Generation and evaluation of forest aboveground biomass and irrigation products for the Southern United States using Earth Observation data

by

Md. Mozahidul Islam

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science in Natural Resources

Auburn, Alabama
December 09, 2023

Keywords:
Aboveground Biomass (AGB), Lidar, NEON, Irrigation, Data Science

Approved by

Dr. Lana L. Narine, Chair, Assistant Professor, Geospatial Analytics, College of Forestry, Wildlife and Environment, Auburn University

Sanjiv Kumar, Assistant Professor, Earth System Modeling and Observation, College of Forestry, Wildlife and Environment, Auburn University

Richard Cristan, Extension Specialist & Asst. Professor, Forest Operations, College of Forestry, Wildlife and Environment, Auburn University

Abstract

The conservation of natural resources in the southern United States plays a pivotal role in addressing climate change, safeguarding its rich forest resources, maintaining ecological integrity, ensuring food sources, and sustaining the national economy. This region, characterized by rapid population growth and agricultural production, demands special attention for its holistic development. In light of this, the study encompasses two interrelated studies: the first deriving high-resolution spatially continuous aboveground biomass (AGB) maps using NEON's airborne lidar, while the second delves into identifying zones with varying irrigation water demands to enhance agricultural productivity.

Chapter 2 underscores the critical need for accurate high-resolution AGB mapping to support ecological integrity, carbon dynamics, and natural resource management. Despite the challenges associated with AGB estimation due to limited data, the National Ecological Observatory Network (NEON) offers a valuable resource with quality-controlled, consistent datasets at various spatial scales. This study leverages NEON's airborne lidar point clouds and plot-level vegetation structure data from Ordway-Swisher Biological Station (OSBS) in Florida, Talladega National Forest (TALL) in Alabama, and Oak Ridge National Laboratory (ORNL) in Tennessee. By utilizing lidar-derived canopy metrics, Sentinel imagery, and ancillary data, AGB is predicted using multiple linear regression (MLR) and Random Forest (RF) models. MLR outperforms RF across all sites, yielding R^2 values of 0.91, 0.52, and 0.63 for OSBS, TALL, and ORNL respectively, compared to RF's 0.67, 0.20, and 0.40. The %RMSE is 31.32%, 32.7%, and 30.75% for MLR, as opposed to 71.44%, 44.21%, and 45.47% for RF, respectively. Sites with steep slopes and complex structures (e.g., ORNL and TALL) exhibit lower accuracy, highlighting the need for an increased number and distribution of field plots to enhance AGB estimation accuracy. In summary, this study sheds light on the potential and limitations of NEON's datasets in generating spatially continuous AGB estimates.

Chapter 3 shifts focus towards optimizing agricultural efficiency and ensuring the sustainable allocation of water resources by comprehending the specific water requirements unique to the

region, particularly the dynamics of root zone soil moisture. This research concentrates on three key southern states- Alabama (AL), Florida (FL), and Georgia (GA), recognizing their substantial contributions to agricultural production, the national economy, and the regional ecological equilibrium. Departing from traditional manual data collection, this study employs a data-driven approach to delineate irrigation potential zones. The methodology entails monthly averaging of daily evapotranspiration, precipitation, and root zone soil moisture during the cropping season (May to October) from 2015 to 2023. A comprehensive correlation analysis unveils the intricate relationship between water demand (expressed as potential evapotranspiration – precipitation) and root zone soil moisture, exhibiting a negative correlation pattern spanning from -0.06 to -0.78. The correlation map was categorized into five distinct classes, highlighting varying water requirements in different regions. Analysis of variance (ANOVA) establishes the high overall significance of the correlation categories and varying irrigation required regions ($F = 38.59$, $p < 0.05$), and Tukey's Honest Significant Difference (HSD) post hoc test confirms the statistical significance of all categories within the $p < 0.05$ range. These research findings equip stakeholders with invaluable insights to conserve resources and promote sustainable agricultural practices. By seamlessly integrating data science, remote sensing datasets, and agricultural resource management, this study emerges as a pivotal catalyst for optimizing irrigation practices and ensuring the sustainability of precision agriculture in the region.

Acknowledgement

First and foremost, I extend my sincere thanks to the Divine, the Creator of all, whose blessings have endowed me with the health and strength required to undertake and accomplish this scholarly endeavor. I am profoundly grateful for the mental serenity and clarity of purpose that have allowed me to navigate the challenges of this intellectual pursuit, even in moments of intense pressure and emotional turbulence.

Dr. Lana L. Narine, my esteemed academic supervisor, deserves my deepest gratitude. Her unwavering support, infinite patience, and remarkable organizational skills have been indispensable to the realization of this work. In every circumstance, her radiant smile has provided solace, even in the face of the most daunting challenges. Without her guidance and encouragement, this journey would have been an insurmountable feat. My heart brims with gratitude for her steadfast belief in me and her pivotal role in my academic voyage.

I must also express my profound appreciation to Dr. Sanjiv Kumar, who graciously supervised me in the pursuit of the second objective of this research. Dr. Kumar's expertise in data science and his insightful comparisons of South Asian agricultural practices with those in the United States have broadened my understanding immeasurably. His guidance and mentorship have been invaluable throughout this endeavor, and for that, I am immensely thankful.

To Dr. Richard Cristian, a respected member of my committee, I extend my gratitude for his unwavering commitment, despite his demanding schedule. Dr. Cristian's constructive feedback and innovative perspectives have been instrumental in shaping this study. I am deeply appreciative of his generous assistance in reviewing my manuscripts, which has significantly enhanced the quality of my work.

In the depths of my heart, I reserve a special place for my family. To my mother, the wellspring of my motivation, my father, a philosopher of life's profound truths, my beloved wife, and my supportive siblings, I owe an immeasurable debt of gratitude. Their unwavering encouragement

and the timeless wisdom of the Quranic verse, "Allah does not charge a soul except [with that within] its capacity" (2:286), have provided me with the strength to persevere, even in moments of exhaustion and doubt.

I would be remiss not to acknowledge the contributions of my colleagues, friends, classmates, well-wishers, and critics. Your collective support and constructive input have been invaluable in shaping my ideas and honing my research. I extend a special word of gratitude to Sckhylar Brown for the invaluable assistance provided during the lidar analysis, as well as to all my lab mates, peers, and writing consultants for their unwavering dedication to meaningful discourse.

In closing, this journey of scholarly pursuit has been challenging, yet immensely rewarding, and it would not have been possible without the support and encouragement of those mentioned above. To all who have played a part in this endeavor, I extend my deepest appreciation and heartfelt thanks.

Table of contents

List of Figures

List of Tables

List of Abbreviations

ABA	-	Area-Based Approaches
AGB	-	Aboveground Biomass
AL	-	Alabama
ANOVA	-	Analysis of Variance
AOP	-	Airborne Platform
CD	-	Crop Density
CDL	-	Cropland Data Layer
CHIRPS	-	Climate Hazards Group Infrared Precipitation with Stations
CV	-	Cross-Validation
DBH	-	Diameter At Breast Height
DD	-	Degree Decimal
ECMWF	-	European Centre for Medium-Range Weather Forecasts
ERA5	-	Fifth Generation of The European Reanalysis for Global Climate Monitoring
ERA5-Land	-	ERA5-Land Is a Subset of ERA5 Data Specifically Focused on Land Based Variables
ET	-	Evapotranspiration
FL	-	Florida
GA	-	Georgia
GEDI	-	Global Ecosystem Dynamics Investigation
GEE	-	Google Earth Engine
GIS	-	Geographic Information System
HSD	-	Honest Significant Difference
ICESat-2	-	Ice, Cloud, And Land Elevation Satellite-2
IEEE	-	Institute Of Electrical and Electronics Engineers
IncMSE	-	Incremental Mean Squared Error
IQR	-	Interquartile Range
Lidar	-	Derived- Information or Data Derived from Lidar Technology
LiDAR	-	Light Detection and Ranging

MIrAD-US	-	Moderate Resolution Imaging Spectroradiometer (MODIS) Irrigated Agriculture Datasets for The Conterminous United States
MLR	-	Multiple Linear Regression
MODIS	-	Moderate Resolution Imaging Spectroradiometer
NASA	-	National Aeronautics and Space Administration
NDVI	-	Normalized Difference Vegetation Index
NEON	-	National Ecological Observatory Network
NIR	-	Near Infrared
NISAR	-	Nasa Isro Synthetic Aperture Radar
NLCD	-	National Land Cover Dataset
NOAA	-	National Oceanic and Atmospheric Administration
OOB	-	Out Of Bag
ORNL	-	Oak Ridge National Laboratory
OSBS	-	Ordway-Swisher Biological Station
PET	-	Potential Evapotranspiration
PET-P	-	Potential Evapotranspiration
PlnCur	-	Plan Curvature
ProCur	-	Profile Curvature
RF	-	Random Forest
RMSE	-	Root Mean Square Error
SAR	-	Synthetic Aperture Radar
SM	-	Soil Moisture
SMAP	-	Soil Moisture Active and Passive
SMOS	-	Soil Moisture of Ocean Salinity
SRI	-	Solar Radiation Index
SWIR	-	Short Wave Infra-Red
TALL	-	Talladega National Forest
TBA	-	Tree-Based Approach
TPI	-	Topographic Position Index
U.S.	-	United States Wave Infrared
US	-	United States

UTM	-	Universal Transverse Mercator (A Map Projection System)
VIF	-	Variance Inflation Factor
VST	-	Vegetation Structural

Chapter 1: Introduction

Sustainable natural resources management in the southern United States is important for ecological integrity, forest production, and food security (Williams et al., 2017). Biomass is one of the key parameters that indicates the amount of vegetation, ecological status, the role of plants, potential productivity of rangeland, and hydrological properties of an area (Ajani et al., 2013). Specifically, forest aboveground biomass (AGB) is the most studied parameter for understanding the ecological status and forest productivity (Duncanson et al., 2020) of an area. More than 50% of the area of the Alabama, Florida, and Georgia states in the Southern United States, is forested, and the cultivated land is the second most predominant land use and land cover (LULC) class (Wickham et al., 2021). To continue ongoing agricultural production in vast areas and sustainable management of water resources in this region, classifying areas with higher to lower irrigation needs to be estimated. Therefore, for sustainable management of natural resources in this region, both high-resolution spatially continuous AGB estimation and identifying different areas with varying irrigation needs will help ensure sustainable economic and ecological growth in this area (J. H. Brown et al., 2005; Kim et al., 2021).

Despite the importance of AGB estimation, it is often challenging to estimate over large areas, especially for dense mixed forested sites (Jenkins et al., 2003; Khati et al., 2020; Qi et al., 2019; Thomas & Famiglietti, 2019). Although manually collecting tree information like measuring each tree's height, diameter at breast height, and using statistical equations to calculate the AGB of each tree is accurate, it is too time and cost-consuming, inaccessible to several locations, and prone to human error (Lu et al., 2016; Wang et al., 2022).

Over large areas, space-based optical remote sensing technology (e.g., satellite imagery) is useful for monitoring forest coverage, but may not be well-suited for deriving forest structural information, such as tree heights (Lu et al., 2016). The study reported that estimating biomass using only optimal remote sensing compromised accuracy due to lack of tree height information (Zeng et al., 2022). Forest heights are often used as surrogates for biomass estimation (Y. Zhang et al., 2019) since taller trees have disproportionately large biomass despite the same dbh values

(J. H. Brown et al., 2005). Therefore, synergistic use of active sensors (i.e., light detection and ranging or lidar) and passive remote sensing data (e.g. imagery) has been found to be more accurate (Nandy et al., 2021).

The airborne lidar is an active remote sensing device that is capable of precisely measuring ground elevation and tree heights by calculating the time required for a pulse of laser energy to reach the ground and reflect back to the sensor (Coops et al., 2021; Popescu, 2007). Airborne and spaceborne are two lidar platforms that differ by preciseness, coverage, and altitude of the flight operation (Nie et al., 2018). Although airborne lidar is the more accurate data source for forest applications, its operation is often limited due to the high operation costs (Silva et al., 2021). As such, airborne lidar is often planned for a specific project with outcomes significantly varying by mission focus and logistic availability (Silva et al., 2021). However, the National Ecological Observatory Network (NEON) is a government-owned platform that feely provides airborne lidar point clouds and up-to-date vegetation structural datasets across 83 sites in the United States. These quality-controlled datasets enable the generation of high-resolution AGB mapping. However, there are currently no available AGB map products for NEON sites. This study was undertaken to demonstrate a methodology to create high-resolution spatially continuous AGB estimates for three NEON sites, that that can be adapted to other NEON sites.

The first part of this study serves to develop a methodology for applying vegetation structural data and airborne lidar point cloud datasets for three National Ecological Observatory Network (NEON) sites in the Southeastern US to develop high-resolution aboveground biomass products. The significance of using NEON datasets includes the availability of periodically acquired up-to-date airborne lidar point clouds (DP1.30003.001) and vegetation structural information (DP1.10098.001). Those manual field data are collected by updated techniques while maintaining a high standard of quality control (García et al., 2018; Khati et al., 2020; Scholl et al., 2020). For this work, NEON's airborne lidar point clouds and vegetation measurements were used to develop reference AGB estimates for:(1) Talladega National Forest (TALL) in Alabama, (2) Oak Ridge National Laboratory (ORNL) in Florida, and (3) Ordway-Swisher Biological Station (OSBS) in Tennessee. The datasets were used for creating spatially continuous 20m resolution AGB maps for each NEON site.

In addition to forest vegetation, the southern US is also significantly renowned for agricultural production and rapid population growth rate (Pervez & Brown, 2010). A recent study shows about 14% of agricultural products are generated from this region and the population growth rate is 10.2% higher than the national average of 7.4% (Zaussinger et al., 2019). The groundwater resources of the south are facing a rapidly declining trend which seems to be due to irrigation, speedy vegetation, and population growth trends (Condon et al., 2020; Condon & Maxwell, 2019; W. Zhang et al., 2020). Sustainable management of water resources is essential for agricultural production and overall natural resources management of this region. This management initiative demands up-to-date irrigation information, but the current county-wide irrigation dataset is almost a decade old and coarser in resolution (1km x 1km) (Pervez & Brown, 2010). Furthermore, very few studies were undertaken to find the irrigation water needs in the plant's root zone and dividing areas based on varying water needs (Bwambale et al., 2022; C. Zhang & Long, 2021; Y. Zhang et al., 2019). Furthermore, an irrigation dataset from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor for the contiguous US (irrigation) and global-scale vegetation information are provided at 250m and 1km grid sizes where the cell value presented in percentage of water applied (Pervez & Brown, 2010).

New science and satellite data have become available in the last decade, providing the potential to identify soil moisture conditions up to 100 cm from the soil surface in depth where most cultivated crop's roots are grown (McDermid et al., 2023; Xu et al., 2019; L. Zhang et al., 2022). For example, NASA launched the Soil Moisture Active and Passive (SMAP) satellite that collects soil moisture observations from 2015 and is capable of providing soil moisture information up to 100 cm, providing a unique opportunity to measure historical data for further analysis (Abbaszadeh et al., 2019; Lawston et al., 2017; W. Zhao et al., 2018). Finally, combining soil moisture with updated climatic datasets would be feasible for identifying varying irrigation demanded zones, while estimating spatially continuous AGB would be helpful for better understanding forest resources and would be useful for managing natural resources of this region.

1.1 Study objectives

The overall objectives were to-

1. Develop a methodology for utilizing NEON datasets to generate spatially continuous high-resolution AGB biomass products.

2. Identify varying irrigation demanded areas using soil moisture, climatic and remotely sensed datasets.

Each objective is addressed as a separate chapter in this thesis. Objective 1 is addressed in Chapter 2, and the irrigation potential zone identification using data science approaches is discussed in Chapter 3. Finally, conclusions and recommendations are presented in the fourth chapter of this thesis.

Chapter 2: A Spatially Continuous AGB Dataset for 3 NEON Sites in the Southeastern US Using NEON Airborne Lidar and Vegetation Data

2.1 Introduction

Forest aboveground biomass (AGB) is a critical parameter for understanding ecological integrity, forest production, and the hydrological properties of an area (Ajani et al. 2013; Williams et al. 2017). In the southern United States (US), forests cover over half of the land area, significantly contributing to the national economy, and carbon storage to mitigate the adverse effects of climate change (Gassert, Burke, and Zimmerman 2020; Wickham et al. 2021). Despite its importance, accurately estimating AGB is often challenging due to limited ground measurements, as well as spatially coincident tree attributes, precise methods, and statistical modeling (Wang et al. 2022). Although optical remote sensing, such as satellite imagery, has the capability to map a forest, it is unable to provide 3D vegetation structural information, such as height, or penetrate the top canopy layer of a tree (Lu et al. 2016). Meanwhile, only considering tree diameter at breast height (DBH) can lead to significant errors in AGB estimation, since taller trees have disproportionately more biomass than shorter trees (Brown, West, and Enquist 2005). Therefore, utilizing active sensors, such as light detection and ranging (lidar) that are capable of measuring tree height, helps improve AGB estimation models, with a combination of accurate field-observed datasets (Nandy, Srinet, and Padalia 2021; NOAA 2012).

Airborne and spaceborne lidar are two types of platforms used for lidar data collection, and they differ in their preciseness, coverage, altitude of operation, and overall cost. Lidar can precisely measure ground elevation and tree height by calculating the time required for a laser pulse to reflect the sensor (Coops et al. 2021; Popescu 2007). While airborne lidar is more accurate, its use may be limited due to cost. However, the availability of airborne lidar data from the National Ecological Observatory Network (NEON) sites across the United States offers opportunities for research (Nie et al. 2018; Silva et al. 2021). The NEON of the National Science Foundation (NSF) is a continental-scale observation facility designed for understanding how ecosystems of the US are changing over time. This facility is operated by Battelle and provides long-term open-access data for ecosystem research. It operates 81 field sites strategically located across 20

ecoclimatic domains across the US, including 47 terrestrial field sites and 34 freshwater aquatic field sites. When logistically possible, aquatic, and terrestrial sites are collocated (i.e., in proximity) to support the understanding of linkages across terrestrial and aquatic ecosystems and their interactions with the atmosphere. Therefore, the information from one site is extensively useful for similar kinds of forest sites in different parts of the US. NEON conducts extensive surveys consisting of structure measurements of individual woody and non-herbaceous perennial plants and maps the position of qualifying woody and non-woody plants based on plot/sub-plots levels. NEON's collected Vegetation Structural (VST) data are quality-controlled from in-situ measurements of live and standing dead woody individuals, shrub groups, and non-herbaceous perennial plants from all terrestrial NEON sites (NEON, 2022a). Alongside those, related remote sensing data products such as airborne lidar point clouds and high-resolution digital terrain and surface model datasets are available. For this study, along with the airborne lidar point cloud datasets, spatially corresponding tree attribute information was also used to model AGB. In the VST dataset, tree attributes including height, DBH, taxonomic information, tree formation, current tree condition, and relative site parameters are all included (Nandy et al. 2021; Venier et al. 2019).

There are few studies on using NEON VST or lidar point cloud datasets for modeling AGB for an area. Among published literature that used NEON's VST data include validation studies, or onsite testing purposes rather than AGB mapping (Duncanson et al. 2020, Wang et al. 2022). Currently, AGB map products are not available from NEON portals. Thus, this study serves to develop a methodology and assess the capability for creating AGB maps using only NEON datasets. Considering the same data collection protocol for every NEON site, written scripts, workflows, and overall methodology should be useful for other NEON sites and similar forest patches for generating spatially comprehensive AGB estimates.

Although airborne lidar data can provide height information with centimeter-level accuracy, the performance of AGB modeling with airborne lidar data depends on various factors, such as tree attributes, forest types, topography, climate, and weather patterns (Wang et al. 2022). Among these, forest types are a particularly important factor because of the differences in accuracies for capturing forest attributes from single-species to mixed-species forests (Popescu 2007). Mixed

forest types are often less accurate than single species since lidar has limitations in penetrating dense forest canopies (Donager, Sánchez Meador, and Blackburn 2021). Considering this, AGB estimation work for this study considered three southern forest sites with dense tree canopies and mixed forest structures with terrain differences. Three NEON sites, namely Ordway-Swisher Biological Station (OSBS) in Florida, Talladega National Forest (TALL) in Alabama and Oak Ridge National Laboratory (ORNL) in Tennessee were considered, where NEON reports quality-controlled tree attributes, including DBH, height, and current conditions of each tree (Kampe 2010; NEON 2022a, 2022b). The objectives of this study are listed below:

a) To develop a precise model training dataset using NEON's vegetation structure and airborne lidar point cloud data.
b) To develop models for spatially continuous AGB estimation using Multiple Linear Regression (MLR) and Random Forest (RF) at three NEON sites.
c) To compare model performance at three NEON sites and develop new insights and recommendations for NEON vegetation data protocol.

2.2 Materials and methods

2.2.1 Study site

This study considered three NEON sites in Florida, Alabama, and Tennessee, USA, where quality controlled VST and airborne lidar point cloud data are available. The Ordway-Swisher Biological Station (OSBS) in Putnam County, Florida, Talladega National Forest (TALL) located in the Bibb, Hale, and Perry counties County of Alabama, and Oak Ridge National Laboratory in Roane and Anderson counties in Tennessee were used (Figure 2-1). The OSBS and TALL sites are under NEON's D08 domain, and ORNL is under the D07 domain geographical centroid of each field site (29.689282 N, -81.993431W; 32.95047 N, -87.393259 W; 35.964128 N, -84.282588 W, respectively).

