REMOTE SENSING OF VEGETATION SPECIES DIVERSITY: THE UTILITY OF INTEGRATED AIRBORNE HYPERSPECTRAL AND LIDAR DATA

by

KEITH STUART KRAUSE

B.A., Rochester Institute of Technology, 1999

M.S., University of Arizona, 2001

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

Of the requirement for the degree of

Doctor of Philosophy

Department of Aerospace Engineering Sciences

2015

This thesis entitled:

Remote Sensing of Vegetation Species Diversity: The Utility of Integrated Airborne Hyperspectral and Lidar Data

written by Keith Stuart Krause

has been approved for the Department of Aerospace Engineering Sciences

_____

William Emery

_____

Jeffrey Thayer

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Krause, Keith Stuart (Ph.D., Aerospace Engineering Sciences)

Remote Sensing of Vegetation Species Diversity: The Utility of Integrated Airborne Hyperspectral and Lidar Data

Thesis directed by Professor William Emery

The change, reduction, or extinction of species is a major issue currently facing the Earth. Efforts are underway to measure, monitor, and protect habitats that contain high species diversity. Remote sensing technology shows extreme value for monitoring species diversity by mapping ecosystems and using those land cover maps or other derived data as proxies to species number and distribution. The National Ecological Observatory Network (NEON) Airborne Observation Platform (AOP) consists of remote sensing instruments such as an imaging spectrometer, a full-waveform lidar, and a high-resolution color camera. AOP collected data over the Ordway-Swisher Biological Station (OSBS) in May 2014. A majority of the OSBS site is covered by the Sandhill ecosystem, which contains a very high diversity of vegetation species and is a native habitat for several threatened fauna species. The research presented here investigates ways to analyze the AOP data to map ecosystems at the OSBS site. The research attempts to leverage the high spatial resolution data and study the variability of the data within a ground plot scale along with integrating data from the different sensors. Mathematical features are derived from the data and brought into a decision tree classification algorithm (rpart), in order to create an ecosystem map for the site. The hyperspectral and lidar features serve as proxies for chemical, functional, and structural differences in the vegetation types for each of the ecosystems. K-folds cross validation shows a training accuracy of 91%, a validation accuracy of 78%, and a 66% accuracy using independent ground validation. The results presented here represent an important contribution to utilizing integrated hyperspectral and lidar remote sensing data for ecosystem mapping, by relating the spatial variability of the data within a ground plot scale to a collection of vegetation types that make up a given ecosystem.

ACKNOWLEDGMENTS

I would like to thank: my committee members Bill Emery, Jeff Thayer, Xinzhao Chu, Peter Pilewskie, and Carol Wessman for their knowledge and guidance; Brian Johnson, Michele Kuester, Joel McCorkel, Tom Kampe, Nathan Leisso, Tristan Goulden, Shelley Petroy, Dave Barnett, Courtney Meier, Eve-Lyn Hinckley, Elena Azuaje, Becky Kao, Ben Chemel, Brian Damiani, the AOP flight crew, and the rest of the NEON field crew and contractors for planning and collection of the airborne and ground data; Steve Coates and John Hayes at the Ordway-Swisher Biological Station; and my wife Joella for her support during the writing of this dissertation.

The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc. This material is based in part upon work supported by the National Science Foundation under the following grants: EF-1029808, EF-1138160, and DBI-0752017. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CONTENTS

CHAPTER

TABLES

Table

FIGURES

Figure

CHAPTER I

SCIENCE BACKGROUND

Human actions are affecting biological diversity, species composition, and ecosystem structure and function (Maestre et al., 2012; NRC, 2003; Vitousek et al., 1997). Land use changes and the spread of invasive species are occurring on unprecedented levels leading to shifts in abundances of species and decreases in biodiversity, which are vital to ecosystem functioning and ecosystem services (Maestre et al., 2012; NRC, 2003; Schimel, Asner, & Moorcroft, 2013). Comprehensive biodiversity observation efforts have been proposed with the desire for standardized observations across taxonomic groups and spatially detailed data collection across a range of ecoclimatic regions in order to collect a baseline global inventory of life on Earth, construct ecological forecasting models, and monitor changes with time due to land use and climate change (Schimel et al., 2013). The first section of this chapter describes the concepts of biodiversity, biodiversity monitoring efforts, and reviews the National Research Council (NRC) Grand Challenges in Environmental Sciences which lists biological diversity and ecosystem functioning as a grand challenge.

Remote sensing is an extremely powerful tool for biodiversity monitoring. Instruments can make global, repeated measurements of land cover and vegetation (Strand et al., 2007a; Schimel et al., 2013; Turner et al., 2003). Remote sensing can measure the environmental and climate parameters (drivers of biological change) necessary to estimate species richness, species distributions, and community structure (Turner et al., 2003; Turner, 2014). Long-term monitoring of climate, ecosystem processes, and degradation or disturbance of ecosystems can also be used to determine the effect on biodiversity (Rose et al., 2015). The second section of this chapter discusses the benefits and limitations of remote sensing of

biodiversity including: mapping of species, land cover, vegetation traits, biodiversity through spectral proxies, and the creation of biodiversity indicators for conservation efforts.

Biodiversity is a major component of the terrestrial and aquatic sampling of the National Ecological Observatory Network (NEON), described in the third section of this chapter. Remote sensing data are collected by another component of NEON, the Airborne Observation Platform (AOP). NEON remote sensing products and ground truth vegetation species lists are used as the data sources for this research.

Importance of Biodiversity and Biodiversity Monitoring

Key climate and environmental parameters drive the distribution and abundance of species across landscapes (Turner et al., 2003). This correlation is used to generate species environmental niche models. Climate change affects ecosystems by causing shifts in species ranges, fostering pests and pathogens, altering fire disturbance regimes, changing migration patterns, and causing species extinctions (Strittholt & Steininger, 2007). In the boreal forests for example, recent observations have seen an increase in the growing season length and annual primary productivity, tree line expansion to the north, and an increase in fire frequency and intensity (Strittholt & Steininger, 2007).

In addition to climate, assemblages of species are also influenced by topography, soils, hydrology, species interactions, dispersal limitation, disturbance history, and human land use patterns (Colwell, 2009). Human actions are affecting biological diversity, species composition, and ecosystem

structure and function (Maestre et al., 2012; NRC, 2003; Vitousek et al., 1997). Human land management such as urbanization, silviculture, agriculture, grazing, fire-suppression, habitat fragmentation, etc. has led to shifts in abundances of species and decreases in biodiversity (NRC, 2003). Destruction and fragmentation of natural habitats is a major driver in the global decline of biodiversity (Strittholt & Steininger, 2007). The spread of invasive species is accelerated by adjacent land uses, roads, edges around forests, and the declining size of forest fragments (Strittholt & Steininger, 2007). Highly fragmented landscapes may also prevent some species from moving with corresponding changes in climate, leading to their ultimate extinction (Schimel et al., 2013). Species diversity is vital to ecosystem functioning and providing ecosystem services to humans (Maestre et al., 2012; NRC, 2003).

Biodiversity Concepts

Biodiversity is a conjunction of biological diversity and was first coined in 1985 by Walter G Rosen during planning of the "1986 National Forum on BioDiversity" (Harper & Hawksworth, 1995, p. 6). The United Nations Environmental Programme definition is:

"Biodiversity is the variety of life on Earth. It includes diversity at the genetic level, such as that between individuals in a population or between plant varieties, the diversity of species, and the diversity of ecosystems and habitats" (UNEP, 2007, p. 160).

Species form characteristic groupings, are associated with particular geographic localities, and are shaped by regional processes (Magurran, 2004). Ecological communities include interacting and interdependent constituent species (Magurran, 2004; Turner et al., 2003). The Florida Natural Areas Inventory (FNAI) defines a natural community as, "a distinct and recurring assemblage of populations of plants, animals, fungi and microorganisms naturally associated with each other and their physical environment" (FNAI,

2010, p. 2). Anne Magurran defines biodiversity in a more relevant definition with regards to this research as, “the variety and abundance of species in a defined unit of study” (Magurran, 2004, p. 8). For subsets of taxa, some experts use the term “assemblage” instead of “community,” where assemblages consist of phylogenetically related members of a community (Fauth et al., 1996; Magurran, 2004). Community is believed to be a natural study unit (Harper & Hawksworth, 1995) and this research focuses on assemblages of vegetation in a community.

Biodiversity is typically measured on three levels (Harper & Hawksworth, 1995; Magurran, 2004; Norse, 1986):

- Genetic diversity (within species)
- Species or organismal diversity (species number)
- Ecological diversity (community)

When it comes to species or organismal diversity, indices or metrics are calculated at three levels of complexity (Gotelli & Chao, 2013; Harper & Hawksworth, 1995):

- Species richness: number of species
- Species diversity: richness and relative abundance (individuals, biomass, or % cover)
- Phylogenetic, taxonomic, and functional diversity

The simplest measure of species diversity is species richness or the number of species of a particular taxon that characterize a particular biological community (Colwell, 2009). This measure allows the comparison of the number of species observed in a unit of study against other locations (of the same

unit size) or as a function of time in the same location. The higher the number of species observed, the higher the diversity. Measurement of species richness follows one of the fundamental rules of ecology, the species-area curve (Gotelli, 2008):

$$S = cA^z,$$

Where S is the number of species, A is the area, and c and z are fitted constants. According to this equation, the larger the observation unit, the more species that will be observed. When a large enough area is sampled, the species-area curves reaches an asymptote of the value of the total number of species in the community. Species richness is often used in efforts to preserve biodiversity due to its simplicity, where species losses can occur from extinction and species increases can reflect introductions or range shifts of species driven by climate change or anthropogenic causes (Gotelli & Chao, 2013).

A more complicated measure of species diversity includes abundance (evenness) along with species richness (Magurran, 2004; Colwell, 2009). Most individuals in a community typically belong to a few common species (Colwell, 2009). These uneven distributions dominated by a few species have less diversity than even distributions with equal abundances of all species present. Species richness alone cannot capture this level of variability. Conservation biologists use relative abundance to identify rare species that are more vulnerable to extinction (Colwell, 2009). Species diversity metrics or indices are mathematical functions that combine richness and evenness into a single measure (Colwell, 2009). Two widely used nonparametric methods are the Shannon entropy

$$H_{Sh} = -\sum_{i=1}^{S} p_i log p_i,$$

and the Gini-Simpson index

$$H_{GS} = 1 - \sum_{i=1}^{S} p_i^2,$$

where S is the number of species in an assemblage and p_i is the ith species relative abundance. For a given pattern of evenness, both measures increase as richness increases and for a given richness level, both increase as evenness increases, though the Simpson index is more sensitive to evenness (Colwell, 2009).

The third and most advanced measure of species diversity takes into account phylogenetic, taxonomic, and/or functional diversity. These measures are more than just species number and abundance but put a weight on how related species are (Harper & Hawksworth, 1995). Many studies care about the variety and number of species that fulfill different functional roles (Colwell, 2009). For instance, an ecosystem may have a uniform distribution of a large number of species but would not be as diverse if they are all functionally similar to each other.

