwall HPI Trajectory Plotter, Headwall Lidar Tools, CloudCompare v2.12.4, ArcGIS Pro 3.1.0, VeloView 4.1.3, and lastools.

4.2.2.2 Multispectral imagery – Sentinel-2 MSI

Freely available multispectral data from Sentinel-2 MSI near the UAS acquisition was used to derive satellite-derived spectral indices. Sentinel-2 MSI provides wide-swath, high-resolution, multispectral images supporting the Copernicus Land Monitoring studies. The sensor samples 13 spectral bands at different resolutions every five days across the visible, near infrared (NIR), and

shortwave infrared (SWIR) portions of the spectrum. Red, green, blue, and NIR bands are at 10 m resolution; red-edge and SWIR bands are at 20 m resolution; and coastal aerosol, water vapor, and SWIR cirrus bands are at 60 m resolution.

4.2.3 Methods

4.2.3.1 Forest canopy metrics from lidar

The Velodyne LiDAR puck-LITE sensor provides the data as pcap (packet capture) files directly from the sensor that can be viewed in VeloView as a video. The Headwall HPI Trajectory Plotter in MATLAB R2023a produces graphs of mission parameters with time stamps, speed, altitude, latitude, longitude, roll, pitch, yaw and flight lines that are used to delineate the time stamps of the data collection mission area. High-performance GPS/IMU post processing was done in POSPac UAV 8.9 applied to the GPS/IMU data from the sensor with a local base station to create a sbet (smoothed best estimate of trajectory) file that is required for the next steps of post-processing. Next, Headwall's LidarTools is used to create a las file by inputting the time stamps and the processing parameters from a lidar.ini file that is specific to our sensor. The las file of the mission area was loaded into CloudCompare to clip out noisy data (random points above trees from sun flecks on the sensor or particles in the sky) using the segmentation tool (Figure 4.2). The segmented las file was loaded into lastools to normalize the heights and derive forest metrics. The lasheight tool was used to normalize the heights by locating ground pixels to set the point values from elevation to heights above ground. The lascanopy tool was used to derive a suite of forest metrics (Table 4.1). Forest metrics were output as 10 m rasters and clipped to the 420 m by 420 m bounding box to get rid of edge effects.

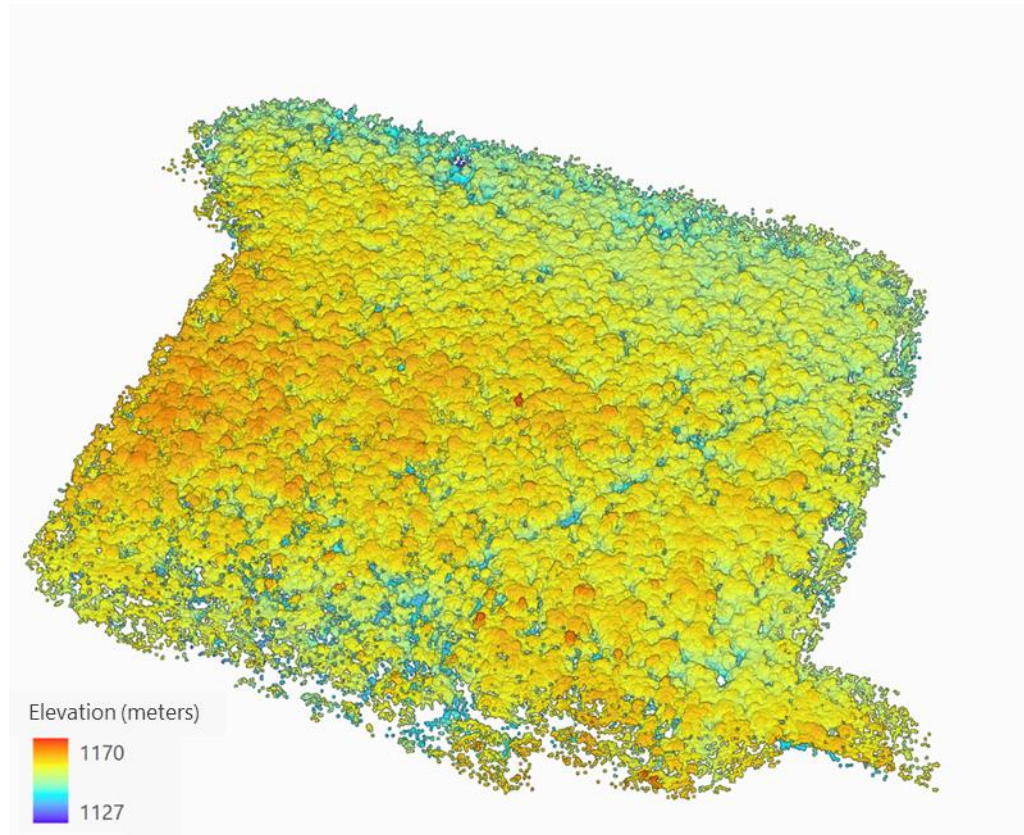


Figure 4.3. Lidar point cloud showing elevation from the UAS acquisition on July 27th, 2023 at MLBS.

Table 4.1. Forest metrics from lidar points calculated in las tools and exported as raster images with 10 m spatial resolution and clipped to 420 m by 420 m bounding box.

Forest Metric Group	Forest Metric Abbrev.	Description
Canopy Statistics	Cavg	for each grid cell compute average
	Cmin	for each grid cell keep lowest value
	Cmax	for each grid cell keep highest value
	Cstd	for each grid cell compute standard deviation
Height Percentiles	Cp99	compute a number of height percentiles ('-p 5 10 25 50 75 90'), giving the average of all heights above the cutoff at breast height (2 m).
	Cp95	
	Cp90	
	Cp75	
	Cp50	
	Cp25	
Height Bincntiles	Cb20	the percentage or fraction of points between the height cutoff at breast height (2 m) and the maximum height.
	Cb50	
	Cb70	
	Cb90	

4.2.3.2 Multispectral indices from S2

Imagery from Sentinel-2 was collected on July 27, 2023, the day after the airborne UAS acquisition over MLBS. Data was downloaded as the surface reflectance product from the Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). Image data processing was done in the Sentinel Application Platforms (SNAP) software package (<https://earth.esa.int/eogateway/tools/snap>). The S2 image was loaded into SNAP and clipped to the 420 m x 420 m study area resulting in an image of 42 pixels by 41 pixels. The reason for the resulting pixels is due to misalignment of S2 images with the study area boundary. Next, the Raster>Geometric>Resampling tool was used to resample the 20 m and 60 m pixels to 10 m pixels. A suite of indices were calculated using SNAP's Thematic Land Processor tools: biophysical processor, vegetation radiometric indices, and water radiometric indices (Table 4.2). Figure 4.3 shows an example of the S2 raster images over the mission area for NDVI.

Table 4.2. Suite of S2 multispectral indices calculated in SNAP's Thematic Land Processor including the name abbreviation, full name, index description, and the equation, if applicable. NIR: near infrared; R = red; G = green; middle IR= middle infrared; R1 = red edge 1; R2 = red edge 3; R3 = red edge 3.

Name	Full Name	Description	Equation
NDVI	Normalized Difference Vegetation Index	This index is a measurement for the photosynthetic activity and is strongly in correlation with density and vitality of the vegetation	$\frac{(NIR - R)}{(NIR + R)}$
DVI	Difference Vegetation Index	This index is sensitive to the amount of vegetation	$(NIR - R)$
RVI	Ratio Vegetation Index	The simplest ratio-based index, it is also called the Simple Ratio (SR). It indicates the amount of vegetation. It also reduces the effects of atmosphere and topography	$\frac{NIR}{R}$
PVI	Perpendicular Vegetation Index	This index could be considered a generalization of the DVI, which allows for soil lines of different slopes.	$\sin(a) * NIR - \cos(a) * R$
IPVI	Infrared Percentage Vegetation Index	This index is functionally equivalent to NDVI and RVI, but it only ranges in value from 0.0 - 1.0	$1/2 * (NDVI + 1)$

WDVI	Weighted Difference Vegetation Index	This index is a mathematically simpler version of PVI, but it has an unrestricted range	$(NIR - g * R)$ <i>where: g is the slope of the soil line.</i>
TNDVI	Transformed Normalized Difference Vegetation Index	This index indicates a relation between the amount of green biomass that is found in a pixel	$\sqrt{(NIR - R) / (NIR + R) + 0.5}$
GNDVI	Green Normalized Difference Vegetation Index	This index is more sensible than NDVI to identify different concentration rates of chlorophyll, which is highly correlated at nitrogen.	$\frac{(NIR - green)}{(NIR + green)}$
ARVI	Atmospherically Resistant Vegetation Index	This index takes advantage of the different scattering responses from the blue and R band to retrieve information regarding the atmosphere opacity	$\frac{(NIR - rb)}{(NIR + rb)}$ <i>where: rb = (red) - gamma * (blue - red), with gamma = 1</i>
NDI45	Normalized Difference Index	This index algorithm is more linear, with less saturation at higher values than the NDVI.	$\frac{(NIR - R)}{(NIR + R)}$
MCARI	Modified Chlorophyll Absorption Ratio Index	This index was developed to be responsive to both leaf chlorophyll concentrations and ground reflectance.	$((R2 - R1) - 0.2 * (R2 - green)) * (R2 / R1)$
REIP	Red-Edge Inflection Point Index	This index was developed for applications in biomass and nitrogen (N) uptake measurement/management in heterogeneous fields.	$700 + 40 * \frac{\left(\frac{(R1 + NIR)}{2} - R2\right)}{R3 - R2}$
S2REP	Sentinel-2 Red-Edge Position Index	This index is based on linear interpolation as presented by Guyot and Baret (1988). The reflectance at the inflexion point is estimated and in turn, the REP is retrieved through interpolation of S-2 band 5 and 6 which are positioned on the RE slope.	$705 + 35 * \frac{\left(\frac{(R1 + NIR)}{2} - R2\right)}{R3 - R2}$
IRECI	Inverted Red-Edge Chlorophyll Index	This index incorporates the reflectance in four bands to estimate canopy chlorophyll content	$\frac{(NIR - R1)}{(R2 / R3)}$
PSSRa	Pigment Specific Simple Ratio (chlorophyll) index	This index was developed to investigate the potential of a range of spectral approaches for quantifying pigments at the scale of the whole plant canopy.	$\frac{NIR}{R}$
NDWI	Normalized Difference Water Index	This index is a measure of liquid water molecules in vegetation canopies that interacted with the incoming solar radiation.	$\frac{(NIR - middle IR)}{(NIR + middle IR)}$
NDWI2	Second Normalized Difference Water Index	This index was developed to detect surface waters in wetland environments and to allow for the measurement of surface water extent.	$\frac{(green - NIR)}{(green + NIR)}$
MNDWI	Modified Normalized	This index was developed to enhance open water features, while efficiently	$\frac{(green - middle IR)}{(green + middle IR)}$

	Difference Water Index	suppressing and even removing built-up land noise as well as vegetation and soil noise.	
SAVI	Soil Adjusted Vegetation Index	This index attempts to be a hybrid between the ratio-based indices and the perpendicular indices	$(1 + L) * \frac{(NIR - R)}{(NIR + R + L)}$ where: <i>L</i> is a correction factor which ranges from 0 for very high vegetation cover to 1 for very low vegetation cover. The most typically used value is 0.5 which is for intermediate vegetation cover.
TSAVI	Transformed Soil Adjusted Vegetation Index	This index assumes that the soil line has arbitrary slope and intercept, and it makes use of these values to adjust the vegetation index	$\frac{s * (NIR - s * R - a)}{(s * NIR + R - a * s + X * (1 + s * s))}$ where: - <i>a</i> is the soil line intercept - <i>s</i> is the soil line slope - <i>X</i> is the adjustment factor to minimize soil noise.
MSAVI	Modified Soil Adjusted Vegetation Index	The basic idea of MSAVI was to provide a variable correction factor <i>L</i> . The correction factor used is based on the product of NDVI and WDV	$(1 + L) * \frac{(NIR - R)}{(NIR + R + L)}$ where: $L = 1 - 2 * s * NDVI * WDV$ and <i>s</i> is the soil line slope
MSAVI2	Second Modified Soil Adjusted Vegetation Index	Basically, it is used as an iterative process and substitute $1 - MSAVI(n-1)$ as the <i>L</i> factor in $MSAVI(n)$. They then inductively solve the iteration where $MSAVI(n) = MSAVI(n-1)$	$\frac{1/2 * (2 * NIR + 1)}{\sqrt{(2 * NIR + 1) * (2 * NIR + 1) - 8 * (NIR - R)}}$
LAI	Leaf Area Index	This index is defined as half the developed area of photosynthetically active elements of the vegetation per unit horizontal ground area.	Calculated using PROSAIL radiative transfer model and neural networks
CAB	Chlorophyll Content	This index estimates the chlorophyll content in the leaf ($\mu g.cm^{-2}$)	
CWC	Canopy Water Content	This index is defined as the mass of water per unit ground area ($g.m^{-2}$)	
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation	This index corresponds to the fraction of photosynthetically active radiation absorbed by the canopy	
FVC	Fraction Vegetation Cover	This index is computed from LAI and used to separate vegetation and soil in energy balance processes, including temperature and evapotranspiration.	

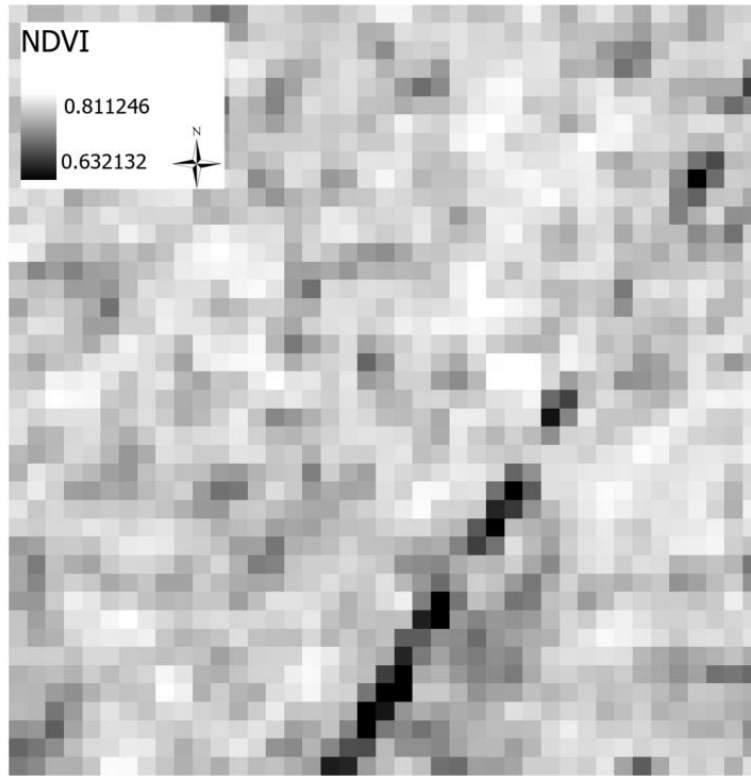


Figure 4.4. S2 image example of the normalized difference vegetation index (NDVI) image.