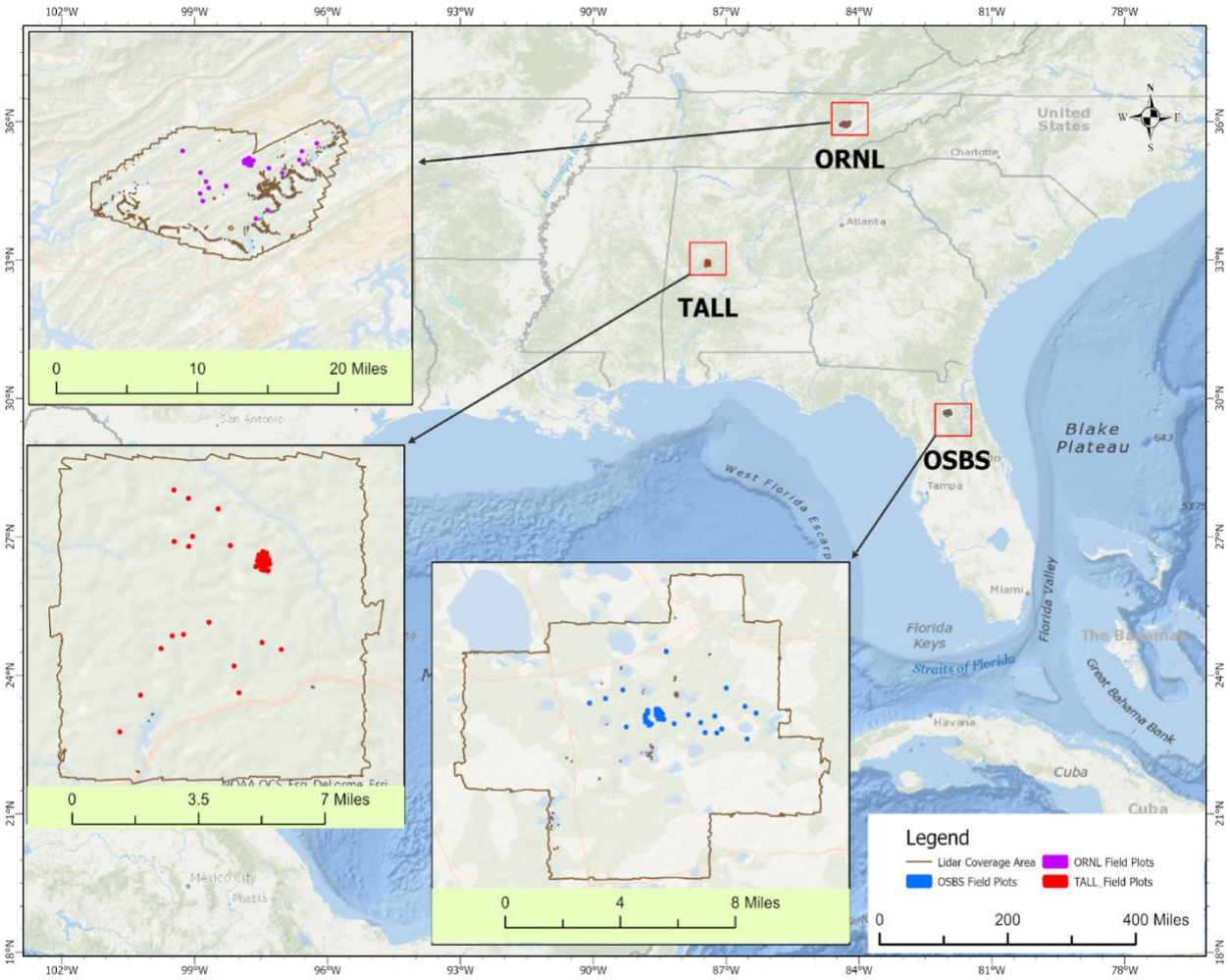

Figure 2-1: Three study sites in Florida, Alabama, and Tennessee, United States. The airborne lidar point data coverage of each site is shown in the line boundary, and polygons present field plots where tree attributes are available. All field plot polygons are 20m x 20 *m in resolution, clustered plots are those in which data is collected by the tower plotting system. Tree available subplots (20m) from 40m x 40m tower plots were extracted and used for clipping to get a 20m lidar extent. The blue, red, and magenta colors present field plots of OSBS, TALL, and ORNL field plots respectively.*

2.2.2 Weather pattern and field characteristics

The field characteristics and weather patterns of the TALL and ORNL are comparable, but OSBS is different. The OSBS site is hot and humid with gentle to flat terrain which has comparatively shorter trees than the TALL and ORNL sites (Battelle, 2023a, 2023c, 2023b). At OSBS, canopy cover ranges from 40-60%, with tree canopy heights between 15-30 m (avg 23 m)

and has a temperature range of 4 to 32 °C (avg 20.9 °C) (Battelle, 2023b). Average precipitation is 1302 mm (Table 2-1), and tree density ranges from 500-2000 trees per hectare (trees/ha). The TALL site has a higher canopy cover range of 50-80% and canopy heights of 20-35 m, with a temperature range of -9 to 36 °C and an elevation range of 63-179 m (Battelle, 2023c). Average precipitation is 1383 mm/year and tree density ranges from 500-3000 trees/ha. The ORNL site has the highest canopy cover of 60-80% and the highest canopy height among the sites with an average of 28 m. This site also has significant elevation ranges of 230-356 m (avg 344 m) with the highest temperature range of -11 to 35 °C (avg 14.4 °C), which is also the lowest average among the sites (Battelle, 2023a). Precipitation is between 1000-1400 mm per year and tree density ranges from 500-2000 trees/ha. The TALL site has a dominant land cover of oak-hickory, mixed hardwood, and pine forests, while the OSBS site is characterized by pine savanna and forest and the ORNL site, by oak-hickory forests mixed with deciduous and evergreen trees (Battelle, 2023a, 2023c, 2023b). The major water bodies in each site are Newnans Lake, Orange Lake, and Prairie Creek for OSBS; Cheaha Creek, Talladega Creek, and Coosa River for TALL; and Melton Hill Lake and Clinch River for ORNL (Table 2-1).

Table 2-1: Field climate condition, topographic differences, and overview of soil and vegetation dynamics of three study sites.

Characteristic	**OSBS**	**TALL**	**ORNL**
Canopy cover (%)	40-60	50-80	60-80
Canopy height (m)	15-30 (23 m)	20-35 (25 m)	20-40 (28.0m)
Temperature range and average (°C)	4 to 32 (20.9 °C)	-9 to 36 (17.2 °C)	-11 to 35 (14.4 °C)
Precipitation (mm/year)	1200-1600 (1302 mm)	1000-1500 (1383 mm)	1000-1400 (1340 mm)
Slope range (degrees)	0-20	0-45	0-45
Tree density (trees/ha)	500-2000	500-3000	500-2000
Elevation (m)	22-49 (46 m)	63-179 (166 m)	230 – 356 m (344 m)
Forest Coverage (ac)	9,500	12,924	14,000
Topography	Flat with gentle slopes	Hilly to mountainous with steep slopes	Rolling to mountainous with steep slopes
Climate zone	Warm and Humid	Humid subtropical	Humid continental
Soil class	Spodosols and Entisols	Fine, loamy, siliceous, thermic. Typic Hapludults	Inceptisols, Alfisols, Ultisols

Soils type	Sandy and well-drained	Well-drained and acidic	Well-drained and acidic
Dominant land cover/vegetation	Pine, savanna, and forest	Oak-hickory, mixed hardwood, and pine forest	Oak-hickory forest, mixed deciduous and evergreen
Dominant understory vegetation	Wiregrass, saw palmetto	Dogwood, greenbrier	Rhododendron, ferns
Major water bodies	Newnans Lake, Orange Lake, Prairie Creek	Cheaha Creek, Talladega Creek, Coosa River	Melton Hill Lake, Clinch River

Source: USFS (2021) and NEON (2022b).

2.2.3 Vegetation compositions and site-specific historical context

The study sites significantly differ in their vegetation compositions in both over and understory trees. Major trees of the OSBS site are characterized by longleaf pine (*Pinus palustris*), slash pine (*Pinus elliottii*), loblolly pine (*Pinus taeda*), live oak (*Quercus virginiana*), water oak (*Quercus nigra*), laurel oak (*Quercus laurifolia*), whereas the TALL site has white oak (Quercus alba), northern red oak (*Quercus rubra*), black oak (*Quercus velutina*), hickory (*Carya spp.*) and four types of pine trees. The ORNL site has oak-hickory forests mixed with deciduous and evergreen trees. The understory vegetation at each site is also unique. The OSBS site has saw palmetto (*Serenoa repens*), wiregrass (*Aristida spp.*), goldenrod (*Solidago spp.*), blackberry (*Rubus spp.*), gallberry (*Ilex glabra*), dogwood (*Cornus spp.*), redbud (*Cercis spp.*), greenbrier (*Smilax spp.*), honeysuckle (*Lonicera spp.*), wild grape (*Vitis spp.*) shrubs and the ORNL site has rhododendron (*Rhododendron spp.*), ferns (*Various species*), wildflowers (*Various species*), raspberry (*Rubus spp.*) (Table 2-1). Each site has undergone land use change in the past, with OSBS used for agricultural land and cattle ranch, TALL being used for timber harvesting and mining purposes, and ORNL being used for uranium enrichment, agriculture, and logging. These sites also support diverse wildlife depending on the specific vegetation characteristics of each site.

2.2.4 Datasets

The availability of accurate and comprehensive data is crucial in ecological research, including the estimation of AGB. Airborne lidar datasets have proven to be valuable sources of

information for AGB modeling due to their ability to provide detailed three-dimensional representations of vegetation structure. However, before these airborne lidar datasets can be effectively used for AGB modeling, systematic pre-processing of the data is necessary to ensure the quality and suitability of the training dataset. AGB estimated from field measurements was used as the dependent variable for this study. Each tree's AGB was calculated using allometric equations (Chojnacky, Heath, and Jenkins 2014; Jenkins et al. 2003). The calculated AGB was used as the dependent variable while variables extracted from airborne lidar, elevation and satellite imagery were used as dependent variables. Datasets associated with processing included a) field-collected tree attributes (VST), b) airborne lidar extracted variables and c) ancillary variables from satellite imagery and related other data sources.

2.2.4.1 Vegetation structural data

Tree attributes are extracted from NEON's data portal namely VST data with ID: DP1.10098.001. VST sampling plots are square in nature with two dimensions 40 m x 40 m and 20 m x 20 m, named 'tower' and 'distributed' plots, respectively. While the tower plot's tree attribute collects and updates data by a high-resolution camera installed on the tower, the distributed plots' tree attributes are collected by field visits. The data collection frequency of tower plot data (clusters plots in Figure 2-1) is 2 years while distributed plot data are collected every five years (NEON 2022b). During the Covid-19 period, data collection was resumed in some plots, so available data varies, but this study utilized the most updated VST dataset available. NEON conducts surveys to measure the structure of woody and non-herbaceous perennial plants and maps their positions based on plots and subplots, including nested plots. The resulting quality-controlled data provides information on live and standing dead woody trees, shrub groups, and non-herbaceous perennial plants at every terrestrial NEON site. NEON's VST data collection protocol is the same across each site.

2.2.4.2 Airborne lidar point cloud data

The classified airborne lidar point cloud data used in this study was collected from NEON's data portal (https://www.neonscience.org/data). NEON's airborne platform (AOP) consists of hyperspectral, multispectral, and lidar instruments mounted on a light aircraft to collect data

products, including VHR imagery, at the meter scale. The lidar instrument used by the AOP is the Optech Gemini system with a wavelength of 1064 nm and a laser footprint of 0.25 m in diameter (0.8 m in wide beam divergence mode) alongside an elevation accuracy of <5-35 cm (1σ). The AOP is flown at an altitude of 1000 m above ground level and at a speed of 100 knots to obtain a mosaic of data products. The NEON lidar point clouds for this study were acquired in May 2021 for TALL and ORNL and May 2018 for OSBS, during a period when NEON terrestrial sites reached at least 90% of peak vegetation greenness. The point cloud data protocols also changed over the years, and OSBS has a significantly higher point density. Overall point density of the lidar data is 1-5 points per square meter with up to 5 returns per footprint (NEON, 2022a). These data specifications make them suitable for use in this research.

2.2.4.3 Satellite imagery and ancillary datasets

Airborne lidar provides a 3-dimensional representation of vegetation structures. Along with these, 2-dimensional vegetation characteristics including vegetation cover, land cover classification, and topographic features are found useful for modeling AGB. Sentinel-2 satellite imagery, the National Land Cover Dataset (NLCD) of 2019, and the National Tree Cover Dataset of 2021 were considered. The image resolution of Sentinel was the same size as the field plots. For vegetation indices generation purposes, Sentinel images in May 2018 for the OSBS, May 2021 for ORNL, and June 2021 for TALL were used, the same months the lidar point cloud data were collected. Cloud coverage for all images was less than 3%.

The National Land Cover Database (NLCD) is a comprehensive geospatial dataset that categorizes and maps land cover types across the US, providing valuable insights into the country's terrestrial landscape. In addition, the Tree Canopy Cover dataset offers specific information about the extent of tree cover, aiding in the assessment of urban and rural green spaces and their ecological contributions. The NLCD products used in this study were provided by the Multi-resolution Land Characteristics Consortium in partnership with the U.S. Geological Survey (USGS), and several federal agencies. The NLCD- 2019 land cover and national tree canopy cover dataset of 2021 (CONUS) was collected in the geographical coordinate system (Wickham et al. 2021).

2.2.5 Data processing

The data processing step combines the activity of cleaning VST data that is required for calculating each tree's AGB. The data were reprocessed from two different sized field plots to a unique size and each plot's tree AGB was summed to derive plot-level AGB. All other variables that were processed from airborne lidar, satellite imagery, and national land and tree cover datasets were used as independent/predictive variables for AGB estimation (Figure 2-2). The methods for processing are intended to support working with NEON datasets to create mapped AGB products with NEON data and contribute to broader-scale high-resolution AGB mapping.

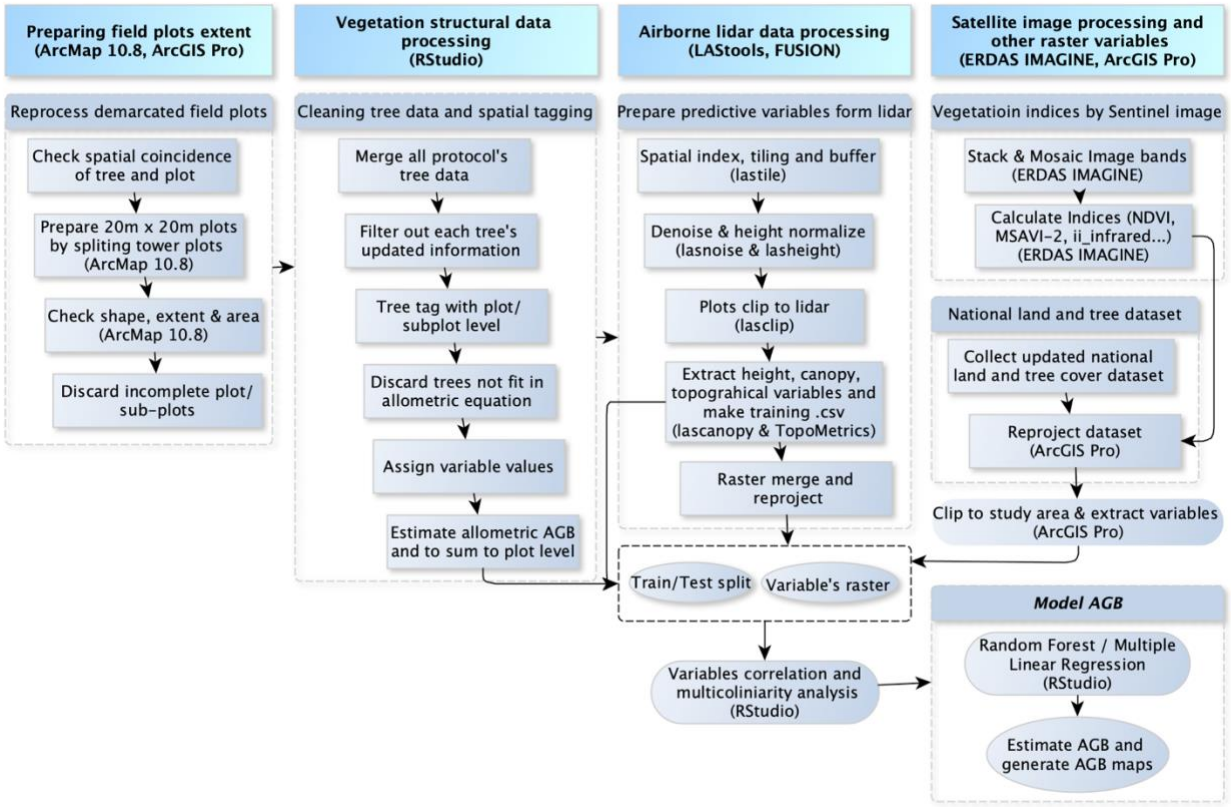

Figure 2-2: Data science workflow for modeling spatially continuous AGB estimation using NEON vegetation data.

2.2.5.1 Cleaning VST dataset for calculating allometric AGB

To prepare VST data for use in this study, NEON suggested “neonUtilities” which is extensively used under the R programming language. It allows downloading all tree attribute data of each site directly from the server by specifying the site name, such as “OSBS”, “TALL” or “ORNL”. The collected tree attributes were stored by NEON in ten different tables, including the “vst_mapping and tagging” table which contains spatial information and the “vst_apparentindividual” table which contains information about each tree's DBH, height, and current condition (Figure 2-3).

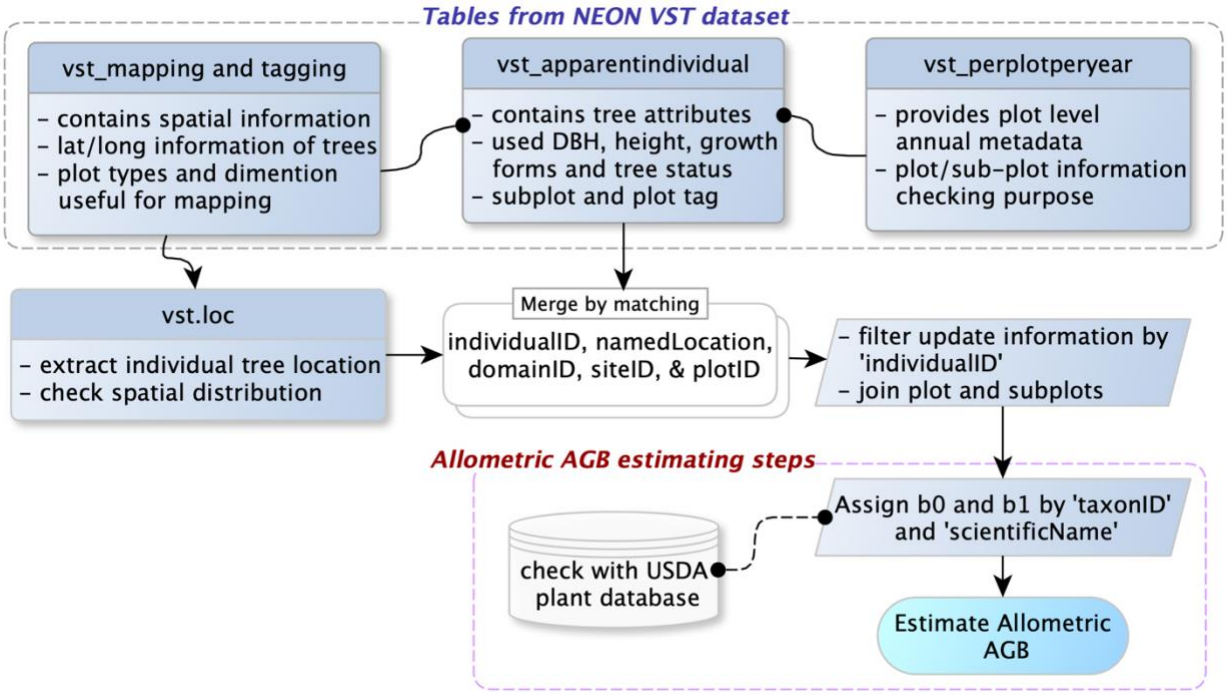

Figure 2-3: Database designing tables, relations, and process flow for cleaning VST data

To merge the required tables, the tree dataset from the previous protocol is matched with the current one based on specific attributes, such as "individualID," "namedLocation," "domainID," "siteID," and "plotID," where “individualID” is a unique ID of each focus table/tree. NEON updates the tower plot’s tree attributes every two years and distributes the plot’s trees every five years. Thus, the merged table may potentially contain each tree's information several times (i.e., found 3 times or more per tree). To filter out each tree's updated information, a loop is used over the data to keep the most updated one from the segment. The "taxonID” of these updated tree

datasets was then used for joining parametric values grouped from the Chojnacky literature (Chojnacky et al. 2014; Nelson et al. 2014).

2.2.5.2 Reprocess VST field plots and check subplot tag

Tree-based and *area-based* are two primary approaches used for modeling AGB of a forest site and this study used area-based approaches. In tree-based approaches require each tree's precise geographic location, and area-based approaches require each plot's centroid geographic location, and all plots should be the same in size (like 20m). NEON divides its 40m x 40m dimension tower plots into four equal subplots namely 21, 23, 39, and 41, each with 20m resolution, meaning each area is 400 sqm. Among these four plots, two randomly selected subplots of a tower plot are used for data collection purposes by NEON (NEON, 2022a). Thus, considering the tower plot extent of NEON, training data generation purposes can lead to significant bias in lidar-based AGB modeling. Therefore, it is necessary to split tower plots and carefully check each tree with subplot tagging (i.e., 21/23/39/41) to ensure spatial coincidence. Each tower plot was split into four equal parts using ArcMap 10.8 where each 400 m^2 polygon is allowed to have 398 ± 2 m (~19.96 m on each side) (Figure 2-1). Despite NEON ensuring high data accuracy, there were only around 20 distributed plots per site, which is less than 10% area of each study site, so considering only distributed plots may not be adequate for AGB model training. In this case, despite sacrificing the sampling distribution, the split tower plots may be needed to train the model (Figure 2-1). As such, all subplots with trees were considered.

Like the tower plot, NEON divides distributed plots into four equal subplots each with 10m x 10m dimensions. Those plots do not potentially need to split but checking trees into subplot scale helps to identify whether each plot's tree information was complete. It was observed that very few distributed plots' trees miss subplot tags, or a subplot has no trees. In this scenario, zooming into the visual extent using the time series feature of Google Earth Pro helped identify the present and past conditions of the subplot location. If the subplot location is originally barren, then the distributed plot is considered complete. If visual observation refers to trees but is not available in NEON's VST data, the plot is marked as incomplete and discarded, as done by previous studies (Duncanson et al. 2020).

2.2.5.3 Calculate plot-level allometric AGB

The AGB of a tree refers to the total dry weight of the living vegetation and its associated products (e.g., leaves, branches, and fruits) that are above ground within a particular area. Each tree AGB is often estimated using statistical equations called allometric equations when tree-specific attributes like DBH, height, and species are available. This study uses DBH-based allometric equations developed and modified by Jenkins and Chojnacky to estimate AGB (Chojnacky, Heath, and Jenkins 2014; Jenkins et al. 2003) (Table 2-2).

It is also worth noting that NEON records all vegetation of a plot including trees, shrubs, sub-trees, sub-trees, and hanging plants (i.e., liana). As such, precise variable selection is a challenge with the available equation, which causes for overestimation of results (Equation-2-Table 2-2). Therefore, single-bole live trees with more than 2.5cm DBH are further considered for tree-based allometric AGB estimation.

Table 2-2: NEON's VST data brief and used allometric equation for estimating AGB.

Name	**Specifications and equations**
Tree attribute used from NEON	Diameter at breast height (DBH), height, status of the tree
Vegetation coverage	All trees and shrubs, liana, fern
Plot type and size	Tower plots (1600 sq. m) with 4 subplots 21, 23, 39 and 41 (randomly sample any two subplot per tower plot) Distributed plots (400 sq. m) with 4 subplots 31, 32, 40, and 41 (usually sampled all subplots per plot)
Data updating frequency	2 years for tower plots; 5 years for distributed plots
Total sampling plots	OSBS – 56, TALL – 57 and ORNL - 55
Allometric equation used	***Eq.1 - AGB (kg) = exp (b_0 +b_1 ln dbh);*** Exp = exponential function b_0, *and* b_1 are statistical parameters; In = log base e (2.718) and DBH = diameter at breast height (cm) (Jenkins et al. 2003) ***Eq. 2 - Y = a X D +b;*** Y = AGB in kg, D is DBH, and a and b are parametric values (Brantley et al. 2016)

After estimating AGB in two datasets, those are summed to the plot level, which all are in the same 20m x 20m size (described in section 2.3.2). Plot-level AGB was used as the dependent variable in the model training dataset. To generate an AGB map over the full study area, AGB

needs to be estimated for each pixel, where tree attributes are not recorded but airborne lidar point cloud, satellite imagery, or other national scale datasets are available.

2.2.5.4 Lidar data processing and comparison with ground measurements

The obtained data contained classified and unclassified airborne lidar point clouds, metadata, shapefiles, and KML files. The classified point clouds were in the compressed lidar file format (.laz) and were converted to the uncompressed lidar file format (.las) for processing using LAStools. Points classified as noise were removed using 'lasnoise' and tree heights were normalized using classified ground points with 'lasheight'. The 'lascanopy' tool was then used to estimate canopy height metrics, including cover, density, and height percentiles (Isenburg, 2013). Topographic variables, including elevation, aspect, slope, and hillshade were estimated using 'CloudMetrics', 'GridMetrics', 'TopoMetrics' in USDA-developed Fusion software (McGaughey, 2016). Since tree height impacts model accuracy, the lidar-estimated canopy heights and field-measured tree heights were also compared and presented for OSBS, TALL, and ORNL (Figure 2-4). To compare each plot's ground observed tree height and lidar height, the 'max' lidar height variable was used. As minimum height was zero for most of the plots in both datasets and the mean height of lidar differed from ground observations due to the limited number of lidar pulse penetrations through the top canopy, max height comparison was useful. The R^2 for TALL was 0.84 and RMSE of 2m, ORNL had an R^2 of 0.78 and RMSE of 2.47m, and OSBS had an R^2 of 0.53 and RMSE of 3.43m.