Biodiversity is a comparative science where an analysis might seek to determine if an area is more diverse than another or if diversity has changed over time (Magurran, 2004). Caution is required when comparing different diversity estimates (such as species number against biomass) or different units of observation (Magurran, 2004). Species area curves need to be adjusted using rarefaction in order to compare species numbers for different measurement areas (Gotelli & Chao, 2013). Diversity comparison measures are commonly categorized into alpha, beta, and gamma diversity. Alpha diversity studies the local diversity in a spatially defined unit (Colwell, 2009; Magurran, 2004). Species richness, Shannon entropy, and the Simpson index are all measures of alpha diversity. Beta diversity compares differences in the compositional diversity between areas of alpha diversity to look for species overlap or distinctness (Gotelli & Chao, 2013; Magurran, 2004). This could be a comparison of two sample areas in the same community or between two different habitats (Colwell, 2009). A more sophisticated set of measures such as biotic similarity is utilized to compare multiple assemblages (Gotelli & Chao, 2013). Finally, gamma diversity analyzes regional diversity for an entire landscape (Colwell, 2009).

Biodiversity Monitoring Efforts

Biodiversity raises many science questions but a few relevant ones are (Dahlin, Asner, & Field, 2013; NRC, 2001; Schimel et al., 2013; Strand, Fosnight, Herkenrath, & Höft, 2007b; Turner et al., 2003):

- What drives species distribution and diversity?
- How and why do communities vary across space and time?
- Why are some areas species rich?
- Why do some have high levels of species endemism?
- How does land use change affect biodiversity?
- What is the sensitivity of species to environmental change?
- What is the relationship between biodiversity and ecosystem structure and functioning?

Observations of biodiversity attempt to answer these questions. However, field surveys can be flawed due to preferential treatment or weight towards certain taxa, rare species, specialists, or species with lots of genetic diversity. Other issues include: gaps in surveys; too small of study areas due to the time consuming and expensive realities of field work; and the need for taxonomical experts (Harper & Hawksworth, 1995; Magurran, 2004). Site based studies are highly variable due to disturbance, competition, predation, etc. so regional and global studies are required to truly understand trends in species ranges (Schimel et al., 2013).

Several biodiversity observation monitoring efforts have been proposed with the desire for standardized observations across taxonomic groups and spatially detailed data collection across a range of ecoclimatic regions. In the US, a series of workshops was held to discuss the concept of a Biodiversity Observation Network (NSF, 1998; NSF, 1999a; NSF, 1999b; NSF, 1999c), which was incorporated as a major component into the science design of the National Ecological Observatory Network (AIBS, 2004; NRC, 2003; Schimel et al., 2011), as discussed further in the third section of this chapter. Internationally, GEO BON – the Group on Earth Observations Biodiversity Observation Network (GEOBON, 2015) is working toward improving, coordinating, and standardizing the collection of biodiversity data and delivery of information to decision makers. GEOBON has identified a minimum candidate list of essential biodiversity variables (EBV) that capture the dimensions of biodiversity and are to be measured as part of the biodiversity monitoring efforts including: genetic composition, species population, species traits, community composition, ecosystem structure, and ecosystem function (Pereira et al., 2013). The EBVs address some of the Aichi Biodiversity Targets (CBD, 2015a) under the United Nations Convention of Biological Diversity (CBD, 2003; CBD, 2015b) as part of the Decade on Biodiversity (CBD, 2015c) to reduce the rate of biodiversity loss. Comprehensive biodiversity observation efforts are required now to collect a baseline global inventory of life on Earth, construct ecological forecasting models, and monitor changes with time due to land use and climate change (Schimel et al., 2013).

NRC Grand Challenges in Environmental Sciences

The National Research Council (NRC) formed a committee on behalf of the National Science Foundation (NSF) to identify grand challenges in environmental sciences (NRC, 2001). The committee expressed concern about the stresses placed on the biosphere and ecosystems by increases in human

population and economic development. Eight grand challenge areas were identified to the NSF to focus future funding to fill gaps in current understanding:

- biogeochemical cycles,
- biological diversity and ecosystem functioning,
- climate variability,
- hydrologic forecasting,
- infectious disease and the environment,
- institutions and resource use,
- land-use dynamics,
- and reinventing the use of materials.

One desire is to better understand the factors affecting biological diversity and ecosystem structure and functioning and what role human activity plays. Many worry that habitat alteration or loss may cause ecosystems to be unable to support human needs or cause extinction of species. Important research areas include:

- improving tools for rapid assessment of diversity at all scales,
- producing a quantitative, process based theory,
- understanding the relationship between diversity and ecosystem structure and functioning,
- and developing and testing techniques for modifying, creating, and managing habitats that can sustain biological diversity.

Land-use dynamics is also relevant with the desire to better understand how changes in land use and land cover may affect ecosystem functioning and services, or more specifically determine the

consequences of spatial changes in land use and land cover on biological diversity. Four primary science questions related to biological diversity can be formulated:

- What factors affect biological diversity and ecosystem structure and functioning?
- How does ecosystem structure and function relate to biodiversity?
- How does biodiversity change with climate change, land use change, or invasive species?
- What tools can we use to rapidly assess diversity at different spatial and temporal scales?

The report identifies remote sensing as a valuable tool for its ability to generate global maps of ecosystem distribution which is critical information required to test and improve diversity models. Specifically remote sensing should be used to assess diversity and examine ecosystem characteristics. Remote sensing of biodiversity is discussed further in the next section. A series of workshops were held around the same time as the formulation of the Grand Challenges to discuss the idea of a National Ecological Observatory Network (NSF, 2010a; NSF, 2010b; NSF, 2010c; NSF, 2012a; NSF, 2012b; NSF, 2012c), which is discussed in third section in this chapter.

Remote Sensing of Biodiversity

Global data is limited on key aspects of biodiversity such as species identity, species numbers, and ecosystem condition (Turner, 2014). Conservation policy calls for accurate and timely information (Rose et al., 2015), but extensive ground sampling takes too long and is too cost-preventative (Schimel et al., 2013). Satellite and airborne remote sensing are powerful technologies to monitor biodiversity as they are a more cost effective means to achieve regional and global coverage (Turner et al., 2003), can make repeated measurements across space and time (Turner, 2014), can measure change over large-scale

ecosystems (Strand et al., 2007a). Remote sensing can measure the environmental and climate parameters (drivers of biological change) necessary to estimate species richness, species distributions, and community structure (Turner et al., 2003; Turner, 2014). Long-term monitoring of climate, ecosystem processes, and degradation or disturbance of ecosystems can also be used to determine the effect on biodiversity (Rose et al., 2015). Remote sensing has been shown as an effective tool for environmental assessment and monitoring (Steininger & Horning, 2007), working across a range of ecosystems (Schimel et al., 2013), while increasing the speed and frequency to analyze landscapes in an objective and comprehensive assessment (Steininger & Horning, 2007). In the past decade, several books (Strand et al., 2007a; Horning, Robinson, Sterling, Turner, & Spector, 2010), journal special issues (Asner, 2013; Rocchini, 2015), and workshops (Rose et al., 2015) have highlighted how remote sensing can be used for conservation of ecosystems and biodiversity.

Multi sensor data has been identified as being extremely useful for ecosystem studies (Rose et al., 2015). Several multi-modal airborne remote sensing platforms combine imaging spectrometers (hyperspectral data) with lidar instruments (3D structure) including the Carnegie Airborne Observatory (Asner et al., 2007; Asner et al., 2012); Goddard's LiDAR, Hyperspectral, and Thermal Imager (Cook & Corp, 2012); the Airborne Prism Experiment (Schaepman et al., 2003, Schaepman et al., 2015); the Airborne Snow Observatory (JPL, 2015a); and the NEON Airborne Observation Platform (Kampe, Johnson, Kuester, & Keller, 2010; Kampe et al., 2011). The Airborne Observation Platform (AOP) is described in detail in the "Remote Sensing with the Airborne Observation Platform" Section.

Direct Versus Indirect Remote Sensing

Remote sensing as applied to biodiversity studies can be grouped into one of two categories: either direct or indirect (Turner et al., 2003; Turner, 2014). With direct remote sensing, individual organisms, species assemblages, or ecological communities are identified in the data. Depending on the spatial resolution of the sensors, this can include detection of individual tree canopies, large mammals, or groups of birds from high resolution data. Alternatively, tree species can be identified by spectral analysis with hyperspectral data or time-series analysis of phenology (Turner et al., 2003). The methods and results presented in this research fall under the category of direct remote sensing of biodiversity through mapping of vegetation communities.

With indirect remote sensing, climatic and environmental parameters are calculated from the remote sensing data and used as inputs into biodiversity models (Turner et al., 2003). Some commonly used environmental parameters include: climate and rainfall, geology, soil moisture, elevation, general land cover, habitat structure and topography, primary productivity, phenology, chlorophyll, and distance to manmade features (Turner et al., 2003; Steininger & Horning, 2007). Analysis of some species can be as simple as determining environmental limiting factors such as temperature, moisture and rainfall, or elevation (Turner et al., 2003). Another common approach is to correlate the variability in reflectance data (or derived indices) with field-based observations of vegetation diversity (White, Gómez, Wulder, & Coops, 2010). Previous studies have shown that areas of higher vegetation species diversity present as higher variability in the remote sensing data (Palmer et al., 2000; Palmer, Earls, Hoagland, White, & Wohlgemuth, 2002; Carlson, Asner, Hughes, Ostertag, & Martin, 2007). Either way, metrics or indices calculated from the remote sensing data are used as proxies for patterns of biodiversity (Turner et al., 2003; Turner, 2014).

Land Cover Mapping

Remote sensing is capable of global mapping of ecosystems, vegetation, and land use. The technology can also be used to analyze trends in extents of ecosystems, coverage of protected areas, threats to biodiversity, connectivity or fragmentation of ecosystems, trends in populations of select species, and human development (Strand et al., 2007a). Land cover mapping is an approach that typically tries to create a standard set of land classes that can be compared over large spatial extent and over time for change detection. A National Land Cover Database (NLCD) was created to "assess wall-to-wall, spatially explicit, national land cover changes and trends across the United States" (MRLC, 2015). NLCD maps are generated from remote sensing data at a spatial resolution of 30 m and contain 16 land cover classes for the continental U.S. (Fry et al., 2011):

- Open Water
- Perennial Ice/Snow
- Developed-Open Space
- Developed-Low Intensity
- Developed-Med. Intensity
- Developed-High Intensity
- Barren Land
- Deciduous Forest
- Evergreen Forest
- Mixed Forest
- Shrub/Scrub
- Grassland/Herbaceous

- Pasture/Hay
- Cultivated Cropland
- Woody Wetlands
- Herbaceous Wetlands

Only 7 of the classes represent natural landscapes and the forest descriptions are extremely generic. A National Vegetation Classification Standard (NVCS) was created and allows descriptions of vegetation cover over broad spatial scales with upper level global to continental, middle level subcontinental to regional, and lower level sub-regional to local descriptions (FGDC, 2008). Specifically the lower level descriptions are designed to capture plant alliance or association. Additionally, NatureServe has created a mid-scale classification scheme to describe ecological systems in the U.S. (Comer et al., 2003). For this research, a regional community classification system, the Florida Natural Areas Inventory (FNAI, 2010) is used and described in Chapter III.