4.2.3.3 Tree mortality

Due to tree mortality this area has experienced, we generated a classification image as our “truth” using visual interpretation to identify dead and live plant material present within each pixel (Figure 4.4). An empty grid with 42 pixels by 41 pixels matching the S2 and lidar raster images was overlain on NEON’s high resolution orthorectified camera imagery mosaics (DP3.30010.00) from 2021 and 2022. Pixels were classified as dead if the pixel had a full dead tree present within the pixel boundary and any dead branches, thus some pixels will have a mix of dead and live material, however the pixels classified as live do not have dead material present in the image so mixed pixels are less likely. Figure 4.5 shows examples of a dead and live 10 m pixel boundary overlain onto NEON’s RGB high resolution orthorectified camera imagery

mosaic from 2022 for visualization purposes. The dead pixel includes an entire dead tree along with shadows and scattered greenness in the understory. The live pixel shows a few healthy trees with shadows in the canopy gaps. Figure 4.6 overlays the 2D point clouds within the pixel boundary (~120,000 total points within each 10 m pixel), showing that the dead pixels have more points at the lower heights due to the understory below the dead branches, and the live pixels have higher points mainly in the upper canopy due to crown closure from healthy trees.

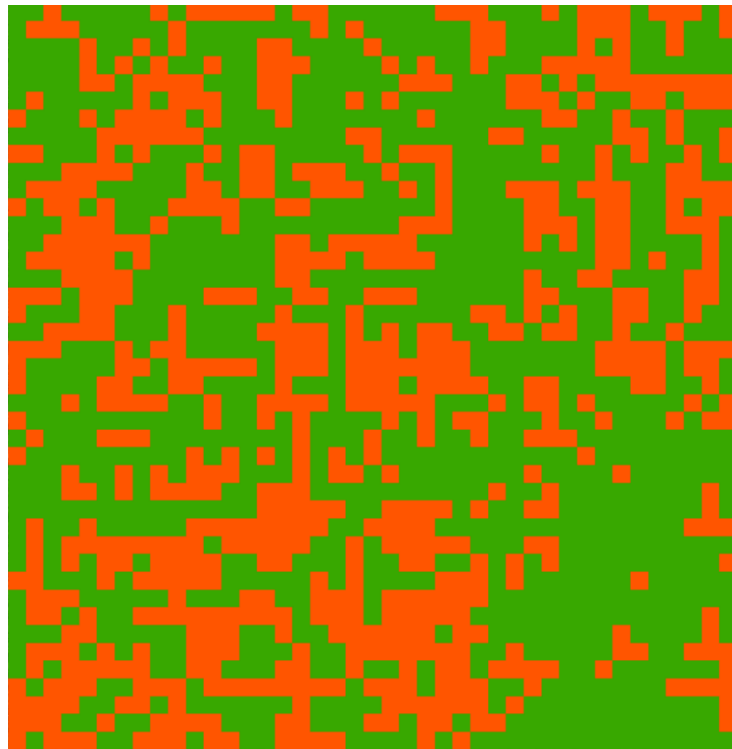


Figure 4.4. Classification image of tree mortality showing pixels with dead material as orange and live material as green.



Figure 4.5. Example of dead and live 10 m pixels overlain on NEON's high resolution orthorectified camera imagery mosaics (NEON product number DP3.30010.00).

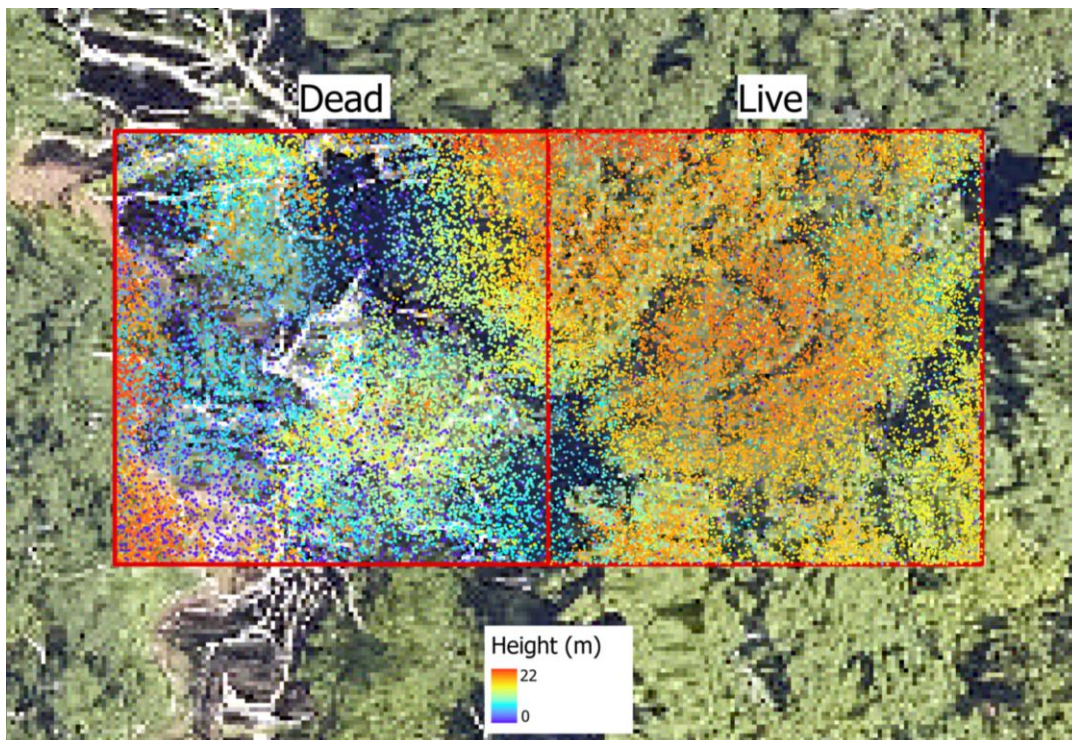


Figure 4.6. Example of dead and live 10 m pixels with 2D point clouds showing height in meters overlain on NEON's high resolution orthorectified camera imagery mosaics (NEON product number DP3.30010.00).

4.2.3.4 Statistical analysis

The Kolmogorov-Smirnov (KS) test is a nonparametric approach to compare distributions of two unpaired groups of data and is generally unaffected by sample size (Fasano and Franceschini, 1987). The KS test is sensitive to any differences in two distributions and computes p-values to test the null hypothesis that the two groups have the same distribution. Substantial differences in shape, spread or median will result in a small p-value and the maximum absolute deviation (D) between the empirical distribution functions (EDFs) for the two groups. The significance of D+ or D- provides information on which distribution's central tendency is lower or higher. We used JMP Pro 16 to calculate the KS statistics for the lidar metrics and S2 indices with the class set as the binary dead and live raster.

Random forest machine learning classification models in Salford Predictive Modeler 8 (SPM) were used in this study. The response variable is set to the live and dead classification and the predictor variables are a variation of input RS variables. To determine the strength of a variation of lidar metrics and S2 multispectral indices as predictor variables for the models, we chose to group into two separate models (lidar only and S2 only) and a combined model (lidar and S2 together). Table 4.3 depicts the groupings used for the remote sensing variables used as model predictors.

Table 4.3. Modeling groups and remote sensing variables used in each machine learning model group. The lidar group has 14 variables, the S2 group has 28 variables, and the combined has 42 variables.

Modeling Group	Remote Sensing Variables
Lidar Metrics	Cmax, Cmin, Cavg, Cstd, Cp25, Cp50, Cp75, Cp90, Cp95, Cp99, Cb20, Cb50, Cb70, Cb90
S2 Multispectral Indices	ARVI, DVI, GEMI, GNDVI, IPVI, IRECI, MCARI, MSAVI, MSAVI2, NDI45, NDPI, NDTI, NDVI, NDWI, PSSRA, PVI, REIP, RVI, S2REP, SAVI, TNDVI, TSAVI, WDV, LAI, CAB, CWC, FAPAR, FVC
Combination of Lidar Metrics & S2 Indices	Cmax, Cmin, Cavg, Cstd, Cp25, Cp50, Cp75, Cp90, Cp95, Cp99, Cb20, Cb50, Cb70, Cb90, ARVI, DVI, GEMI, GNDVI, IPVI, IRECI, MCARI, MSAVI, MSAVI2, NDI45, NDPI, NDTI, NDVI, NDWI, PSSRA, PVI, REIP, RVI, S2REP, SAVI, TNDVI, TSAVI, WDV, LAI, LAI, LAI_CWC, FAPAR, FVC

4.2 Results

4.3.1 Kolmogorov-Smirnov tests

If the KS statistics are low and p-value is high, then the variable is considered to have an insignificant distribution for dead and live pixels. For this specific dataset, if Prob > D+ has a significant p-value, then the dead trees have significantly lower value compared to live trees, while if Prob > D- is significant, then the dead trees have significantly higher values compared to live trees. For example, NDVI has lower values for dead trees compared to the values for the live trees. Table 4.4 depicts which lidar and S2 metrics are insignificant, significant in the positive direction (Prob > D+), or significant in the negative direction (Prob > D-).

Table 4.4. KS statistics showing which lidar and S2 multispectral indices were insignificant, significant in the positive direction (Prob >D+), or significant in the negative direction (Prob >D-).

KS Statistics	Lidar	S2 Spectral Indices
Insignificant	Cmax Cp95, Cp99	REIP, S2REP
Prob > D-	Cstd, Cb20, Cb50, Cb70, Cb90	NDPI, NDTI
Prob > D+	Cmin, Cavg, Cp25, Cp50, Cp75, Cp90	ARVI, DVI, GEMI, GNDVI, IPVI, IRECI, MCARI, MSAVI, MSAVI2, NDI45, NDPI, NDTI, NDVI, NDWI, PSSRA, PVI, REIP, RVI, S2REP, SAVI, TNDVI, TSAVI, WdVI, LAI, CAB, CWC, FAPAR, FVC

4.3.2 Random forest classification

For the lidar metrics only group, we used the 14 lidar metrics as the predictor variables for the dead/live random forest classification model. Table 4.5 shows the confusion matrix of model success resulting in an overall accuracy of 71.1%. The top three most important variables for model predictions were Cb50, Cstd, and Cb70 (Figure 4.7).

Table 4.5. Confusion matrix for the lidar metrics only random forest classification model results.

Predicted Class	Actual Class		Total	User's Acc. (%)
	Dead	Live		
Dead	573	364	937	61.1
Live	134	651	785	82.9
Total	707	1015	1722	
Prod. Acc. (%)	81.1	64.1		
			Overall Accuracy:	71.1%

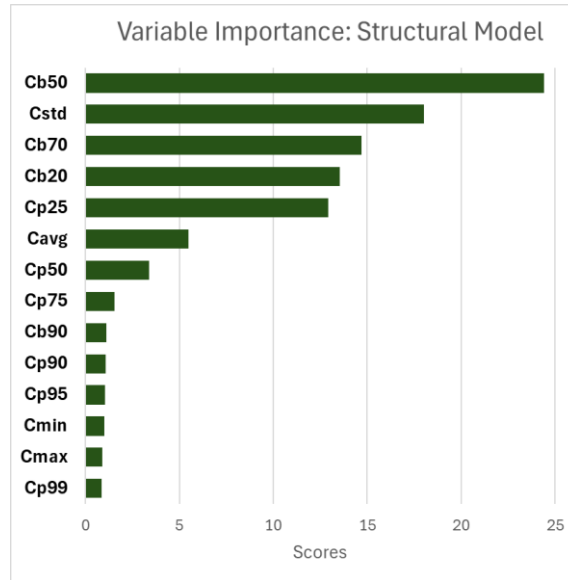


Figure 4.7. Variable importance scores for the lidar metrics only model.

For the S2 only group, we input the 28 S2 multispectral indices as the predictor variables for the dead/live random forest classification model. Table 4.6 shows the confusion matrix of model success resulting in an overall accuracy of 66.3%. The top three most important variables for model predictions were TNDVI, NDVI, and IPVI, which are all variations of NDVI (Figure 4.8).

Table 4.6. Confusion matrix for the S2 indices only random forest classification model results.

Predicted Class	Actual Class		Total	User's Acc. (%)
	Dead	Live		
Dead	577	450	1027	56.2
Live	130	565	695	81.3
Total	707	1015	1722	
Prod. Acc. (%)	81.6	55.7		
Overall Accuracy:				66.3%

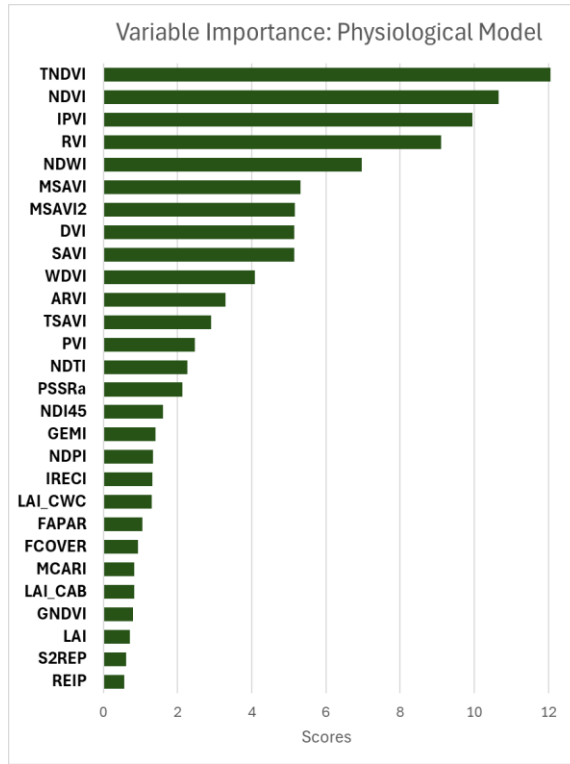


Figure 4.8. Variable importance scores for the S2 multispectral indices only model.