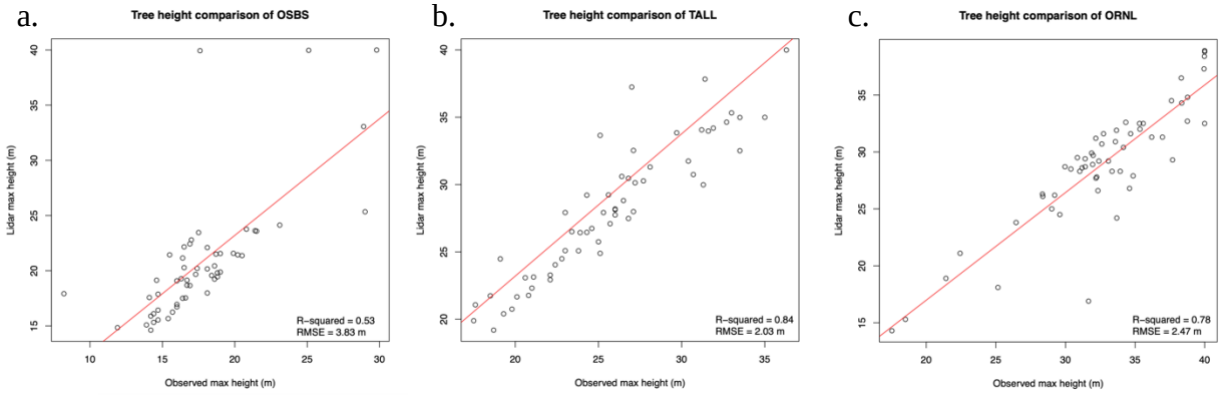

Figure 2-4: Tree height comparison for the three study sites with lidar-derived maximum height and field observed max height at the plot level for OSBS(a), TALL(b) and ORNL (c).

2.2.5.5 Processing satellite imagery and ancillary datasets

The Sentinel imagery was available at the same spatial resolution as our study field plots (20m), so no resampling was required to match lidar variables and spatial coinciding plots. Like the Sentinel satellite imagery, NEON's airborne lidar point clouds for TALL and ORNL sites were in projected coordinate system's (PCS) UTM zone 16N and the OSBS site in UTM zone 17N. The collected satellite imagery's cloud cover was less than 10% which was suitable for indices calculation. Bands 2-8 were stacked to make multi-spectral composite imagery and clipped into the study extent. The composite image was used to calculate the vegetation indices and were processed using ERDAS IMAGINE 2014 (Table 2-3). After getting the vegetation indices maps, 'Extract Multi Values to Points' features were used to extract pixel values using ArcMap 10.8.

Table 2-3: Vegetation indices used in the study

Abbreviation	Full name (common use)	Calculating equation
ndvi	Normalized Difference Vegetation Index (measure of vegetation health)	NDVI = (NIR-RED) / (NIR+RED).
msavi2	Modified Soil-Adjusted Vegetation Index 2 (a measure of vegetation greenness)	MSAVI2 = (2 * NIR + 1 - sqrt ((2 * NIR + 1) ^2 - 8 * (NIR - Red))) / 2
ii_infrared	Infrared Index;	ii_infrared = NIR - Red
mcari2	Modified Chlorophyll Absorption in Reflectance Index 2;	mcari2 = (1.5 * (2.5 * (NIR - Red) - 1.3 * (NIR - Green))) / sqrt((2 * NIR + 1)^2 - (6 * NIR - 5 * sqrt(Red) - 0.5))
msr705	Modified Simple Ratio at 705 nm:	msr705 = NIR / Red
mtvi2	Modified Triangular Vegetation Index 2;	mtvi2 = 1.5 * (1.2 * (NIR - Green) - 2.5 * (Red - Green)) / sqrt ((2 * NIR + 1)^2 - (6 * NIR - 5 * sqrt(Red) - 0.5))
ndii	Normalized Difference Infrared Index;	ndii = (NIR - SWIR) / (NIR + SWIR)

Ancillary datasets included NLCD landcover and tree cover which are produced at a 30 m resolution and were then resampled from 30 m pixels to 20m raster. The 'nearest neighbor'

method was used for land cover datasets and the bilinear interpolation method was used for tree cover datasets. The 'Extract Multi Values to Points' tool was again used to extract pixel values.

2.2.5.6 Predictive variables generation for modeling AGB

The lidar point cloud-derived variables (also called lidar metrics) used as independent variables in the AGB models were height percentiles, coverage, density metrics, and topographic metrics (Table 2-4). Variables from satellite imagery included vegetation indices and ancillary variables included national land and tree datasets. Variables were input in the RF model in two forms, as comma-separated values (.csv) for the training set and each variable's gridded raster format to generate a spatially continuous AGB map over each site.

Table 2-4: Lidar-derived, satellite imagery-derived, and NLCD variables used for AGB modeling

Dependent Variable	**Independent Variables**		**Descriptions**
Plot-based Above ground Biomass (AGB) in Megagram/hectare (Mg/ha) Calculated from VST data	**Canopy Cover** (lidar derived)	cov_2m	Percentage of all returns above 2 meters
		cov_4_6	Percentage of all returns above 4.6 meter
	Density metrics (lidar derived	d00	The proportion of returns in the stratum 0-5 m
		d01	The proportion of returns in the stratum 5-10 m
		d02	The proportion of returns in the stratum 10-15 m
		d03	The proportion of returns in the stratum 15-20 m
		d04	The proportion of returns in the stratum 20-25 m
		d05	The proportion of returns in the stratum above 25 m
	Height percentiles (lidar-derived)	max	Maximum height
		avg	Average height (mean height)
		p05	5th percentile height
		p10	10th percentile height
		p25	25th percentile height
		p50	50th percentile height (median)
		p75	75th percentile height
		p90	90th percentile height
		p95	95th percentile height
		p99	99th percentile height
		elev	Elevation of the terrain

	Topographic metrics (lidar derived)	aspect	Aspect of the terrain (direction the slope faces)
		slope	Slope of the terrain
		hillshade	Hill shade of the terrain
		TPI	Topographic Position Index. *TPI = elev - mean(elev_neighborhood)*
		SRI	Solar Radiation Index based on Keating et al. (2007) *SRI = 1.0+cos(latitude)*cos(slope)+sin(latitude)* sin(slope)*cos(aspect)*
		ProCur	Profile curvature (curvature along the slope)
		PlnCur	Plantation curvature (curvature across the slope)
		OvrCur	Longitudinal curvature (curvature in the direction of slope)
	Vegetation indices (satellite imagery based)	ndvi	Normalized Difference Vegetation Index
		msavi2	Modified Soil-Adjusted Vegetation Index 2
		ii_infrared	Infrared Index;
		mcari2	Modified Chlorophyll Absorption in Reflectance Index 2
		msr705	Modified Simple Ratio at 705 nm
		mtvi2	Modified Triangular Vegetation Index 2;
		ndii	Normalized Difference Infrared Index;
	NLCD	nlcd_tc	National Tree Cover Database
		nlcd_lc	National Land Cover Database

2.2.6 Modeling Aboveground biomass

This study considered two modeling approaches for modeling AGB, namely MLR and RF. To model AGB with RF, this study used the ModelMap package under the R environment which is widely used due to user-friendly modeling, validation, and mapping over large geographic areas (Brown, Narine, and Gilbert 2022; Narine, Popescu, and Malambo 2020). It also enables a graphical user interface under an R environment and is capable of modeling with both continuous or discrete responses using RF and gradient boosting approaches (Freeman et al., 2016). As discussed, each field site has significant differences in tree composition, weather patterns, and land use, so separate models for each NEON site were developed for estimating AGB instead of building a single model with all datasets. Separate models helped to assess different sites' model performances and find each site's contributing variables. Selection of the significant variables was also important, without it, the model might overfit or underfit the model, leading to overestimation or underestimation of AGB. Variable selection was carried out using linear regression through several iterations to ensure higher R^2 and lower RMSE and %

RMSE. To address multicollinearity, the variance inflation factor (VIF) was assessed through linear regression to ensure VIFs less than 5 (Thompson et al., 2017). After assessing multicollinearity, variables with values <5 were used for MLR-based AGB estimation.

For RF modeling, the initial significant set of variables was also further processed and input for RF model training and subsequent map generation. To select the final set of variables, iterative check correlation and multicollinearity analysis were carried out to select the best independent variables. In the ModelMap package, figures for identifying negative and positive correlations among variables were generated which helped discard strongly correlated variables. After that, variables were tested for VIF values where a VIF less than 5 means no collinearity. After discarding correlated variables and keeping the most significant variables, the final variables were selected for modeling AGB using the RF algorithm.

2.2.6.1 Distribution of data in modeling approaches

To examine AGB data density, ranges, and overall distribution, violin box plots with associated error bars are presented in Figure 2-5. A violin plot combines aspects of a box plot and a kernel density plot and is used to display the distribution and summary statistics of a dataset, providing insights into the central tendency, spread, and shape of the data. As seen in Figure 2-5, VST represents observed AGB calculated from VST data by allometric equations and lidar denotes modeled AGB from lidar metrics and other spatial variables. Reference data distribution varied site-by-site, like ORNL, followed a normal distribution, OSBS has several skewed data distributions and TALL is moderately skewed. Among CV and OOB validation approaches, CV indicated slightly better results than OOB. As seen in Figure 2-5, ORNL has higher ranges of AGB, since it has significantly taller trees. The TALL site also has a mixture of taller and smaller trees and OSBS shows significantly smaller tree sites. Considering the train-test split, the maximum AGB estimation for OSBS was observed to be higher than the predicted values, while the minimum AGB estimation for the same site was lower than the predicted values. TALL had the highest maximum AGB estimation for observed values and predicted values, while its minimum AGB estimation was higher in the predicted values of the test set shown in Figure 2-5.

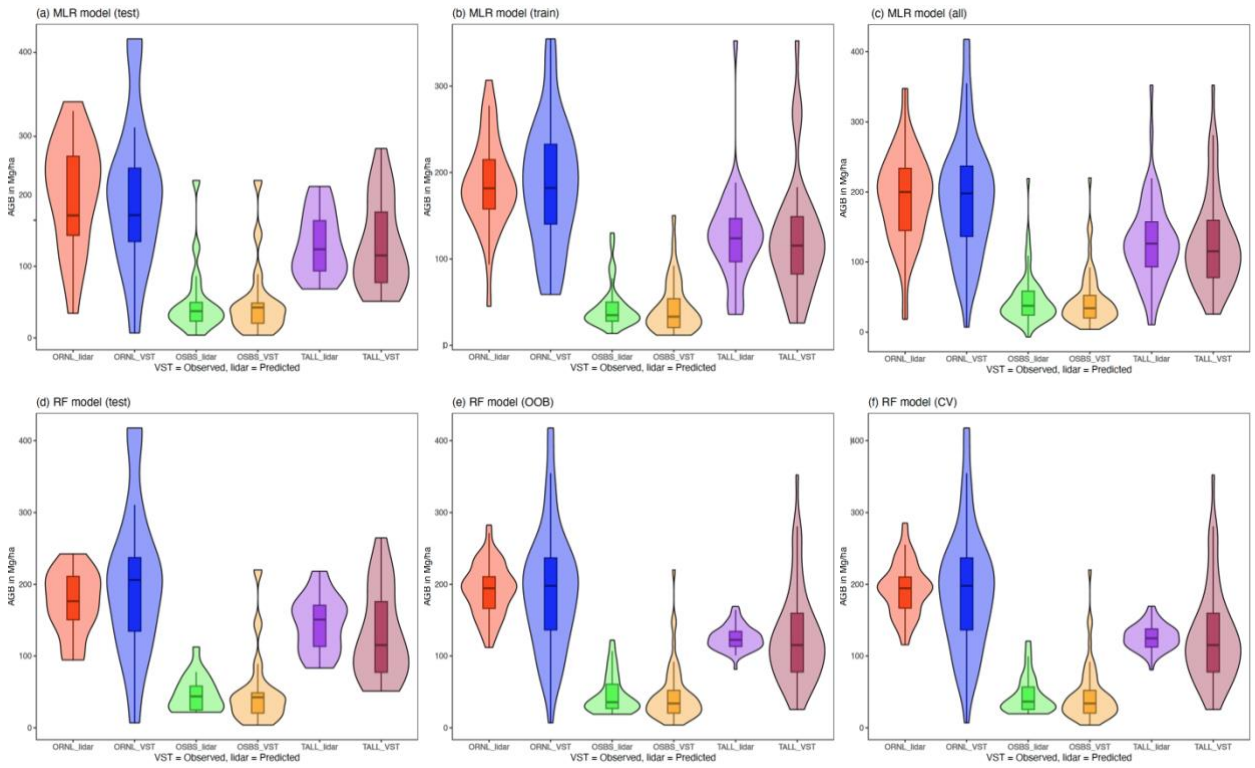

Figure 2-5: Field plots and modeled AGB distribution, kernel density of data, ranges, and model performances in different validation approaches. Figures a-c represent MLR model performances for (a) separate test set, (b) separate train set, (c) full dataset and modeled outcomes. Figures d-f represent RF model performances for (d) separate test set, (e) out-of-bag error validation method with full set and (f) 10-fold cross-validation for full dataset.

2.2.6.2 Improving model performances

To improve model performance, additional variables were tested and included topographic position index (TPI), Profile curvature (ProCur), Plan curvature (PlnCur), Overall curvature (OvCur), and Solar radiation index (SRI) (Keating et al. 2007; McGaughey 2016; Roy and Das 2021; Weiss 2001), were used. The TPI compares the elevation of each cell on a surface to the mean elevation of a specified neighborhood around the cell, and the ProCur combines along the slope while PlnCur related perpendicular to slope relations. The SRI was a single index combining the effects of aspect, slope, and latitude and described the amount of solar radiation theoretically striking an arbitrarily oriented surface during the hour surrounding noon on the equinox (Keating et al. 2007; McGaughey 2016).

2.2.6.3 Model validation

The validation of models is also an essential step to measure model performance and to improve model accuracy. This study used an independent test set, cross-validation, and Out OF Bag (OOB) methods to evaluate model performance. As discussed in Table 2-2, there was limited field data for training models, so Cross-Validation (CV) and Out OF Bag (OOB) methods were tested by full sets of training data using the RF model. For the separate test set, one-third of the dataset was set aside for model validation. Model performances were reported using the coefficient of determination (R^2), root mean square error (RMSE), and %RMSE that are presented in scatter plots. The ranges of VST-estimated AGB, RF, and MLR-estimated AGB were also presented in data density with an error box plot graph (Figure 2-5) and the final spatially continuous 20m resolution AGB maps generated for each NEON site and presented in a subsequent section.

2.3 Results

2.3.1 Model performances

2.3.1.1 Multiple linear regression (MLR) model

The models performed differently in each site in both training and test samples. The R^2 values were 0.91, 0.52, and 0.63 for OSBS, TALL, and ORNL respectively with the test dataset (Figure 2-6: d, e, f). These values were lower for the training dataset, which was 0.69, 0.62, and 0.46 respectively, with an opposite trend following the %RMSE for each site in the MLR models' training dataset (Figure 2-6: a, b, c). Although test datasets usually exhibit lower accuracies than training datasets, the separate test set of OSBS shows a significantly higher R^2. In these cases, another modeling approach i.e., RF was tested which revealed some complex relationships among the field observed dataset.

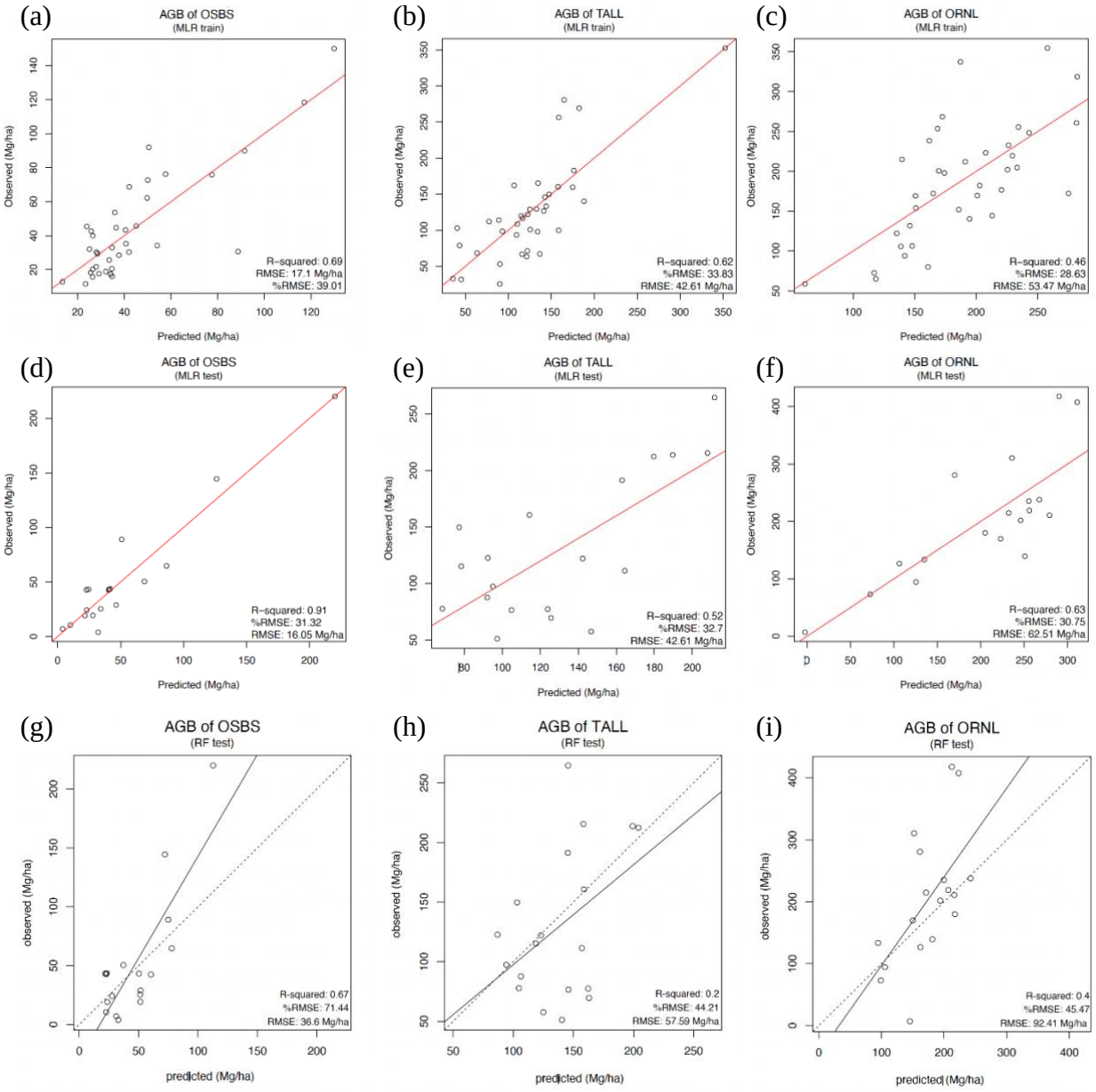

Figure 2-6: MLR and RF-based model performances over three study sites where a-c and d-f showed train and test sets R2, RMSE and %RMSE value using MLR and g-i RF model performances. The unit of RMSE values was in Mg/ha.

A slightly higher %RMSE was found for the TALL sites than OSBS and ORNL in the individual train and test datasets. Considering the R^2 and RMSE values, the model struggled with taller tree sites due to higher topographic differences in mixed forests. The SRI and PlnCur variables contributed to understanding the complex nature of the ORNL and TALL sites and improving

model performances. The MLR equations presented in Eq. 3, 4 and 5 for the OSBS, TALL, and ORNL, respectively.

Eq. 3 - AGB (Mg/ha) = 40.26+0.24*cov_2m+10.48*p25+2.57*p90+0.2*d04+2.33*d05+-1.05*elev
Eq. 4 - AGB (Mg/ha) = -437.71+509.05*p05-1.89*d00-2.81*d02+1.03*elev+163.47*SolRI+2.43*nlcd_tc
Eq. 5 - AGB (Mg/ha) = 379.16-3.2*cov_2m+1.73*p05+1.43*d03+2.72*d05+1.65*ProCur-0.78*PlnCur

2.3.1.2 Random forest (RF) model

Although RF is a demonstrated approach for modeling AGB with remotely sensed data (Freeman, Frescino, and Moisen 2016), this approach provided lower accuracy in previous studies over mixed forest sites in the southern United States, likely due to the heterogeneity of tree structures (Brown et al. 2022; Freeman, Frescino, and Moisen 2016). This statement also holds true for NEON datasets; the RF model performed poorly for all three sites in comparison to the MLR model's individual test set. The RF model's individual test set R^2-the in test-train-based modeling approaches. Among CV and OOB, the CV validation approach provided slightly better results than OOB. The R^2 ranges for CV validation were 0.62, 0.17, and 0.17, and for the OOB it was 0.55, 0.14, and 0.15 for OSBS, TALL, and ORNL, respectively. The RMSE and %RMSE values for the CV were 24.91 Mg/ha (54%), 61.92 Mg/ha (49%), and 77.63 Mg/ha (40%), respectively. The OOB validation based RMSE and %RMSE for these three sites were 26.76 Mg/ha (57%), 62.75 Mg/ha (49%), and 78.61 Mg/ha (41%), respectively. The %RMSE was also higher in the RF model for the three sites when compared to the MLR model.

2.3.2 Predictive variables performances over study sites

The importance of predictive variables in the RF model varied for each site based on their characteristics (section 3.1.2). Tree height percentile variables were found important in modeling AGB of OSBS in where most trees were smaller in height. For taller tree sites, density metrics, topographic metrics, and tree structures, namely plantation curvature and profile curvature, were found useful in modeling AGB (i.e., ORNL and TALL sites), consistent with previous studies (Brantley et al. 2016; Donager et al. 2021; Zhou et al. 2018). The 75% tree height (p75) and site elevation were the most significant variables in the OSBS site where tree density in the height range of 20-25m was found less impactful in the dataset presented in Figure 2-7 (left). In the

TALL site, 10-15m and 0-5m taller tree density carry valuable information in estimating AGB while solar radiation index (SRI) helps to predict the model but is lower in impact, presented in Figure 2-7 (mid). Similar to the TALL site, the density of trees above 25m in height was the most important variable for estimating the AGB of ORNL. Tree texture such as plantation curvature (PlnCur) and profile curvature (PlnCur) also contributed significantly where tree cover above 2m height less contributed to model estimation (Figure 2-7, right). The influence is estimated based on the p-values; lower p-values (typically below 0.05) suggest a stronger influence on prediction power in the MLR of variables.

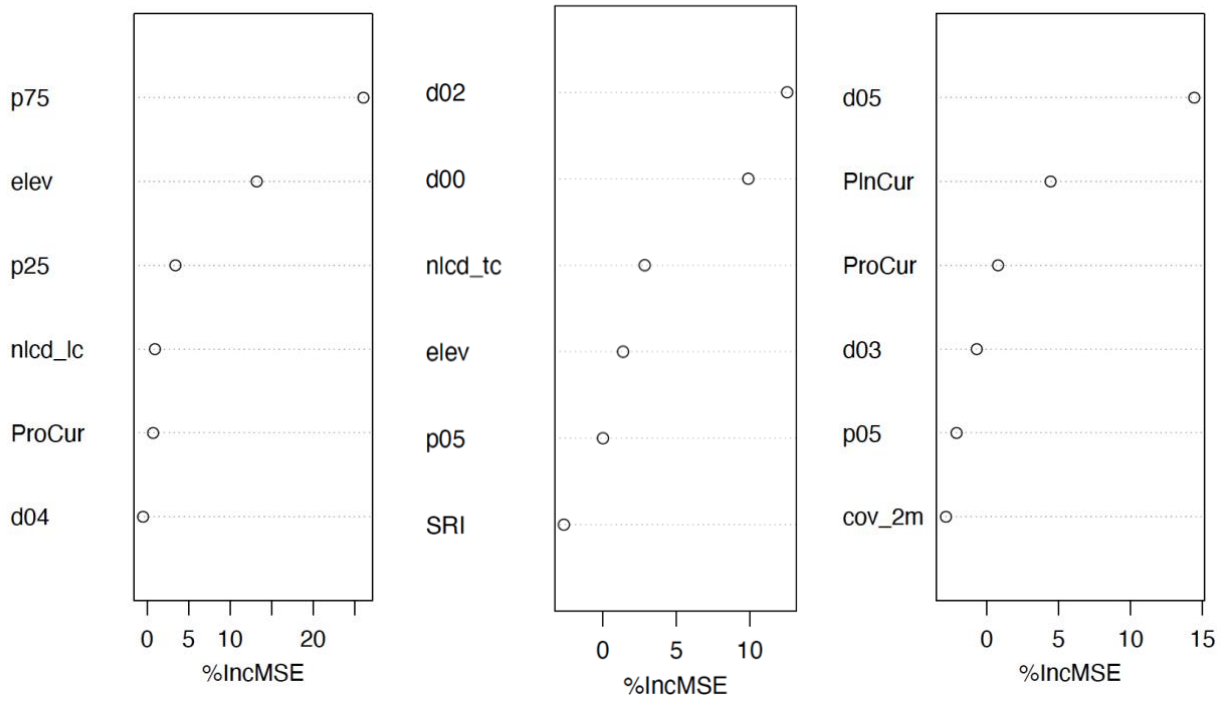

Figure 2-7: Predictive variables' importance in plotting OSBS, TALL, and ORNL (left to right) on y-axis. The %IncMSE in the x-axis refers to the percentage increase in Mean Squared Error with each predictor variable.