Even with specific ecosystem classes, land use land cover maps may not be adequate for answering certain biodiversity questions. For instance, other descriptive classes, individual species ids, and derived remote sensing data are required to identify natural forests versus industrial pine plantations (Strand et al., 2007b), differentiate natural open canopy versus degraded forests (Strittholt & Steininger, 2007), and detect invasive species. Advances in remote sensing technology and algorithms have moved beyond basic mapping of vegetation distribution and have allowed the identification of individual species; characterization of structural, chemical, and functional traits; measurement of above ground biomass and carbon stocks; and detailed analyses of plant community composition, vegetation diversity (Asner, 2013; Schimel et al., 2013; Strittholt & Steininger, 2007; Turner et al., 2003).

Observations of Vegetation Chemical, Structural, and Functional Traits

Reflectance as a function of wavelength has a distinct shape for vegetation. The reflectance values for each wavelength in the solar-reflective region are a combination of absorption and scattering. Airborne hyperspectral instruments measure the radiance at-sensor, which includes the sun as a light source, transmission and scattering in Earth's atmosphere, and the reflectance of an object on the ground that is viewed by a detector element. The raw sensor data is calibrated to units of radiance and then an atmospheric correction algorithm can be applied to solve for the surface reflectance. Plant compound absorption occurs in discrete portions of the electromagnetic spectrum (Ollinger, 2011). Chlorophyll (Chl a and b) absorption in the blue and red wavelengths in the human visual range is the reason why plants are green (Ustin et al., 2009). Other plant pigments such as carotenoids and anthocyanins also absorb light at visible wavelengths (Ollinger, 2011). Scattering by leaves is dependent on the number and arrangement of cellular surfaces, and leaf area plus canopy structure influences multiple scattering (Asner, 1998; Ollinger, 2011). Absorption is low and scattering is high at near infrared wavelengths, which together with chlorophyll absorption in the red presents the classic "red edge" of vegetation curves. Healthy leaves are mostly water, where major water absorption features present in the shortwave infrared bands and moderately in the near infrared (Kokaly, Asner, Ollinger, Martin, & Wessman, 2009; Ollinger, 2011). Non-pigment leaf constituents such as nitrogen, proteins, lignin, and cellulose affect light absorption and also scattering indirectly through their influence on leaf/plant structure (Kokaly et al., 2009; Ollinger, 2011). Figure 1 shows some reflectance curves of ground targets at Ordway-Swisher Biological Station.

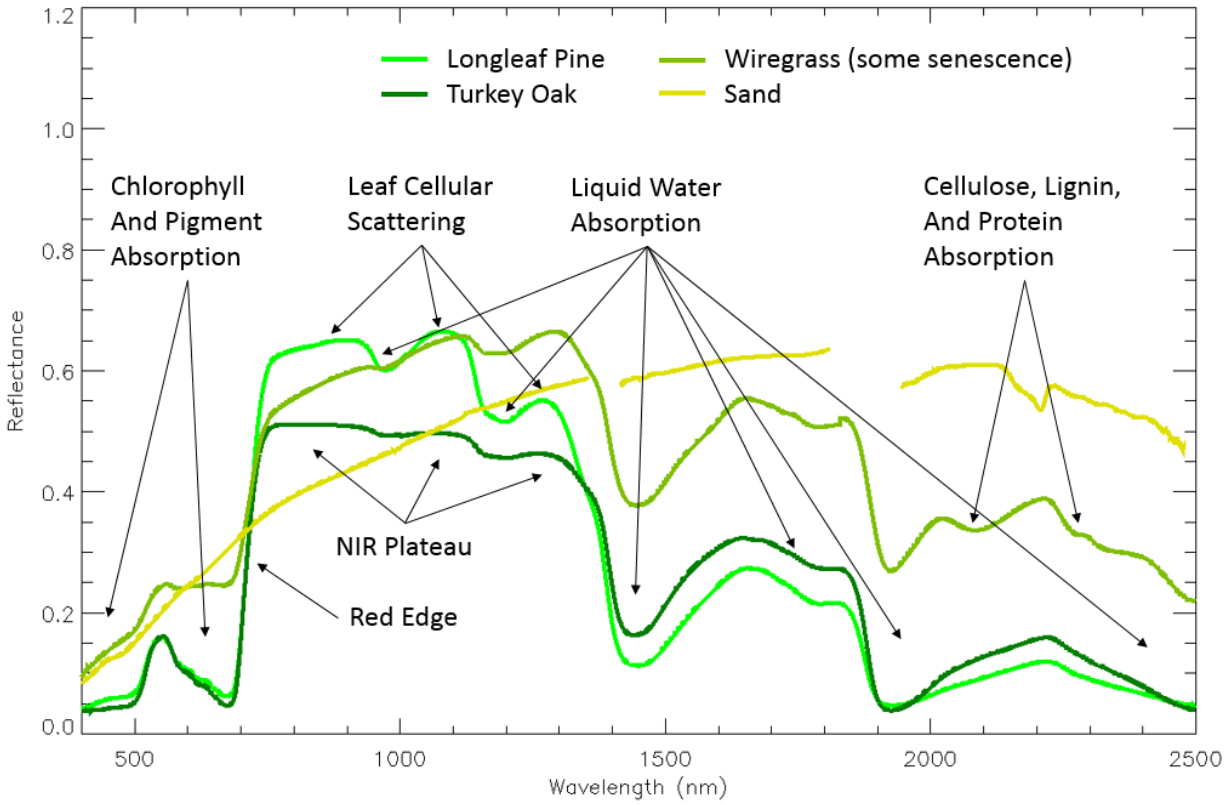

Figure 1: Reflectance curves of ground targets showing absorption and scattering spectral traits

Wavelengths with significant vegetation absorption and scattering are highlighted. The reflectance of bare sand is shown for comparison.

Remote sensing can also be used to assess plant functional types such as conifers, broadleaved evergreens, deciduous trees, shrubs, and herbs (Ollinger, 2011; Schimel et al., 2013; Ustin & Gamon, 2010). Many ecosystem process models rely on plant functional types (Dahlin et al., 2013). Spectral traits that can be derived from hyperspectral data include: chlorophyll, carotenoids, xanthophyll, anthocyanin, water content, carbon, nitrogen, phosphorus, lignin, cellulose, leaf area index, leaf mass per unit area (Asner, 1998; Asner & Martin, 2009; Asner, Martin, Ford, Metcalfe, & Liddell, 2009; Asner, Martin,

Anderson, & Knapp, 2015; Curran, 1989; Dahlin et al., 2013; Gamon, Peñulas, & Field, 1992; Gamon & Surfus, 1999; Kokaly et al., 2009; Martin, Plourde, Ollinger, Smith, & McNeil, 2008; Ollinger, 2011; Townsend, Foster, Chastain, & Currie, 2003; Townsend, Asner, & Cleveland, 2008; Ustin, Roberts, Gamon, Asner, & Green, 2004; Ustin et al., 2009; Ustin & Gamon 2010; Wessman, Aber, Peterson, & Melillo, 1988; Wessman, 1990).

Lidar is an active remote sensing instrument that directly measures the 3D distribution of vegetation canopies and ground topography by emitting a laser pulse and then recording the received return signal as a function of time. Using lidar as a range-finder, structural traits that can be derived from lidar data include: vegetation height, canopy cover, canopy horizontal and vertical structure, topography, leaf area index, and above ground biomass (Lefsky, Cohen, Parker, & Harding, 2002). Different species or communities have different physical canopy and understory structures. Invasive species have been shown to alter the vegetation structure in areas they have expanded into (Asner et al., 2008).

Spectral Variability Hypothesis

The spectral variability hypothesis (SVH) attempts to eliminate the need to perform a land cover classification in order to assess biodiversity from remote sensing data. Specifically, SVH proposes that biodiversity is directly correlated to an objective measure of the variation in reflectance or intensity in an image (Palmer et al., 2000; Palmer et al., 2002). Put another way, the spatial variation in reflectance is correlated with the spatial variation in the environment, which is correlated to species richness. The original study used panchromatic imagery in a prairie ecosystem.

The hypothesis has been proven in other studies, with one notable example using hyperspectral imagery of a tropical lowland forest in Hawaii (Carlson et al., 2007). Carlson et al. conducted foliar sampling of identified woody species and then had biogeochemical measurements made on the leaves, specifically chlorophyll, nitrogen, and equivalent water thickness. If the inter-species variation in leaf biochemical properties exceeds the intra-species variation, then the SVH should hold. A Monte-Carlo analysis showed that the range (max-min) of values of each of the biogeochemical parameters increased with the number of individuals included in a sample (similar to a species-area curve). Based on the success with the foliar data, Carlson et al. next analyzed the hyperspectral imagery. Four spectral bands (530, 720, 1201, and 1523 nm) were chosen due to their correlation with chlorophyll, nitrogen, and water (either absorption bands or strong scattering wavelengths due to canopy structure). The first-derivative or slope of reflectance was calculated for each wavelength, then the range of the first-derivative was calculated for the spatial plot under study. Finally a regression was performed on the range in first-derivative versus species richness for all the study plots. The final regression model was then applied to all of the hyperspectral imagery to create a spatial map of species richness. While this study successfully demonstrated the SVH, two issues are evident. The first is that the study sites used in Hawaii contain very tightly packed canopies which provides excellent results for those areas regressed against woody canopy species diversity but cannot speak to understory or ground cover species richness. The second issue is evident in the resulting maps of species richness which display roads as having the highest species richness (Asner, 2015). This highlights the need for pre-processing to remove pixels of man-made features and also suggests that variability in pixel reflectance due to shadowing or directional reflectance illumination differences could skew the resulting species richness maps.

In a different study, foliar measurements of nitrogen mass, carbon mass, leaf water concentration, and canopy water content were performed in a Mediterranean-type ecosystem (Dahlin et al., 2013). Dahlin et al. proposed that variation in plant traits across an ecosystem is a proxy for variation in

functional diversity. A high variance in a trait within a community indicates that the community is capable of supporting a wide range of plant functional types and a narrow variance could be due to environmental filtering of functional types. The study performed an ordinary least squares analysis of plant traits versus environmental gradients, known land use history, plant community, and spatial autocorrelation. Plant community was found to be the strongest predictor of chemical trait variation independent of spatial autocorrelation. This important study establishes a correlation between plant communities and plant traits that can be measured by remote sensing data (either directly or through spectral proxies), supporting the spectral variability hypothesis and lending credibility to the research presented here to map vegetation communities using remote sensing data.