For the combined group, we input the 42 lidar metrics and S2 multispectral indices as the predictor variables for the dead/live random forest classification model. Table 4.7 shows the confusion matrix of model success resulting in an overall accuracy of 72.1%. The top three most important variables for model predictions were Cstd, Cb50, and Cb70, which are all lidar metrics. NDVI was the first S2 index and sixth in order of variable importance (Figure 4.9).

Table 4.7. Confusion matrix for the combined lidar and S2 random forest classification model results.

Predicted Class	Actual Class		Total	User's Acc. (%)
	Dead	Live		
Dead	593	367	960	61.8
Live	114	648	762	85.1
Total	707	1015	1722	
Prod. Acc. (%)	83.9	63.8		
Overall Accuracy:				72.1%

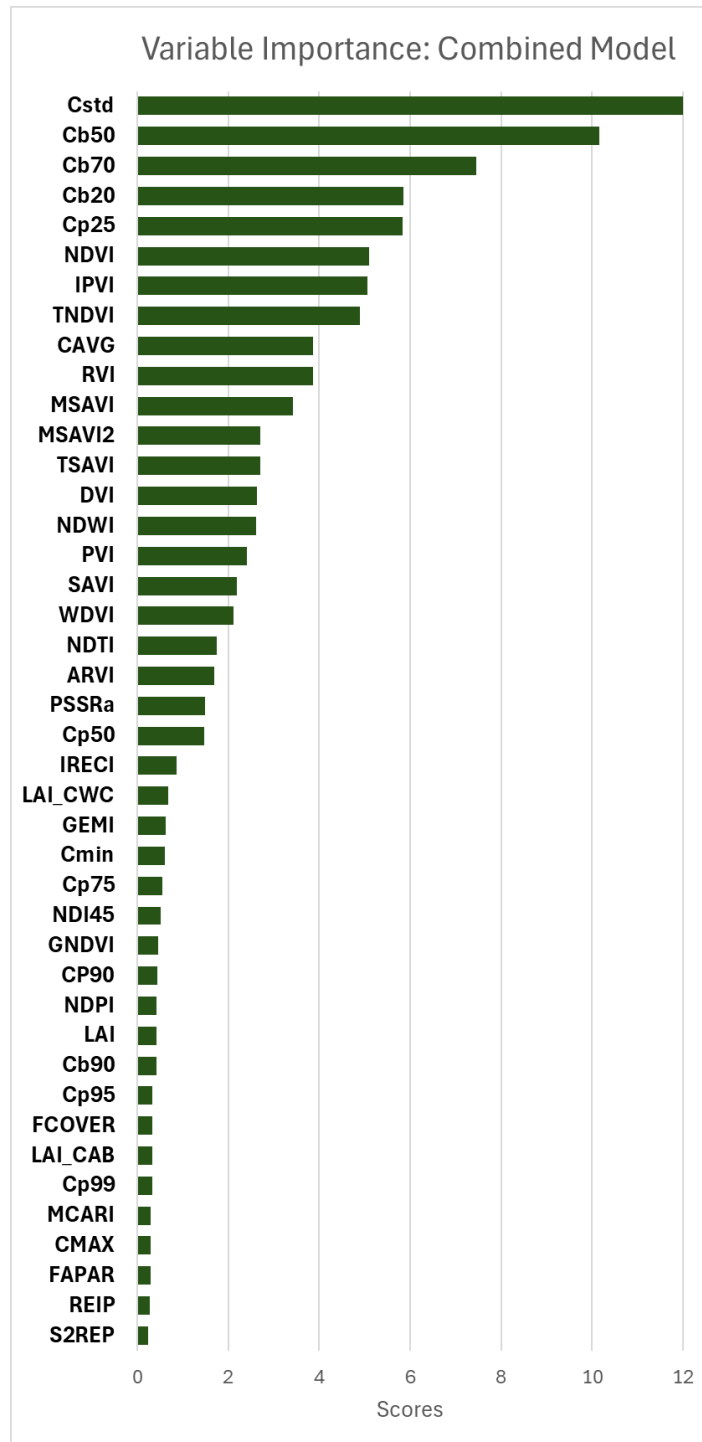


Figure 4.8. Variable importance scores for the combined model.

4.3.4 Vertical profile of 10 m dead and live examples

The 3D visualizations of the dead and live pixel examples reveal the vertical structure within each 10 m pixel. With the canopy bincentile 50% (Cb50) lidar metric being an important variable for all the random forest classification models, we extracted the point cloud from within each dead and live 10 m pixel and classified the Cb50 range. To determine where the Cb50 height was located in the point cloud for each pixel, we subtracted 2 m (ground to breast height) from 22 m (maximum height) and divided that value in half to delineate 50% up the canopy and above 2 m. Figure 4.10 shows the point clouds heights above ground on the first column and the second column with the Cb50 height classification with ground (0 m – 2 m) as brown, Cb50 delineation (2 m - 10 m) as light green, and above the Cb50 delineation (10 m – 22 m) as dark green for the dead and live pixel examples.

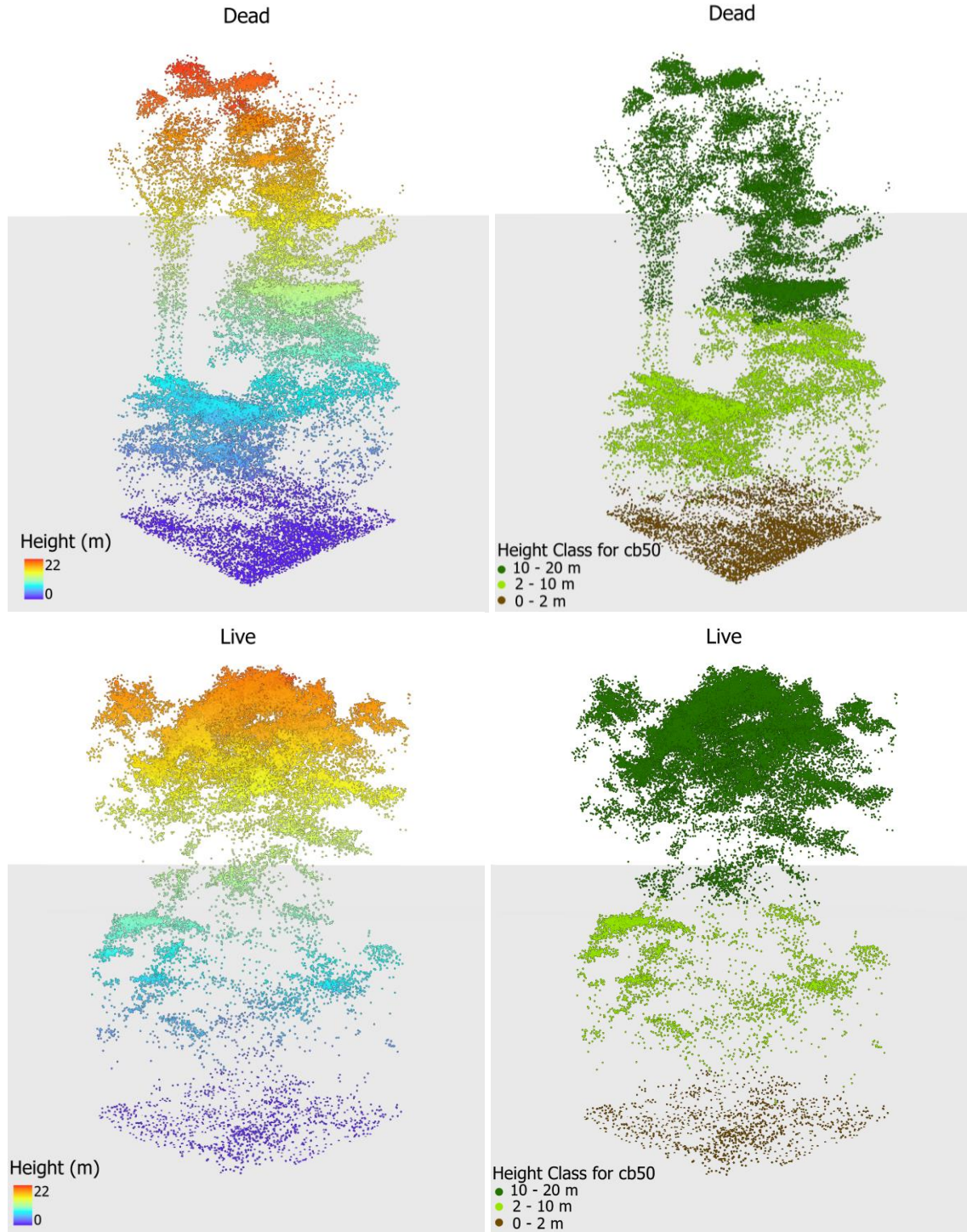


Figure 4.10. Example of dead and live 10 m pixels with 3D point clouds showing a gradient of heights in meters in the first column and the Cb50 height classification showing ground as 0 – 2 m in brown, the Cb50 delineation as 2 – 10 m in light green, and above the Cb50 delineation as 10 – 20 m in dark green.

4.4 Discussion

The random forest classification model results show that a combination of S2 multispectral indices and lidar metrics or lidar metrics alone can be used to detect the impacts of insect-induced tree mortality at this spatial scale. The model utilizing only S2 multispectral indices achieved the lowest performance, with an overall accuracy of 66%. In contrast, the highest performing model, which combined both lidar metrics and S2 multispectral indices, attained an overall accuracy of 72%. The lidar only model demonstrated an accuracy of 71%, which is only 1% less than using the lidar metrics combined with S2 indices. These results demonstrate the importance of structure-related measurements on improving classification accuracy for tree mortality impact investigations.

In Tables 4.7-4.9, there is a consistent overestimation of live pixels being classified as dead for all models tested. However, the S2 only model misclassified more live pixels as dead than the lidar only and combined models, which is also reflected in the user's (56% for S2 only and 61% for lidar only and combined) and producer's accuracies (56% for S2 only and 64% for lidar only and combined). These statistical metrics suggest that spectral measurements overestimate tree mortality compared to integration of lidar structural measurements. The misclassification could be due to mixed pixels or shadows at the 10 m resolution. The pixels were classified with visual interpretation, where a dead branch or tree within the pixel would be considered a dead pixel, however the pixel may still include live healthy green material around the dead branch or tree. The live pixels have less likelihood of having dead plant material within the pixel, thus a pixel with a small portion of dead material and majority live material may be misclassified as dead. Also, if there were shadows present within the 10 m pixel boundary due to the structure of the canopy, the overall spectral reflectance would be lower, thus affecting the spectral indices.

Furthermore, species-related variations in sub pixel fractions of reflectance are driven by crown architecture (Ferreira et al., 2018).

For the S2 indices only random forest classification model, the most important variables to predict tree mortality were all variations of NDVI. NDVI is used to quantify vegetation greenness, suggesting that this index is well-suited for distinguishing between dead and live plant material at the canopy scale. The KS statistics in Table 4.4 show that live pixels tend to have higher NDVI than dead pixels. NDVI has been shown to increase slowly after tree mortality as the surrounding trees recover or the understory emerges (Huang et al., 2019).

Post-mortality, there are evident canopy gaps that increase both in size and quantity in mature forest areas, therefore benefitting understory growth within the canopy gaps from increased illumination (Li et al., 2013). Figure 4.5 shows examples of lidar points from a 2D view for a dead and live pixel where the dead pixel shows a wider variation of heights, and the live pixel shows a smaller range of heights mainly in the upper canopy. The dead pixels will have more canopy gaps allowing for increased access to solar illumination in the understory due to loss of the upper story from mortality.

The high point density of our lidar acquisition provided detailed information of the vertical structure of the forest within the 10 m pixel boundary (~120,000 points per 10 m pixel). The penetration of the dual return lidar system at 90 m flight altitude was sufficient to locate the ground and understory in most cases, however with denser forest canopies, the penetration to the ground and understory is not as consistent. For the live pixel in Figure 4.10, the majority of the points in the point cloud are in the upper range of the canopy due to the canopy closure from a healthy tree blocking the laser from penetrating deeper into the canopy. Contrarily, the dead pixel in Figure 4.10 reveals the understory and even the trunks and branches of the dead trees within

the point cloud. In addition to the pulses penetrating farther, there is potential for more understory leaf area due to increased light penetration.

The canopy bincenile metrics from lidar provide a way to quantify the fraction of points within a specified vertical range of the canopy. The Cb50 metric, showing the fraction of points at about halfway (50%) up the canopy, was significant to model success for the lidar metrics only and combined models. The classification of the Cb50 for a dead and live pixel example in Figure 4.10 reveals that there are more points within this specified height range for the dead pixel compared to the live pixel. As mentioned above, the density of the points in the point cloud for the live pixel occurs mainly at the top of the canopy, whereas the dead pixel has more points within the mid-range due to lidar penetration. Furthermore, the canopy standard deviation (Cstd), which measures the amount of variation in heights within the 10 m pixel, was also an important variable for differentiation of dead and live pixels. The KS statistics in Table 4.4 show that dead pixels have higher Cstd than the live pixels, which is indicative of more variation of heights within the canopy of the dead pixels. These metrics provide a sense of how to capture more complex structural characteristics within the forest canopy post-mortality.

Higher mortality rates for larger canopy trees create substantial gaps in the canopy after defoliation and mortality (Francis et al., 2023). In our study area, the older oaks were the primary trees affected by the cankerworm infestation, where the lidar acquisition was able to penetrate deeper into the canopy and provide more detailed information on the morphological structure of dead trees and revealing the understory. This penetration depth concept is consistent with literature using lidar during leaf-off and leaf-on conditions for deciduous forests and delineation of conifers versus deciduous due to canopy closure affecting the lidar penetration depth (Kim et al., 2011; Lu et al., 2014).

Using visual interpretation on high resolution imagery to create a “truth” classification image at 10 m resolution was the optimal approach due to the high percentage of dead trees within this small area. Locating dead stems with GPS is very time consuming and not as precise due to GPS error and the high frequency of dead trees and branches that can fall within a 10 m pixel. NEON provides ground measurements of plant cover and vegetation structure, however only one plot is available within the study area and was collected in the fall of 2022 when leaves are senescing.

This study demonstrates the impacts on the morphological and physiological changes in a forest from insect infestation after substantial tree mortality. Our research findings suggest that multispectral indices, like NDVI, are sufficient for mortality investigation, however the canopy structural variables were stronger predictors of tree mortality impacts and provided vital information on vertical differences for live and dead trees.