2.3.3 Mapped AGB

Vegetation structure, weather patterns, topographical features, and vegetation characteristics varied in the three mixed forestry sites (Figure 2-8). The dark green represents the highest amount

of AGB in the range of 238.9-295.4 Mg/ha, whereas the lighter green represents the lowest amount of AGB in the range of 12.5 - 69.1 Mg/ha.

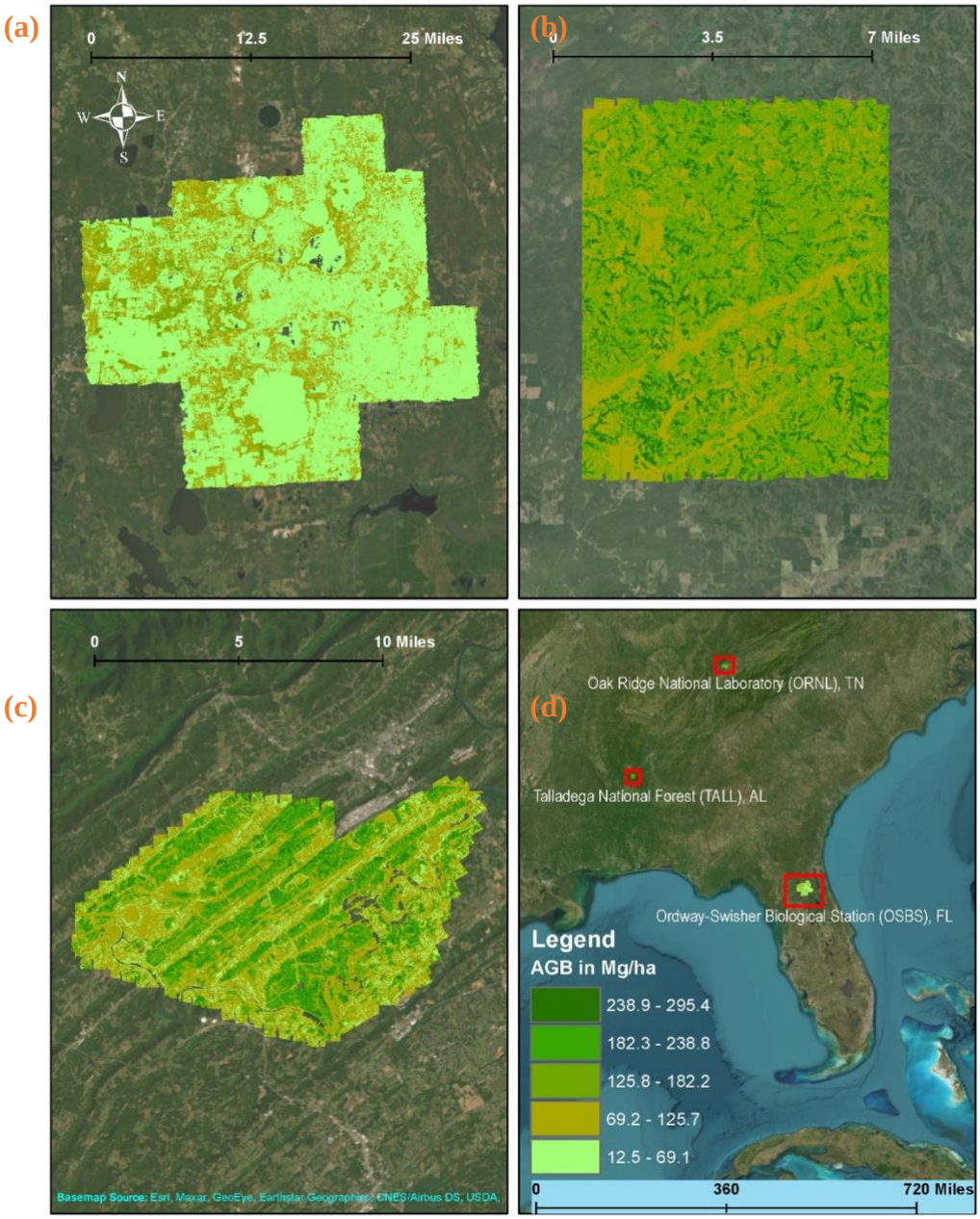

Figure 2-8: The estimated above-ground biomass (AGB) map of the three sites. The ORNL (c) has taller trees with higher ranges of AGB. TALL (b) is comparatively lower, and OSBS (a) has the lowest AGB ranges. Inset shows three study site boundaries on the Bing mapping database.

As expected, the small tree site (OSBS) had the lowest aboveground biomass (AGB) range, whereas ORNL had the highest range due to being taller. The AGB ranges were found to vary for the observed and predicted AGB of each site in the train-test set as presented in the density plotting (Figure 2-5).

2.4 Discussion

This study generated fine-scale AGB maps for mixed forestry sites using NEON-provided airborne lidar and ground measurements. Three southern forest sites with varying field characteristics and weather patterns were selected and two modeling approaches: Multiple linear regression (MLR) and Random Forest (RF) were examined for estimating AGB. To prepare datasets, unique data cleaning approaches were suggested to utilize the best available ground measurements and discard incomplete data for modeling AGB with NEON's VST dataset. The study showed stable performances among MLR and RF over three sites with higher accuracies and lower root mean square errors (RMSEs) for the MLR model compared to the RF model. This finding is comparable with those reported for similar heterogeneous forests of the southern United States (S. Brown et al., 2022). Based on the literature, different modeling approaches also perform differently in estimating AGB, which was found true in this study (Q. Zhao et al., 2019). In MLR model validation with a 1:3 test: train split shows overall higher accuracies that may indicate that underfit data means an oversimplicity tendency. This oversimplicity tendency refers to MLR's failed to assess the complex relationship among field vegetation characteristics of OSBS. This might be because, of the limited number of field samples and the skewed data distribution in some sites as shown in Figure 2-5, where linear regression assumes the dataset is normally distributed. In this regard, the RF model was utilized. Based on the performance of RF presented in Figure 2-6 (g-i), it might have been an indication of RF's ability to capture some complex relationships among variables for the taller tree sites, namely TALL and ORNL, than the MLR model. Alternatively, RF may have also needed more training samples to learn, model, and predict AGB (Freeman, Frescino, and Moisen 2016). This model learning approach often depends on selected variable sets and previous similar studies include satellite imagery-based indices and ancillary datasets for improving model performances (Wang et al. 2022, Brown et al. 2022). These variables are particularly useful to understand forest structure and model AGB in

steep terrain sites with taller trees. So, the identification of significant factors influencing AGB estimation including tree height percentiles, land cover variables, tree density, and topographical variables were a significant outcome that would be useful for similar studies and for the use of NEON datasets.

Based on the study outcomes, all topographic variables were not significant; for instance, slope was assumed to be a significant variable, but it negatively contributed to the model. Some other variables like the solar radiation index (SRI), a combination of slope and aspect (Table 2-3), were found useful in the TALL site (Figure 2-5). Likewise, satellite imagery-based variables were also found useful specifically the NDVI, MSAVI-2 vegetation indices and NLCD's land and tree cover dataset helps to differentiate tree and vacant areas. The vegetation indices are important for differentiating forests from non-forested areas based on the greenness value and were found useful in modeling AGB (Hasmadi et al. 2010; Islam et al. 2020; Schmidt and Karnieli 2001). These freely available and readily available products help to improve model accuracy along with NEON VST and airborne lidar datasets (García et al. 2018; Khati et al. 2020). The consideration of further variables and sensitivity analysis by extensive field survey would be an option (Coops et al. 2021; Popescu and Wynne 2004) with increasing field sampling points, and quality-controlled datasets (Pau et al. 2022). Additionally, field characteristics and forest types (homogeneity, mixed forest type, and tree species composition) were significant parameters that need to be considered along with slope, canopy coverage, and heterogeneity of tree species which were also found useful in previous studies (Wang et al. 2022).

NEON datasets and studied field sites possess unique, unavoidable limitations in creating training samples for developing AGB maps, like heterogeneity of the study site, the significantly higher number of substory/understory trees, the limited number of field plots, and clustered field sampling plots. Tree species composition and terrain differences contribute to the heterogeneity of the TALL and ORNL sites. Lidar has limitations in penetrating a thick canopy layer and steep terrain site's attenuations contributed to lower accuracy in similar site characteristics (Wang et al. 2022) which might have an issue too in this study. Moreover, some of the field plots in each site have substory/understory trees (DBH < 2.5 cm), and multi-bole trees where allometric AGB was not possible based on currently available literature but were presented in lidar metrics.

Similarly, NEON collects only ~20 well-distributed tree plots per site which may not have been a sufficient sample size for modeling AGB. In this scenario, subplots that were prepared from the tower plot would be needed. However, these additional field plots increase each site's training sample, but tower plot extracted subplots are clustered since it samples were adjacent to a tower location. NEON considered one tower plot in each of our study sites, which means if the tree composition surrounding the tower plot does not represent all tree types of sample, and distributed plots also do not capture those, that it would lead to further bias in the model's understanding. Furthermore, NEON considers only a fixed dimension like 40m X 40m for tower plots and 20m X 20m for distributed plots in every forest site, which might capture tree composition and heterogenicity for a forest site (mostly in single forest tree sites) but might fail for mixed forestry sites. In this case, to capture tree composition, sampling dimension or initiate some additional field visits may need to be considered. This study's findings also suggested the collection of additional field data considering spatial distribution is recommended for potentially obtaining higher accuracies in modeling AGB. It is worth noting there were some potential limitations, but NEON observation provided useful information to estimate AGB as this study found on the OSBS site. The TALL and ORNL also estimated AGB successfully where model %RMSE was higher. These findings contribute to understanding the applicability of the NEON dataset for creating spatially continuous AGB model into different forest sites. Finally, this study only considered area-based approaches (ABA) instead of tree-based approach (TBA), which would be future avenues for checking the NEON data potential for AGB estimation using TBA. In this case, if each tree of the NEON consist of very accurate geographical location tagging would be possible to estimate but not tested on this study.

2.4.1 Applicability of study outcomes in non-NEON sites

This study used NEON's small-footprint airborne lidar observations and VST dataset for making spatially continuous high-resolution AGB modeling using MLR and RF approaches. This methodology of preparing a spatially continuous AGB map will be useful for mapping NEON sites AGB which is not yet available. Additionally, study outcomes will be useful for mapping non-NEON site areas AGB by combining with similar NASA products. For example, NASA developed ICESat-2 lidar's simulated data compared with NEON's airborne lidar dataset and

found it useful for retrieving plant structural traits (Purslow et al. 2023). In this evaluation study, NEON's data played a crucial role in the evaluation and found close similarity in plant attributes such as canopy height, canopy cover, and plant area volume density (Battelle, 2023). The previous study by Narine et al. 2019 demonstrated the methodology for estimating a spatially continuous AGB model using ICESat-2 data and discreet return lidar data (Narine et al. 2019). In this study, an airborne lidar point cloud was used as a reference dataset, and simulated ICESat-2 and associated ancillary products were used for upscaling data for getting a spatially continuous AGB map over an area. From the demonstrated methodology of this study, creating a reference dataset from NEON and upscale using ICESat-2 and ancillary matrices like indices from Sentinel imagery, land cover and tree cover dataset from NLCD dataset will be useful for broad scale mapping using area-based approach (ABA). In ABA, a series of grid-based metrics from airborne lidar (density, coverage, height percentile, topographic) were used at a grid size consistent with that of field plots i.e., 20m x 20m (Næsset, 2002). The spatial resolution was therefore the VST plot size of NEON which minimizes error due to incomplete capture of large trees or edge effects (Frazer et al., 2011, White et al., 2013). Another methodology demonstrated by Pascual et al. (2010) extrapolated summaries of airborne lidar height percentile metrics with Landsat 30-m pixels in a demonstration area in Spain. For using NEON's airborne lidar point cloud and VST data, the 20m X 20m scale is appropriate for AGB map generation (Coops et al. 2021, Margolis et al., 2015; Nelson et al., 2009; Boudreau et al., 2008). This grid size should be useful for developing new models between these predictions and spatially comprehensive satellite data (Sentinel) and ancillary information like terrain and climatic variables (Coops et al. 2021). This will allow for AGB over the broader landscape where NEON-collected VST and airborne lidar point clouds are not available.

Conclusion

NEON airborne lidar point cloud and VST data were found to be useful for high-resolution spatially continuous AGB maps of NEON sites. Considering the same data protocols for every NEON site, the method may be adapted for other sites. This study also found insights into the limitations of NEON data for creating high-accuracy AGB maps. Careful processing and modeling approaches can help get high-resolution AGB maps over sites that are not yet

available. Although, NEON collects many datasets in demarcated plots, which may reflect tree composition over homogenous sites, but not specifically for southern heterogeneous mixed forest sites where dense tree canopy cover significant terrain features. In order to improve AGB model accuracy increasing the distribution and number of field observations per site is recommended. Additionally, every model has its own limitations, such as the RF needing a large number of datasets to understand site heterogeneity and tree species. This study had 55 plots (train sample) that were usable for OSBS, ORNL site and 57 plots for the TALL site. The MLR model demonstrated higher accuracy and better performance than the RF model in all three study sites, as evidenced by higher R^2 values and lower root mean square error (RMSE). It is important to note for similar studies that the MLR model may struggle to capture the complex nature of the data, especially in sites with steep terrain and mixed forest types. In addition, the RF model tends to overcomplicate the estimation process, resulting in underfitting and lower accuracy. However, study findings evolve important insights for the NEON data producibility in high resolution spatially continuous AGB mapping. This research opens avenue for spatially continuous AGB mapping into NEON sites and based on this reference datasets, future research for fusing with ICESat-2 demonstrates way for broad scale AGB mapping into non-NEON sites.

References

Ajani, Judith I., Heather Keith, Margaret Blakers, Brendan G. Mackey, and Helen P. King. 2013. "Comprehensive Carbon Stock and Flow Accounting: A National Framework to Support Climate Change Mitigation Policy." *Ecological Economics* 89:61–72. doi: 10.1016/j.ecolecon.2013.01.010.

Battelle. (2023). *July 2023 What's New with NEON?* Inside Battelle. Retrieved August 15, 2023, from https://inside.battelle.org/blog-details/july-2023-what-s-new-with-neon

Brantley, Steven T., Morgan L. Schulte, Paul V. Bolstad, and Chelcy F. Miniat. 2016. "Equations for Estimating Biomass, Foliage Area, and Sapwood of Small Trees in the Southern Appalachians." *Forest Science* 62(4):414–21. doi: 10.5849/forsci.15-041.

Brown, J. H., Geoffrey B. West, and B. J. Enquist. 2005. "Yes, West, Brown and Enquist's Model of Allometric Scaling Is Both Mathematically Correct and Biologically Relevant." *Functional Ecology* 19(4):735–38. doi: 10.1111/j.1365-2435.2005.01022. x.

Brown, Schyler, Lana L. Narine, and John Gilbert. 2022. "Using Airborne Lidar, Multispectral Imagery, and Field Inventory Data to Estimate Basal Area, Volume, and Aboveground Biomass in Heterogeneous Mixed Species Forests: A Case Study in Southern Alabama." *Remote Sensing* 14(11):2708. doi: 10.3390/rs14112708.

Chojnacky, David C., Linda S. Heath, and Jennifer C. Jenkins. 2014. "Updated Generalized Biomass Equations for North American Tree Species." *Forestry* 87(1):129–51. doi: 10.1093/forestry/cpt053.

Coops, Nicholas C., Piotr Tompalski, Tristan R. H. Goodbody, Martin Queinnec, Joan E. Luther, Douglas K. Bolton, Joanne C. White, Michael A. Wulder, Oliver R. van Lier, and Txomin Hermosilla. 2021. "Modelling Lidar-Derived Estimates of Forest Attributes over Space and Time: A Review of Approaches and Future Trends." *Remote Sensing of Environment* 260(May):112477. doi: 10.1016/j.rse.2021.112477.

Donager, Jonathon J., Andrew J. Sánchez Meador, and Ryan C. Blackburn. 2021. "Adjudicating Perspectives on Forest Structure: How Do Airborne, Terrestrial, and Mobile Lidar-Derived Estimates Compare?" *Remote Sensing* 13(12):1–18. doi: 10.3390/rs13122297.

Duncanson, Laura, Amy Neuenschwander, Steven Hancock, Nathan Thomas, Temilola Fatoyinbo, Marc Simard, Carlos A. Silva, John Armston, Scott B. Luthcke, Michelle Hofton, James R. Kellner, and Ralph Dubayah. 2020. "Biomass Estimation from Simulated GEDI, ICESat-2 and NISAR across Environmental Gradients in Sonoma County, California." *Remote Sensing of Environment* 242(November 2019):111779. doi: 10.1016/j.rse.2020.111779.

Freeman, Elizabeth A., Tracey S. Frescino, and Gretchen G. Moisen. 2016. "ModelMap : An R Package for Model Creation and Map Production." 1–43.

García, Mariano, Sassan Saatchi, Susan Ustin, and Heiko Balzter. 2018. "Modelling Forest Canopy Height by Integrating Airborne LiDAR Samples with Satellite Radar and Multispectral Imagery." *International Journal of Applied Earth Observation and Geoinformation* 66(November 2017):159–73. doi: 10.1016/j.jag.2017.11.017.

Gassert, Francis, Sharon Burke, and Rachel Zimmerman. 2020. "UPTEMPO: The United States and Natural Disasters in the Pacitic." (March):1–107.

Glenn, Nancy F., Amy Neuenschwander, Lee A. Vierling, Lucas Spaete, Aihua Li, Douglas J. Shinneman, David S. Pilliod, Robert S. Arkle, and Susan K. McIlroy. 2016. "Landsat 8 and ICESat-2: Performance and Potential Synergies for Quantifying Dryland Ecosystem

Vegetation Cover and Biomass." *Remote Sensing of Environment* 185:233–42. doi: 10.1016/j.rse.2016.02.039.
Hasmadi, Mohd, Che Ku Akmar, U. Norsaliza, and I. &. Kasawani. 2010. "Comparison of Several Vegetation Indices for Mangrove Mapping Using Remotely Sensed Data." *International Geoinformatics Research and Development Journal* 1(3):2010.
Islam, Md Mozahidul, Ruhul Mohaiman Chowdhury, A. K. M. Mostafa Zaman, Md Saifur Rahman, Md Nazrul Islam, Ajit Kumar Rudra, and Md Salim Azad. 2020. "Spatiotemporal Mapping Mangroves of Tengragiri Wildlife Sanctuary under Barguna District of Bangladesh Using Freely Available Satellite Imagery." *Modeling Earth Systems and Environment* 6(2):917–27. doi: 10.1007/s40808-020-00728-7.
Jenkins, Jennifer C., David C. Chojnacky, Linda S. Heath, and Richard A. Birdsey. 2003. "National-Scale Biomass Estimators for United States Tree Species." *Forest Science* 49(1):12–35.
Kampe, Thomas U. 2010. "NEON: The First Continental-Scale Ecological Observatory with Airborne Remote Sensing of Vegetation Canopy Biochemistry and Structure." *Journal of Applied Remote Sensing* 4(1):043510. doi: 10.1117/1.3361375.
Khati, Unmesh, Marco Lavalle, Gustavo H. X. Shiroma, Victoria Meyer, and Bruce Chapman. 2020. "Assessment of Forest Biomass Estimation from Dry and Wet SAR Acquisitions Collected during the 2019 UAVSAR AM-PM Campaign in Southeastern United States." *Remote Sensing* 12(20):1–18. doi: 10.3390/rs12203397.
Lu, Dengsheng, Qi Chen, Guangxing Wang, Lijuan Liu, Guiying Li, and Emilio Moran. 2016. "A Survey of Remote Sensing-Based Aboveground Biomass Estimation Methods in Forest Ecosystems." *International Journal of Digital Earth* 9(1):63–105. doi: 10.1080/17538947.2014.990526.
McGaughey, Robert J. 2016. "FUSION – Software for Lidar Data Analysis and Visualization." (January):1–4.
Næsset, E., Bollandsas, O.M., Gobakken, T., Gregoire, T.G., Stahl, G., 2013. Modelassisted estimation of change in forest biomass over an 11 year period in a sample survey supported by airborne LiDAR: a case study with post-stratification to provide "activity data". Remote Sens. Environ. 128, 299–314.
Nandy, Subrata, Ritika Srinet, and Hitendra Padalia. 2021. "Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, -1 and SentinelUsing Random Forest Algorithm in Northwest Himalayan Foothills of India." *Geophysical Research Letters* 48(14):1–10. doi: 10.1029/2021GL093799.
Narine, Lana L., Sorin Popescu, Amy Neuenschwander, Tan Zhou, Shruthi Srinivasan, and Kaitlin Harbeck. 2019. "Estimating Aboveground Biomass and Forest Canopy Cover with Simulated ICESat-2 Data." *Remote Sensing of Environment* 224(June 2018):1–11. doi: 10.1016/j.rse.2019.01.037.
Narine, Lana L., Sorin C. Popescu, and Lonesome Malambo. 2020. "Using ICESat-2 to Estimate and Map Forest Aboveground Biomass: A First Example." *Remote Sensing* 12(11):1–16. doi: 10.3390/rs12111824.
National Ecological Observatory Network (NEON). 2022a. "Discrete Return LiDAR Point Cloud (DP1.30003.001)."
National Ecological Observatory Network (NEON). 2022b. "Vegetation Structure (DP1.10098.001)."
Nelson, Andrew S., Aaron R. Weiskittel, Robert G. Wagner, and Michael R. Saunders. 2014.

"Development and Evaluation of Aboveground Small Tree Biomass Models for Naturally Regenerated and Planted Species in Eastern Maine, U.S.A." *Biomass and Bioenergy* 68(September):215–27. doi: 10.1016/j.biombioe.2014.06.015.

Nie, Sheng, Cheng Wang, Xiaohuan Xi, Shezhou Luo, Guoyuan Li, Jinyan Tian, and Hongtao Wang. 2018. "Estimating the Vegetation Canopy Height Using Micro-Pulse Photon-Counting LiDAR Data." *Optics Express* 26(10):A520. doi: 10.1364/oe.26.00a520.

NOAA, National Oceanic and Atmospheric Administration NOAA Coastal Services. 2012. "Lidar 101 : An Introduction to Lidar Technology , Data , and Applications." *NOAA Coastal Services Center* (November):76.

Pau, Stephanie, Jesse B. Nippert, Ryan Slapikas, Daniel Griffith, Seton Bachle, Brent R. Helliker, Rory C. O'Connor, William J. Riley, Christopher J. Still, and Marissa Zaricor. 2022. "Poor Relationships between NEON Airborne Observation Platform Data and Field-Based Vegetation Traits at a Mesic Grassland." *Ecology* 103(2):1–10. doi: 10.1002/ecy.3590.