Biodiversity Indicators

An indicator is a measure that is used to evaluate the performance of functions, processes, and outcomes over time (Strand et al., 2007b). For this instance, indicators will monitor the effectiveness of biodiversity policies and conservation efforts. Successful indicators are tied to key questions, problem oriented (ex: human caused change), simple to understand, and cost-effective for long term implementation (Strand et al., 2007b). Rates of change are common indicators (ex: reduction in loss of natural forest extent) but require data collection at three or more points in time (Strand et al., 2007b). Analysts must understand what remote sensing data represent and translate that into a useful indicator (Strand et al., 2007a). A sample of commonly used indicators for ecosystem and vegetation monitoring are listed (Loucks & Leimgruber, 2007; Steininger & Horning, 2007; Strand et al., 2007b; Strand, Leimgruber, & Mueller, 2007c; Strittholt, 2007; Strittholt & Steininger, 2007; Underwood & Ustin, 2007):

- Extent of biomes, ecosystems, and habitats
- Area and location of old-growth forests, plantations, and sustainable forestry
- Coverage of protected areas
- Isolation of, levels of encroachment on, and degradation of protected areas
- Landscape condition adjacent to protected areas
- Connectivity/ fragmentation of forests and ecosystems
- Area and number of large forest blocks
- Ecosystem integrity and condition
- Forest degradation
- Forest change
- Rate of deforestation/reforestation
- Trends in abundance and distribution of select species
- Location and extent of species aggregations
- Species endemism
- Rate of change in biological diversity
- Carbon storage
- Nitrogen deposition
- Trends in invasive alien species
- Grazing pressure
- Fire location and frequency
- Water quality

The indicators listed come directly from, or are adapted from, the Convention on Biological Diversity (CBD, 2003). In general, maps are not as good as indicators because they require interpretation.

Statistical tables or graphs are ideal (Strand et al., 2007b). Depending on the application and user group, biodiversity indicators can be focused on local scale, regional to national, or global scale.

Remote Sensing Limitations

Several issues exist with respect to remote sensing for biodiversity studies including instrument spatial resolution, species or habitat detectability, accuracy of derived data products, and the need for in-situ ground data. Remote sensing typically requires a tradeoff of resolution versus coverage (Turner, 2014). High spatial resolution instruments that can observe fine detail can only do so over a small area and instruments that can acquire continental wide swaths are only able to do so by having coarse spatial resolution pixels. One challenge with global remote sensing is that the coarse pixel spatial resolution is larger than resolution that many taxa interact with (Rose et al., 2015). Remote sensing may not be able to detect certain individuals (Magurran, 2014). Features of interest must be large with respect to the resolution and land classification maps derived from the remote sensing data must contain sufficient thematic detail (Steininger & Horning, 2007; Strittholt & Steininger, 2007). Determination of forest extent is better with high resolution sensors (Strittholt & Steininger, 2007). More research is required to study pixel spatial resolution on species id and diversity detection (Schimel et al., 2013).

Most remote sensing instruments are mounted on-board aircraft or spacecraft which means they can only monitor features that can be viewed from above (Steininger & Horning, 2007). Passive solar-reflective instruments cannot see through clouds and even if the clouds are not between the sensor and the ground, atmosphere (including haze and thin clouds) and sensor artifacts that affect the data must be corrected (Steininger & Horning, 2007). Furthermore, features of interest must have sufficiently unique

spectral signatures to be discriminated from other targets (Steininger & Horning, 2007). Finally, passive solar-reflective instruments cannot see the understory directly in forests with closed canopies (Steininger & Horning, 2007).

The accuracy of derived land cover and change data sets may not be high enough for certain applications (Strand et al., 2007a; Strand et al., 2007b; Steininger & Horning, 2007). Specific land cover classes are needed because some forest communities are more threatened than others and ideally, natural open canopy should be differentiated from degraded forests (Strittholt & Steininger, 2007). Also, change detection using two images collected in different seasons (dry versus wet or leaf-on versus leaf-off) can lead to errors (Strittholt & Steininger, 2007).

For remote sensing to reach its full potential, in-situ ground sampling is required to improve accuracy and relate the remote sensing data to ecosystem components (Rose et al., 2015 and Turner, 2014). Additionally, regional expertise of land use and land cover, species trends and habitat usage, and ecological communities and systems is required to properly interpret the remote sensing data (Steininger & Horning, 2007). Studies must take into account historical anthropological impacts as they impact ecosystems for a considerable length of time (Dahlin et al., 2013).

The National Ecological Observatory Network

The National Research Council (NRC) formed a committee to evaluate the National Ecological Observatory Network (NEON) workshop recommendations with respect to the Grand Challenges (NRC,

2001). The general opinion was that "environmental change and its influence on biological processes occurs at regional, continental, and global scales" (NRC, 2003, p. 23), however, previous ecological research was typically conducted by single investigators or small groups working at local scales and not capable of addressing the Grand Challenges (NRC, 2003). The advances in scientific knowledge required justifies a new national-level research infrastructure to study the direct effects and feedbacks between environmental change and biological processes. The over-arching goals of NEON are to:

- measure the pace and nature of biological change,
- improve mechanistic understanding of the environment,
- develop a predictive capability (spatio-temporal analyses and forecasts),
- and identify solutions to ecological problems.

NEON would provide the capability to apply results from one site to other regions and ecosystems plus address questions regarding biodiversity, species composition, and ecosystem functioning. Terrestrial observations of plant species richness and airborne remote sensing data collected by NEON are the primary data used in this research.

Overview of NEON

The National Ecological Observatory Network (NEON) is "a continental-scale observation system for examining ecological change over time" (Neon, 2015a). More specifically, NEON is a continental-scale research platform to "enable understanding and forecasting of the impacts of climate change, land-use change, and invasive species on aspects of continental-scale ecology such as biodiversity, biogeochemistry, infectious diseases, and ecohydrology" (Keller et al., 2008; Schimel et al.,

2011, p. 8). NEON science focuses on questions related to six of the eight Grand Challenges (NRC, 2001) plus science questions regarding invasive species. The NEON design strives to address questions relevant to large regions and not easily addressed by traditional ecological research. NEON will observe both the causes and effects to allow researchers to understand the underlying processes and better study the interactions and feedbacks. NEON construction is funded by the National Science Foundation under a Major Research Equipment and Facilities Construction (MREFC) account. Once completed in 2017, the first continental-scale observatory will operate for a period of 30 years and data products will be freely available. NEON's open access approach to its data and information products will enable scientists, educators, planners, decision makers and the public to map, understand and predict the effects of human activities on ecology and effectively address critical ecological questions and issues.

NEON has partitioned the U.S., including Alaska, Hawaii and Puerto Rico, into 20 eco-climatic domains, each of which represents different regions of vegetation, landforms, climate, and ecosystem performance, as depicted in Figure 2:

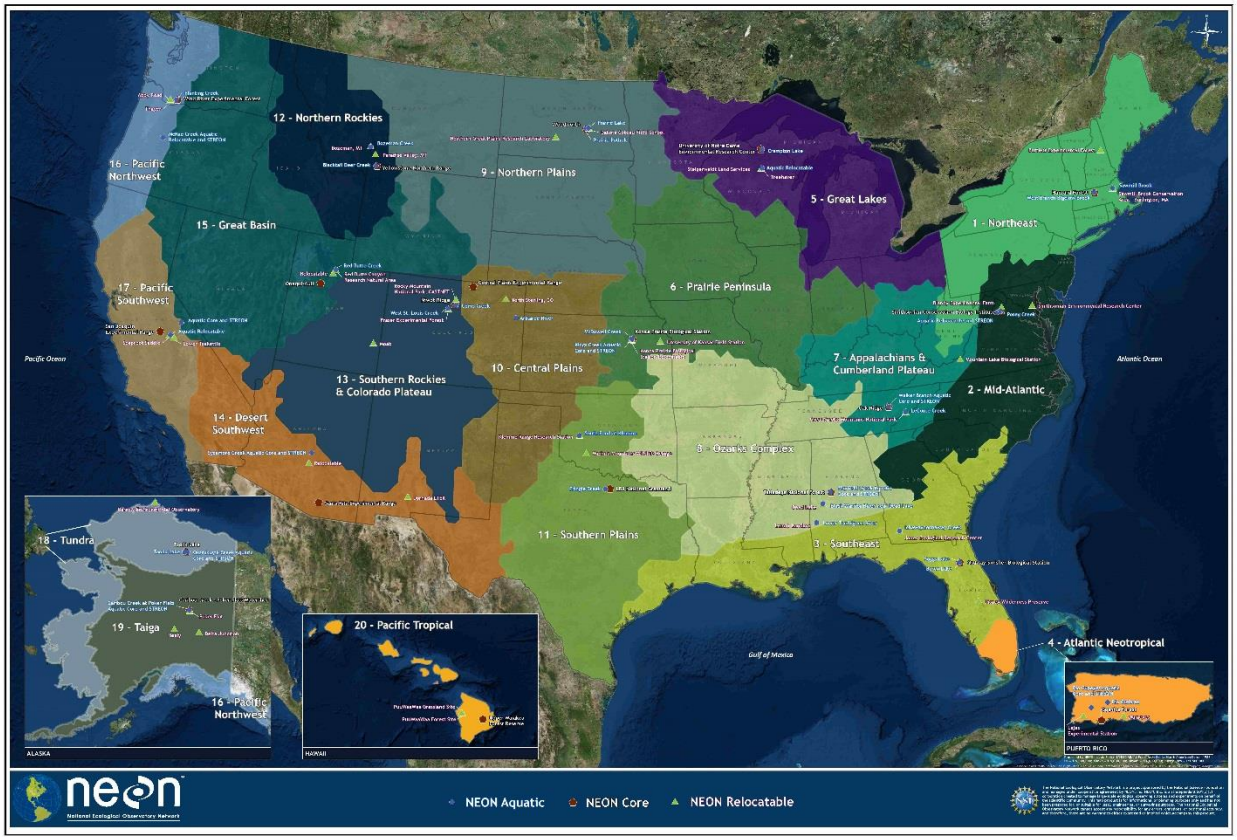

Figure 2: Map of the 20 NEON Domains and proposed site locations. Figure credit: NEON

The statistical design is based upon algorithms for multivariate geographic clustering using national data sets of ecoclimatic variables (Hargrove & Hoffman, 1999; Hargrove & Hoffman, 2005). This ensures that NEON is able to systematically sample the U.S. in a design that covers the environmental variability in an approximately uniform way (Schimel et al., 2011). Each domain contains one core site representative of minimally managed wildland conditions within that domain and will be sampled for 30 years. In addition, relocatable sites are located in each domain to collect data over shorter temporal time spans and will focus key science questions specific to each region such as human land management effects on ecosystems. These comparisons at both the domain and national levels will

provide critical information that can be used to test ecological models and to identify the impacts of land use change and invasive species on ecology.

The observatory consists of five data collection components: Terrestrial Instrument System (TIS), Terrestrial Observation System (TOS), Aquatic Instrument System (AIS), Aquatic Observation System (AOS), and the Airborne Observation Platform (AOP) (NEON, 2015b). Collocation of measurements associated with each of these components will allow for linkage and comparison of data products.