4.5 Conclusion

This study shows the influence of combining physiological functional traits from S2 multispectral indices with morphological function traits from airborne lidar to study the impact of insect-induced tree mortality. Our random forest classification results indicate that inclusion of structural-related lidar metrics improves the classification of dead trees by 6%. Conversely, it shows that spectral information is rather minimal - only contributing an extra 1% over the classification from lidar alone. The S2-only model relied heavily on NDVI, which effectively distinguishes between live and dead plant material but may not fully account for post-mortality canopy gaps and recovery of surrounding vegetation within the understory at this spatial scale. Lidar metrics provided detailed vertical structure information, with the canopy bincenile 50% (Cb50) metric and canopy standard deviation (Cstd) being crucial for differentiating between

dead and live pixels. Despite all models consistently overestimating tree mortality impacts, there is substantial evidence supporting the efficacy of including vertical measurements of the forest for enhanced model predictions at this spatial scale. High lidar point density and penetration depth also contributed to capturing detailed structural features of the forest. The structural characteristics of the forest can provide a deeper context on the underlying morphological effects of tree mortality within the canopy.

References

- Bergmüller, K. O., & Vanderwel, M. C. (2022). Predicting tree mortality using spectral indices derived from multispectral UAV imagery. *Remote Sensing*, 14(9), 2195.
- Bozzini, A., Francini, S., Chirici, G., Battisti, A., & Faccoli, M. (2023). Spruce bark beetle outbreak prediction through automatic classification of Sentinel-2 imagery. *Forests*, 14(6), 1116.
- Brodrick, P. G., & Asner, G. P. (2017). Remotely sensed predictors of conifer tree mortality during severe drought. *Environmental Research Letters*, 12(11), 115013.
- Campbell, J. B., Wynne, R. H., Thomas, V. A. (2023). *Introduction to remote sensing*. United Kingdom: Guilford Publications.
- Chamberlin, L., & Mooneyham, K. (2020). Forest Health Highlights. *Virginia Department of Forestry*.
- Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1416–1427.

- Darr, M. N., & Coyle, D. R. (2021). Fall cankerworm (lepidoptera: geometridae), a native defoliator of broadleaved trees and shrubs in North America. *Journal of Integrated Pest Management*, 12(1).
- Fasano, G., & Franceschini, A. (1987). A multidimensional version of the Kolmogorov–Smirnov test. *Monthly notices of the Royal Astronomical Society*, 225(1), 155–170.
- Fernández-Alonso, J. M., Llorens, R., Sobrino, J. A., Ruiz-González, A. D., Alvarez-González, J. G., Vega, J. A., & Fernández, C. (2022). Exploring the potential of lidar and Sentinel-2 Data to model the post-fire structural characteristics of gorse shrublands in NW Spain. *Remote Sensing*, 14(23), 6063.
- Ferreira, M. P., Féret, J.-B., Grau, E., Gastellu-Etchegorry, J.-P., Shimabukuro, Y. E., & de Souza Filho, C. R. (2018). Retrieving structural and chemical properties of individual tree crowns in a highly diverse tropical forest with 3D radiative transfer modeling and imaging spectroscopy. *Remote Sensing of Environment*, 211, 276–291.
- Francis, E. J., Lutz, J. A., & Farrior, C. E. (2023). Elevated mortality rates of large trees allow for increased frequency of intermediate trees: A hypothesis supported by demographic model comparison with plot and LiDAR data. *Forest Ecology and Management*, 540, 121035.
- Goldbergs, G., & Upenieks, E. M. (2024). Hierarchical integration of UAS and Sentinel-2 imagery for spruce bark beetle grey-attack detection by vegetation index thresholding approach. *Forests*, 15(4), 644.
- Hardiman, B., Bohrer, G., Gough, C., & Curtis, P. (2013). Canopy structural changes following widespread mortality of canopy dominant trees. *Forests*, 4(3), 537–552.

- Huang, C., Anderegg, W. R. L., & Asner, G. P. (2019). Remote sensing of forest die-off in the Anthropocene: From plant ecophysiology to canopy structure. *Remote Sensing of Environment*, 231, 111233.
- Kim, S., Hinckley, T., & Briggs, D. (2011). Classifying individual tree genera using stepwise cluster analysis based on height and intensity metrics derived from airborne laser scanner data. *Remote Sensing of Environment*, 115(12), 3329–3342.
- Li, Q., Bonebrake, T. C., Michalski, J. R., Kit Wong, F. K., & Fung, T. (2023). Combining multi-temporal airborne LiDAR and Sentinel-2 multispectral data for assessment of disturbances and recovery of mangrove forests. *Estuarine, Coastal and Shelf Science*, 291, 108444.
- Lu, X., Guo, Q., Li, W., & Flanagan, J. (2014). A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 94, 1–12.
- Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Zald, H. S. J. (2018). Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sensing of Environment*, 209, 90–106.
- Meng, R., Dennison, P. E., Zhao, F., Shendryk, I., Rickert, A., Hanavan, R. P., Cook, B. D., et al. (2018). Mapping canopy defoliation by herbivorous insects at the individual tree level using bi-temporal airborne imaging spectroscopy and LiDAR measurements. *Remote Sensing of Environment*, 215, 170–183.

Mountain Lake Biological Station NEON | NSF NEON | Open Data to Understand our

Ecosystems. (n.d.). Retrieved August 2, 2024, from <https://www.neonscience.org/field-sites/mlbs>

Navarro-Cerrillo, R. M., Varo-Martínez, M. Á., Acosta, C., Palacios Rodriguez, G., Sánchez-Cuesta, R., & Ruiz Gómez, F. J. (2019). Integration of WorldView-2 and airborne laser scanning data to classify defoliation levels in *Quercus ilex* L. Dehesas affected by root rot mortality: Management implications. *Forest Ecology and Management*, 451, 117564.

NEON (National Ecological Observatory Network). High-resolution orthorectified camera imagery mosaic (DP3.30014.001). <https://data.neonscience.org> (accessed July 30, 2023)

NEON (National Ecological Observatory Network). Plant presence and percent cover (DP1.10058.001). <https://data.neonscience.org> (accessed July 29, 2024)

Rock, B. N., Hoshizaki, T., & Miller, J. R. (1988). Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline. *Remote Sensing of Environment*, 24(1), 109–127.

Romero Ramirez, F. J., Navarro-Cerrillo, R. M., Varo-Martínez, M. Á., Quero, J. L., Doerr, S., & Hernández-Clemente, R. (2018). Determination of forest fuels characteristics in mortality-affected *Pinus* forests using integrated hyperspectral and ALS data. *International Journal of Applied Earth Observation and Geoinformation*, 68, 157–167.

White, J. C., Coops, N. C., Hilker, T., Wulder, M. A., & Carroll, A. L. (2007). Detecting mountain pine beetle red attack damage with EO-1 Hyperion moisture indices. *International Journal of Remote Sensing*, 28(10), 2111–2121.

Wilbur, R. B., Wilbur, H. M., & Peterson, N. J. (2021). Flora of the Forest at Mountain Lake Biological Station, Giles County, Virginia. *Castanea*, 85(2), 376-403.

- Wulder, M. A., Dymond, C. C., White, J. C., Leckie, D. G., & Carroll, A. L. (2006). Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. *Forest Ecology and Management*, 221(1–3), 27–41.
- Zhao, F., Meng, R., Gu, H., & Serbin, S. (2019). Assessing post-fire tree mortality and biomass change by integrating lidar and hyperspectral data. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 7346–7349). Presented at the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE.
- Xu, C., Morgenroth, J., & Manley, B. (2015). Integrating data from discrete return airborne lidar and optical sensors to enhance the accuracy of forest description: a review. *Current Forestry Reports*, 1(3), 206–219.

Chapter 5 - Conclusion

Summary and Conclusions

Results from this work illustrate the effectiveness of combining physiological function from spectral data and morphological structure from lidar to investigate foliage efficiencies vertically and diurnally, estimate productivity for a range of forest types, and detect disturbances that effect forest productivity like insect-induced tree mortality. An over-arching theme of this work is that structure measurements from lidar were essential in every chapter.

Specifically, this work addressed each of the objectives described earlier.

Objective 1: The Photochemical Reflectance Index (PRI) of forest canopies changes throughout the day and at different vertical levels, influenced by varying light conditions. This study highlights a new approach using simulated panchromatic bands to classify illumination levels. The findings reveal that PRI values differ between sunlit and shaded parts of the canopy leading to variations in diurnal foliage efficiency, and the canopy structure affects light interception.

Objective 2: Integrating remotely sensed data with *in situ* flux tower measurements can effectively capture variations in functional traits that influence forest productivity. A combination of all three types of forest functional traits (i.e., morphological, environmental, and physiological) enhances predictions of GPP. The best fit GPP model was the 4-term combination of functional traits model using CHM mean, DSM standard deviation, PRI standard deviation, and WBI mean producing a R^2 of 0.87, an adjusted R^2 of 0.84, a PRESS R^2 of 0.75 and RMSE of $3.48 \mu\text{mol CO}_2 / \text{m}^2/\text{s}$.

Objective 4: The incorporation of lidar derived forest metrics with S2 multispectral indices enhanced tree mortality classification accuracy by 6%. The S2 model relied heavily on NDVI, however at the spatial scale of 10 m it may not fully account for post-mortality canopy gaps and recovery of surrounding vegetation within the understory. Lidar alone can detect tree mortality with a 70% accuracy, using the variation in the canopy height (Cstd) and the canopy bincentile 50% (Cb50) within a 10 m pixel proved to provide essential details on vertical structure information essential for accurately identifying dead trees. High lidar point density and penetration depth also contributed to capturing detailed structural features of the forest.

Future Directions

Recommendations for future analysis are increased concentration on leveraging multiple sources of remote sensing technologies with *in situ* measurements. Many stages of the analysis would have benefited from more data – more frequent acquisitions, consistent ground measurements, or a greater number of study sites.

Chapter 2 only provided diurnal data from a single clear-sky day late in the growing season. It is not easy to acquire diurnal data due to changing cloud conditions affecting the illumination calibrations and creating shadows in different portions of the canopy throughout the day (learned through first-hand experience). A useful dataset for diurnal PRI investigations would need to be acquired on clear-sky days before, during, and after peak greenness to evaluate the seasonal drivers of this index.

Chapter 3 would have benefitted from more study sites to provide examples for the wider range of forest characteristics within the US. The study sites selected were constrained by the availability of flux tower data at the sites, which were very spotty and lacking full year data in

most cases. Furthermore, how would these methods be applicable for forest productivity investigations outside of the US, like in the Amazon rain forest or forest plantations in Southeast Asia? A wider range of functional traits would enhance our understanding of global forested ecosystems. Also, does species-specific information enhance our understanding of productivity or are the functional traits sufficient to describe plant function in different environments?

Chapter 4 was intended to collect diurnal lidar and hyperspectral data, but due to gimbal leveling issues on the UAS, the hyperspectral data was not optimal for diurnal investigations due to sensor view angles exceeding 10-20 degrees. With the gimbal now stabilized on the UAS, diurnal flights remain feasible; however, atmospheric conditions and cloud cover continue to pose significant challenges. To gain a more thorough understanding of tree mortality in this area, it would be valuable to analyze high-resolution multitemporal lidar data and S2 multispectral indices from before and after disturbances. This would help identify the key changes resulting from the disturbances.

Appendix A – MATLAB Code for Chapter 2 and 3

Chapter 2 MATLAB 2021b code

```
%%
% name:    composite_stats.m
% date:    Oct 7, 2020
% author:  Val Thomas & Paige Williams
% purpose: generates composite thresholds and saves as a text files- these thresholds
%          need to be
%          called by the main program.
%          %this program should only run once per plot.

%Setup initial parameters for the file.
%filename = '/Users/thomasv/Dropbox (VTFRS)/G-LIHT-
SAFE/Parker_Tract_AM_NS_26Oct2013_mosaicked_refl_VIs/Parker_Tract_AM_NS_26Oct2013_at-
sensor_refl_L1G.hdr';

cd 'C:\Users\Paige Williams\Dropbox (VTFRS)\paige\gliht\'
MyFolderInfo = dir;

monster_pan = double.empty;

out_path='C:\Users\Paige Williams\Dropbox
(VTFRS)\paige\gliht\matlab_outputs\composites\';
out_thresholds_file= strcat(out_path,'composites_thresholds_plot3.txt');

rootdir = 'C:\Users\Paige Williams\Dropbox (VTFRS)\paige\gliht\'
filelist = dir(fullfile(rootdir, '**\*.hdr')); %get list of hdr files in any
subfolder
filelist = filelist(~[filelist.isdir]);

ref_files = extractfield(filelist,'name');
ref_paths = extractfield(filelist,'folder');
str_names = string(ref_files);
str_paths = string(ref_paths);

len=length(str_names);
count=1;
for k=1:len
    Y = strfind(str_names(k),'refl_L1G');
    if Y > 0
        D(count)=str_names(k);
        gliht_name(count)=strcat(str_paths(k),'\',str_names(k)); %Thes are only the
L1G reflectance files
        count=count+1;
    end
end

len2 = length(gliht_name);
%%
for l=1:len2
```

```

%%
%Set the output file names, based on the input names (that way you only
%need to change one line of code each time

%%
% Read the envi file and make a hypercube

%Read the envi header.

info = enviinfo(gliht_name(1))

%create a corresponding hypercube.

% namecheck = extractfield(info,'Filename');
% str_namecheck = string(namecheck);
% hcube = hypercube(info.Filename);

%the images is now in hcube.DataCube
%%
% Crop the hypercube, so that it only contains the area of interest
%The study are for Parker Tract has the following coordinates (this can be
%changed)
%plot1:
%xlimits = [349030 349450];
%ylimits = [3963169 3963589];
%
%plot2:
%xlimits = [349135 349555];
%ylimits = [3964380 3964800];
%
%plot3:
%xlimits = [350397 350817];
%ylimits = [3964253 3964673];
%What row does that make it? Well, we need to pull the extent from the hcube

MapTies = extractfield(hcube.Metadata.MapInfo,'MapTiePoints');
PixelSize = extractfield(hcube.Metadata.MapInfo,'PixelSize');
%the top of the image is at MapTies(2)
% 3973739 and the pixels are 2 m, so it would be
rowstart = int16((MapTies(2)-3964673)/2);
rowend = rowstart+209;
%Note: the total number of rows must be divisible by 15 so that we can
%rescale to 30m pixels later.
%so make it 210x210
colstart = int16(abs(350397-MapTies(1))/2);
colend = colstart+209;

%Because not all files are at Parker Tract, double check
%to see if the study area is within the image.
maxrows = extractfield(hcube.Metadata,'Height');
maxcols = extractfield(hcube.Metadata,'Width');