Popescu, Sorin C. 2007. "Estimating Biomass of Individual Pine Trees Using Airborne Lidar." *Biomass and Bioenergy* 31(9):646–55. doi: 10.1016/j.biombioe.2007.06.022.

Popescu, Sorin C., and Randolph H. Wynne. 2004. "Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable Window Size for Estimating Tree Height." *Photogrammetric Engineering and Remote Sensing* 70(5):589–604. doi: 10.14358/PERS.70.5.589.

Schmidt, H., and A. Karnieli. 2001. "Sensitivity of Vegetation Indices to Substrate Brightness in Hyper-Arid Environment: The Makhtesh Ramon Crater (Israel) Case Study." *International Journal of Remote Sensing* 22(17):3503–20. doi: 10.1080/01431160110063779.

Silva, Carlos Alberto, Laura Duncanson, Steven Hancock, Amy Neuenschwander, Nathan Thomas, Michelle Hofton, Lola Fatoyinbo, Marc Simard, Charles Z. Marshak, John Armston, Scott Lutchke, and Ralph Dubayah. 2021. "Fusing Simulated GEDI, ICESat-2 and NISAR Data for Regional Aboveground Biomass Mapping." *Remote Sensing of Environment* 253(December 2020):112234. doi: 10.1016/j.rse.2020.112234.

Sun, Tong, Jianbo Qi, and Huaguo Huang. 2020. "Discovering Forest Height Changes Based on Spaceborne Lidar Data of ICESat-1 in 2005 and ICESat-2 in 2019: A Case Study in the Beijing-Tianjin-Hebei Region of China." *Forest Ecosystems* 7(1). doi: 10.1186/s40663-020-00265-w.

Vaglio Laurin, Gaia, Nicola Puletti, William Hawthorne, Veraldo Liesenberg, Piermaria Corona, Dario Papale, Qi Chen, and Riccardo Valentini. 2016. "Discrimination of Tropical Forest Types, Dominant Species, and Mapping of Functional Guilds by Hyperspectral and Simulated Multispectral -2 Data." *Remote Sensing of Environment* 176:163–76. doi: 10.1016/j.rse.2016.01.017.

Venier, Lisa A., Tom Swystun, Marc J. Mazerolle, David P. Kreutzweiser, Kerrie L. Wainio-Keizer, Ken A. McIlwrick, Murray E. Woods, and Xianli Wang. 2019. "Modelling Vegetation Understory Cover Using LiDAR Metrics." *PLoS ONE* 14(11):1–17. doi: 10.1371/journal.pone.0220096.

Wang, Cangjiao, Andrew J. Elmore, Izaya Numata, Mark A. Cochrane, Lei Shaogang, Jiu Huang, Yibo Zhao, and Yuanyuan Li. 2022. "Factors Affecting Relative Height and Ground Elevation Estimations of GEDI among Forest Types across the Conterminous USA." *GIScience and Remote Sensing* 59(1):975–99. doi: 10.1080/15481603.2022.2085354.

Wickham, James, Stephen V. Stehman, Daniel G. Sorenson, Leila Gass, and Jon A. Dewitz.

2021. “Thematic Accuracy Assessment of the NLCD 2016 Land Cover for the Conterminous United States.” *Remote Sensing of Environment* 257(February):112357. doi: 10.1016/j.rse.2021.112357.

Williams, Marcus D., Christie M. S. Hawley, Marguerite Madden, and J. Marshall Shepherd. 2017. “Mapping the Spatio-Temporal Evolution of Irrigation in the Coastal Plain of Georgia, USA.” *Photogrammetric Engineering and Remote Sensing* 83(1):57–67. doi: 10.14358/PERS.83.1.57.

Xie, Bo, and Zhibin Huang. 2020. “Estimates of Forest Canopy Height Using a Combination of ICESat-2 / ATLAS Data.” *Remote Sensing*.

Zhao, Qingxia, Shichuan Yu, Fei Zhao, Linghong Tian, and Zhong Zhao. 2019. “Comparison of Machine Learning Algorithms for Forest Parameter Estimations and Application for Forest Quality Assessments.” *Forest Ecology and Management* 434(December 2018):224–34. doi: 10.1016/j.foreco.2018.12.019.

Zhou, Tan, Sorin Popescu, Lonesome Malambo, Kaiguang Zhao, and Keith Krause. 2018. “From LiDARWaveforms to Hyper Point Clouds: A Novel Data Product to Characterize Vegetation Structure.” *Remote Sensing* 10(12):1–23. doi: 10.3390/rs10121949.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Qu´ebec. Remote Sens. Environ. 112 (10), 3876–3890.

Frazer, G.W., Magnussen, S., Wulder, M.A., Niemann, K., 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDARderived estimates of forest stand biomass remote Sens. Environ. 115 (2011), 636–649.

Margolis, H.A., Nelson, R.F., Montesano, P.M., Beaudoin, A., Sun, G., Andersen, H.-E., Wulder, M.A., 2015. Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. Can. J. For. Res. 45, 838–855.

Nelson, R., Boudreau, J., Gregoire, T.G., Margolis, H.A., Næsset, E., Gobakken, T., Ståhl, G., 2009. Estimating Quebec provincial forest resources using ICESat/GLAS. Can. J. For. Res. 39, 862–881.

Pascual, C., García-Abril, A., Cohen, W.B., Martín-Fern´andez, S., 2010. Relationship between LiDAR-derived Forest canopy height and Landsat images. Int. J. Remote Sens. 31 (5), 1261–1280.

White, J.C., Tompalski, P., Vastaranta, M., Wulder, M.A., Saarinen, S., Stepper, C., Coops, N.C., 2017. A model development and application guide for generating an enhanced forest inventory using airborne laser scanning data and an area-based approach. In: CWFC Information Report FI-X-018, Canadian Forest Service. Pacific Forestry Centre, Victoria, BC, Canada, p. 38.

Chapter 3: Detecting irrigation potential zone of the southern United States using data science and remotely sensed data

3.1 Introduction

A crucial aspect of modern agricultural practices involves efficient irrigation management, which plays a pivotal role in enhancing crop productivity and ensuring water resource sustainability. Over time, researchers have dedicated their efforts to crafting models to accurately estimate irrigation statistics and map irrigation patterns and dynamics, which are required for effective planning (L. Zhang et al., 2022). These models have evolved from labor-intensive manual data collection and sampling methods to cutting-edge remote sensing techniques. For example, previous root zone soil moisture estimate by digging holes at depths of 6, 12, or 24 inches based on crop types, collecting samples and analysis in the laboratory, for water stress estimation by visually observing plant leaves, and field interview with farmers to know their irrigation practices and statistics, possibility of human error and bias (Babaeian et al., 2021). Remote sensing has gained prominence due to its efficiency in providing detailed insights at lower costs and with higher accuracy (Ambika et al., 2016; Pervez & Brown, 2010). The distinction between passive and active remote sensing approaches forms a critical part of this evolution.

Satellite-based remote sensing sensors like Landsat and Sentinel referred to as passive remote sensing can measure the spectral reflectance for mapping and calculating indices for extracting information from an area, but are limited to collecting soil moisture information to a certain depth from the soil surface (Chance et al., 2017; Rahman et al., 2011). Additionally, the spectrum used by satellite imagery is incapable of penetrating cloudy skies, and dense vegetation cover for extracting information on soil surface (Dari et al., 2021; Ishtiaque et al., 2020; Kim et al., 2021; Wakigari & Leconte, 2022). In contrast, active remote sensors or a combination of active-passive sensors like Soil Moisture Active Passive (SMAP) emit microwave energy that is capable of penetrating the soil surface into a certain depth (Dash & Sinha, 2022; Kolassa et al., 2020). The (SMAP) mission is equipped with radar and radiometer instruments, two active-passive instruments (Dash & Sinha, 2022; El Sharif et al., 2015). The radar emits microwave

pulses toward the earth's surface and measures the time it takes for the signals to return, providing information on soil moisture (Rabiei et al., 2021). The radiometer, on the other hand, measures natural microwave emissions from the earth's surface, which are influenced by soil moisture and freeze/thaw state required for predicting carbon and water cycle (Dash & Sinha, 2022). Therefore, a combination of active and passive instruments would provide more precise information about soil surface, drought and flood scenarios, and overall water and carbon cycles (Kisekka et al., 2022; Lawston et al., 2017; Zhu & Zhu, 2021)

The SMAP sensor was launched on January 31, 2015, with a mission to provide accurate and high-resolution soil moisture freeze-thaw cycle data into variable climatic conditions where passive sensors cannot achieve this information (Wind & Penman-monteith, 2003; L. Zhang et al., 2022). This advancement is a pivotal Earth-observing satellite program designed to offer near-global coverage (Platonov & Vincent, n.d.). Along with the primary mission, the SMAP mission is capable of supporting agriculture by monitoring irrigation practices, aiding in weather forecasting and disaster management, and contributing to climate studies by providing critical insights (Lawston et al., 2017). This also plays a crucial role in enhancing water balance research due to the very high temporal resolution (3-hour interval) and open accessibility policy (Zappa et al., 2022; Zaussinger et al., 2019). In this pursuit, SMAP along with other datasets like Soil Moisture of Ocean Salinity (SMOS) and Synthetic Aperture Radar (SAR) data of Sentinel has been used in irrigation signal detection purposes, mapping irrigated areas, and study outputs used for the land surface model (LSM) (Dari et al., 2021). Dari et al. show that instead of SMAP's native resolution of the L-band microwave, the downscaled 1km version gave more accurate values when combined with other datasets, like SMOS and ASCAT (Advanced Scatterometer).

Current irrigation practices in the earth system model showed that SMAP is capable of addressing root zone soil moisture conditions up to ~100 cm in which depth most cultivated crop's roots are growing (McDermid et al., 2023). SMAP's enhanced L2 SM products at 9km resolution, integration of different time series satellite image-based indices like Normalized Difference Vegetation Index (NDVI), and use of machine learning approaches also provide actual evapotranspiration (ET) and soil moisture information with higher accuracies in a range of 90.1 ± 2.7% (Paolini et al., 2022). Like soil moisture estimation, for irrigation water use (IWU)

purposes, the passive sensor applications are also limited where a combination of data is needed (Bwambale et al., 2022). Along with SMAP, those studies utilized reanalysis datasets which prepare by combining various sources of observational data, such as satellite measurements, ground-based measurements, and weather models, to produce a more accurate and consistent representation of historical weather or environmental conditions. Combining SMAP with reanalysis datasets ASCAT that were found useful for getting better accuracy (Zaussinger et al., 2019). Therefore, root zone soil moisture information from SMAP, potential evapotranspiration (PET) from the reanalysis dataset, and precipitation information can be utilized to find complex interplay in the water balance model of this study area.

The PET stands as a pivotal process representing the combined loss of water to the atmosphere through evaporation and plant transpiration. This estimates water loss from the system while precipitation plays the opposite factor of adding water to the system (Paolini et al., 2022; Safaeian et al., 2023). The latest advancements in PET estimation techniques are similarly important in leveraging both ground-based and remote-sensing data to address complex challenges (Ougahi & Mahmood, 2022). The updated PET information in the reanalysis dataset at higher resolution from the Fifth generation of the European Centre for Medium-Range Weather Forecasts Reanalysis Land (ECMWF-ERA5) that has incorporated those updated methods makes it considerable for this study (Paolini et al., 2022). Furthermore, its native resolution has a similar resolution to the SMAP dataset which should reduce the potential bias that is often created when any remote sensing pixels go through the transformation process (Shahzaman et al., 2021). The amount of precipitation, on the other hand, is important to add in the modeling approach since it indicates how much water is added to the soil (Wu & Zhao, 2023). The 5km grid pixel resolution precipitation datasets from Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) provide precipitation information that is useful in the use of the water balance model (Kolluru et al., 2020). These three data (ERA5 Land, CHIRPS, SMAP) from different sources processing, calculating, and integration of remote sensing and data science method is needed to better understand of water needed for croplands areas of the AL, FL, and GA.

3.1.1 Objectives

This scientific endeavor strives to estimate the irrigation level required in croplands in the AL, FL, and GA states based on plant needs. To investigate irrigation areas, this study utilizes SMAP, ERA5-Land, and CHIRPS datasets and advanced data science techniques to -

a. Conduct correlation analysis between atmospheric water demand and root zone soil moisture
b. Delineate irrigated and non-irrigated areas based on croplands
c. Find varying irrigation demand zones in AL, FL, and GA

3.2 Methodology

3.2.1 Study area

The study considered three southern states, Alabama (AL), Florida (FL), and Georgia (GA), in the United States considering their challenging climatic pattern, agricultural production, ecological integrity, and contribution to the national economy (Figure 3-1). Within this diverse expanse, numerous environmental instances assume significance, like the Appalachian Mountains exert a profound influence on local climates and ecosystems across states such as Alabama and Georgia. In Florida, the exceptional biodiversity, and ecological functions of the Everglades ecosystem act as a natural water filter and support various species. By collectively examining these diverse environmental and geographical instances, this study differentiates irrigation potential zones based on the plant's root zone level water required that are crucial part of sustainable agricultural production and water balance of an area. This water balance dynamic sheds light on the complexities of various ecosystems and landscapes and finds the amount of irrigation needed for different regions to continue ongoing agricultural production.

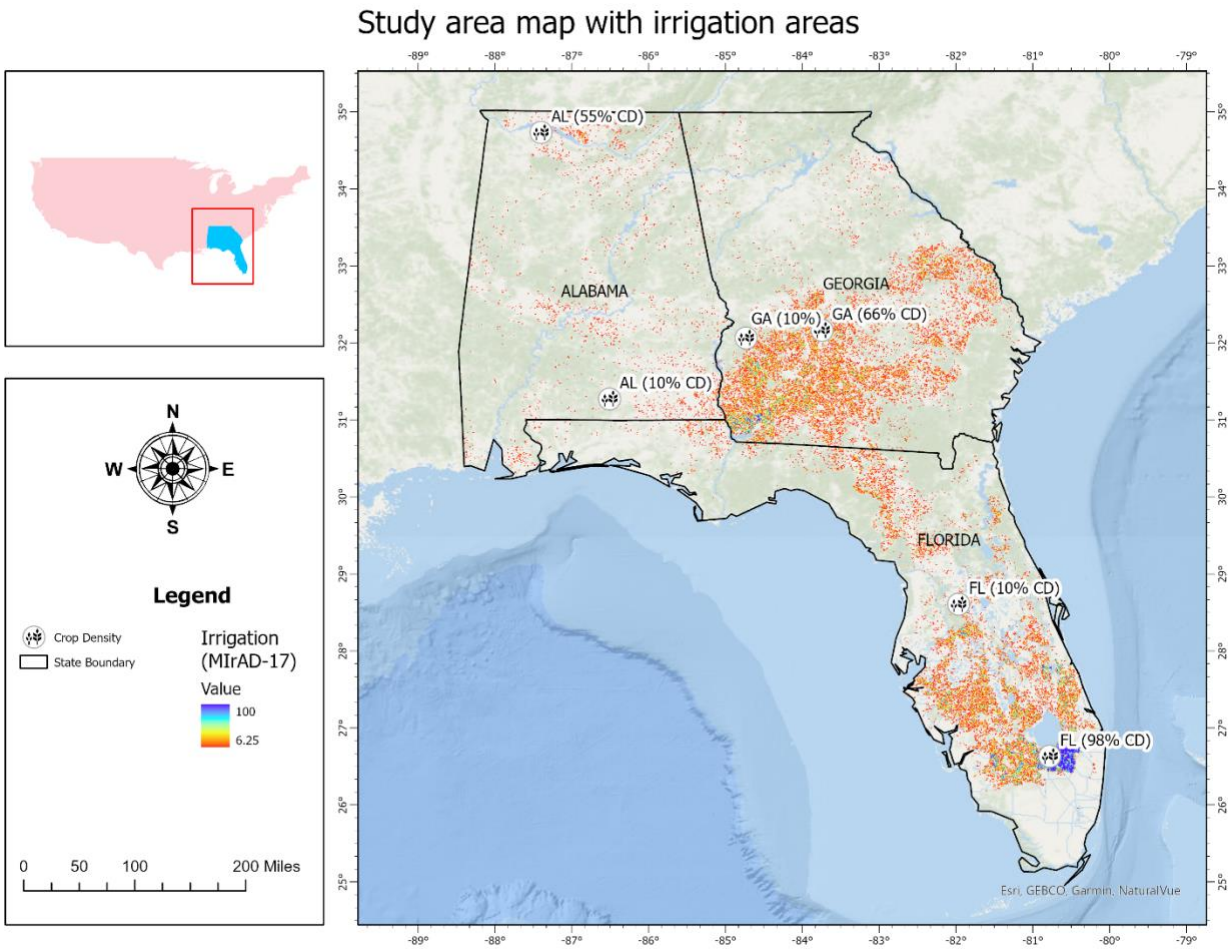

Figure 3-1: Study area map, reference high and low crop density reference points with irrigation (%) raster distribution over the study area. The 1-km resolution irrigation Moderate Resolution Imaging Spectroradiometer (MODIS) Irrigated Agriculture Datasets for the Conterminous United States (MIrAD-US) were used for extracting the percent of irrigation applied in a location.

Source: (United States Geological Survey, 2019)

3.2.2 Weather and climate

In the southeastern United States, Alabama (AL), Florida (FL), and Georgia (GA) exhibit distinct climates and agricultural characteristics (Table 3-1). Alabama boasts a humid subtropical climate with hot summers and mild winters. Rainfall is evenly distributed throughout the year, with an annual average of approximately 55 inches (140 cm). The summer temperature ranges from 90-95°F (32-35°C), while winters see lows of 30-40°F (0-4°C). The state enjoys around 220-240

days of sunshine annually but experiences high humidity, particularly in the summer (National Weather Service, 2017). Various soil types, including clay, loam, and sandy soils, support a diverse agriculture system, including the cultivation of cotton, soybeans, peanuts, corn, and livestock. Alabama's abundant water resources, including rivers and aquifers, are managed sustainably through water usage regulation, efficient irrigation practices, and water source protection (County et al., 1966).

Table 3-1: Weather and climatic pattern of the Southern United States

Aspect	Alabama (AL)	Florida (FL)	Georgia (GA)
Climate	Humid subtropical, hot summers, mild winters	Diverse climates: tropical in the south, temperate in the north	Humid subtropical, hot summers, mild winters
Rainfall Patterns	Even distribution, slight increase in summer	The wet season in summer, the potential for drought	Varied, more rain in north, seasonal patterns
Annual Rainfall	~ 55 inches (140 cm)	~ 54 inches (137 cm)	~ 50 inches (127 cm)
Temperature	Summer: 90-95°F (32-35°C), Winter: 30-40°F (0-4°C)	Summer: 85-95°F (29-35°C), Winter: 50-60°F (10-16°C)	Summer: 90-95°F (32-35°C), Winter: 30-40°F (0-4°C)
Avg. Annual Temperature	Approximately 64°F (18°C)	Varies by region, southern part around 75°F (24°C)	Approximately 63°F (17°C)
Sunshine Days	Approx. 220-240 days of sunshine annually	Approx. 230-250 sunny days per year	Approx. 210-230 days of sunshine per year
Humidity	High, especially in summer	High, particularly during the wet season	High, especially during the summer
Cloud Cover	Partly cloudy to cloudy conditions, more cloud cover during summer	Partly cloudy to cloudy, especially during the wet season	Varies with seasonal conditions
Topography	Varied terrain	Predominantly flat	Varied terrain
Slopes	sloped in the north	high-water table	mountains in the north
Water Resources	Abundant rivers, reservoirs, aquifers	Aquifers, lakes, extensive coastal waters	Rivers, reservoirs, groundwater sources
Water Management Practices	Regulation of water usage, efficient irrigation, protection of water sources	Water conservation, aquifer recharge, monitoring water use	Equitable water allocation, conservation, efficient irrigation, responsible groundwater use

Aspect	Alabama (AL)	Florida (FL)	Georgia (GA)
Soil Types	Various soil types including clay, loam, and sandy soils	Predominantly sandy soils, clay in some areas	Range of soil types, including clay, loam, and sandy soils
Soil Water Drainage Capacity	Moderate to good drainage capacity	Varies by location, can be well-draining to poorly draining	Varies by region, with different drainage capacities
Main Cultivation Pattern	Diverse agriculture, including row cropping and mixed farming	Large-scale commercial agriculture and diversified farming	Diverse agriculture with row crops and livestock
Main Crop Cultivated	Cotton, soybeans, peanuts, corn, poultry, cattle	Citrus, sugarcane, tomatoes, vegetables, cattle, poultry	Peaches, peanuts, cotton, corn, poultry, cattle
Main Irrigation System	Center pivot, drip irrigation, furrow irrigation	Drip irrigation, center pivot, furrow irrigation	Center pivot, furrow irrigation, drip irrigation
Geographical Location	Southeastern U.S., bordered by MS, GA, FL	Southeastern U.S., Gulf of Mexico, Atlantic Ocean	Southeastern U.S., bordered by AL, SC, FL

Sources: (Babaeian et al., 2021; County et al., 1966; Dower, 1965; Hermsen, 2022; National Weather Service, 2017; Saul, 2018; Schmidt & Karnieli, 2001; Shaw, 2018)

Florida, with its diverse climate, experiences tropical conditions in the south and temperate conditions in the north (Table 3-1). Rainfall, averaging approximately 54 inches (137 cm) annually, occurs predominantly during the wet season in the summer, posing the potential for drought. Florida's summer temperatures range from 85-95°F (29-35°C), while winters are milder, averaging 50-60°F (10-16°C) in the southern regions (Shaw, 2018). The state enjoys abundant sunshine, with around 230-250 sunny days annually, but is characterized by high humidity, especially during the wet season. Predominantly sandy soils and some clay areas support large-scale agriculture, featuring citrus, sugarcane, vegetables, and cattle. Florida utilizes drip irrigation, center pivot systems, and furrow irrigation for efficient water management, complementing its water resources, including aquifers and coastal waters (Hermsen, 2022; National Weather Service, 2017; Shaw, 2018).

Georgia, situated in the southeastern U.S., shares a humid subtropical climate with Alabama (Table 3-1). Its annual rainfall averages around 50 inches (127 cm), with more rain in the northern

regions and seasonal patterns. Georgia experiences summer temperatures ranging from 90-95°F (32-35°C) and mild winter temperatures around 30-40°F (0-4°C) (National Weather Service, 2017; Saul, 2018). The state enjoys approximately 210-230 days of sunshine annually and high humidity levels, primarily during the summer. With various soil types, including clay, loam, and sandy soils, Georgia's agriculture encompasses row crops and livestock such as peaches, peanuts, cotton, and poultry. The state's water resources, including rivers, reservoirs, and groundwater sources, are managed equitably with conservation, efficient irrigation, and responsible groundwater use practices (Hermsen, 2022; National Weather Service, 2017; Shaw, 2018).

3.2.3 Characteristics of data visualizing reference points

In the study area, six reference points were considered for data visualization, which served to identify crop density and their locations (Figure 3-1). Two points from each state were considered, one for the highest crop density location and another for the lowest crop density location. Table 3-2 highlights key features of each reference point, including the location name, state, and county, as well as geographical coordinates (latitude and longitude in decimal degrees). Moreover, it presents the presence of irrigation percentage on a 9km pixel scale, and average irrigation application (%) associated with these regions. From the high crop density of FL (98% crop density (CD) in Palm Beach County to the AL (10% CD) in Covington County, these reference points collectively illustrate the wide-ranging agricultural and environmental characteristics present in the study area, which are pivotal for understanding the intricate dynamics of irrigation potential and soil moisture.

Table 3-2: Characteristics of result representing reference points.