The TIS will observe key physical and chemical climate causes of ecological change, as well as ecosystem-level responses to those causes, in terrestrial ecosystems to advance our understanding of how the structure and function of these systems are changing (Schimel et al., 2011). The TIS includes a flux tower at each site that measures key climate inputs such as air temperature, precipitation, solar and net radiation, aerosol optical depth, photosynthetically active radiation, and wind speed and direction (NEON, 2015c; Schimel et al., 2011). Bioclimate variables are captured with vertical profiles of measurements into the canopy and near the ground including cameras to monitor phenology and measurements of latent heat exchange to determine evapotranspiration (Schimel et al., 2011). The towers also capture carbon cycling using micrometeorological techniques measuring 3D winds, CO_2, soil respiration, and nutrient use efficiency (NEON, 2015c; Schimel et al., 2011). The TIS also includes soil arrays to measure temperature, moisture, CO_2 concentration (NEON, 2015d) and a soil characterization is performed at each site. The TIS measurements cover the following ecological themes: soil structure physics, atmosphere and air quality, biogeochemistry, phenology, bioclimate, and energy balance (Schimel et al., 2011).

The TOS will quantify the impacts of climate change, land use, and biological invasions on terrestrial populations and processes by sampling key groups of organisms (sentinel taxa), infectious disease, soil, and nutrient fluxes across system interfaces (Kao et al., 2012; Schimel et al., 2011). The sentinel taxa include organisms with varying life spans and generation times, and wide geographic distributions to allow for standardized comparisons across the continent and include the following taxa: plants, birds, small mammals, ground beetles, mosquitos and ticks, and soil microbes (NEON, 2015e). The TOS samples cover the following ecological themes: diversity, abundance, phenology, biomass and productivity, biogeochemistry, metabolism, and pathogens (NEON, 2015b; Schimel et al., 2011). The vegetation sampling is discussed in more detail in the next section.

Aquatic measurements and observations observe key physical, biological, and chemical drivers of ecological change, as well as responses to those factors, in freshwater systems. The AIS makes measurements of the physical and chemical properties of streams, rivers, and lakes. Physical measurements include flow rates and depth, ice cover, and temperature (Schimel et al., 2011). Chemical measurements of isotopes, dissolved oxygen, pH, specific conductivity, fluorescence dissolved organic matter, chlorophyll a, and nitrates are made (NEON, 2015f). Lakes include buoys to measure depth profiles of water quality. Each NEON aquatic site also measures the microclimate with meteorological sensors. The AIS measurements cover the following ecological themes: biogeochemistry, hydrology, ecohydrology, biodiversity, invasive species, genomics, habitat and landscape structure, biomass, and stream metabolism (Schimel et al., 2011).

The AOS performs sampling on the following taxa: fish, macro-invertebrates, aquatic plants, algae, microbes (NEON, 2015g). Chemical analyses of sediment and water samples are performed and habitat observations are made such as: stream and lake morphology, stream/lake bed composition, habitat

type, riparian vegetation, and dead or down wood in stream corridor (Schimel et al., 2011). The AOS samples cover the following ecological themes: diversity, abundance, metabolism, biomass, and biogeochemistry (NEON, 2015g).

The NEON airborne remote sensing instrumentation is designed to bridge scales from organism and stand scales, as captured by plot and tower observations, to the scale of satellite based remote sensing (Kampe et al., 2010; Schimel et al., 2011). Each AOP contains an instrument payload containing a hyperspectral imaging spectrometer, a full-waveform lidar, and a high spatial resolution color digital camera (NEON, 2015h). Airborne remote sensing plays a critical role in the scaling strategy of the Observatory by making measurements at the scale of individual shrubs (1-3 meters) and that of larger plants (over hundreds of square kilometers). The AOP data collections cover the following ecological themes: canopy chemistry, vegetation structure, land use and land cover, and topography (NEON, 2015h; Schimel et al., 2011). The AOP is discussed in detail in the "Remote Sensing with the Airborne Observation Platform" Section.

The data products derived from the data collection components fall into five broad themes: atmosphere; biogeochemistry; ecohydrology; land use, land cover, and land processes; and organisms, populations, and communities (NEON, 2015i). Products from the instruments and observations are listed in detail in the NEON data products catalog (NEON, 2015j). NEON will combine site-based data with remotely sensed data and existing continental-scale data sets (e.g. satellite data) to provide a range of scaled data products that can be used to describe changes in the nation's ecosystem through space and time. The land use analysis package provides access to satellite and national data sets to extend coverage to areas not sampled by the Observatory.

Vegetation Sampling With the Terrestrial Observation System

The goal of the terrestrial sampling design is to direct the observation of terrestrial organisms and biogeochemistry endpoints for long-term trend detection within specific NEON sites, facilitate comparability across sites as well as with other ecological investigations, and contribute to the understanding of the cause and consequence of ecological change (Barnett et al., 2015). A site-level spatial sampling design that can be applied consistently across NEON domains has been developed for the Terrestrial Observation System (TOS). Many of the specific locations of the TOS sampling elements are collocated with each other and with environmental sensors (i.e., within the flux zone of the tower) to allow comparison of the data streams. Within a site, organismal and soil sampling for the TOS has been collocated to the extent possible to optimize linkages between data products and provide a more complete picture of processes associated with targeted observations and trends across the groups to be sampled (Fancy, Gross, & Carter, 2009). TOS sampling occurs using three different plot types (Barnett et al., 2015):

- Tower plots are located in the 90% flux area of the primary and secondary airsheds and observe the dominant vegetation type at the site. These plots allow linkages between TOS and TIS tower climate and flux measurements. A spatially balanced random design is used to establish ground plot coordinates in an unbiased and consistent fashion.
- Distributed plots are located across the entire site in order to sample the variability of organisms and processes across the site and capture the heterogeneity. These plots allow analysis of gradients and enable scaling from ground measurements to airborne remote sensing. A stratified (by land cover type) random, spatially balanced design is used to establish ground plot coordinates and efficiently capture landscape-scale patterns and trends.

- Gradient plots are used if distributed plots fail to fully capture site-level gradients in vegetation structure, leaf area index (LAI), or plant canopy chemistry. A targeted, non-random approach informed by NEON's Aerial Observation Platform remote-sensing data is used to establish ground plot coordinates.

A general example of TOS plot locations at a site is given in Figure 3:

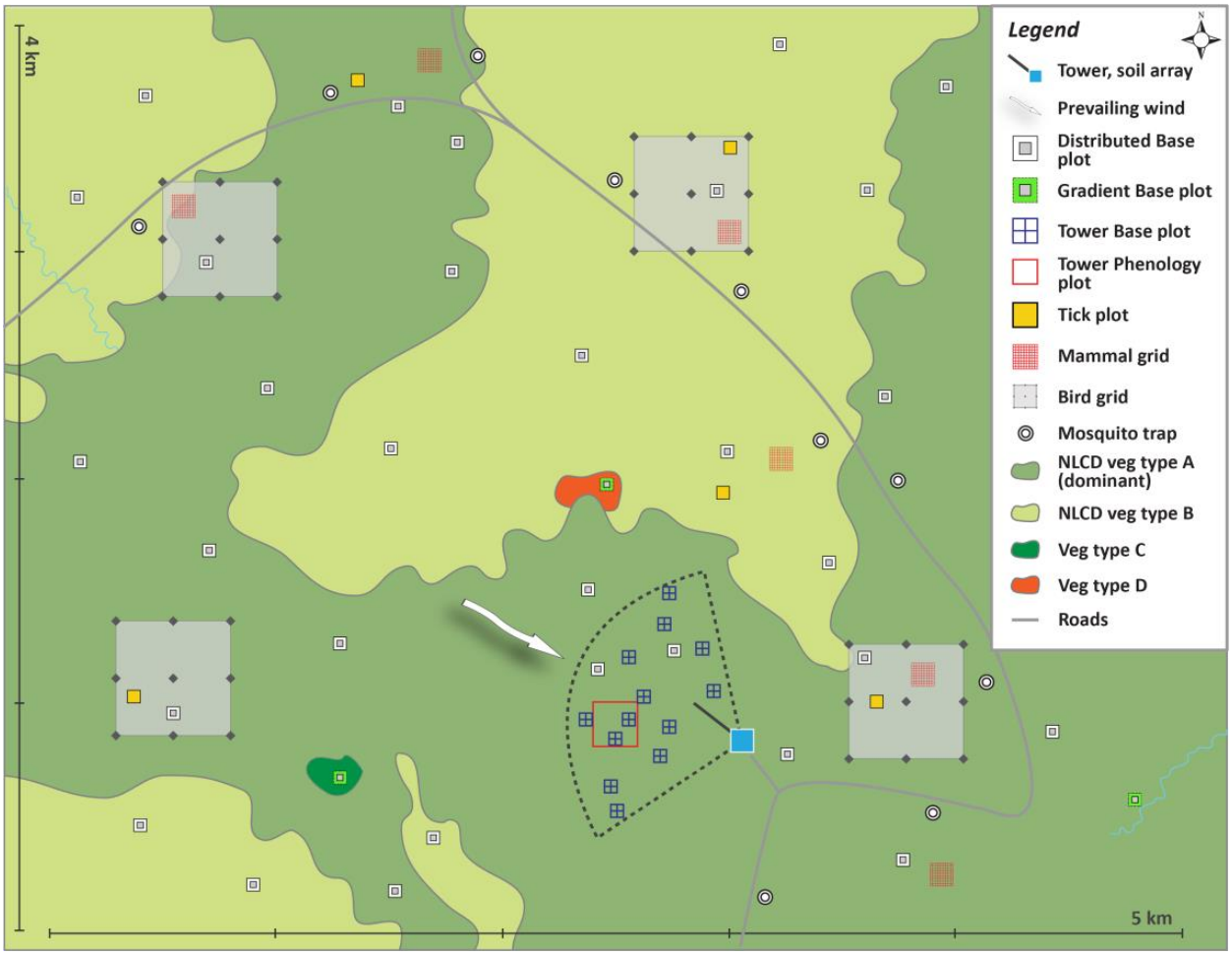

Figure 3: Generalized site-scale schematic for the NEON Terrestrial Observation System (TOS). Figure credit: NEON

The plot locations are determined using a stratified, spatially balanced random design (Barnett et al., 2015). The spatially balanced random design establishes varied spatial distances between plots, allowing for better analysis of spatial patterns associated with various ecological processes. Stratification by vegetation type allows efficient observations of the variability across a given site. National Land Cover Database (NLCD) data (Fry et al., 2011) are used to stratify by vegetation land cover classes. NLCD provides a high-level land cover classification scheme that is appropriate for the entire continent, allowing a consistent approach for selecting TOS plot locations for all NEON sites. Next, a Reversed Random Quadrat-Recursive Raster (RRQRR) algorithm (Theobald et al., 2007) identifies randomly located potential plots within each vegetation class. The final plot locations are determined by field crews based on if they are the correct vegetation type and if they are logistically feasible to sample. An example of the RRQRR algorithm for plot location selection is shown in Figure 4:

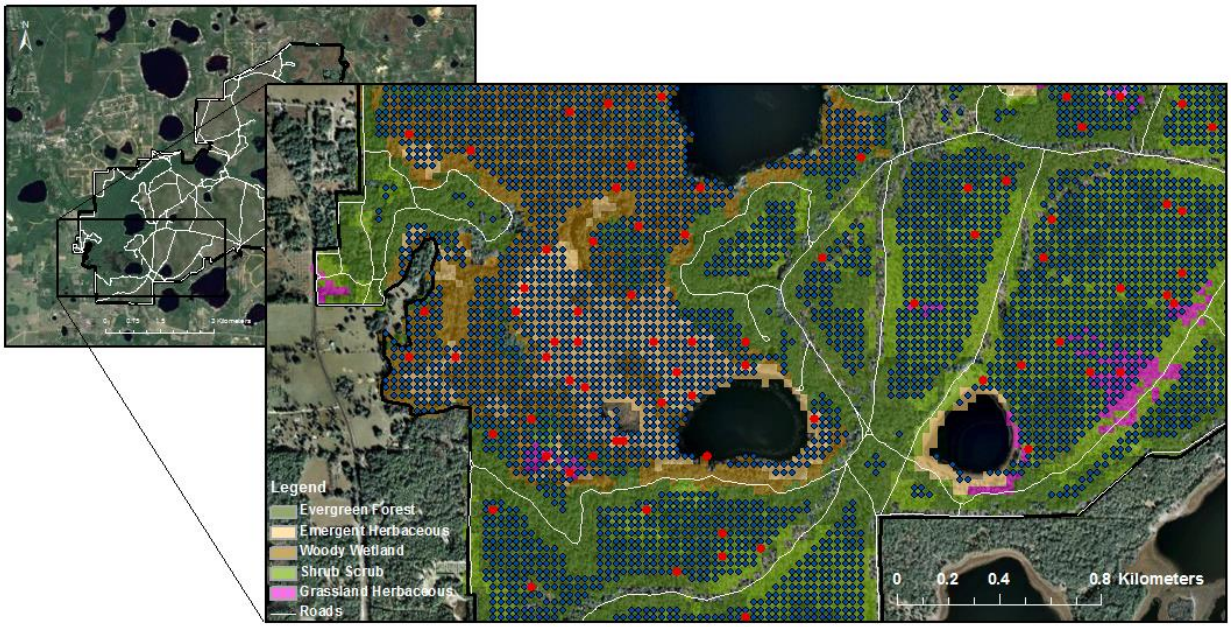

Figure 4: The spatially balanced and randomized sample at the Ordway-Swisher Biological Station. Red points indicate plots selected from the complete sample (blue points). Plots are selected from a spatially-balanced one dimensional list that is filtered by vegetation type. Figure credit: NEON

NEON has chosen the plot design to allow sampling at multiple spatial scales in order to improve understanding of processes operating at different spatial extents, address problems associated with detectability of species, facilitate an understanding of local and regional patterns of diversity, and provide data to validate airborne observations at multiple grain (i.e. pixel) sizes (Barnett, 2015). A multi-scale plot design adapted from the Carolina Vegetation Survey (CVS, 2015; Peet, Wentworth, & White, 1998) is used to sample plant species in the tower and distributed plots and consists of a 20 m x 20 m square plot comprised of four modules with nested subplots in each module. This plot design allows scalability such as in the case of sampling vegetation structure over a 40 m x 40 m area. A schematic of the plot design is shown in Figure 5:

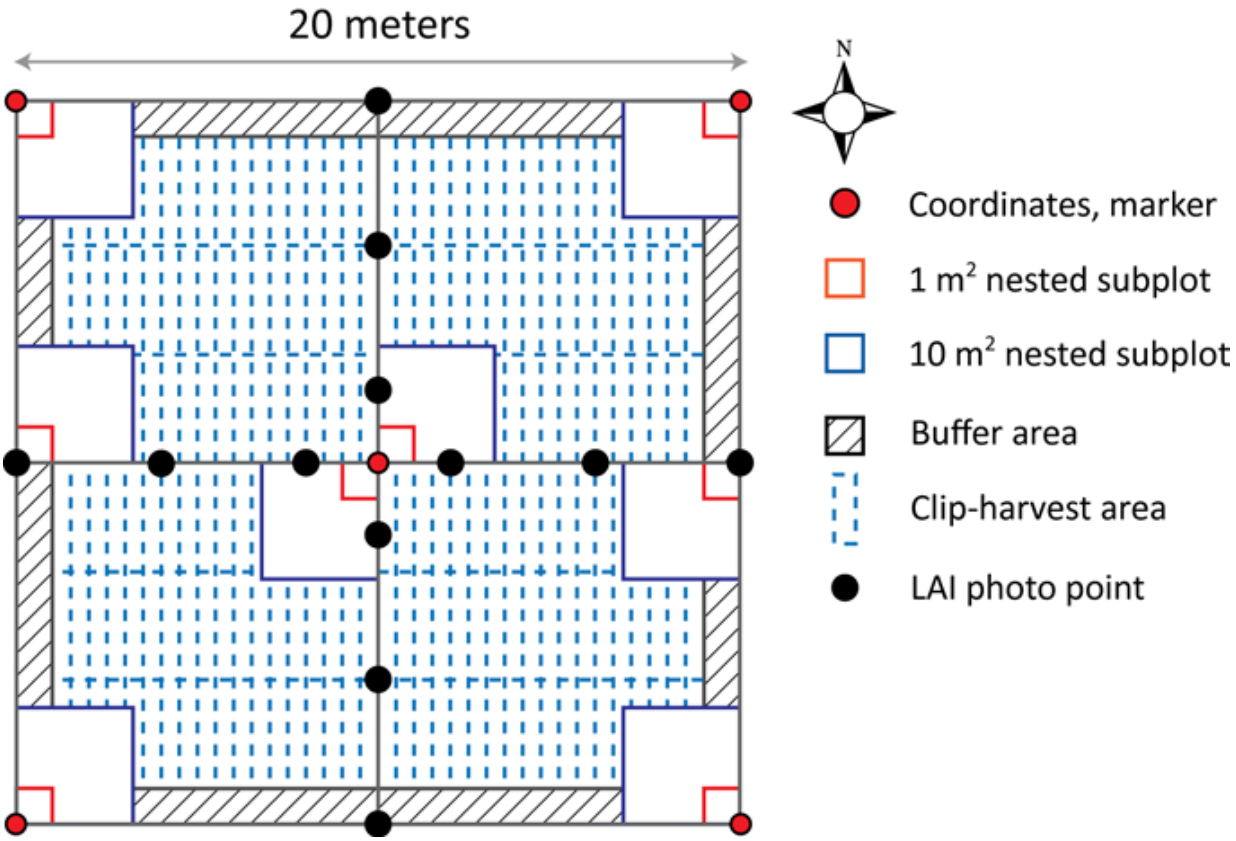

Figure 5: Schematic of the TOS tower and distributed plot design showing the sampling area in which plant biogeochemistry, biodiversity, above-ground biomass, productivity, and leaf area index (LAI) are sampled. Figure credit: NEON

Nested subplots are used to estimate percent cover and richness of herbaceous species. Herbaceous clip-harvesting is excluded from nested subplots in order to minimize the impact of clipping on plant biodiversity data products. Vegetation biogeochemistry (Hinckley, 2014), plant phenology (Elmendorf & Jones, 2015), plant diversity (Barnett, 2014), and plant biomass, productivity, and leaf area index (Meier & Jones, 2014) are sampled by the TOS. Except for plant penology, all of the other measurements occur at both the tower and distributed plot locations.

NEON will quantify variation in plant tissue chemistry to provide valuable information about nutrient uptake and storage in terrestrial ecosystems, both above- and belowground (Hinckley, 2014). Samples of sun-lit canopy foliar tissues will be collected in tower and distributed plots at NEON sites during peak biomass and coincident with data collection by the airborne team. Sun-lit canopy foliar samples will be collected from 10-20 plots per site (i.e., four tower plots and 6-16 distributed plots), depending on the variability of the vegetation communities at each site. Ground-based collections for biogeochemistry will include sampling and analysis of total C and nutrient concentrations (N, P, S, Ca^{2+}, Mg^{2+}, K^{+}, and Al^{3+}) in sun-lit foliage of dominant and co-dominant canopy species, and C and N of litter and roots. Stable isotopic composition ($\delta^{13}C$ and $\delta^{15}N$) and other chemical characteristics such as chlorophyll content and lignin plus leaf mass per area (LMA) will be measured. At grassland sites, samples (species bulked) will be obtained during aboveground biomass clip harvests. The ground-based plot-level assessments of plant chemical characteristics will be used to scale from plot-level to site wide maps of biogeochemistry derived from the hyperspectral airborne remote sensing data.

Phenology is defined as the seasonal timing of life cycle events (Elmendorf & Jones, 2015). Phenology is a simple way to track changes in ecology in response to climate change, for example, monitoring the annual timing of breaking bud leafs each spring. During the first three years of operations, NEON will conduct intensive monitoring of the phonological stages of the three most dominant species at the site. This includes monitoring approximately 10 individuals of each species 2-3 times per week during the canopy development stages (spring during green-up and in the fall during senescence). Less frequent observations will take place during the flowering stages of plants. After the first three years of operations, limited sampling of up to 20 species per site will take place. In order to facilitate linkages between ground-based measurements and landscape greening metrics, NEON's most intensive measurements will occur at the Tower and surrounding airshed. Phenology observations will be performed on marked individuals that are evenly situated along a fixed, 800 meter square 'loop' transect (200 meters on a side)

within the tower airshed. A phenocam is located on each TIS tower and pointed in the North direction. In the case where the in-situ phenology transect loop is not located within the view of the phenocam, an additional three individuals of each of the dominant species that are visible from the phenocam but offset from the phenology transect will also be monitored in order to make explicit linkages between phenocam greenness metrics and in-situ phenophase observations.

Species composition drives ecosystem structure and function and species richness has been shown to be important to functional diversity and ecosystem multifunctionality (Maestre et al., 2012). NEON observations of plant diversity will focus on plant species presence, diversity, and abundance and the collection of plant material for genetic analyses (Barnett, 2014). Co-location of species composition and abundance with other NEON measurements and observations will allow analysis of the causes and consequences of changes in plant diversity and the relationship with ecosystem structure and function. Plant diversity information will be collected in the tower and distributed plots on:

- the species present and species composition in each subplot across each module,
- estimates of herbaceous abundance at the 1 m^2 and 10 m^2 subplots,
- and the abundance of woody species as estimated from the location and size of each woody species in each 10 x 10 m module.