```

```

%crop the study area
hchip = cropData(hcube, rowstart:rowend, colstart:colend);

%%
%NDVI calculation
%picking a threshold to mask out non veg.... tesing .5
band640 = (double(hchip.DataCube(:,:,54))+double(hchip.DataCube(:,:,55)))/2;
%find bands
band700 = (double(hchip.DataCube(:,:,88))+double(hchip.DataCube(:,:,89)))/2;
%find bands

%Calculate NDVI
NDVItop = band700-band640;
NDVIbottom = band700+band640;
NDVI = NDVItop./NDVIbottom;

NDVIthres =double(NDVI > 0.5);
NDVIthres(NDVIthres==0)=NaN;

%%
% Calculate SAFE PRI
%Assuming your spectral cgtube A is m x n x p (with p spectral bands),
%you would pick out the first band with A(:,:,1),
%and the pth band with A(:,:,p).
band531 = (double(hchip.DataCube(:,:,25))+double(hchip.DataCube(:,:,26)))/2;
band570 = (double(hchip.DataCube(:,:,36))+double(hchip.DataCube(:,:,37)))/2;

%Calculate the equivalent of SAFE PRI
%PRI = (b531-b570)/(b531+b570)
PRItop = band531-band570;
PRIbottom = band531+band570;
PRICALC = PRItop./PRIbottom;

PRI = PRICALC.*NDVIthres;

%Save PRI as a Geotiff

%First, tell it the coodtinate space and cellsize
%R = maprefcells(xlimits,ylimits,xcellextent,ycellextent)
%R = maprefcells(xlimits,ylimits,extent,extent, ...
% 'ColumnsStartFrom','north')
%that the upper left corner is north and west
%Note that the EPSG code is 32618 for UTM Zone 18N, wgs84

%pull from map tie points *****
%plot1:
%xlimits = [349030 349450];
%ylimits = [3963169 3963589];
%
%plot2:
%xlimits = [349135 349555];
%ylimits = [3964380 3964800];
%plot3:

```

```

xlimits = [350397 350817];
ylimits = [3964253 3964673];
%
%xlimits = [350397 350817];
%ylimits = [3964253 3964673];
xcellextent = PixelSize(1); %I might have these mixed up.
ycellextent = PixelSize(2);
R = maprefcells(xlimits, ylimits, xcellextent, ycellextent,...
    'ColumnsStartFrom','north',...
    'RowsStartFrom','west');

% geotiffwrite(outfile_PRI,PRI,R, 'CoordRefSysCode',32618);

%%
% Simulate the SAFE panchromatic band and save as Geotiff.
% create a simulated SAFE panchromatic band (18 bands from 525nm-600nm)
%in band math for GLiHT B25+B26+B27+....+B42/18.

Addbands =
hchip.DataCube(:,:,25)+hchip.DataCube(:,:,26)+hchip.DataCube(:,:,27)...
    +hchip.DataCube(:,:,28)+hchip.DataCube(:,:,29)+hchip.DataCube(:,:,30)...
    +hchip.DataCube(:,:,31)+hchip.DataCube(:,:,32)+hchip.DataCube(:,:,33)...
    +hchip.DataCube(:,:,34)+hchip.DataCube(:,:,35)+hchip.DataCube(:,:,36)...
    +hchip.DataCube(:,:,37)+hchip.DataCube(:,:,38)+hchip.DataCube(:,:,39)...
    +hchip.DataCube(:,:,40)+hchip.DataCube(:,:,41)+hchip.DataCube(:,:,42);

panchromatic = (double(Addbands))/18;
panchromatic(panchromatic==0)=NaN;

%%
% Make a giant panchromatic composite to calculate composite statistics

monster_pan = [monster_pan panchromatic];

end

%%
%5. Create masks for sunlit, shadowed, and mixed only pixels from
panchromatic band
% a. mixed pixels - the pan values between the std above and below the mean
% for the whole little chip

pan_mean = mean(monster_pan(:), 'omitnan');
pan_std = std(monster_pan(:), 'omitnan');
lower_diff = pan_mean-pan_std;
higher_diff = pan_mean+pan_std;

out_matrix = [pan_mean pan_std lower_diff higher_diff];
writematrix(out_matrix,out_thresholds_file);

```

```

%%
% name:    composites.m
% date:    Oct 7, 2020
% author:  Val Thomas & Paige Williams
% purpose: To read gliht hyperspectral reflectance data from parker tract.
% 1) simulated band 531 (B26+B27)/2,
% 2) simulated band 570 (B35+B36)/2,
% combine bands 1 and 2 into a single file and save as envi format
% 3) a simulated SAFE panchromatic band
%%

cd 'C:\Users\Paige\Dropbox (VTFRS)\paige\gliht\'
MyFolderInfo = dir;

%composite_thresholds =readmatrix('/Users/thomasv/Dropbox (VTFRS)/paige
(1)/gliht/matlab_outputs/composites/composites_thresholds.txt');
composite_thresholds =readmatrix('C:\Users\Paige\Dropbox
(VTFRS)\paige\gliht\matlab_outputs\composites\composites_thresholds_plot1.txt');

%rootdir = '/Users/thomasv/Dropbox (VTFRS)/paige (1)/gliht/';
rootdir = 'C:\Users\Paige\Dropbox (VTFRS)\paige\gliht\'
filelist = dir(fullfile(rootdir, '**\*.hdr')); %get list of hdr files in any
subfolder
filelist = filelist(~[filelist.isdir]);

ref_files = extractfield(filelist,'name');
ref_paths = extractfield(filelist,'folder');
str_names = string(ref_files);
str_paths = string(ref_paths);

len=length(str_names);
count=1;
for k=1:len
    Y = strfind(str_names(k),'refl_L1G');
    if Y > 0
        D(count)=str_names(k);
        gliht_name(count)=strcat(str_paths(k),'\',str_names(k)); %Thes are only the
L1G reflectance files
        count=count+1;
    end
end

len2 = length(gliht_name);
%%
for l=1:len2

    %%
    %Set the output file names, based on the input names (that way you only
    %need to change one line of code each time
    %out_path='/Users/thomasv/Dropbox (VTFRS)/paige
(1)/gliht/matlab_outputs/composites/prisplot1/';

```

```

    %out_path_2='/Users/thomasv/Dropbox (VTFRS)/paige
(1)/gliht/matlab_outputs/composites/componentsplot1/';
    out_path='C:\Users\Paige\Dropbox
(VTFRS)\paige\gliht\matlab_outputs\composites\prisplot3_testing\';
    out_path_2='C:\Users\Paige\Dropbox
(VTFRS)\paige\gliht\matlab_outputs\composites\componentsplot3\';
    outfile_PRI = strcat(out_path,D(1), '_PRI_2m.tif');
    outfile_panchromatic = strcat(out_path,D(1), '_PAN_2m.tif');
    out_PRI_mixed_2m = strcat(out_path,D(1), '_PRI_mixed_2m.tif');
    out_PRI_sunlit_2m = strcat(out_path,D(1), '_PRI_sunlit_2m.tif');
    out_PRI_shadow_2m = strcat(out_path,D(1), '_PRI_shadow_2m.tif');

    out_class_map = strcat(out_path,D(1), '_class_map.tif');

    outfile_NDVI = strcat(out_path,D(1), '_NDVI.tif');

    out_PRI_30m = strcat(out_path,D(1), '_PRI_30m.tif');
    out_PRI_mixed_30m = strcat(out_path,D(1), '_PRI_mixed_30m.tif');
    out_PRI_sunlit_30m = strcat(out_path,D(1), '_PRI_sunlit_30m.tif');
    out_PRI_shadow_30m = strcat(out_path,D(1), '_PRI_shadow_30m.tif');

    out_shadow_fractions = strcat(out_path_2,D(1), '_shadowfractions_30m.tif');

    %the stats files
    PRI_30_stat_file = strcat(out_path,D(1), '_PRI_30m_statistics.txt');
    PRI_2_stat_file = strcat(out_path,D(1), '_PRI_2m_statistics.txt');

    %%
    % Read the envi file and make a hypercube

    %Read the envi header.

    info = enviinfo(gliht_name(1));
    gliht_name(1)

    %create a corresponding hypercube.

    % namecheck = extractfield(info,'Filename');
    % str_namecheck = string(namecheck);
    hcube = hypercube(info.Filename);

    MapTies = extractfield(hcube.Metadata.MapInfo,'MapTiePoints');
    PixelSize = extractfield(hcube.Metadata.MapInfo,'PixelSize');
    %the top of the image is at MapTies(2)
    % 3973739 and the pixels are 2 m, so it would be
    rowstart = int16((MapTies(2)-3963589)/2);
    rowend = rowstart+209;
    %Note: the total number of rows must be divisible by 15 so that we can
    %rescale to 30m pixels later.
    %so make it 210x210
    colstart = int16(abs(349030-MapTies(1))/2);
    colend = colstart+209;

```

```

%Because not all files are at Parker Tract, double check
%to see if the study area is within the image.
maxrows = extractfield(hcube.Metadata,'Height');
maxcols = extractfield(hcube.Metadata,'Width');
%if ((rowstart < 0) | (rowstart > maxrows))
    % this means the study area is either north or south of the gliht image
    % message1 = 'study area is either north or south of the gliht image'
    % D(1)
%elseif (colstart > maxcols)
    % this means the area is either west or east of the gliht image
    % message2 = 'study area is either west or east of the gliht image'
    % D(1)
%else
%%
    % the gliht image contains the study area

%crop the study area
hchip = cropData(hcube, rowstart:rowend, colstart:colend);

%%
%NDVI calculation
%picking a threshold to mask out non veg.... tesing .5
band640 = (double(hchip.DataCube(:,:,54))+double(hchip.DataCube(:,:,55)))/2;
%find bands
band700 = (double(hchip.DataCube(:,:,88))+double(hchip.DataCube(:,:,89)))/2;
%find bands

%Calculate NDVI
NDVItop = band700-band640;
NDVIbottom = band700+band640;
NDVI = NDVItop./NDVIbottom;

NDVithres =double(NDVI > 0.6);
NDVithres(NDVithres==0)=NaN;

%%
% Calculate SAFE PRI
%Assuming your spectral cgtube A is m x n x p (with p spectral bands),
%you would pick out the first band with A(:,:,1),
%and the pth band with A(:,:,p).
band531 = (double(hchip.DataCube(:,:,26))+double(hchip.DataCube(:,:,27)))/2;
band570 = (double(hchip.DataCube(:,:,35))+double(hchip.DataCube(:,:,36)))/2;

%Calculate the equivalent of SAFE PRI
%PRI = (b531-b570)/(b531+b570)
PRItop = band531-band570;
PRIbottom = band531+band570;
PRICALC = PRItop./PRIbottom; % there is probably a cleaner way two write these
three commands.

PRI = PRICALC.*NDVithres;

%Save PRI as a Geotiff

```

```

%First, tell it the coodtinat space and cellsize
%R = maprefcells(xlimits,ylimits,xcellextent,ycellextent)
%R = maprefcells(xlimits,ylimits,extent,extent, ...
    % 'ColumnsStartFrom','north')
%that the upper left corner is north and west
%Note that the EPSG code is 32618 for UTM Zone 18N, wgs84

%pull from map tie points *****
%plot1:
xlimits = [349030 349450];
ylimits = [3963169 3963589];
%
%plot2:
%xlimits = [349135 349555];
%ylimits = [3964380 3964800];
%plot3:
%xlimits = [350397 350817];
%ylimits = [3964253 3964673];
%
xcellextent = PixelSize(1); %I might have these mixed up.
ycellextent = PixelSize(2);
R = maprefcells(xlimits, ylimits, xcellextent, ycellextent, ...
    'ColumnsStartFrom','north',...
    'RowsStartFrom','west');
%geotiffwrite(outfile_PRI,hchip.DataCube,R, 'CoordRefSysCode',32618); %this lines
it up good, except it is upside down

geotiffwrite(outfile_PRI,PRI,R, 'CoordRefSysCode',32618);

%output NDVI image if needed
geotiffwrite(outfile_NDVI,NDVI,R, 'CoordRefSysCode',32618);

%%
% Simulate the SAFE panchromatic band and save as Geotiff.
%he create a simulated SAFE panchromatic band (18 bands from 525nm-600nm)
%in band math for GLiHT B25+B26+B27+....+B42/18.

Addbands =
hchip.DataCube(:,:,25)+hchip.DataCube(:,:,26)+hchip.DataCube(:,:,27)...
+hchip.DataCube(:,:,28)+hchip.DataCube(:,:,29)+hchip.DataCube(:,:,30)...
+hchip.DataCube(:,:,31)+hchip.DataCube(:,:,32)+hchip.DataCube(:,:,33)...
+hchip.DataCube(:,:,34)+hchip.DataCube(:,:,35)+hchip.DataCube(:,:,36)...
+hchip.DataCube(:,:,37)+hchip.DataCube(:,:,38)+hchip.DataCube(:,:,39)...
+hchip.DataCube(:,:,40)+hchip.DataCube(:,:,41)+hchip.DataCube(:,:,42);

panchromatic = (double(Addbands))/18;
panchromatic(panchromatic==0)=NaN;

%write out the panchromatic band as a geotiff file
geotiffwrite(outfile_panchromatic,panchromatic,R, 'CoordRefSysCode',32618);

```

```

%%
%5.      Create masks for sunlit, shadowed, and mixed only pixels from
panchromatic band
% a.      mixed pixels - the pan values between the std above and below the mean
% for the whole little chip

pan_mean = composite_thresholds(1);
pan_std = composite_thresholds(2);
lower_diff = composite_thresholds(3);
higher_diff = composite_thresholds(4);
mixed_mask = (panchromatic>lower_diff) & (panchromatic<higher_diff);

%out the pan statistics to csv - PAIGE ADDITION

%b.      Sunlit pixels - the pan values ABOVE the 2st std (the std above mean)
sunlit_mask = (panchromatic>=higher_diff);

%c.      Shadowed pixels - the pan values BELOW the 1st std (the std below mean)
shadow_mask = (panchromatic<=lower_diff);

%%
% This section makes a map of the illumination classes.
% Sunlit = 1
% Mixed = 2
% Shadow = 3

Sunlit = double(sunlit_mask); %converts logical array to double
Mixed = double(mixed_mask);
Shadow = double(shadow_mask);
Mixed2 = Mixed*2;
Shadow3 = Shadow*3;
Class_map = Sunlit + Mixed2 + Shadow3;
Class_map_masked = Class_map.*NDVIthres;

Class_map_masked(isnan(Class_map_masked))=0;

geotiffwrite(out_class_map,Class_map_masked,R, 'CoordRefSysCode',32618);

%%
% Apply masks to PRI
mixed_mask2 = double(mixed_mask); % converts the logical mask to a float
mixed_mask2(mixed_mask2==0)=NaN; % takes the 0's of the logical (false) to NaN
(so the 0's are not confused with a PRI of 0
PRI_mixed = PRI.*mixed_mask2;

sunlit_mask2 = double(sunlit_mask);
sunlit_mask2(sunlit_mask2==0)=NaN;
PRI_sunlit = PRI.*sunlit_mask2;

shadow_mask2 = double(shadow_mask);
shadow_mask2(shadow_mask2==0)=NaN;
PRI_shadow = PRI.*shadow_mask2;