Name with state	County Name	Latitude Degree Decimal (DD)	Longitude Degree Decimal (DD)	Crop Density (CD in %)	Irrigated on 9km Pixel	Average irrigation applied (%)
FL (98% CD)	Palm Beach	26.630	-80.780	98.1	99.7	52.4
GA (66% CD)	Dooly	32.163	-83.746	65.8	67.6	25
AL (55% CD)	Lawrence	34.731	-87.401	54.8	11.6	23.9
FL (10% CD)	Sumter	28.606	-81.966	10	13.5	12.5
AL (10% CD)	Covington	31.273	-86.510	10	0.3	0.3
GA (10% CD)	Stewart	32.064	-84.733	10	1.2	8.33

3.2.4 Data

3.2.4.1 SMAP dataset and processing

The Soil Moisture Active Passive (SMAP) mission employs a dual-instrument approach to collect soil moisture data. The SMAP mission, launched in 2015, has been a vital source of soil moisture data (Rabiei et al., 2021). Its active radar instrument, operating at L-Band frequencies, emits microwave pulses toward Earth and measures their return times to determine surface soil moisture content. Complementing this, the L-Band passive radiometer captures natural microwave emissions from the Earth's surface. This is influenced by soil moisture which enables SMAP sensors to capture soil moisture information at deeper root zone levels. This combination of active and passive instruments facilitates the collection of accurate soil moisture data across varying conditions, such as cloud cover and dense vegetation conditions which enhances global understanding of soil moisture dynamics (Lawston et al., 2017).

To harness its valuable information, the high-resolution 3-hour *'root zone soil moisture'* band from SMAP was processed in Google Earth Engine (GEE) into the daily average dataset. Subsequently, temporal aggregations were performed to calculate monthly averages for the dry months, May through October of each year, spanning the years 2015 to 2023. This monthly data aggregation allows for a comprehensive understanding of soil moisture dynamics during the growing seasons over this extended time frame (Kumar et al., 2019). It aids in a wide range of applications, from agricultural planning to hydrological modeling and climate research (Hao et al., 2019). SMAP's contributions, when processed and aggregated in this manner, provide essential insights into long-term soil moisture trends and variations that can have significant implications for various Earth and environmental sciences (Jimma et al., 2023; McDermid et al., 2023; Shahzaman et al., 2021).

3.2.4.2 Reanalysis ERA5 Land and processing

The ECMWF Reanalysis version-5 land, also known as ERA5-Land, represents a significant advancement in atmospheric and global climate change understanding from time series and

global scale reanalysis datasets (Kolluru et al., 2020). Developed by the Copernicus Climate Change Service (C3S) at ECMWF, ERA5 stands as the fifth-generation reanalysis, encompassing data from January 1940 to the present day (Kolluru et al., 2020; Ougahi & Mahmood, 2022). This comprehensive dataset offers hourly assessments of a diverse array of climatic variables, spanning the realms of atmosphere, land, and oceans. This dataset covers the total globe with different grid resolution approaches. The original dataset is composed of a 30km grid of intricated atmospheric structure with revolving 137 levels extending from the surface to a height of 80km. However, the reanalysis data for the Land surface of the contiguous US provides 11km pixels. In this study, among 12 bands of the ERA5 Land dataset available from the Google Earth Engine (GEE) platform, the daily 'potential evapotranspiration sum' (PEV) band was used. Although the native resolution is 30km, the reanalysis dataset provides information in 11km, which was further averaged to make a monthly average dataset for the months of May to October of each year.

3.2.4.3 CHIRPS dataset and processing

The Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset was employed for this study. CHIRPS combines daily satellite infrared data with ground-based observations which produces 5km pixels with a very high temporal resolution suitable for climatic research. These are often useful for assessing hydrological conditions, drought monitoring, and climate research due to its high-quality precipitation data for the time span 1981 to the present, providing long-term historical records (Fassoni-Andrade et al., 2021; Fernandez-Palomino et al., 2022).

In this study, the CHIRPS dataset was also processed in Google Earth Engine (GEE) to create monthly average precipitation data for the period spanning from May 2015 to October 2023, resulting in a total of 52 data points matching the months of SMAP and ERA5 Land datasets. The process involved filtering the CHIRPS images to select specific months and years, calculating monthly averages, and converting the original resolution to SMAP resolution. Since the precipitation amount recorded in the CHIRPS dataset is in millimeters and the calculation of SMAP and ERA5 Land is in meters, the precipitation measures were converted to meters per day

(Wu & Zhao, 2023). The precipitation data works as an input whereas the evapotranspiration denotes how much water is being lost from the source (W. Zhang et al., 2020).

3.2.4.4 Removing Seasonal Variability

The seasonal variation creates unusual bias between the real ground observation and the observed dataset (Kolassa et al., 2020). Extreme evapotranspiration, precipitation, or soil moisture information leads to unrealistic assumptions that need to be removed before further analysis. This study removes seasonal variable data that was applied to the monthly average of the ERA5, CHIRPS and SMAP datasets described earlier. Firstly, for the SMAP dataset, the 3-hourly data was converted to daily data, ensuring the consistency of daily composites for further analysis. Subsequently, a defined climatology time range was used to create daily composite images (Felfelani et al., 2018; Zaussinger et al., 2019). The monthly climatology was calculated by averaging these daily composites, with images tagged by their respective months. Long-term monthly averages are then computed, further aiding in the anomaly detection process (Kumar et al., 2019). By subtracting the monthly climatology from the original data, this section provides a critical foundation for seasonal variability removal, allowing for the isolation of anomalies, trends, or unique deviations from the climatological norm in subsequent analyses.

3.2.5 Data Analysis

The identification of irrigation potential zones depends on assessing water deficit areas, which can be achieved through a water balance model (Paolini et al., 2022). The equilibrium condition of the water balance model states that there is no need to supply water from other sources when the natural condition is sufficient for plant growth. This is opposite for the water deficit condition, in that, additional supply of water is needed to grow plants (Anurag & Kumar, 2016). The root zone soil moisture deficit happens when the evapotranspiration from the soil surface exceeds the amount of water received from precipitation in an area (El Sharif et al., 2015). Consequently, estimating the water balance model of the study area is required to understand the underlying situations (Rahimi et al., 2015). The PET-P is defined as the difference between evapotranspiration (ERA5 Land) and precipitation amount (CHIRPS). The resulting value reveals whether water loss due to evapotranspiration is higher than the precipitation or not. If

evapotranspiration is higher, then artificial irrigation is needed to cultivate crops (Wu & Zhao, 2023). The estimated PET-P indicates the soil surface condition which has a relation with the root zone soil moisture condition of the study area.

3.2.5.1 Preparing correlation map

The SMAP root zone soil moisture data provides information about soil moisture levels up to 100 cm below the soil surface. Most of the cultivated crop roots are within this level so analyzing moisture conditions can determine the actual water requirements for the plants. Hence, correlation analysis between PET-P vs. SMAP can reveal where water is needed, not just based on the soil surface, but also on the type of plant. To find the relationship between PET-P and SMAP the Pearson correlation matrix (R) was used considering its capability to indicate both positive and negative directional relations. Pearson correlation ranges from -1 to +1, where a negative correlation value means variables have opposite directional relationships, by increasing one variable value another value will decrease or vice versa for the positive correlation value. For this study, the relationship should be negatively correlated since both variables PET-P and soil moisture from SMAP indicate opposite features of each other. Thus, the strongly negative value areas in this study should mean an increase of PET-P over atmospheric water demand is extreme means root zone soil moisture should be minimal.

3.2.5.2 Masking for croplands

The cropland areas are the most fluctuated class in terms of root zone soil moisture since artificial irrigation is used to provide for cultivating crops. Therefore, cropland areas' features and trends would be different than other land cover areas like forest, built-up, and wetland areas. Aligning with the study focus, cropland areas are masked from the state-level correlation map Figure 3-6. To refine our analysis, we overlaid this map with the Cropland Data Layer (CDL) obtained from the United States Department of Agriculture (USDA), which was at a finer 30-meter resolution. Our objective was to isolate and select only those pixels in the correlation map that intersected with CDL-defined cropland regions. Following this masking process, we proceeded to extract boundary information and pixel values for the selected areas. Subsequently, we overlaid the dimensions of these extracted regions with the CDL layer to quantify the extent

to which each pixel was covered by cropland, providing insights into the areas suitable for cultivation. To further refine our analysis, the study established a criterion of selecting areas with a crop density exceeding 10%. Such regions, where cropland covered more than 10% of the pixel area, were earmarked for subsequent investigation. This meticulous methodology ensured that our analysis focused on areas with significant agricultural potential, as supported by the CDL data.

3.2.5.3 Identification of irrigation potential zone

Following the selection of areas with crop density exceeding 10% from the correlation map, the study proceeded to calculate the percentage of irrigation area and the average irrigation amount for each of these selected pixels. These calculations were performed through overlay operations using the ArcGIS Pro software, which gave insights into the spatial distribution of irrigated land and the corresponding average irrigation intensity. Subsequently, the study aimed to categorize the correlation pixels into five distinct classes to explore potential relationships between irrigation water quantities and the correlation ranges, which were derived from the map generated by comparison of potential evapotranspiration (PET-P) and SMAP data.

In this endeavor, the irrigation area and average irrigation amounts were derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) Irrigated Agriculture Datasets for the Conterminous United States (MIrAD-US). This dataset furnishes comprehensive information on irrigated agriculture, encompassing various aspects of irrigation practices in the United States. By integrating this valuable resource, the study aimed to gain further insights into the correlation patterns and their relevance to irrigation patterns and practices, enabling us to identify potential zones for irrigation with a refined understanding of the associated factors.

3.2.5.4 Statistical analysis

The study tried to find if there is any group that has significant differences among groups obtained from the correlation map. The study also finds, if there is any pattern among the correlation ranges and the amount of irrigation applied using statistical methods. The RStudio (visual interpreter of R software) was used for this purpose where ranges were considered as

categorical variables and the average percent of irrigation applied collected from MIrAD data was considered a continuous variable. After getting an overall significant p-value ($\alpha = 0.05$), the study tried to find if there is any group that has significant differences. Therefore, the analysis of variance (ANOVA) was conducted. The study further investigated to see pairwise significance results using Tukey's Honest Significant Differences (HSD) post hoc test.

3.3 Result

3.3.1 Potential evapotranspiration trends

In the analysis of the potential evapotranspiration (PET), the ERA5 Land reanalysis data was used in the 9km resolution grid cell during the period spanning May to October, which constitutes the six driest months crucial for understanding the irrigation requirements in crop cultivation. After masking the areas with more than 10% croplands, further processing was undertaken. Scenarios should be different in high and low crop density areas, attributable to shifts in precipitation, irrigation practices, and atmospheric conditions. However, to gain a comprehensive visual perspective, two specific geographical ranges for each state were presented, considering both the lowest and highest crop density. The corresponding geographical coordinates indicated in degree decimal format (DD) for AL, FL, and GA, provide a clear reference point for the data visualization. The same location and subsequent graphical demarcation also followed for the next graphs and figures for precipitation, PET-P, and the root zone soil moisture data for presentation purposes.

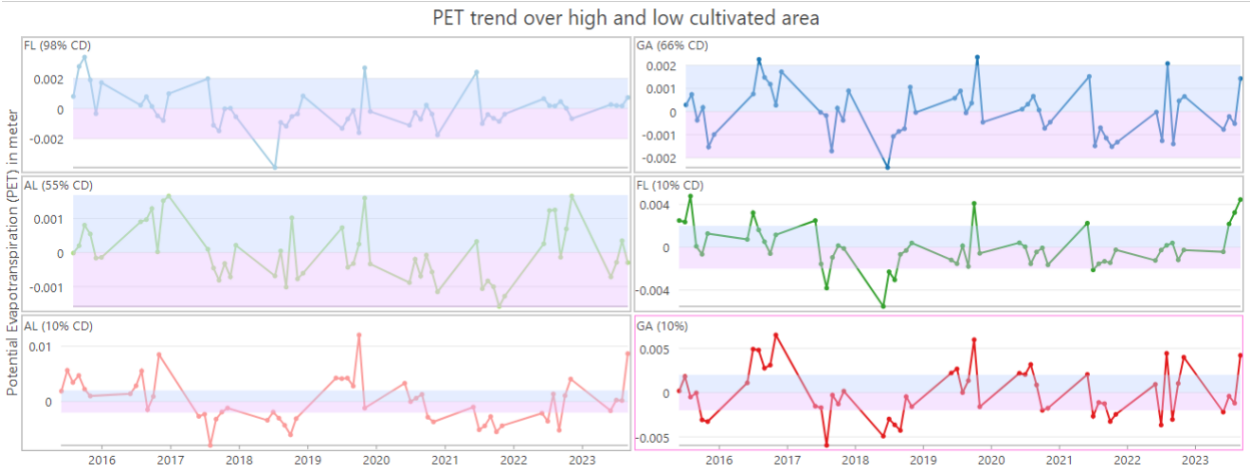

Figure 3-2: Potential evapotranspiration trend over low and high-cultivated areas from May to October from 2015 to 2023. The positive guide 0.002 (blue shaded) and negative guide -0.002 (pink shaded) help to focus on extreme PET values in different reference points. Overall, low cropland areas have higher positive and negative directional trends than cultivated croplands. Analyzing the six-month period and connecting it with the next month shows a sharp line in the graph.

Study findings reveal that in areas characterized by higher crop density, such as GA (66% CD) at (32.163, -83.745), the range of PET values was comparatively lower, reflecting the impact of irrigation practices (Figure 3-2). This phenomenon aligns with the negative values depicted in Figure 3-2, indicating that precipitation exceeded PET during the specified months. Conversely, in regions with the lowest crop density, the PET consistently exceeded precipitation levels. For instance, in Florida with 10% crop density, coordinates (28.61, -81.97), PET significantly exceeded precipitation (0.004) in the years 2018 and 2020. Notably, sharp lines between the previous year's October to the next year's May signify the absence of data which is not considered in this study's scope.

3.3.2 Precipitation trends

The study expanded the analysis to incorporate precipitation data collected and processed from the CHIRPS dataset's precipitation band. This integration gains a more complete understanding of the impact of precipitation on the atmospheric water balance. Study findings reveal that regions with flood-based irrigation systems show a higher amount of precipitation trend than other areas mostly located in FL. Figure 3-3 indicated that precipitation for FL was higher than

AL and GA. The bluish-shaded background in the positive direction called the guide considered 2 millimeters (mm) which helps to identify that May and October had less precipitation than other months (Figure 3-3). It is important to note that zero values on the graph represent the baseline, signifying the absence of data for those months. These results emphasize the trend of precipitation in both forms from natural and artificial sources. This helps to assess the atmospheric water balance, incorporating insights from CHIRPS precipitation data over the study area.

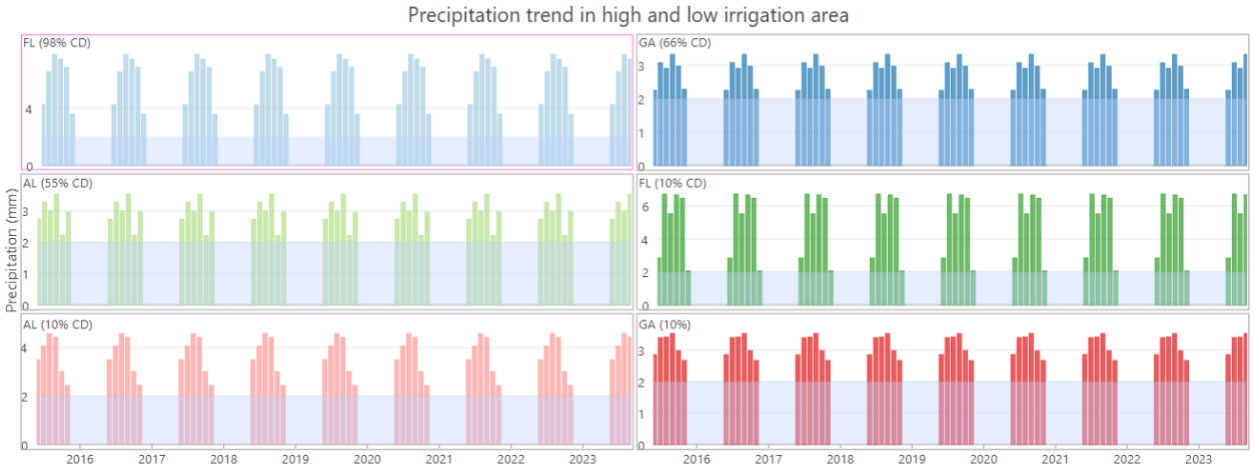

Figure 3-3: Precipitation trend over the study area from 2015 to 2023 for the driest month of May to October of each year. The 2mm indicates, that other than flood irrigation, other manual forms of irrigation show similar kinds of amounts provided in AL and GA.

3.3.3 Atmospheric water balance and trends

This study observed intriguing trends in the atmospheric water balance across different cropland areas, which are denoted by percentages of crop density. The data, as shown in Figure 3-4, revealed distinctive patterns in potential evapotranspiration (PET) minus precipitation values over time. Positive values Figure 3-4 indicated a water deficit. In the high-density cropland areas, such as FL with 98% crop density and GA with 66% crop density, there were no months where precipitation exceeded PET amounts (Figure 3-4). In contrast, for low crop density areas, like AL with 10% crop density, mainly in July of 2018 and 2019 precipitation exceeded plant requirements. Notably, FL and AL consistently exhibited more extreme positive values.

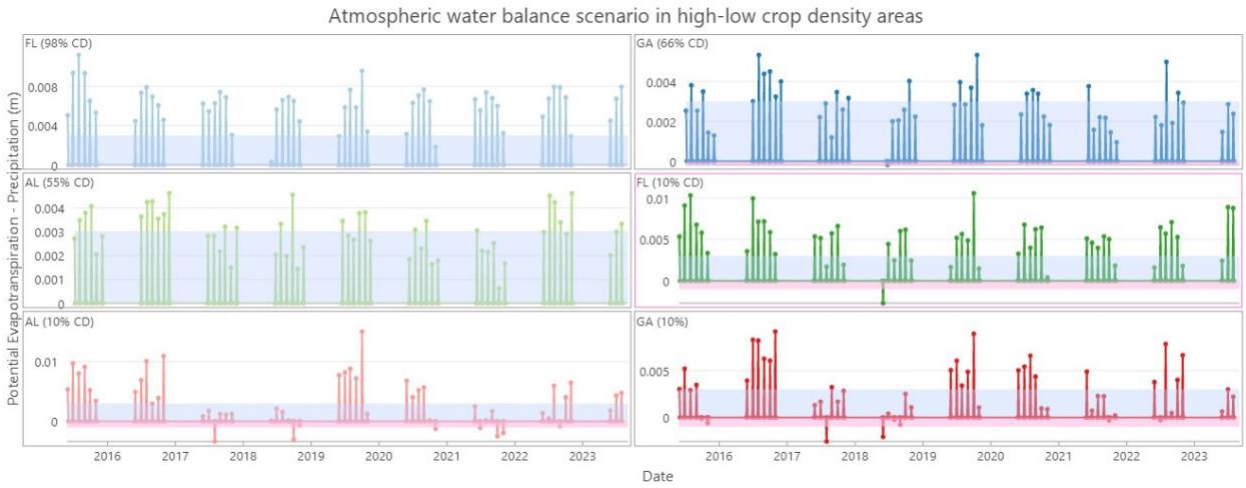

Figure 3-4: PET-P scenario for the six driest months from May to October in the highest and lowest crop density areas in AL, FL, and GA. There are two guides considered for clear visual observation in positive and negative directions. The positive guide value is 0.003 and the negative is -0.001clearly shows, PET-P value is prominent in FL with 98% crop density (CD) among others. In the lowest crop density areas, AL point's PET-P is more prominent in both positive and negative directions.

3.3.4 Root zone soil moisture trend

The potential evapotranspiration trend refers mostly to surface condition while root zone soil moisture focuses on below the soil surface up to 100 cm using the SMAP dataset. This depth is considered as the useful depth in which most crop roots develop (C. Zhang & Long, 2021). The study reveals both positive and negative soil moisture values, each conveying vital information about the study area's soil moisture state. Positive root zone soil moisture values indicate relatively higher soil moisture content compared to a reference condition, suggesting adequate soil moisture to support plant growth without additional irrigation. In contrast, negative values signify lower soil moisture levels relative to the reference state, signaling potential water stress and the need for irrigation or water management interventions to maintain healthy vegetation. The less crop-density areas in AL are the most water-stressed zone shown in Figure 3-5, which ranges from -0.1 m^3/m^3 in 2020 since there was a minuscule amount of irrigation applied. The bluish shade in the positive direction and pink shade in the negative direction which helps for visual enhancement called guide selected positive 0.03 and negative -0.02 shows very narrow

bands in AL's 10% CD area. The second most water-stressed zone is in the FL's 98% CD area which needed more artificial irrigation as observed (provided irrigation was only 52.4%). The soil moisture conditions in GA are somehow between these two regions of FL and AL.

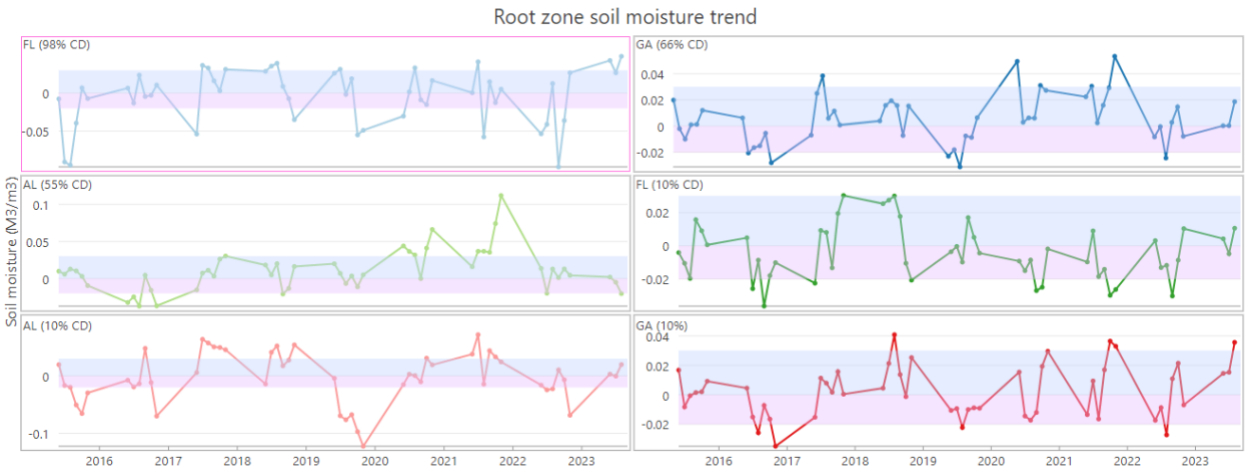

Figure 3-5: Root zone soil moisture trend over the high and low crop density areas of three states. A positive value indicates higher than reference soil moisture while negative indicates water stress. Guides used 0.03 in positive direction and -0.02 in negative directions.

3.3.5 Correlation analysis

The correlation analysis between PET-P and SMAP's root zone soil moisture indicate soil moisture dynamics across different regions. Initially, the unmasked correlation values exhibited a wide range from -0.798 to -0.003, emphasizing the inverse relationship between atmospheric water demand and soil moisture levels. However, when the analysis was refined by masking more than 10% cultivated areas, the correlation values ranged from -0.059 to -0.776, indicating that the presence of croplands significantly influenced the relationship between PET-P and soil moisture (Figure 3-6). The upper part of AL exhibited a more extreme negative correlation. Similarly, the mid-part of GA showed medium to high-level negative correlations (Figure 3-6). Notably, FL, with its abundant wetland areas, exhibited less distinctive correlation values, highlighting the complex interplay between land use and soil moisture dynamics.