```

```

%Save the 2m files as geotiffs so can look in arcgis
geotiffwrite(out_PRI_mixed_2m,PRI_mixed,R, 'CoordRefSysCode',32618);
geotiffwrite(out_PRI_sunlit_2m,PRI_sunlit,R, 'CoordRefSysCode',32618);
geotiffwrite(out_PRI_shadow_2m ,PRI_shadow,R, 'CoordRefSysCode',32618);

%%

%compute the shadow fractions

% Define the sum function to be used with blockproc for shadowfraction.
mysumFunction = @(block_struct) sum(block_struct.data(:), 'omitnan');
% Get the sums by each 15-by-15 block.
mixed_30m = blockproc(mixed_mask, [15 15], mysumFunction);
shadow_30m = blockproc(shadow_mask, [15 15], mysumFunction);
sunlit_30m = blockproc(sunlit_mask, [15 15], mysumFunction);
% convert to fractions
mixed_fract_30m = mixed_30m./225;
shadow_fract_30m = shadow_30m./225;
sunlit_fract_30m = sunlit_30m./225;

%combine these 3 images into a single, 3-band image

fractions_30m = cat(3,mixed_fract_30m, shadow_fract_30m, sunlit_fract_30m);

%save 3-band image to geotiff out_shadow_fractions
xcellextent_30 = 30;
ycellextent_30 = 30;
R30 = maprefcells(xlimits, ylimits, xcellextent_30, ycellextent_30,...
    'ColumnsStartFrom','north',...
    'RowsStartFrom','west');

geotiffwrite(out_shadow_fractions,fractions_30m,R30, 'CoordRefSysCode',32618);

%%

%Aggregate to 30m and calculate statistics.
% Define the mean function to be used with blockproc.
myMeanFunction = @(block_struct) mean(block_struct.data(:), 'omitnan');
% Get the means by each 1-by-20 block.
PRI_30m = blockproc(PRI, [15 15], myMeanFunction);
PRI_mixed_30m = blockproc(PRI_mixed, [15 15], myMeanFunction);
PRI_shadow_30m = blockproc(PRI_shadow, [15 15], myMeanFunction);
PRI_sunlit_30m = blockproc(PRI_sunlit, [15 15], myMeanFunction);

xcellextent_30 = 30;
ycellextent_30 = 30;
R30 = maprefcells(xlimits, ylimits, xcellextent_30, ycellextent_30,...
    'ColumnsStartFrom','north',...
    'RowsStartFrom','west');

```

```

geotiffwrite(out_PRI_30m,PRI_30m,R30, 'CoordRefSysCode',32618);
geotiffwrite(out_PRI_mixed_30m,PRI_mixed_30m,R30, 'CoordRefSysCode',32618);
geotiffwrite(out_PRI_shadow_30m,PRI_shadow_30m,R30, 'CoordRefSysCode',32618);
geotiffwrite(out_PRI_sunlit_30m,PRI_sunlit_30m,R30, 'CoordRefSysCode',32618);

%%
% Calculate statistics and save as csv
% count, min, max, std dev, mean
% Create the output file names

pan_mean = mean(panchromatic(:), 'omitnan');
pan_std = std(panchromatic(:), 'omitnan');
%calculate the statistics and put them into arrays:
PRI_30m_stats = [nnz(~isnan(PRI_30m))...
    min(PRI_30m,[],'all') max(PRI_30m,[],'all') std(PRI_30m(:),'omitnan')...
    mean(PRI_30m(:),'omitnan') median(PRI_30m(:),'omitnan')];
% PRI_30m_diffuse_stats
PRI_30m_mixed_stats = [nnz(~isnan(PRI_mixed_30m))...
    min(PRI_mixed_30m,[],'all') max(PRI_mixed_30m,[],'all')...
    std(PRI_mixed_30m(:),'omitnan')...
    mean(PRI_mixed_30m(:),'omitnan') median(PRI_mixed_30m(:),'omitnan')];
% PRI_30m_sunlit_stats
PRI_30m_sunlit_stats = [nnz(~isnan(PRI_sunlit_30m))...
    min(PRI_sunlit_30m,[],'all') max(PRI_sunlit_30m,[],'all')...
    std(PRI_sunlit_30m(:),'omitnan')...
    mean(PRI_sunlit_30m(:),'omitnan') median(PRI_sunlit_30m(:),'omitnan')];
% PRI_30m_shadow_stats
PRI_30m_shadow_stats = [nnz(~isnan(PRI_shadow_30m))...
    min(PRI_shadow_30m,[],'all') max(PRI_shadow_30m,[],'all')...
    std(PRI_shadow_30m(:),'omitnan')...
    mean(PRI_shadow_30m(:),'omitnan') median(PRI_shadow_30m(:),'omitnan')];

%create/open the output csv files.
fileID = fopen(PRI_30_stat_file, 'w');

StdDev      col_header = 'Variable          Count          Min          Max
Mean      Median';
formatSpec = 'PRI_30m          %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';
formatSpec1 = 'PRI_30m_diffuse %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';
formatSpec2 = 'PRI_30m_sunlit  %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';
formatSpec3 = 'PRI_30m_shadow  %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';

fprintf(fileID,'% -50s\n', col_header);
fprintf(fileID,formatSpec,PRI_30m_stats); %this prints the first line
fprintf(fileID,formatSpec1,PRI_30m_mixed_stats);
fprintf(fileID,formatSpec2,PRI_30m_sunlit_stats);
fprintf(fileID,formatSpec3,PRI_30m_shadow_stats);

fclose(fileID);

%%
%2m stats

```

```

%calculate the statistics and put them into arrays:
PRI_2m_stats = [nnz(~isnan(PRI))...
    min(PRI,[],'all') max(PRI,[],'all') std(PRI(:),'omitnan')...
    mean(PRI(:),'omitnan') median(PRI(:),'omitnan')];
% PRI_2m_mixed_stats
PRI_2m_mixed_stats = [nnz(~isnan(PRI_mixed))...
    min(PRI_mixed,[],'all') max(PRI_mixed,[],'all')...
    std(PRI_mixed(:),'omitnan')...
    mean(PRI_mixed(:),'omitnan') median(PRI_mixed(:),'omitnan')];
% PRI_2m_sunlit_stats
PRI_2m_sunlit_stats = [nnz(~isnan(PRI_sunlit))...
    min(PRI_sunlit,[],'all') max(PRI_sunlit,[],'all')...
    std(PRI_sunlit(:),'omitnan')...
    mean(PRI_sunlit(:),'omitnan') median(PRI_sunlit(:),'omitnan')];
% PRI_2m_shadow_stats
PRI_2m_shadow_stats = [nnz(~isnan(PRI_shadow))...
    min(PRI_shadow,[],'all') max(PRI_shadow,[],'all')...
    std(PRI_shadow(:),'omitnan')...
    mean(PRI_shadow(:),'omitnan') median(PRI_shadow(:),'omitnan')];

%create/open the output csv files.
fileID2 = fopen(PRI_2_stat_file, 'w');

col_header = 'Variable          Count          Min          Max
StdDev      Mean      Median';
formatSpec_2m = 'PRI_2m          %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';
formatSpec1_2m = 'PRI_2m_mixed %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';
formatSpec2_2m = 'PRI_2m_sunlit %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';
formatSpec3_2m = 'PRI_2m_shadow %10.5f %10.5f %10.5f %10.5f %10.5f
%10.5f\n';

fprintf(fileID2,'%50s\n', col_header);
fprintf(fileID2,formatSpec_2m,PRI_2m_stats); %this prints the first line
fprintf(fileID2,formatSpec1_2m,PRI_2m_mixed_stats);
fprintf(fileID2,formatSpec2_2m,PRI_2m_sunlit_stats);
fprintf(fileID2,formatSpec3_2m,PRI_2m_shadow_stats);

fclose(fileID2);

end

%end

```

Chapter 3 MATLAB Code

```
%%
% name:    NEON_par_ptw.m
% date:    Sept. 10, 2021
% author:  Val Thomas & Paige Williams
% purpose: to read the NEON PAR data and do some basic calculations

%Setup initial parameters for the file.

%%
clear

%%
%to read in the list of NEON sites to loop through them
sites = 'E:\Ch2_NEON\sites_list_testing.txt';
sitesread = readtable(sites);
numsites=height(sitesread)

for i = 1:numsites

%rootdir3 = 'E:\Ch2_NEON\NEON_allsites_Tair\ABBY';
in_path=strcat('E:\Ch2_NEON\NEON_allsites_PAR\',sitesread{i,:});
rootdir2 = in_path{1}
cd(rootdir2)

%out_path='/Users/thomasv/Dropbox
(VTFRS)/Randy_Val_Share/Silvilaser2021/funcdiv/Tair/';
out_path='E:\Ch2_NEON\NEON_processing_outputs\PAR_stats\';

%%

MyFolderInfo2 = dir;

filelist2 = dir(fullfile(rootdir2, '**\*.csv')); %get list of csv files in any
subfolder
filelist2 = filelist2(~[filelist2.isdir]);

ref_files2 = extractfield(filelist2,'name');
ref_paths2 = extractfield(filelist2,'folder');
str_names2 = string(ref_files2);
str_paths2 = string(ref_paths2);

len=length(str_names2);
count=1;
clear PAR_name_and_path
clear PAR_name2
clear Yo
%Get a list of the .060.030 csv files, which I believe are the 30 min par
%for top of the tower, at least at mountain lake. For other towers, you
%need to change this to the highest on they have. The lowest I saw was 030

for k=1:len
```

```

    Yo = strfind(str_names2(k),'.050.030. ');
    if Yo > 0
        PAR_name_and_path(count)=strcat(str_paths2(k),'\',str_names2(k)); %These are
the names and paths
        PAR_name2(count)=str_names2(k);
        count=count+1;
    end
end

k=0;
count=1;

%%

%do the first file
l=1;
outfile_par = strcat(out_path,PAR_name2(l), '_par.xlsx');
filename2 = PAR_name_and_path(l);

pardata=readtable(filename2);
timestart=table2array(pardata(:,1));
PARMean=pardata(:,3);
outPARMean=pardata(:,11);

%%
%read the start time and convert date stamps
tstart = convertCharsToStrings(timestart{1,:}); % the {} are because this is a cell
array - use instead of parenthesis
% tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
'TimeZone', 'UTC');

time_blocks=strsplit(tstart(1,:), '-');
Y = double(time_blocks(:,1));
M = double(time_blocks(:,2));
day_blocks = strsplit(time_blocks(:,3),'T');
D = double(day_blocks(:,1));
hms_blocks=strsplit(day_blocks(:,2), ':');
H=double(hms_blocks(:,1));
MI=double(hms_blocks(:,2));
sec_blocks=strsplit(hms_blocks(:,3),'Z');
S=double(sec_blocks(:,1));

tstamp2=datetime(Y,M,D,H,MI,S);

time_start = [tstamp2];
lentime = length(timestart);
for m=2:lentime

    %deal with the character string that contains the timestamps
    tstart = convertCharsToStrings(timestart(m,:));
    %tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
'TimeZone', 'UTC');
    time_blocks=strsplit(tstart(1,:), '-');
    Y = double(time_blocks(:,1));
    M = double(time_blocks(:,2));

```

```

    day_blocks = strsplit(time_blocks(:,3),'T');
    D = double(day_blocks(:,1));
    hms_blocks=strsplit(day_blocks(:,2),':');
    H=double(hms_blocks(:,1));
    MI=double(hms_blocks(:,2));
    sec_blocks=strsplit(hms_blocks(:,3),'Z');
    S=double(sec_blocks(:,1));

    tstamp2=datetime(Y,M,D,H,MI,S);

    time_start = [time_start;tstamp2];

end

bigPARmean = PARMean;
bigoutPARmean = outPARMean;
bigtimePAR = time_start;
%%

num_parfiles=length(PAR_name2);

for l=2:num_parfiles

    outfile_par = strcat(out_path,PAR_name2(l), '_par.xlsx');
    filename2 = PAR_name_and_path(l);

    pardata=readtable(filename2);
    timestart=table2array(pardata(:,1));
    PARMean=pardata(:,3);
    outPARMean=pardata(:,11);

    %%
    %read the start time and convert date stamps
    tstart = convertCharsToStrings(timestart{1,:}); % the {} are because this is a
cell array - use instead of parenthesis
    % tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
'TimeZone', 'UTC');

    time_blocks=strsplit(tstart(1,:), '-');
    Y = double(time_blocks(:,1));
    M = double(time_blocks(:,2));
    day_blocks = strsplit(time_blocks(:,3),'T');
    D = double(day_blocks(:,1));
    hms_blocks=strsplit(day_blocks(:,2),':');
    H=double(hms_blocks(:,1));
    MI=double(hms_blocks(:,2));
    sec_blocks=strsplit(hms_blocks(:,3),'Z');
    S=double(sec_blocks(:,1));

    tstamp2=datetime(Y,M,D,H,MI,S);

    time_start = [tstamp2];
    lentime = length(timestart);
    for m=2:lentime

```

```

        %deal with the character string that contains the timestamps
        tstart = convertCharsToStrings(timestart(m,:));
        %tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
        TimeZone', 'UTC');
        time_blocks=strsplit(tstart(1,:), '-');
        Y = double(time_blocks(:,1));
        M = double(time_blocks(:,2));
        day_blocks = strsplit(time_blocks(:,3), 'T');
        D = double(day_blocks(:,1));
        hms_blocks=strsplit(day_blocks(:,2), ':');
        H=double(hms_blocks(:,1));
        MI=double(hms_blocks(:,2));
        sec_blocks=strsplit(hms_blocks(:,3), 'Z');
        S=double(sec_blocks(:,1));

        tstamp2=datetime(Y,M,D,H,MI,S);

        time_start = [time_start;tstamp2];

    end

    bigPARmean = [bigPARmean;PARMean];
    bigoutPARmean = [bigoutPARmean;outPARMean];
    bigtimePAR = [bigtimePAR;time_start];

end
%%
%Plot incident PAR over time

%Get information from the file names about site
names2 = (convertCharsToStrings(str_names2)).';
name_blocks2=strsplit(names2(1,:), '.');
sz = 10;

figure

scatter(bigtimePAR,bigPARmean,sz,'filled');
title('Incident PAR for:',name_blocks2(3))
xlabel('time')
ylabel('PAR PPFD (umol/m2/s)')

drawnow;

%% make into a giant table and write to a file.

T=table(bigtimePAR, bigPARmean,...
    'VariableNames',{'timeBgn','PARmean'});

out_name = strcat(out_path,name_blocks2(3), '_meanPAR.csv');

%write the table to an excel file.
writetable(T,out_name);

end

```

```

%%
% name:      NEON_temp_ptw.m
% date:      Sept. 10, 2021
% author:    Val Thomas & Paige Williams
% purpose:   to read the NEON temp data and do some basic calculations

%%
clear
%%
%to read in the list of NEON sites to loop through them
sites = 'E:\Ch2_NEON\sites_list_testing.txt';
sitesread = readtable(sites);
numsites=height(sitesread)

for i = 1:numsites

%rootdir3 = 'E:\Ch2_NEON\NEON_allsites_Tair\ABBY';
in_path=strcat('E:\Ch2_NEON\NEON_allsites_Tair\ ',sitesread{i,:});
rootdir3 = in_path{1}
cd(rootdir3)

%outfile_par = strcat(out_path,temp_name2(1), '_temp.csv');
%filename3 = temp_name_and_path(1);

%out_path='/Users/thomasv/Dropbox
(VTFRS)/Randy_Val_Share/Silvilaser2021/functdiv/Tair/';
out_path='E:\Ch2_NEON\NEON_processing_outputs\Tair_stats\';
%%
MyFolderInfo3 = dir;
filelist3 = dir(fullfile(rootdir3, '**\*.csv')); %get list of csv files in any
subfolder
filelist3 = filelist3(~[filelist3.isdir]);

ref_files3 = extractfield(filelist3,'name');
ref_paths3 = extractfield(filelist3,'folder');
str_names3 = string(ref_files3);
str_paths3 = string(ref_paths3);

len=length(str_names3);
count=1;
clear temp_name_and_path
clear temp_name2
clear Yt

for k=1:len
    Yt = strfind(str_names3(k),'.050.030. '); %020 is the level on the flux tower
height & 030 means 30 mint interval,
    if Yt > 0
        temp_name_and_path(count)=strcat(str_paths3(k), '\ ',str_names3(k)); %These
are the names and paths
        temp_name2(count)=str_names3(k);
        count=count+1;
    end
end

```

```

end

k=0;
count=1;

%%
%do the first file
l=1;
outfile_par = strcat(out_path,temp_name2(l), '_temp.csv');
filename3 = temp_name_and_path(l);

tempdata=readtable(filename3);
timestart3=table2array(tempdata(:,1));
tempMean=tempdata(:,3);

%%
%read the start time and convert date stamps
tstart = convertCharsToStrings(timestart3{1,:}); % the {} are because this is a cell
array - use instead of parenthesis
% tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
'TimeZone', 'UTC');

time_blocks=strsplit(tstart(1,:), '-');
Y = double(time_blocks(:,1));
M = double(time_blocks(:,2));
day_blocks = strsplit(time_blocks(:,3), 'T');
D = double(day_blocks(:,1));
hms_blocks=strsplit(day_blocks(:,2), ':');
H=double(hms_blocks(:,1));
MI=double(hms_blocks(:,2));
sec_blocks=strsplit(hms_blocks(:,3), 'Z');
S=double(sec_blocks(:,1));

tstamp3=datetime(Y,M,D,H,MI,S);

time_start = [tstamp3];
lentime = length(timestart3);
    for m=2:lentime

        %deal with the character string that contains the timestamps
        tstart = convertCharsToStrings(timestart3(m,:));
        %tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
'TimeZone', 'UTC');
        time_blocks=strsplit(tstart(1,:), '-');
        Y = double(time_blocks(:,1));
        M = double(time_blocks(:,2));
        day_blocks = strsplit(time_blocks(:,3), 'T');
        D = double(day_blocks(:,1));
        hms_blocks=strsplit(day_blocks(:,2), ':');
        H=double(hms_blocks(:,1));
        MI=double(hms_blocks(:,2));
        sec_blocks=strsplit(hms_blocks(:,3), 'Z');
        S=double(sec_blocks(:,1));

        tstamp3=datetime(Y,M,D,H,MI,S);
    end

```

```

        time_start = [time_start;tstamp3];

    end

    bigtempmean = tempMean;
    bigtimetemp = time_start;
%%

    num_tempfiles=length(temp_name2);

    for l=2:num_tempfiles

        outfile_par = strcat(out_path,temp_name2(l), '_temp.xlsx');
        filename3 = temp_name_and_path(l);

        tempdata=readtable(filename3);
        timestart3=table2array(tempdata(:,1));
        tempMean=tempdata(:,3);

        %%
        %read the start time and convert date stamps
        tstart = convertCharsToStrings(timestart3{1,:}); % the {} are because this is a
cell array - use instead of parenthesis
        % tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
'TimeZone', 'UTC');

        time_blocks=strsplit(tstart(1,:), '-');
        Y = double(time_blocks(:,1));
        M = double(time_blocks(:,2));
        day_blocks = strsplit(time_blocks(:,3), 'T');
        D = double(day_blocks(:,1));
        hms_blocks=strsplit(day_blocks(:,2), ':');
        H=double(hms_blocks(:,1));
        MI=double(hms_blocks(:,2));
        sec_blocks=strsplit(hms_blocks(:,3), 'Z');
        S=double(sec_blocks(:,1));

        tstamp3=datetime(Y,M,D,H,MI,S);

        time_start = [tstamp3];
        lentime = length(timestart3);
        for m=2:lentime

            %deal with the character string that contains the timestamps
            tstart = convertCharsToStrings(timestart3(m,:));
            %tstamp=datetime(tstart, 'InputFormat', 'uuuu-MM-dd'T'HH:mm:ss.SSS'Z'',
'TimeZone', 'UTC');
            time_blocks=strsplit(tstart(1,:), '-');
            Y = double(time_blocks(:,1));
            M = double(time_blocks(:,2));
            day_blocks = strsplit(time_blocks(:,3), 'T');
            D = double(day_blocks(:,1));

```

```

hms_blocks=strsplit(day_blocks(:,2),':');
H=double(hms_blocks(:,1));
MI=double(hms_blocks(:,2));
sec_blocks=strsplit(hms_blocks(:,3),'Z');
S=double(sec_blocks(:,1));

tstamp3=datetime(Y,M,D,H,MI,S);

time_start = [time_start;tstamp3];

end

bigtempmean = [bigtempmean;tempMean];
bigtimetemp = [bigtimetemp;time_start];

end
%%
%Plot Temp over time

%Get information from the file names about site
names3 = (convertCharsToStrings(str_names3)).';
name_blocks3=strsplit(names3(1,:),'.');
sz = 10;

figure

scatter(bigtimetemp,bigtempmean,sz,'filled');
title('Single Aspirated Air Temperature for:',name_blocks3(3))
xlabel('time')
ylabel('Temperature (celsius)')

drawnow;

%% make into a giant table and write to a file.

T=table(bigtimetemp, bigtempmean,...
'VariableNames',{'timeBgn','meanTair'});

out_name = strcat(out_path,name_blocks3(3), '_meanTair.csv');

%write the table to an excel file.
writetable(T,out_name);
end

```

```

%%
% name:    NEON_eddyflux_ustar_timestamp_ptw.m
% date:    Sept. 10, 2021
% author:  Val Thomas & Paige Williams
% purpose: to read all of the NEON eddyflux data (from multiple files),
%          graphs the net carbon exchange, and writes to a file.
%          The assumption is that you have downloaded the eddy flux package
%          from the data folder, and unzipped the file (and any subsequent
%          zipped files in the subfolders). The eddy flux package is
%          DP4.00200.001

% DP4.00067 Carbon dioxide flux, qfqm | fluxCo2 | 30-min

%%
clear

%%
%to read in the list of NEON sites to loop through them
sites = 'E:\Ch2_NEON\sites_list_testing.txt';
sitesread = readtable(sites);
numsites=height(sitesread)

for i = 1:numsites
in_path=strcat('E:\Ch2_NEON\NEON_allsites-eddyflux\',sitesread{i,:});
rootdir = in_path{1}
cd(rootdir)
site=string(sitesread{i,:})
out_path='E:\Ch2_NEON\NEON_processing_outputs\eddyflux_stats\';

%%
% Get the zip file names and unzip all of the files
MyFolderInfo = dir;
filelist = dir(fullfile(rootdir, '**\*.h5.gz')); %get list of h5 files in any
subfolder
filelist = filelist(~[filelist.isdir]);

ref_files = extractfield(filelist,'name');
ref_paths = extractfield(filelist,'folder');
str_names = string(ref_files);
str_paths = string(ref_paths);

%skip this unless you haven't processed/unzipped the data before.
%It takes too long otherwise
%leny=length(str_names);
%for i=1:leny
%    filename = strcat(str_paths(i),'\',str_names(i));
%    gunzip(filename);
%end

%Now get a list of the .h5 files
filelist = dir(fullfile(rootdir, '**\*.h5')); %get list of h5 files in any subfolder
filelist = filelist(~[filelist.isdir]);

```

```

ref_files = extractfield(filelist,'name');
ref_paths = extractfield(filelist,'folder');
str_names = string(ref_files);
str_paths = string(ref_paths);

len=length(str_names);

%%

%do the first file
l=1;
filename = strcat(str_paths(l), '/', str_names(l));

eddy_info = h5info(filename);
h5_path_co2=strcat('/',site,'/dp04/data/fluxCo2/nsae');
eddy_co2 = h5read(filename, h5_path_co2);
co2 = eddy_co2.flux; %this is the flux
timestart = eddy_co2.timeBgn; %this is the time stamp for the beginning of the
halfhour

%date character formatting, yes it has a weird space at the end
%time begin as a date vector, these are easy to index
ts = datevec(eddy_co2.timeBgn, 'yyyy-mm-ddTHH:MM:SS.FFF ');
%but if you'd rather have it as a datetime object, you can do it here
ts_as_dt = datetime(ts);

h5_path_ustar=strcat('/',site,'/dp04/data/fluxMome/turb');
eddy_ustar = h5read(filename, h5_path_ustar);
ustar = eddy_ustar.veloFric; %this is the flux
%Note. by this point we have read the co2 and time stamp, and could
%simply write them to a file.

%%

bigco2 = co2;
bigtime = ts_as_dt;
bigustar=ustar;
%%
for l=2:len

    %%
    %Set input names

    filename = strcat(str_paths(l), '/', str_names(l));

    %%
    % Read the NEON eddy flux data

    eddy_info = h5info(filename);
    %eddy_co2 = h5read(filename, '/HARV/dp04/data/fluxCo2/nsae');
    h5_path_co2=strcat('/',site,'/dp04/data/fluxCo2/nsae');
    eddy_co2 = h5read(filename, h5_path_co2);

```

```

co2 = eddy_co2.flux;
%date character formatting, yes it has a weird space at the end
%time begin as a date vector, these are easy to index
ts = datevec(eddy_co2.timeBgn','yyyy-mm-ddTHH:MM:SS.FFF ');
%but if you'd rather have it as a datetime object, you can do it here
ts_as_dt = datetime(ts);

h5_path_ustar=strcat('/',site,'/dp04/data/fluxMome/turb');
eddy_ustar = h5read(filename, h5_path_ustar);
ustar = eddy_ustar.veloFric; %this is the flux

bigco2=[bigco2;co2];
bigtime=[bigtime;ts_as_dt];
bigustar=[bigustar;ustar];
end

% Make a graph for a gut check!

%Get information from the file names about site. The third element is
%the site name.
names = (convertCharsToStrings(str_names)).';
name_blocks=strsplit(names(1,:), '.');

sz = 10; %marker size for the graph
figure
scatter(bigtime,bigco2,sz,'filled');
title('Net surface atmosphere exchange for:',name_blocks(3))
xlabel('time')
ylabel('CO2 flux (umol/m2/s)')
ylim([-100 100])
drawnow;

figure
scatter(bigtime,bigustar,sz,'filled');
title('Frictional Velocity for:',name_blocks(3))
xlabel('time')
ylabel('U* (m/s)')
drawnow;

%%
%make into a giant table and write to a csv file. Simply change the extension to
change the type.

T=table(bigtime, bigco2, bigustar,...
    'VariableNames',{'timeBgn','CO2flux', 'U*'});

%out_path='/Users/thomasv/Dropbox
(VTFRS)/Randy_Val_Share/Silvilaser2021/functdiv/flux/';
out_name = strcat(out_path,name_blocks(3), '_net_CO2_fluxes_ustar_tstamp.csv');

%write the table to an excel file.
writetable(T,out_name);

end

```

```

%%
%%
% name:     NEE_to_GPP_ptw.m
% date:     Sept. 23, 2021
% author:   Val Thomas & Paige Williams
% purpose:  The purpose is calculate mean midday GPP, NEE, and Reco for the 2
%           week period centered around the NEON AOP flights.
%           To do this:
%           - open the CO2, U*, PAR and Temp Data
%           - combine into a big array with all of these variables
%           - Determine the U* threshold using JJA nighttime data
%           - Develop the respiration temperature relationship
%           - Subtract respiration data from NEE
%
%%
clear
%Change this section
site='PUUM';

out_path='E:\Ch2_NEON\NEON_processing_outputs\Flux_partitioned_stats\';
out_path2='E:\Ch2_NEON\NEON_processing_outputs\Flux_middaymean_stats\';

%%
% Open and read the 3 data files
inpath1='E:\Ch2_NEON\NEON_processing_outputs\eddyflux_stats\';
filename1=strcat(inpath1,site,'_net_CO2_fluxes_ustar_tstamp.csv');

inpath2='/Users/thomasv/Dropbox
(VTFRS)/Randy_Val_Share/SilviLaser2021/funcdiv/PAR/';
inpath2='E:\Ch2_NEON\NEON_processing_outputs\PAR_stats\';
filename2=strcat(inpath2,site,'_meanPAR.csv');

inpath3='E:\Ch2_NEON\NEON_processing_outputs\Tair_stats\';
filename3=strcat(inpath3,site,'_meanTair.csv');

CO2ustardata = readtable(filename1);
PARdata=readtable(filename2);
Tairdata=readtable(filename3);

figure
plot(CO2ustardata{:,1},CO2ustardata{:,2})
%ylim([-50,50])
title('Net surface atmosphere exchange for:',site)
    xlabel('time')
    ylabel('CO2 flux (umol/m2/s)')

figure
plot(PARdata{:,1},PARdata{:,2})
title('Incident PAR for:',site)
    xlabel('time')
    ylabel('Mean PAR PPFD (umol/m2/s)')