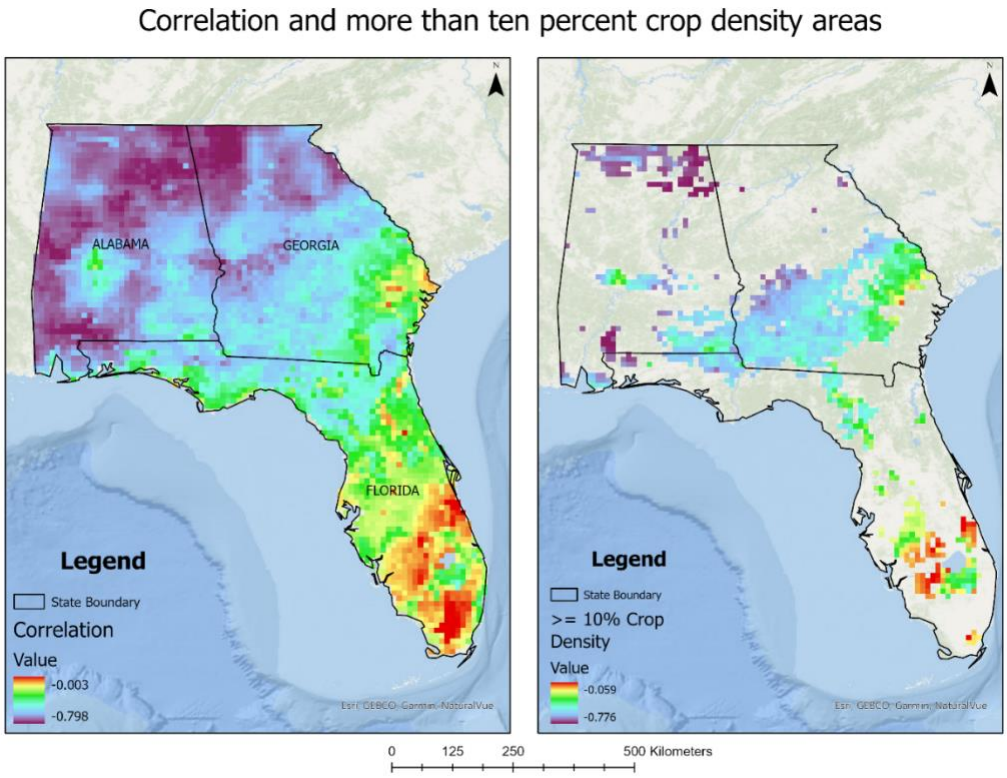

Figure 3-6: The correlation among PET-P and SM of the AL, FL, and GA states. The right side provides the masked cultivated areas, that are obtained from the cropland data layer of 2022.

3.3.5.1 Correlation analysis in high-density croplands

Florida (98% CD)

In the high cropland density region of Florida, where 98% of the 9km grid area is dedicated to cropland, the Pearson correlation value of -0.37 suggests a moderate negative relationship between potential evapotranspiration (PET-P) and soil moisture (Figure 3-7, red and black rectangle in upper left). This result is intriguing, as one might expect that intensive irrigation in such high-crop-density areas would mitigate negative correlations but as Table 3-2 shows, there was only 52.4% of irrigation was applied.

Figure 3-7: Correlation in high-density cropland reference points. The read box indicated specific locations scatter plots and corner connected black rectangles indicate Pearson correlation value on the specific points. The diagonal histogram shows the distribution of the X and Y axis data.

Georgia (66% CD)

The high cropland density region in Georgia, with 66% of the 9km grid area dedicated to cropland, exhibited a Pearson correlation of -0.56 (Figure 3-7, mid-red and black rectangle). This correlation value indicates a more pronounced negative relationship between PET-P and soil moisture compared to Florida since there was only 25% irrigation applied compared to 52.4% in FL.

Alabama (55% CD)

In the high cropland density area of Alabama, with 55% of the 9km grid covered by cropland, the Pearson correlation value of -0.65 reflects a strong negative relationship between PET-P and soil moisture (Figure 3-7). The correlation in Alabama is the most negative among the three states since there was only 12.5% of irrigation was applied.

3.3.5.2 Correlation analysis in low density croplands

Florida (10% CD):

In the low cropland density in Sumter County of Florida (10% CD), the Pearson correlation of -0.32 reflects a relatively low correlation between potential evapotranspiration (PET-P) and soil moisture. This outcome is consistent with the characteristics of areas with limited agricultural activity, where soil moisture is less influenced by human-managed irrigation. In this study, there was only 13.5% of areas irrigated, and average irrigation applied in 12.5%, revealing this relationship (Figure 3-8).

Figure 3-8: Correlation in low-density cropland reference points. The red box indicates specific locations scatter plots and the corner connected black rectangle indicates the Pearson correlation value on the specific points. The diagonal histogram shows the distribution of the X and Y axis data.

Alabama (10% CD)

The low cropland density region in Alabama (10% CD) exhibits a more negative correlation of -0.65. This stronger negative correlation suggests that in areas with sparse cropland coverage, the sparse irrigation practices lead to the relationship between PET-P and soil moisture. This area, under Covington County irrigated areas, was about 0.3, so the percentage was also about 0.0001% (~0.3) (Figure 3-8).

Georgia (10% CD)

Similarly, the low cropland density region in Georgia (10% CD) displays a notable negative correlation of -0.66. This suggests that despite the limited cropland coverage and irrigated areas was about 1.2% (8.33% irrigation applied) as shown in Table 3-2, other aspects like land use and land cover might impacted the relationship between PET-P and soil moisture. The presence of water bodies, forests, or consistent natural land cover types can lead to a more pronounced negative correlation.

3.3.6 Relationship between correlation map and applied irrigation

In the context of irrigation water demand estimation, the Moderate Resolution Imaging Spectroradiometer (MODIS) Irrigated Agriculture Datasets for the Conterminous United States (MIrAD-US) 1-km data was considered as an applied irrigation data source (United States Geological Survey, 2019). This 1-km grid raster was prepared from tabular statistics data, proving irrigation percent value in each percent and indicating an accuracy level of more than 90% considering its field survey tabular datasets (Pervez & Brown, 2010; L. Zhang et al., 2022).

After preparing irrigation applied datasets, the correlation map was categorized into five distinct classes namely -0.77 - -0.74, -0.74 - -0.58, -0.58 - -0.42, -0.42 - -0.25, and -0.25 - -0.05 for statistical analysis in keeping with the null hypothesis (there is no relation between classified class with irrigation applied). However, the analysis of variance (ANOVA) results for the 'Correlation ranges' variable in relation to applied 'Irrigation percentage' reveals a highly significant overall effect ($F = 38.59$, $p < 0.05$), indicating that the correlation ranges significantly influence the percent of irrigation applied. Post hoc Tukey HSD pairwise comparisons further elucidate the significance, showing distinct differences between correlation range categories. For instance, the comparison between the negative range (-0.25 to -0.05) and the others demonstrates a significant impact, with a mean difference of approximately -10.31 ($p < 0.05$), signifying lower irrigation applies as correlation becomes more negative. This significant observation is also clearly denoted in the Figure 3-9 box plot, by applying more irrigation percentage the negative correlation range is reduced.

The first range, from -0.77 to -0.74, displays an extremely negative correlation range with a minimum irrigation percentage of 0.3 and a maximum of 10.3. The interquartile range (IQR), measuring the range within which the middle 50% of values fall, spans from 1.2 to 5.6, with a median correlation of 3.1. This suggests a set of areas where the relationship between potential evapotranspiration (PET-P) and soil moisture is strong, alternatively denotes there was a very small percent of irrigation was applied to those regions (Figure 3-9).

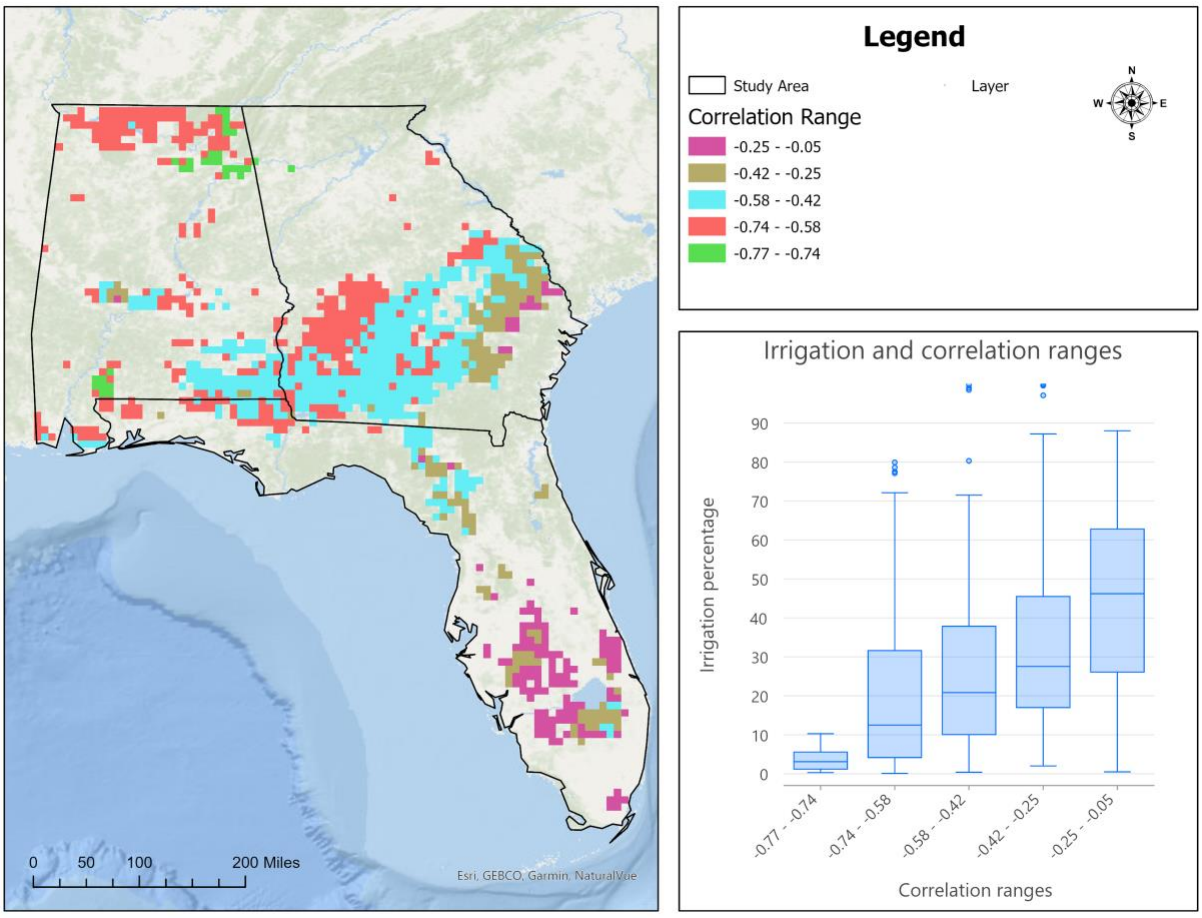

Figure 3-9: Relation between correlation ranges and percentage of irrigation applied.

Moving to the second range, from -0.74 to -0.58, the irrigation percentage extends from a minimum of 0.1 to a maximum of 79.9, with a substantial IQR spanning from 4.18 to 31.63. The median value within this range is 12.5, indicative of areas where the relationship between PET-P and soil moisture and irrigation exhibits a pronounced negative relation. The higher variability in

values and few outliers suggests that these areas may be subjected to more diverse irrigation practices.

The third range, spanning from -0.58 to -0.42, features a minimum irrigation percentage of 0.4 and a maximum of 99.8. The IQR ranges from 10.08 to 37.85, with a median irrigation percent of 20.85. These values signify regions with relatively strong negative correlations between PET-P and soil moisture, suggesting that irrigation practices play a substantial role in shaping soil moisture dynamics (Figure 3-9).

In the fourth range, -0.42 to -0.25, the percentage of irrigation exhibits a minimum of 2 and a maximum of 99.8, with an IQR spanning 17 to 45.5. The median value is 27.55, indicating a strong relation and trend with increasing irrigation and the negative correlation between PET-P and soil moisture-reducing trend. These areas are characterized by robust irrigation practices, leading to significant fluctuations in soil moisture.

The fifth and final range, -0.25 to -0.05, features an irrigation percentage starting from 0.5 and a maximum of 88. The IQR spans from 26.1 to 62.8, with a median correlation of 46.2. In this category, the strong relation shows that increasing applied irrigation percent in higher ranges can significantly reduce the correlation ranges of the PET-P and soil moisture.

3.4 Discussion

A distinct trend in potential evaporation (PET), precipitation, and soil moisture over the study period was observed in this study. Several factors may contribute to this observation including variations in crop types, soil characteristics, temperature patterns, and local practices can lead to uneven soil moisture patterns (Rabiei et al., 2021). Additionally, weather patterns and seasonal changes can influence the timing and effectiveness of irrigation, impacting the correlation. These findings underscore the dynamic nature of environmental conditions through observing PET and their impact on water resources. The trend in precipitation patterns throughout the study period indicates variations play a crucial role in the atmospheric water balance and finding extreme water-scarce months (McDermid et al., 2023). It's crucial to consider the interplay of these local

variables, which can override the general assumption of a straightforward inverse relationship between PET-P and soil moisture in high-density cropland regions (L. Zhang et al., 2022).

This study can provide insights into the availability of natural water resources and artificial irrigation applied for various regions (Zaussinger et al., 2019). Referring to this, the analysis of the atmospheric water balance revealed important information about the interplay between potential evaporation and precipitation (C. Zhang & Long, 2021). This balance is a critical component of the hydrological cycle and influences soil moisture conditions. The findings highlight the significance of atmospheric water demand and availability in shaping soil moisture dynamics. Additionally, investigation on root zone soil moisture trends unveiled essential insights into soil moisture dynamics over time which is invaluable for assessing the water requirements of crops and vegetation (Babaeian et al., 2021; Kisekka et al., 2022). The observed trends point out the influence of climate and land use on soil moisture levels (Jimma et al., 2023). As such, they provide a foundation for informed decision-making in irrigation water needed in agriculture and ecosystem management, allowing for more precise irrigation practices and resource allocation.

The correlation analysis between potential evapotranspiration (PET-P) and root zone soil moisture revealed a clear negative correlation, suggesting that areas with increased atmospheric water demand, as indicated by lower PET-P values, tend to exhibit higher soil moisture (Condon et al., 2020). The higher negative correlation could be attributed to various factors, including differences in crop distribution, irrigation practices, and the interaction between soil types and vegetation cover (IPCC, 2007; Zaussinger et al., 2019). The choice of crops, with varying water requirements and growth cycles, can significantly influence the soil moisture-ET correlation in such regions (El Sharif et al., 2015). Moreover, the timing and efficiency of irrigation activities, which are subject to climate and water availability, can contribute to the observed correlation patterns. This finding is particularly relevant for precision agriculture, as it indicates areas with water requirements for optimal crop growth (Babaeian et al., 2021; Kisekka et al., 2022). Understanding the correlation patterns allows for more efficient water resource allocation. Furthermore, the correlated map provides crucial insights into the connection between correlation patterns and applied irrigation. The post hoc Tukey analysis indicated significant

differences in irrigation water demand among different correlation ranges (Thompson et al., 2017). As the correlation between PET-P and soil moisture becomes more negative, irrigation water tends to be less applied, but demand is high. This information is of paramount importance for crop cultivation and increasing yield by synthesizing root zone-level plant water needed (Meier et al., 2018; Shahzaman et al., 2021). A pathway for planning and resource allocation, as it offers guidance on areas where irrigation practices can be optimized.

The observed correlations underscore the importance of considering local and regional factors when analyzing the relationship between potential evapotranspiration, precipitation, and soil moisture in high cropland density areas. These factors, including crop diversity, irrigation methods, and environmental conditions, can lead to variations in correlation patterns that challenge conventional assumptions and highlight the need for site-specific investigations. For the higher range detection or outliers shown in different correlation ranges in the box plots, possibly there are external factors, like less updated datasets or insignificant land use and land cover mapping over time (Paolini et al., 2022). It is discernible, built up, or similar land use classes where ground variable classes might not show changes very frequently and might show similar relations like the croplands. This is because, if one value of a correlation does not change, it might show a near-zero relation. Likewise, other open areas, like waterbodies or everglade regions where soil moisture is always surplus, near zero correlation would be common (Hargrove & Hoffman, 2004; Kirui et al., 2013). Apart from this, forest covers where artificial irrigation is not common should show a very high negative correlation among those data. These are very important underlying issues that should be considered when estimating water demands in agricultural lands.

The observed correlation values in low cropland density areas, particularly those resembling forest cover where artificial irrigation is not common, provide valuable insights into the impact of land use and land cover types on soil moisture dynamics (Xie et al., 2019; Xu et al., 2019). These regions exhibit near-zero correlations, suggesting that artificial irrigation practices are not significantly influencing soil moisture (Kwon et al., 2022; Pervez & Brown, 2010). This phenomenon extends to areas characterized by consistent land use and land cover types, were soil moisture experiences minimal fluctuations. Such low correlations are not exclusive to

croplands; they can also manifest in built-up urban areas and open regions like water bodies or everglade areas, where soil moisture tends to remain relatively stable or consistently high (Xie et al., 2019). It is crucial to recognize that these underlying factors have substantial implications for the interpretation of the PET-P and soil moisture relationship. In this context, the process of masking these low cropland density areas with the cropland layer should be approached with consideration of the specific study objectives, strongly suggested for similar studies. This is especially true in regions where agricultural practices are not the primary driver of soil moisture dynamics (Amdihun, 2006).

Finally, there might be a few uncertainties in these datasets such as seasonal anomalies. In this case, removing seasonal anomalies is the best option (Kumar et al., 2019, 2020), which was also used in this study, to minimize outliers or any extreme values. The CHIRPS provides valuable precipitation data, but there may be inherent uncertainties associated with satellite-based precipitation estimates that can arise from factors such as cloud cover and calibration errors. The datasets used were collected from reanalysis and ground-based observation that accumulated in ERA5 Land, Precipitation in CHIRPS, and soil moisture in SMAP under the GEE platform. It is essential to note that every dataset used here has been widely validated and used in numerous scientific applications, demonstrating its utility despite these uncertainties.

Conclusion

The study highlights the interplay between atmospheric, climatic, and environmental variables for estimating irrigation water demands in the southern United States. Instead of field-based data collection procedures, which are labor intensive, cost, and time-consuming, and prone to human error, remotely sensed datasets were utilized to find varying water demand locations. Rather than designing specific models for each state, implying the same resolution or ground collected data, this study utilized freely available data sources that were in different resolutions for the model buildup that can be replicated by the science community for sustainable decision-making. The classification of correlation ranges and relation identification between soil moisture and irrigation water demands in areas with varying land use characteristics allows for the identification of potential irrigation zones based on the strength and variability of the correlation,

providing a nuanced understanding of the factors shaping these correlation patterns. The overall findings from this study shed light on the intricate dynamics of atmospheric water balance, soil moisture, and their impact on irrigation practices. The results emphasize the significance of understanding these relationships for efficient resource management, sustainable precision agriculture, and ecological conservation. This knowledge can empower decision-makers to make informed choices for adapting to changing environmental conditions and ensuring water sustainability in various regions.

References

Brandt, J.T., Caldwell, R.R., Haynes, J.V., Painter, J.A., and Read, A.L., 2021, Verified Irrigated Agricultural Lands for the United States, 2002–17: U.S. Geological Survey data release, https://doi.org/10.5066/P9NAWU1U

Abbaszadeh, P., Moradkhani, H., & Zhan, X. (2019). Downscaling SMAP Radiometer Soil Moisture Over the CONUS Using an Ensemble Learning Method. *Water Resources Research*, *55*(1), 324–344. https://doi.org/10.1029/2018WR023354

Ajani, J. I., Keith, H., Blakers, M., Mackey, B. G., & King, H. P. (2013). Comprehensive carbon stock and flow accounting: A national framework to support climate change mitigation policy. *Ecological Economics*, *89*, 61–72. https://doi.org/10.1016/j.ecolecon.2013.01.010

Ambika, A. K., Wardlow, B., & Mishra, V. (2016). Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015. *Scientific Data*, *3*, 1–14. https://doi.org/10.1038/sdata.2016.118

Amdihun, A. (2006). *GIS and Remote Sensing Integrated Environmental Impact Assessment of Irrigation Project in Finchaa Valley Area (Issue July)*. 400–417.

Anurag, & Kumar, A. (2016). Remote Sensing and GIS Applications in Agriculture. *Climate Resilient Farming (Training Manual)*, 37–48.

Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V. K., & Tuller, M. (2021). Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. *Remote Sensing of Environment*, *260*(January), 112434. https://doi.org/10.1016/j.rse.2021.112434

Battelle, N. (2023a). *Oak Ridge NEON / ORNL*. https://www.neonscience.org/field-sites/ornl

Battelle, N. (2023b). *Ordway-Swisher Biological Station NEON / OSBS*. https://www.neonscience.org/field-sites/osbs

Battelle, N. (2023c). *Talladega National Forest NEON / TALL*. https://www.neonscience.org/field-sites/tall

Brown, J. H., West, G. B., & Enquist, B. J. (2005). Yes, West, Brown and Enquist's model of allometric scaling is both mathematically correct and biologically relevant. *Functional Ecology*, *19*(4), 735–738. https://doi.org/10.1111/j.1365-2435.2005.01022.x

Brown, S., Narine, L. L., & Gilbert, J. (2022). Using Airborne Lidar, Multispectral Imagery, and Field Inventory Data to Estimate Basal Area, Volume, and Aboveground Biomass in Heterogeneous Mixed Species Forests: A Case Study in Southern Alabama. *Remote Sensing*, *14*(11), 2708. https://doi.org/10.3390/rs14112708

Bwambale, E., Naangmenyele, Z., Iradukunda, P., Agboka, K. M., Houessou-Dossou, E. A. Y., Akansake, D. A., Bisa, M. E., Hamadou, A. A. H., Hakizayezu, J., Onofua, O. E., & Chikabvumbwa, S. R. (2022). Towards precision irrigation management: A review of GIS, remote sensing and emerging technologies. *Cogent Engineering*, *9*(1). https://doi.org/10.1080/23311916.2022.2100573

Chance, E. W., Cobourn, K. M., Thomas, V. A., Dawson, B. C., & Flores, A. N. (2017). Identifying irrigated areas in the Snake River Plain, Idaho: Evaluating performance across composting algorithms, spectral indices, and sensors. *Remote Sensing*, *9*(6), 1–23. https://doi.org/10.3390/rs9060546

Condon, L. E., Atchley, A. L., & Maxwell, R. M. (2020). Evapotranspiration depletes groundwater under warming over the contiguous United States. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14688-0

Condon, L. E., & Maxwell, R. M. (2019). Simulating the sensitivity of evapotranspiration and streamflow to large-scale groundwater depletion. *Science Advances*, *5*(6). https://doi.org/10.1126/sciadv.aav4574

Coops, N. C., Tompalski, P., Goodbody, T. R. H., Queinnec, M., Luther, J. E., Bolton, D. K., White, J. C., Wulder, M. A., van Lier, O. R., & Hermosilla, T. (2021). Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sensing of Environment*, *260*(May), 112477. https://doi.org/10.1016/j.rse.2021.112477

County, M., County, M., Co, B., & County, M. (1966). *Alabama ' s Climate*. 29–32.

Dari, J., Quintana-Seguí, P., Escorihuela, M. J., Stefan, V., Brocca, L., & Morbidelli, R. (2021). Detecting and mapping irrigated areas in a Mediterranean environment by using remote sensing soil moisture and a land surface model. *Journal of Hydrology*, *596*(December 2020). https://doi.org/10.1016/j.jhydrol.2021.126129

Dash, S. K., & Sinha, R. (2022). A Comprehensive Evaluation of Gridded L-, C-, and X-Band Microwave Soil Moisture Product over the CZO in the Central Ganga Plains, India. *Remote Sensing*, *14*(7). https://doi.org/10.3390/rs14071629

Dower, A. L. G. (1965). *Mr. D. St. A. Butcher - Obituary*. *12*(1), 3.

Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Silva, C. A., Armston, J., Luthcke, S. B., Hofton, M., Kellner, J. R., & Dubayah, R. (2020). Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sensing of Environment*, *242*(November 2019), 111779. https://doi.org/10.1016/j.rse.2020.111779

El Sharif, H., Wang, J., & Georgakakos, A. P. (2015). Modeling regional crop yield and irrigation demand using SMAP type of soil moisture data. *Journal of Hydrometeorology*, *16*(2), 904–916. https://doi.org/10.1175/JHM-D-14-0034.1

Fassoni-Andrade, A. C., Fleischmann, A. S., Papa, F., Paiva, R. C. D. de, Wongchuig, S., Melack, J. M., Moreira, A. A., Paris, A., Ruhoff, A., Barbosa, C., Maciel, D. A., Novo, E., Durand, F., Frappart, F., Aires, F., Abrahão, G. M., Ferreira-Ferreira, J., Espinoza, J. C., Laipelt, L., … Pellet, V. (2021). Amazon Hydrology From Space: Scientific Advances and Future Challenges. *Reviews of Geophysics*, *59*(4), 1–97. https://doi.org/10.1029/2020RG000728

Felfelani, F., Pokhrel, Y., Guan, K., & Lawrence, D. M. (2018). Utilizing SMAP Soil Moisture Data to Constrain Irrigation in the Community Land Model. *Geophysical Research Letters*, *45*(23), 12,892-12,902. https://doi.org/10.1029/2018GL080870

Fernandez-Palomino, C. A., Hattermann, F. F., Krysanova, V., Lobanova, A., Vega-Jácome, F., Lavado, W., Santini, W., Aybar, C., & Bronstert, A. (2022). A Novel High-Resolution Gridded Precipitation Dataset for Peruvian and Ecuadorian Watersheds: Development and Hydrological Evaluation. *Journal of Hydrometeorology*, *23*(3), 309–336. https://doi.org/10.1175/JHM-D-20-0285.1

Freeman, E. A., Frescino, T. S., & Moisen, G. G. (2016). *ModelMap : an R Package for Model Creation and Map Production*. 1–43.

García, M., Saatchi, S., Ustin, S., & Balzter, H. (2018). Modelling forest canopy height by integrating airborne LiDAR samples with satellite Radar and multispectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, *66*(November 2017), 159–173. https://doi.org/10.1016/j.jag.2017.11.017

Hao, Z., Zhao, H., Zhang, C., Zhou, H., Zhao, H., & Wang, H. (2019). Correlation analysis

between groundwater decline trend and human-induced factors in Bashang region. *Water (Switzerland)*, *11*(3). https://doi.org/10.3390/w11030473
Hargrove, W. W., & Hoffman, F. M. (2004). Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental Management*, *34*(1 SUPPL.), 39–60. https://doi.org/10.1007/s00267-003-1084-0
Hermsen, E. J. (2022). Climate of the Southeastern United States. *Earth@Home*, 1–45. https://earthathome.org/hoe/se/climate/
IPCC. (2007). Mitigation of climate change: Contribution of working group III to the fourth assessment report of the Intergovernmental Panel on Climate Change. In *Intergovernmental Panel on Climate Change*.
Isenburg, M. (2013). *LAStools | rapidlasso GmbH*. 9–10. http://rapidlasso.com/lastools/
Ishtiaque, A., Masrur, A., Rabby, Y. W., Jerin, T., & Dewan, A. (2020). Remote sensing-based research for monitoring progress towards SDG 15 in Bangladesh: A review. *Remote Sensing*, *12*(4). https://doi.org/10.3390/rs12040691
Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (2003). National-scale biomass estimators for United States tree species. *Forest Science*, *49*(1), 12–35.
Jimma, T. B., Demissie, T., Diro, G. T., Ture, K., Terefe, T., & Solomon, D. (2023). Spatiotemporal variability of soil moisture over Ethiopia and its teleconnections with remote and local drivers. *Theoretical and Applied Climatology*, *151*(3–4), 1911–1929. https://doi.org/10.1007/s00704-022-04335-7
Khati, U., Lavalle, M., Shiroma, G. H. X., Meyer, V., & Chapman, B. (2020). Assessment of forest biomass estimation from dry and wet SAR acquisitions collected during the 2019 UAVSAR AM-PM campaign in southeastern United States. *Remote Sensing*, *12*(20), 1–18. https://doi.org/10.3390/rs12203397
Kim, H., Lee, S., Cosh, M. H., Lakshmi, V., Kwon, Y., & McCarty, G. W. (2021). Assessment and Combination of SMAP and Sentinel-1A/B-Derived Soil Moisture Estimates with Land Surface Model Outputs in the Mid-Atlantic Coastal Plain, USA. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(2), 991–1011. https://doi.org/10.1109/TGRS.2020.2991665
Kirui, K. B., Kairo, J. G., Bosire, J., Viergever, K. M., Rudra, S., Huxham, M., & Briers, R. A. (2013). Ocean & Coastal Management Mapping of mangrove forest land cover change along the Kenya coastline using Landsat imagery. *Ocean and Coastal Management*, *83*, 19–24. https://doi.org/10.1016/j.ocecoaman.2011.12.004
Kisekka, I., Peddinti, S. R., Kustas, W. P., McElrone, A. J., Bambach-Ortiz, N., McKee, L., & Bastiaanssen, W. (2022). Spatial–temporal modeling of root zone soil moisture dynamics in a vineyard using machine learning and remote sensing. *Irrigation Science*. https://doi.org/10.1007/s00271-022-00775-1
Kolassa, J., Reichle, R. H., Liu, Q., Cosh, M., Bosch, D. D., Todd, G., Colliander, A., Collins, C. H., Jackson, T. J., & Stan, J. (2020). *NASA Public Access*. *9*(11), 1–32. https://doi.org/10.3390/rs9111179.Data
Kolluru, V., Kolluru, S., & Konkathi, P. (2020). Evaluation and integration of reanalysis rainfall products under contrasting climatic conditions in India. *Atmospheric Research*, *246*(July), 105121. https://doi.org/10.1016/j.atmosres.2020.105121
Kumar, S., Newman, M., Lawrence, D. M., Lo, M. H., Akula, S., Lan, C. W., Livneh, B., & Lombardozzi, D. (2020). The GLACE-Hydrology Experiment: Effects of Land-Atmosphere Coupling on Soil Moisture Variability and Predictability. *Journal of Climate*, *33*(15), 6511–

6529. https://doi.org/10.1175/JCLI-D-19-0598.1
Kumar, S., Newman, M., Wang, Y., & Livneh, B. (2019). Potential reemergence of seasonal soil moisture anomalies in North America. *Journal of Climate*, *32*(10), 2707–2734. https://doi.org/10.1175/JCLI-D-18-0540.1
Kwon, Y., Kumar, S. V., Navari, M., Mocko, D. M., Kemp, E. M., Wegiel, J. W., Geiger, J. V., & Bindlish, R. (2022). Irrigation characterization improved by the direct use of SMAP soil moisture anomalies within a data assimilation system. *Environmental Research Letters*, *17*(8). https://doi.org/10.1088/1748-9326/ac7f49
Lawston, P. M., Santanello, J. A., & Kumar, S. V. (2017). Irrigation Signals Detected From SMAP Soil Moisture Retrievals. *Geophysical Research Letters*, *44*(23), 11,860-11,867. https://doi.org/10.1002/2017GL075733
Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., & Moran, E. (2016). A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, *9*(1), 63–105. https://doi.org/10.1080/17538947.2014.990526
McDermid, S., Nocco, M., Lawston-Parker, P., Keune, J., Pokhrel, Y., Jain, M., Jägermeyr, J., Brocca, L., Massari, C., Jones, A. D., Vahmani, P., Thiery, W., Yao, Y., Bell, A., Chen, L., Dorigo, W., Hanasaki, N., Jasechko, S., Lo, M. H., … Yokohata, T. (2023). Irrigation in the Earth system. *Nature Reviews Earth and Environment*. https://doi.org/10.1038/s43017-023-00438-5
McGaughey, R. J. (2016). *FUSION – Software for Lidar Data Analysis and Visualization*. *January*, 1–4.
Meier, J., Zabel, F., & Mauser, W. (2018). A global approach to estimate irrigated areas - A comparison between different data and statistics. *Hydrology and Earth System Sciences*, *22*(2), 1119–1133. https://doi.org/10.5194/hess-22-1119-2018
Nandy, S., Srinet, R., & Padalia, H. (2021). Mapping Forest Height and Aboveground Biomass by Integrating ICESat-2, Sentinel-1 and Sentinel-2 Data Using Random Forest Algorithm in Northwest Himalayan Foothills of India. *Geophysical Research Letters*, *48*(14), 1–10. https://doi.org/10.1029/2021GL093799
National Weather Service, U. (2017). *NWS Southern Region Headquarters*. 35. http://www.srh.noaa.gov/rfcshare/ffg.php
Nie, S., Wang, C., Xi, X., Luo, S., Li, G., Tian, J., & Wang, H. (2018). Estimating the vegetation canopy height using micro-pulse photon-counting LiDAR data. *Optics Express*, *26*(10), A520. https://doi.org/10.1364/oe.26.00a520
Ougahi, J. H., & Mahmood, S. A. (2022). Evaluation of satellite-based and reanalysis precipitation datasets by hydrologic simulation in the Chenab river basin. *Journal of Water and Climate Change*, *13*(3), 1563–1582. https://doi.org/10.2166/wcc.2022.410
Paolini, G., Escorihuela, M. J., Merlin, O., Sans, M. P., & Bellvert, J. (2022). Classification of Different Irrigation Systems at Field Scale Using Time-Series of Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*(Ml), 10055–10072. https://doi.org/10.1109/JSTARS.2022.3222884
Pervez, M. S., & Brown, J. F. (2010). Mapping irrigated lands at 250-m scale by merging MODIS data and National Agricultural Statistics. *Remote Sensing*, *2*(10), 2388–2412. https://doi.org/10.3390/rs2102388
Platonov, A. E., & Vincent, B. (n.d.). *5 - Remote Sensing Applied To Irrigation*. *from 1998*, 63–78.
Popescu, S. C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass*

and Bioenergy, *31*(9), 646–655. https://doi.org/10.1016/j.biombioe.2007.06.022
Qi, W., Saarela, S., Armston, J., Ståhl, G., & Dubayah, R. (2019). Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. *Remote Sensing of Environment*, *232*(June), 111283. https://doi.org/10.1016/j.rse.2019.111283
Rabiei, S., Jalilvand, E., & Tajrishy, M. (2021). A method to estimate surface soil moisture and map the irrigated cropland area using sentinel-1 and sentinel-2 data. *Sustainability (Switzerland)*, *13*(20). https://doi.org/10.3390/su132011355
Rahimi, S., Gholami Sefidkouhi, M. A., Raeini-Sarjaz, M., & Valipour, M. (2015). Estimation of actual evapotranspiration by using MODIS images (a case study: Tajan catchment). *Archives of Agronomy and Soil Science*, *61*(5), 695–709. https://doi.org/10.1080/03650340.2014.944904
Rahman, I. M. M., Islam, M. M., Hossain, M. M., Hossain, M. S., Begum, Z. A., Chowdhury, D. A., Chakraborty, M. K., Rahman, M. A., Nazimuddin, M., & Hasegawa, H. (2011). Stagnant surface water bodies (SSWBs) as an alternative water resource for the Chittagong metropolitan area of Bangladesh: Physicochemical characterization in terms of water quality indices. *Environmental Monitoring and Assessment*, *173*(1–4), 669–684. https://doi.org/10.1007/s10661-010-1414-7
Safaeian, S., Falahatkar, S., & Tourian, M. J. (2023). Satellite observation of atmospheric CO2 and water storage change over Iran. *Scientific Reports*, *13*(1), 1–15. https://doi.org/10.1038/s41598-023-28961-x
Saul, P. (2018). Key points. *Practitioner*, *262*(1812), 24.
Schmidt, H., & Karnieli, A. (2001). Sensitivity of vegetation indices to substrate brightness in hyper-arid. *International Journal of Remote Sensing*, *22*(17), 3503–3520.
Scholl, V. M., Cattau, M. E., Joseph, M. B., & Balch, J. K. (2020). Integrating national ecological observatory network (NEON) airborne remote sensing and in-situ data for optimal tree species classification. *Remote Sensing*, *12*(9), 1–19. https://doi.org/10.3390/RS12091414
Shahzaman, M., Zhu, W., Ullah, I., Mustafa, F., Bilal, M., Ishfaq, S., Nisar, S., Arshad, M., Iqbal, R., & Aslam, R. W. (2021). Comparison of multi-year reanalysis, models, and satellite remote sensing products for agricultural drought monitoring over south asian countries. *Remote Sensing*, *13*(16). https://doi.org/10.3390/rs13163294
Shaw, R. H. (2018). Climate of the United States. *Handbook of Soils and Climate in Agriculture*, 1–102. https://doi.org/10.1201/9781351073073
Silva, C. A., Duncanson, L., Hancock, S., Neuenschwander, A., Thomas, N., Hofton, M., Fatoyinbo, L., Simard, M., Marshak, C. Z., Armston, J., Lutchke, S., & Dubayah, R. (2021). Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. *Remote Sensing of Environment*, *253*(December 2020), 112234. https://doi.org/10.1016/j.rse.2020.112234
Thomas, B. F., & Famiglietti, J. S. (2019). Identifying Climate-Induced Groundwater Depletion in GRACE Observations. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-40155-y
Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the Variance In flation Factor and Other Multicollinearity Diagnostics from Typical Regression Results. *Basic and Applied Social Psychology*, *39*(2), 81–90. https://doi.org/10.1080/01973533.2016.1277529

United States Geological Survey. (2019). Moderate Resolution Imaging Spectroradiometer (MODIS) Irrigated Agriculture Dataset for the United States (MIrAD-US). *US Irrigation and Early Warning and Environmental Monitoring Program*, 56–58. https://doi.org/10.5066/P9NA3EO8

Wakigari, S. A., & Leconte, R. (2022). Enhancing Spatial Resolution of SMAP Soil Moisture Products through Spatial Downscaling over a Large Watershed: A Case Study for the Susquehanna River Basin in the Northeastern United States. *Remote Sensing*, *14*(3), 776. https://doi.org/10.3390/rs14030776

Wang, C., Elmore, A. J., Numata, I., Cochrane, M. A., Shaogang, L., Huang, J., Zhao, Y., & Li, Y. (2022). Factors affecting relative height and ground elevation estimations of GEDI among forest types across the conterminous USA. *GIScience and Remote Sensing*, *59*(1), 975–999. https://doi.org/10.1080/15481603.2022.2085354

Wickham, J., Stehman, S. V., Sorenson, D. G., Gass, L., & Dewitz, J. A. (2021). Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. *Remote Sensing of Environment*, *257*(February), 112357. https://doi.org/10.1016/j.rse.2021.112357

Williams, M. D., Hawley, C. M. S., Madden, M., & Shepherd, J. M. (2017). Mapping the spatio-temporal evolution of irrigation in the coastal plain of Georgia, USA. *Photogrammetric Engineering and Remote Sensing*, *83*(1), 57–67. https://doi.org/10.14358/PERS.83.1.57

Wind, R., & Penman-monteith, F. A. O. (2003). *Chapter 3 - Meteorological data Chapter 3 - Meteorological data Meteorological factors determining ET Chapter 3 - Meteorological data Ò³Âë £¬*. 1–30.

Wu, X., & Zhao, N. (2023). Evaluation and Comparison of Six High-Resolution Daily Precipitation Products in Mainland China. *Remote Sensing*, *15*(1), 1–24. https://doi.org/10.3390/rs15010223

Xie, Y., Lark, T. J., Brown, J. F., & Gibbs, H. K. (2019). Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *155*(July), 136–149. https://doi.org/10.1016/j.isprsjprs.2019.07.005

Xu, T., Deines, J. M., Kendall, A. D., Basso, B., & Hyndman, D. W. (2019). Addressing challenges for mapping irrigated fields in subhumid temperate regions by integrating remote sensing and hydroclimatic data. *Remote Sensing*, *11*(3). https://doi.org/10.3390/rs11030370

Zappa, L., Schlaffer, S., Brocca, L., Vreugdenhil, M., Nendel, C., & Dorigo, W. (2022). How accurately can we retrieve irrigation timing and water amounts from (satellite) soil moisture? *International Journal of Applied Earth Observation and Geoinformation*, *113*(August), 102979. https://doi.org/10.1016/j.jag.2022.102979

Zaussinger, F., Dorigo, W., Gruber, A., Tarpanelli, A., Filippucci, P., & Brocca, L. (2019). Estimating irrigation water use over the contiguous United States by combining satellite and reanalysis soil moisture data. *Hydrology and Earth System Sciences*, *23*(2), 897–923. https://doi.org/10.5194/hess-23-897-2019

Zeng, P., Zhang, W., Li, Y., Shi, J., & Wang, Z. (2022). Forest Total and Component Above-Ground Biomass (AGB) Estimation through C-and L-band Polarimetric SAR Data. *Forests*, *13*(3), 1–19. https://doi.org/10.3390/f13030442

Zhang, C., & Long, D. (2021). Estimating Spatially Explicit Irrigation Water Use Based on Remotely Sensed Evapotranspiration and Modeled Root Zone Soil Moisture. *Water Resources Research*, *57*(12), 1–27. https://doi.org/10.1029/2021WR031382

Zhang, L., Zhang, K., Zhu, X., Chen, H., & Wang, W. (2022). Integrating remote sensing, irrigation suitability and statistical data for irrigated cropland mapping over mainland China. *Journal of Hydrology*, *613*(PA), 128413. https://doi.org/10.1016/j.jhydrol.2022.128413

Zhang, W., Zhao, L., Yu, X., Zhang, L., & Wang, N. (2020). Estimation of Groundwater Evapotranspiration Using Diurnal Groundwater Level Fluctuations under Three Vegetation Covers at the Hinterland of the Badain Jaran Desert. *Advances in Meteorology*, *2020*. https://doi.org/10.1155/2020/8478140

Zhang, Y., Liang, S., & Yang, L. (2019). A Review of Regional and Global Gridded Forest. *Remote Sensing*, *11*, 2744.

Zhao, Q., Yu, S., Zhao, F., Tian, L., & Zhao, Z. (2019). Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *Forest Ecology and Management*, *434*(December 2018), 224–234. https://doi.org/10.1016/j.foreco.2018.12.019

Zhao, W., Sánchez, N., Lu, H., & Li, A. (2018). A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *Journal of Hydrology*, *563*(April), 1009–1024. https://doi.org/10.1016/j.jhydrol.2018.06.081

Zhu, L., & Zhu, A. X. (2021). Extraction of irrigation signals by using smap soil moisture data. *Remote Sensing*, *13*(11). https://doi.org/10.3390/rs13112142

Chapter 4: Conclusion and Recommendations

4.1 Conclusion

This collective research efforts, as presented in both Chapter 2 and Chapter 3, have not only expanded scientific knowledge of environmental dynamics but have also unveiled key insights and implications for future investigations in this domain. These studies, while addressing different facets of environmental science, are interlinked in their pursuit of understanding, and harnessing remote sensing data to enhance our comprehension of critical environmental processes.

The first part of this study presents valuable insights into the utilization of NEON airborne lidar point cloud and VST data for the development of high-resolution spatially continuous aboveground biomass (AGB) maps within NEON sites. The adoption of consistent data protocols among 83 NEON sites across the United States demonstrates the potential for adapting this methodology to other ecological locations, offering a standardized approach for AGB mapping. Additionally, NEON updates airborne lidar point cloud data every year, opening a window for time series forest health and resources analysis. Furthermore, considering the high resolution and accuracy of this dataset, it should be useful for validating coarser resolution AGB products like GEDI L4B AGB. However, the research also underscores the limitations of NEON data, particularly in the context of southern heterogeneous mixed forest sites characterized by dense tree canopy cover and significant terrain variations. While both Random Forest (RF) and Multiple Linear Regression (MLR) models were employed, it was observed that the MLR model outperformed the RF model in all three study sites, as evidenced by higher R^2 values and lower root mean square error (RMSE). Nonetheless, the study acknowledges that the MLR model may struggle to capture the complexity of data in sites with steep terrain and mixed forest types, while the RF model's propensity to overcomplicate the estimation process could result in underfitting and reduced accuracy. Nevertheless, the findings of this research hold promise for advancing the producibility of high-resolution spatially continuous AGB mapping within NEON sites and pave

the way for future research that seeks to integrate these reference datasets with ICESat-2 data for broader-scale AGB mapping beyond NEON sites.

Moving to the second study, it illuminates the intricate relationships among atmospheric, climatic, and environmental variables in the estimation of irrigation water demands in the southern United States. Departing from labor-intensive, time-consuming, and error-prone field-based data collection methods, this research harnesses remotely sensed datasets to pinpoint locations with varying water demand, offering a more efficient and accurate alternative. By not limiting the modeling process to specific states and resolutions but instead employing openly accessible data sources with varying resolutions, this study establishes a replicable model construction framework that can be adopted by the scientific community. The classification of correlation ranges and identification of relationships between soil moisture and irrigation water demands in diverse land use areas bring forth the potential to discern irrigation zones based on the strength and variability of these correlations. This nuanced understanding of the factors influencing these patterns carries significant implications for the management of resources, the promotion of sustainable precision agriculture, and the preservation of ecological systems. Consequently, this knowledge equips decision-makers with the information necessary to adapt to evolving environmental conditions and to ensure the sustainable management of water resources in diverse regions. These findings collectively underscore the importance of a holistic approach to understanding the complex interplay of atmospheric water balance, soil moisture, and their impact on irrigation practices in the southern United States.

4.2 Recommendations

4.2.1 Objective 1: In spatially continuous AGB estimation (Chapter Two)

Diversify Field Sampling:

To improve AGB model accuracy, it is recommended to increase the distribution and number of field observations per site. This includes considering spatial distribution, subplots accuracy, and additional field visits to capture tree composition, especially in mixed forestry sites.

Consider Additional Variables:

The study suggests considering further variables, including tree composition map, terrain, and climatic variables, to test for enhancing model performance. Incorporating variables like NDVI, MSAVI-2 vegetation indices, and NLCD land and tree cover datasets can help improve accuracy by considering holistically. Recognize the significance of field characteristics, forest types (homogeneity, mixed forest type, and tree species composition), canopy coverage, and heterogeneity of tree species in modeling AGB. Tailor modeling approaches accordingly.

Explore Tree-Based with Area based Approaches:

Consider exploring tree-based approaches (TBA) for AGB estimation using NEON data. This approach may require precise geographical location tagging for each tree within the NEON dataset. Utilize NEON's airborne lidar point cloud and VST data for developing new models and spatially comprehensive satellite data to map AGB over broader landscapes where NEON data is not available.

Continuous Data Improvement:

As NEON datasets and field sites have inherent limitations, continuous data improvement efforts, such as increasing field sampling points and quality control, are vital for enhancing AGB modeling accuracy.

Standardize Sampling Dimensions:

Standardize the sampling dimensions in forest sites, especially in mixed forestry sites, to capture tree composition and heterogeneity effectively. Consider the use of variable grid sizes for different site characteristics.

4.2.2 Objective 2: Identifying varying irrigation areas (Chapter three)

Monitor Environmental Trends:

Continue monitoring trends in potential evaporation (PET), precipitation, and soil moisture to gain insights into the dynamic nature of environmental conditions and their impact on water

resources. Utilize the negative correlation between potential evapotranspiration (PET-P) and root zone soil moisture to identify areas with higher water demand.

Further Examine Low Correlations:

Recognize that low correlations in areas with consistent land use and land cover types, like forests or everglade regions, may not be influenced by artificial irrigation although range may look like similar. This understanding is essential when interpreting PET-P and soil moisture relationships.

Consider Masking Low Cropland Density Areas:

In regions where agricultural practices are not the primary driver of soil moisture dynamics, consider masking low cropland density areas with the cropland layer to refine the analysis based on specific study objectives.

Address Data Uncertainties and ground validations:

Acknowledge potential uncertainties in datasets, such as seasonal anomalies, and employ data cleaning techniques like removing outliers or extreme values to enhance data quality and accuracy. Although datasets used in this study are widely validated and have a strong history of utility despite inherent uncertainties, such as satellite-based precipitation estimates. Before implementing result outcomes, it is needed further ground verifications.

These recommendations aim to enhance the understanding and applicability of the study outcomes in both AGB mapping and water resource management, emphasizing the need for continuous data improvement and precision in data analysis and interpretation.