```

```

%Parry=CO2ustardata(50020:68000,:); %this just trims the CO2 data a bit because it
doesn't
%Party=PARdata(71000:92830,:);
Parry=CO2ustardata(:, :); %if using single year, these should match up
Party=PARdata(:, :);
TJ=join(Parry, Party);
%TJ=join(CO2ustardata, PARdata);

TJ2=join(TJ, Tairdata);
tjt=table2timetable(TJ2);
%%
%Filter to get night data, where PAR is less than or equal to zero
%night=tjt{:,3}<=0;
night=tjt{:,3}<1; %0 is too strict for some sites
TJnight=tjt(night,:);

%Filter to get rid of negative CO2 (because respiration is positive)
n2=TJnight{:,1}>0;
TJnight2=TJnight(n2,:);

%Filter for low U*. Let's just get rid of the bottom 5% and call it good.
Y = quantile(TJnight2{:,2},20);
n3=TJnight2{:,2}>Y(1);
TJnight3=TJnight2(n3,:);

%%
%Temp
%Get rid of NAN temperatures at cold sites (alaska BONA and DELA) (assume no
respiration less than 0celsius
%goodT=TJnight3{:,4};
%testgoodT=~isnan(goodT)
%validR2 = TJnight3(testgoodT,:);

%Get rid of cold temperatures (assume no respiration less than 0celsius
% maybe we should just extract JJA instead with just cold T the R2 is
% horrid
goodT=TJnight3{:,4}>0;
validR2=TJnight3(goodT,:);

%%
%respiration
%Get rid of outlier respiration, where the value is ridiculous.
reasonableR=validR2{:,1}<100;
validR4=validR2(reasonableR,:);

%%
%can we now select only JJA?
%lets start with April 01 - Sept 30 which is what the data range for AOP is

S = timerange('2017-01-01','2017-12-30');
S1 = timerange('2018-01-01','2018-12-30');
S2 = timerange('2019-01-01','2019-12-30');
S3 = timerange('2020-01-01','2020-12-30');
S4 = timerange('2021-01-01','2021-12-30');

```

```

tt=validR4(S,:);
tt1=validR4(S1,:);
tt2=validR4(S2,:);
tt3=validR4(S3,:);
tt4=validR4(S4,:);

summer=tt;
summer=[summer;tt1];
summer=[summer;tt2];
summer=[summer;tt3];
summer=[summer;tt4];

%%
% let
%lets look at relationship between respiration and air temperature
%scatter(validR2{:,5},validR2{:,2});

nightco2=summer{:,1}; %by separating them out like this, you can use the curve
nightT=summer{:,4}; %fit app. Otherwise, not needed

[REfit,gof]=fit(nightT,nightco2,'exp1'); %f(x) = a*exp(b*x)

figure
plot(REfit,nightT,nightco2)
title('Model Fit for Summer Temp & Resp (nighttime co2 flux):',site)

%%
% Get the daytime values, apply the model, and separate NEE components

%Filter to get daytime data, where PAR is greater than zero
day=tjt{:,3}>1; %again, zero may be too strict
TJday=tjt(day,:);

S = timerange('2017-01-01','2017-12-30');
S1 = timerange('2018-01-01','2018-12-30');
S2 = timerange('2019-01-01','2019-12-30');
S3 = timerange('2020-01-01','2020-12-30');
S4 = timerange('2021-01-01','2021-12-30');

tt=TJday(S,:);
tt1=TJday(S1,:);
tt2=TJday(S2,:);
tt3=TJday(S3,:);
tt4=TJday(S4,:);

summer2=tt;
summer2=[summer2;tt1];
summer2=[summer2;tt2];
summer2=[summer2;tt3];
summer2=[summer2;tt4];

```

```

%Apply the model to get daytime respiration
Reco=REfit.a*exp(REfit.b*summer2{:,4});
GEE=summer2{:,1}-Reco;
GPP=-GEE;

Resp=array2table(Reco);
GEP=array2table(GPP);
outtable=[summer2 Resp GEP];

% Get rid of any values below zero

%Filter to get rid of negative CO2 (because respiration is positive)
qual1=outtable{:,6}>0;
outtable2=outtable(qual1,:);

%graph for a gut check
figure
sz = 30;
scatter(outtable2.timeBgn,outtable2.GPP,sz,'filled')
title('Summer Daytime GPP',site)
    xlabel('time')
    ylabel('GPP ( $\mu\text{mol CO}_2/\text{m}^2/\text{s}$ )')
ylim([0,50])

%%
%output to table

out_name = strcat(out_path,site, '_partitioned_GPP_TEST.csv');

%write the table to a csv file.
writetimetable(outtable2,out_name);

%%
%get Midday GPP
%Assume midday is between 14:00 and 16:00 (because UTC)

afternoon = outtable2.timeBgn.Hour>=14;
pm=outtable2(afternoon,:);

mid=pm.timeBgn.Hour<=16;
midday=pm(mid,:);

% Use 'daily' timestep with 'mean' method to get the midday GPP average

%midday_means=retime(midday,'daily','mean');
midday_means=retime(midday,'daily','median');

%output to table
out_name2 = strcat(out_path2,site, '_partitioned_middaymeanTEST.csv');

%write the table to a csv file.
writetimetable(midday_means,out_name2);

```

```

%%
% Name: NEON_mosaic_SRs_ewstrips_ptw.m
% Author: Val Thomas & Paige Williams
% Date: Jan. 27, 2023
% Purpose: To mosaic the individual SR tiles for each NEON site for each
% year.
% Note: Run one site at a time so as not to get too much confusion.
% Note: Run in mac environment, assuming external drive attached.
%%
clear
%Change this line to specify the site
site='JERC';

index='NMDI';

%Set the relevant paths for input and output files and working directory. Change as
needed.

%Paige
in_path='E:\Ch2_NEON\NEON_RS\JERC\';
out_path='E:\Ch2_NEON\NEON_RS\JERC\mosaic_strips\';
cd 'E:\Ch2_NEON\NEON_RS\JERC\'

%%
%read all of the files in SRMs into a list array and filter for site

rootdir = in_path;
MyFolderInfo = dir;
filelist = dir(fullfile(rootdir, '**/*.tif')); %get list of tif files in any
subfolder
filelist = filelist(~[filelist.isdir]);
ref_files = extractfield(filelist,'name');
ref_paths = extractfield(filelist,'folder');
str_names = string(ref_files);
str_paths = string(ref_paths);

%Filter for _SR, in case something else gets stored there
len=length(str_names);
count=1;
for k=1:len
    Y = strfind(str_names(k),'_NMDI.tif');
    if Y > 0
        SR_name(count)=str_names(k); %This is just the name of only _DSM files
        SR_namepath(count)=strcat(str_paths(k),'/',str_names(k)); %name with path
        count=count+1;
    end
end

%filter by site
len2=length(SR_name);
count=1;
for k=1:len2
    Y = strfind(SR_name(k),site);
    if Y > 0
        site_name(count)=SR_name(k);
    end
end

```

```

        site_namepath(count)=SR_namepath(k); %These are only the _DSM files
        count=count+1;
    end
end

len_site=length(site_name);

%%
%Use the years col to figure out which rows are the year of interest
for i=1
    years(i)
    %year_index=yearcol==years(i);

    %extract the files for this year.
    site_year=site_name; %just the name
    sitepath_year=site_namepath; %name and path

    %get the number of files in this year
    len3=length(site_year);
    %because the tiles are organized by rows/cols, we can mosaic the east-
    %west strips first

    %so mine the names to pull the coordinates out of the name.
    %In the name, the easting are first, then the northings.
    %
    split_year=split(site_year, '_');
    eastings=split_year(:,5);
    northings=split_year(:,6);

    %get the number of east-west strips by determing how many unique
    %northings there are
    ew_strips=unique(northings);
    num_strips=length(ew_strips);

    %get the number of north-south by determing how many unique
    %easting there are
    ns_strips=unique(eastings);
    num_nsstrips=length(ns_strips);

    %loop through the data from south to north and mosaic the east-west tiles and
    %save them as east-west strips.

    for j=1:num_strips
        %search the files for this year, and extract all files with this
        %particular (j) northing.
        count=1;
        for l=1:len3
            Y = strfind(site_year(l),ew_strips(j));
            if Y > 0
                strip_name(count)=site_year(l);
                strip_namepath(count)=sitepath_year(l); %These are only the _DSM
files
                count=count+1;
            end
        end
    end
end

```

```

len_strip=length(strip_name);

if len_strip >1
    %note: I am not sure we can assume that they are ordered east
    % to west, which they were in my test.
    ew_split=split(strip_name,'_');
    eastings_small=str2double(ew_split(:,5));
    B=sort(eastings_small);
    sortcheck=eastings_small-B;
    S=sum(sortcheck);
    M=strcat('Warning - strip unordered or has gaps: ',ew_strips(j));
    minS=min(eastings_small);
    maxS=max(eastings_small);
    check = (maxS-minS)/len_strip;
else
    S=1;
end

if (S>0) | check>1000
    M
else
    [westA,westR] = readgeoraster(strip_namepath(1));
    [eastA,eastR] = readgeoraster(strip_namepath(2));
    mosaicA = [westA eastA];
    xlimits = [westR.XWorldLimits(1) eastR.XWorldLimits(2)];
    ylimits = westR.YWorldLimits;
    xcellextent= 1;
    ycellextent=1;
    mosaicR = maprefcells(xlimits,ylimits,size(mosaicA));
    % if the files are at the edge - there might be gaps in the
    % strip. So check to see if all of the names in len_strip are
    % 1000 apart.

    for m=3:len_strip
        %Read in the georasters for this strip and mosaic them.
        [eastA,eastR] = readgeoraster(strip_namepath(m));
        mosaicA = [mosaicA eastA];
        xlimits = [westR.XWorldLimits(1) eastR.XWorldLimits(2)];

    end

    % Extract the first band.
    band1=mosaicA; %we don't need the other 2 bands - they are the same as
band 1

outfile=strcat(out_path,site,'_',ew_strips(j),'_EWstrip_',index,'_',num2str(xlimits(1
)), '_', num2str(xlimits(2)),'.tif');
    R = maprefcells(xlimits, ylimits, size(band1),...
        'ColumnsStartFrom','north',...
        'RowsStartFrom','west');
    geotiffwrite(outfile, band1,R, 'CoordRefSysCode',32616);

```

```
        end
    clear strip_name strip_namepath eastings northings
end

%%

    b='got to here'
end
```

```

%%
% Name: NEON_mosaic_SRs_big_ptw.m
% Author: Val Thomas
% Date: Jan. 27, 2023
% Purpose: To mosaic the ew SR strips for each NEON site for each
% year.
% Note: Run one site at a time so as not to get too much confusion.
% Note: Run in mac environment, assuming external drive attached.
%%
clear
%Change this line to specify the site and change the projection
site='JERC';
EPSG = 32616; %UTM zone of site

index = 'NMDI'

in_path='E:\Ch2_NEON\NEON_RS\JERC\mosaic_strips';
out_path='E:\Ch2_NEON\NEON_RS\JERC\mosaic_final\';
cd 'E:\Ch2_NEON\NEON_RS\JERC\mosaic_strips'

%%
% Get the names and paths of the files
rootdir = in_path;

filelist = dir(fullfile(rootdir, '**/*.tif')); %get list of tif files in any
subfolder
filelist = filelist(~[filelist.isdir]);
ref_files = extractfield(filelist,'name');
ref_paths = extractfield(filelist,'folder');
str_names = string(ref_files);
str_paths = string(ref_paths);

%Filter for _site, in case something else gets stored there
len=length(str_names);
count=1;
for k=1:len
    Y = strfind(str_names(k),index);
    if Y > 0
        SR_name(count)=str_names(k); %This is just the name of only _DSM files
        SR_namepath(count)=strcat(str_paths(k), '/',str_names(k)); %name with path
        count=count+1;
    end
end

len_site=length(SR_name);

%Get the maximum EW extent across all years
pieces = split(SR_name,'_');
farwest = pieces(:,5);
fareasttif = split(pieces(:,6),'.tif');
fareast=fareasttif(:,1);

farwest2=str2double(farwest);
fareast2=str2double(fareast);

```

```

westextent=min(farwest2);
eastextent=max(fareast2);
xlimits=[westextent eastextent]; %this is the full east west extent

notherly=pieces(:, :, 2);
notherly2=str2double(notherly);

% I think we need to add 1000 to north extent because the file name is the southern
coordinate
northextent=max(notherly2)+1000;
southextent=min(notherly2);
ylimitsmax=[southextent northextent];

%%
%Loop by year. This is the main loop

for i=1

    %filter by year(i)
    count=1;

    %read the southern most strip
    [southA,southR]=readgeoraster(SR_namepath(1));
    x1=southR.XWorldLimits;
    y1=southR.YWorldLimits;

    %read one strip to the north
    [northA,northR]=readgeoraster(SR_namepath(2));
    x1=northR.XWorldLimits;

    ylimits = [southR.YWorldLimits(1) northR.YWorldLimits(2)];
    mosaicN = [northA; southA]; %you need a semicolon for N-S

    %make it a float?
    mosaic = single(mosaicN);

    % save the strip as a geotif. Add the set azimuth and elevation angle
    outfilebig=strcat(out_path,site,'_mosaic_', index, '.tif');
    R = maprefcells(xlimits, ylimitsmax, size(mosaic),...
        'ColumnsStartFrom','north',...
        'RowsStartFrom','west');
    geotiffwrite(outfilebig, mosaic,R, 'CoordRefSysCode',EPSG,
        'TiffType','bigtiff');

    years(i)

    clear mosaic mosaicN mosaicNb mosaicNRb southA southB northA northB
    clear name namepath
end

```