One high-level component of the plant biomass and productivity sampling design is to measure above-ground biomass (AGB) stocks, vegetation structure, and LAI across the range of variability at each NEON site. Developing a better understanding of the spatial distribution and magnitude of above- and below-ground plant biomass stocks and fluxes is critical to reducing uncertainty in large-scale models of the carbon cycle (Meier & Jones, 2014). Measurements of plant biomass and productivity in the tower plots complement the TIS measurements of net ecosystem exchange. Vegetation structure is measured in

a 40 m x 40 m plot in forests to ensure enough individuals are sampled to get an accurate biomass estimate. The following measurements are made:

- Herbaceous biomass: 20-30 tower plots sampled 1-2 times per year. Mixed C3/C4 grasslands will be clipped 2X per year, other veg types clipped 1X per year, and grazed sites clipped every 4 weeks. Sampling of up to 20 distributed plots clipped once every three years (only clip plots with NLCD class ≠ Forest).
- Leaf area index (LAI): 3 tower plots sampled every 2 weeks during the growing season and 20 distributed plots sampled in coordination with AOP flights once every 5 years. A digital hemispherical photo system is used. Photos are taken along the center axes of the plot.
- Vegetation structure of woody stems: 20-30 tower plots measured once per year. Sampling of up to 20 distributed plots once every three years. These measurements include: stem diameter(s), canopy diameter(s), stem height, stem location (for stems that meet certain criteria), stem species identification, and stem status (i.e. healthy, dead, or damaged).

Above-ground woody biomass will be estimated from allometrical equations using the structure data and productivity will be calculated from changes in above-ground biomass with time (Meier & Jones, 2014). Herbaceous biomass and productivity will be assessed by measurements of the clip harvests.

This research uses TOS plant diversity data and ground photos collected at the Ordway-Swisher Biological Station in 2010 and 2011. Refer to Chapter III for a description of the vegetation communities at the site and Chapter IV for a detailed description of the NEON data used in this research.

Remote Sensing with the Airborne Observation Platform

One of NEON's goals is to provide detailed aerial data about regional landscapes and vegetation. This will be accomplished via the remote sensing arm of NEON, called the Airborne Observation Platform (Kampe et al., 2010). Airborne remote sensing plays a critical role in the scaling strategy of the Observatory by making measurements at the scale of individual shrubs (1–3 meters), larger plants and forest canopies (over hundreds of square kilometers). The NEON airborne remote sensing instrumentation is designed to bridge scales from organism and stand scales, as captured by plot and tower observations, to the scale of satellite based remote sensing. The airborne instrumentation will support research on a range of important themes in ecology in response to grand challenges in the study of biodiversity, biogeochemistry, climate change, ecohydrology, infectious disease, invasive species and land use change. Each aircraft will fly a suite of integrated remote sensing instruments consisting of an imaging spectrometer, a waveform light detection and ranging (lidar) instrument and a high-resolution digital camera. The AOP instrument suite will provide regional observations of: land use, vegetation structure, biochemical and biophysical properties of vegetation, and ecosystem responses to changes in land use, climate and the movement of invasive species.

The NEON Imaging Spectrometer (NIS) is a hyperspectral imager designed and built by NASA JPL and is part of the next generation class of AVIRIS instrument (JPL, 2015b). The NIS is a pushbroom sensor that collects data in 424 spectral bands covering the visible to shortwave-infrared (380-2510 nm) portion of the electromagnetic spectrum. A single 640 x 480 HgCdTe focal plane is used to collect data over the entire spectral range and a specially designed diffraction grating disperses the wavelengths of light (Kampe et al., 2011). The 34 degree field-of-view is covered by 600 effective across-track pixels. The telescope views a single row on the ground by imaging a slit onto the focal plane. The instrument's

telescope, spectrometer, and focal plane are enclosed in a vacuum chamber that is cooled to 140 K for increased performance. A schematic of the instrument is shown on the left side of Figure 6 and a picture of the NIS on a test fixture in the lab is shown on the right. The instrument specifications are listed in Table 1:

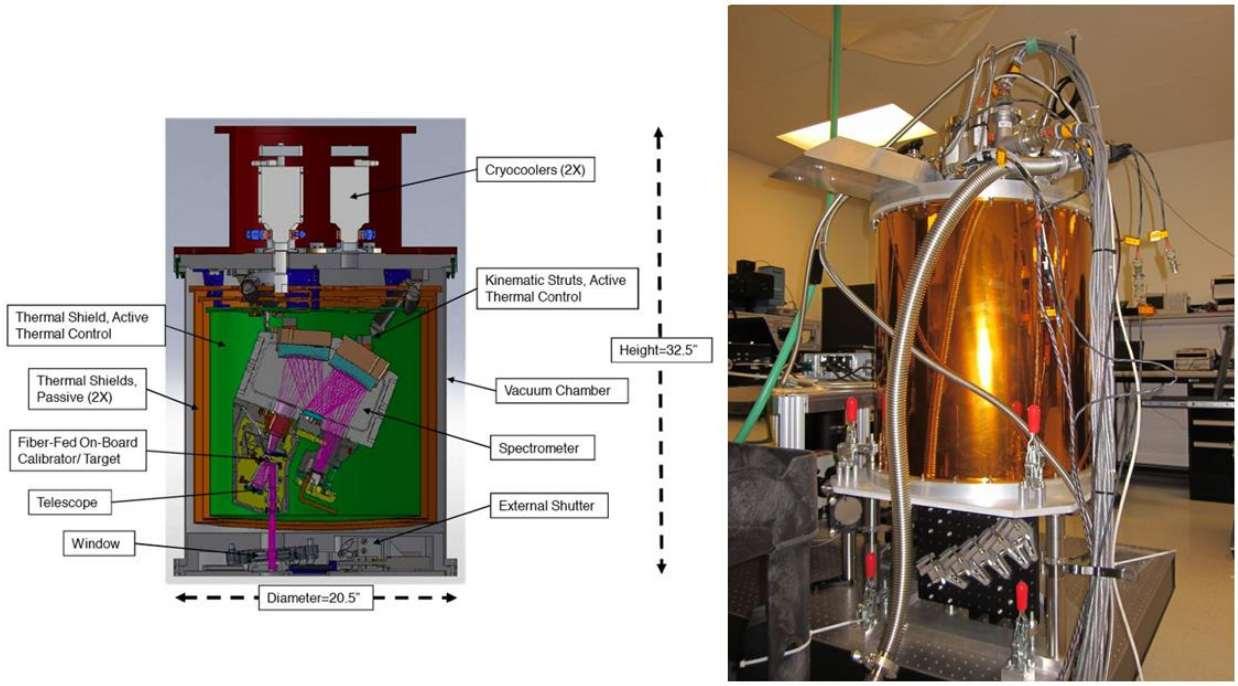

Figure 6: Schematic of the NEON Imaging Spectrometer on the left (figure courtesy of NASA JPL) and a picture of the instrument in the lab on a test fixture on the right

Table 1: Specifications of the NEON Imaging Spectrometer

Parameter	Value
Wavelength Range	380 – 2510 nm
Spectral Sampling (band center spacing)	5 nm
Spectral Resolution (FWHM)	5.7 – 6.9 nm
Field of View	34 degrees with 600 resolved elements
Instantaneous Field of View	1.0 mrad
Spectral and Spatial Response Uniformity	> 95%
Signal to Noise Ratio	> 1000 at 550 and 2100 nm
Focal Plane Array	640 (across-track) x 480 (spectral)
Detector Size	27 µm x 27 µm
Frame Rate	10 – 100 frames per second
Data Resolution	14 bits
Absolute Radiometric Calibration Uncertainty	< 5% (1σ)

The lidar instrument is an Optech Gemini which is a pulsed Nd:YAG laser operating at a wavelength of 1064 nm (Optech, 2010a). The Gemini uses a scanning mirror to sweep out the laser pulses in the across-track direction in a saw tooth pattern on the ground. The mirror scan angle can be configured to set the full field of view of the instrument. The pulse spacing on the ground can be tuned by setting the values for the pulse repetition frequency (PRF), the mirror scan angle, and the mirror scan frequency. The system simultaneously extracts discrete returns and records full digital waveforms by digitizing the analog signal. The Optech Gemini is shown in Figure 7 and the instrument specifications are listed in Table 2:

Figure 7: Optech Gemini full waveform lidar (photo courtesy of Optech)

Table 2: Specifications of the Optech Gemini Full Waveform Lidar

Parameter	Value
Laser Type	Pulsed Nd:YAG
Laser Wavelength	1064 nm
Laser Power	10 W
Laser Pulse Energy	250 μJ
Pulse Repetition Frequency	Programmable: 33, 50, 70, 100, 125, 142 or 166 kHz
Outgoing Pulse FWHM	16 nsec at 100 kHz PRF
Beam Divergence	Dual Divergence: 0.25 mrad (1/e) and 0.8 mrad (1/e)
Mirror Scan Angle (Field of View)	Programmable: 0-50 degrees
Mirror Scan Frequency	Programmable: 0-70 Hz
Discrete Return Range Capture	Up to 4 range measurements, including 1^{st}, 2^{nd}, 3^{rd}, and last returns
Discrete Return Intensity Capture	Up to 4 intensity returns for each pulse, including last (12 bit)
Waveform Digitization	12 bit
Waveform Sampling	1 nsec bins
Operation Altitude	150-4000 m AGL
Horizontal Geometric Accuracy	1/5500 * altitude (AGL)
Elevation Accuracy	< 5-35 cm (1σ)

The high spatial-resolution color camera is an Optech DiMAC D8900 and is a 60 megapixel medium format digital camera that captures true color images (Optech, 2010b). The camera provides detailed aerial views of the regional landscape and can be used to determine land cover and provide context as to what ground targets are being measured by the imaging spectrometer and lidar instruments. The Optech D8900 is shown in Figure 8 and the instrument specifications are listed in Table 3:

Figure 8: Optech DiMAC D8900 high spatial-resolution digital camera

Table 3: Specifications of the High Spatial-Resolution Digital Camera

Parameter	Value
Total Pixels	60 MP Full Frame CCD
2D Array Pixel Dimensions	8984 x 6732 pixels (effective)
Pixel Size	6 µm x 6µm
Array Physical Dimensions	53.9 x 40.4 mm (effective)
Lens	70 mm
Shutter	Electro-mechanical: 1/250 to 1/500 sec exposure
Aperture	Electro-mechanical iris: f-stops: 4, 5.6, 8, 11, 16
Filter	Standard size RGB
Image Output	8 or 16 bits per channel
Capture Rate	1 shot every 2.5 sec
Forward Motion Compensation	Solid-state electro-mechanical

As part of the AOP, the airborne remote sensing instruments will observe ecosystem properties in the region (100-300 km^2) surrounding NEON sites. The high cost of aircraft operations will limit the frequency of standard airborne observations of each site to once per year to detect inter-annual trends. To minimize the phenological contribution to the signal, flights will be planned to occur at each site during a period of peak greenness. Higher temporal frequency data on vegetation function is available from satellite measurements at a coarser spatial resolution. Two payloads are necessary to cover all of the NEON sites based on the deployment and data collection requirements. An example of the flight deployment is shown in Figure 9:

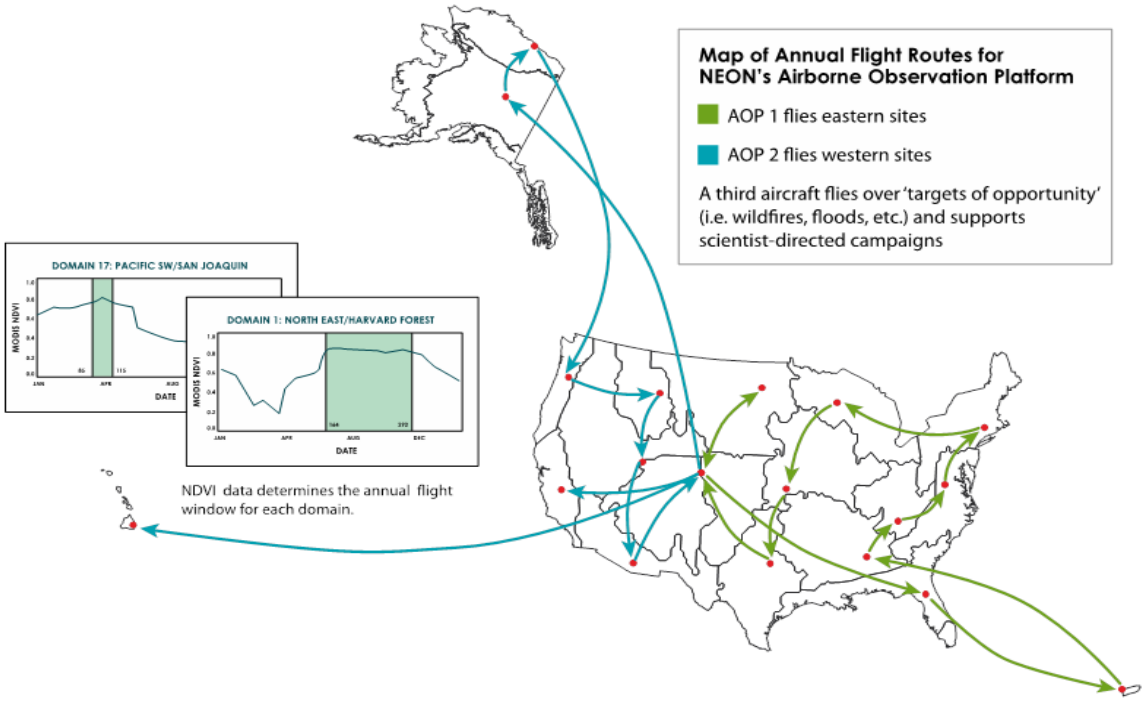

Figure 9: Deployment of two Airborne Observation Platforms to cover each of the NEON sites once per year during peak greenness. Figure credit: NEON

A third dedicated payload will also be deployed in response to extreme events to monitor the ecological forcing (e.g., hurricane damage) and response (regrowth after fire), as well as other science directed requests. The remote sensing instruments are mounted in a common rigid frame to improve co-registration between the data sets. Figure 10 shows a side view of the instruments in the payload mount on the left and a view from underneath the aircraft on the right.

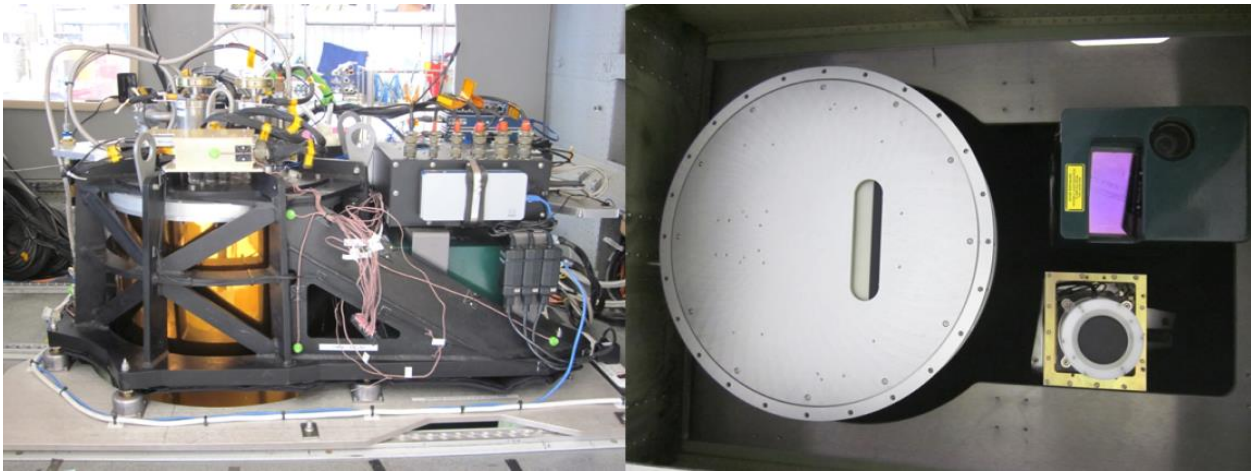

Figure 10: The instruments are mounted in a common rigid frame to improve data co-registration. A side view shown on the left and the view of the instruments from underneath the plane is sown on the right

The payload is installed in a DeHavilland DHC-6 Twin Otter aircraft (Figure 11) flying at 1000-2000 m above ground level (AGL) at a speed of around 97.2 knots (50 m/s).

Figure 11: The instrument payload is installed in a DeHavilland DHC-6 Twin Otter shown on the left and data are collected at a nominal flying altitude of 1000 m above ground level as shown on the right

The flight parameters are largely determined by the requirement to achieve meter-scale spectroscopy measurements and simultaneously achieve a sufficient signal-to-noise ratio to retrieve vertical structure from the waveform lidar and biogeochemical properties from measured reflectance spectra.

This research uses both hyperspectral and lidar data collected by AOP at the Ordway-Swisher Biological Station in May 2014. Refer to Chapter III for a description of the vegetation communities at the site and Chapter IV for a detailed description of the NEON data used in this research.

CHAPTER II

OBJECTIVES AND EXPECTED SIGNIFICANCE

This research falls under the general science themes of biological diversity and land-use with specific research themes of observation of pattern, new technology evaluation, and development of analysis tools. The focus is on techniques for mapping vegetation communities which are an appropriate natural study unit for biodiversity (Harper & Hawksworth, 1995). The vegetation community map serves as a proxy for species richness, and when combined with field data, allows the generation of a species richness map. The result is a description of species pattern (Magurran, 2004) and not a process study.

Remote sensing is a useful technology for observing vegetation biodiversity over large spatial scales (Turner et al., 2003). Imaging spectrometers measure the spectral variability across the landscape. Hyperspectral data from these instruments can be correlated to chemical, functional, or structural characteristics of the vegetation being imaged (Asner, 1998; Asner & Martin, 2009; Asner et al., 2009; Asner et al., 2015; Curran, 1989; Dahlin et al., 2013; Gamon et al., 1992; Gamon & Surfus, 1999; Kokaly et al., 2009; Lefsky et al., 2002; Martin et al., 2008; Ollinger, 2011; Townsend et al., 2003; Townsend et al., 2008; Ustin et al., 2004; Ustin et al., 2009; Ustin & Gamon, 2010; Wessman et al., 1988; Wessman, 1990). Lidar instruments measure the intensity of each laser return pulse as a function of time/distance. The lidar data represents the 3D structural variability across the landscape (Lefsky et al., 2002). Integration of both hyperspectral and lidar data will improve the ability to characterize biodiversity using remote sensing. The NEON Airborne Observation Platform (AOP) consists of cutting edge hyperspectral and lidar instruments collecting data at a ground sample distance of 1 m or less. Access to integrated hyperspectral and lidar data at this high spatial resolution has been very limited in the past and so too has been corresponding algorithm development. While AOP has the potential to significantly advance the

field of remote sensing, the technologies are new and pose some challenges requiring extra processing in order to effectively work with the data.

The Ordway-Swisher Biological Station (OSBS) is chosen as the study site for this research. OSBS has a wide variety of vegetation communities including closed canopy hardwood forests, savannahs of widely spaced pines, areas of altered land use, hardwood and marsh wetlands, plus lakes representing a variety of chemical, functional, and structural types. The OSBS landscape is dominated by the Sandhill vegetation community, which is a variant of the Longleaf Pine ecosystems that once dominated the Southeastern U.S. OSBS maintains GIS data of the vegetation communities at the site, but the accuracy of this data is not known. The goal of this research is to map the vegetation communities using the AOP remote sensing data collected in May of 2014. The NEON Terrestrial Observation System (TOS) also conducted observations of species richness at ground plots across the site in 2010 and 2011. The TOS data serves as independent ground truth validation points to assess the accuracy of the vegetation community maps generated from the AOP data. The OSBS vegetation communities are described in detail in Chapter III including the relevance of spectral or structural differences that can be measured by the remote sensing data. The NEON AOP and TOS data collection, the AOP Level-1 processing steps, and the classification algorithm used in this research are described in Chapter IV.

The spectral variability hypothesis proposes that variability in brightness or reflectance in remote sensing imagery is correlated to the variability of the species in the scene (Palmer et al., 2000; Palmer et al., 2002). While this theory has been demonstrated in practice in a diverse closed-canopy tropical ecosystem (Carlson et al., 2007), it is not clear if their methods can be universally applied to other temperate and savannah systems due to illumination geometry and shadowing, pixel size, and inability of the imaging spectrometer to see the understory and ground vegetation (in the case of a closed-canopy).

Further investigation of the approach of Carlson et al. to biodiversity mapping is required but is not performed in this research and instead is listed as future work. This research takes a more direct approach to the use of remote sensing to study biodiversity. A classification and regression tree (CART) algorithm (Breiman, Friedman, Olshen, & Stone, 1984; Therneau, Atkinson, & Ripley, 2015) is used to generate a vegetation community map. A series of metrics are calculated from the remote sensing data which characterize the spectral and structural variations across the landscape of the study site. The metrics serve as the input feature layers to the classification tree but can support future tools to analyze spatial patterns and inform ground sampling of vegetation diversity. Where possible, inferences are made linking the differences in the remote sensing data (between different vegetation communities) to ecologically meaningful chemical or structural traits. Classification results for hyperspectral data only, lidar data only, and integrated hyperspectral plus lidar data are given in Chapters V-VII respectively. The vegetation community map serves as a proxy for species richness, and when combined with field data, allows the generation of a species richness map (a local scale biodiversity indicator) as discussed in Chapter VIII.

In summary, the research is an evaluation of remote sensing technologies and development of analysis tools to describe the spatial pattern of biodiversity in the form of vegetation communities as proxies for species richness. The project objectives are described next in the first section, knowledge gaps are highlighted in second section, and the expected significance of the research are given in third section of this chapter.

Overview of Objectives

Seven focused objectives that fall within the high-level research themes have been identified for this research:

- Objective 1: Investigate remote sensing technologies to map ecosystems.
- Objective 2: Generate a vegetation community map from remote sensing data. Validate the map with in-situ ground truth data.
- Objective 3: Show that integrated hyperspectral (spectral, chemical, species composition and function) and lidar (vertical structure) data improves the accuracy of the landscape characterization algorithm compared to data from a single instrument.
- Objective 4: Start to relate derived remote sensing features to structure, chemistry, and function.
- Objective 5: Leverage the collection of pixels within a ground plot.
- Objective 6: Analyze the key spectral and structural differences between vegetation communities as measured in the remote sensing data.
- Objective 7: Relate the classification results to a local scale biodiversity indicator.

The objectives are described in detail in the following paragraphs.

Objective 1: Investigate remote sensing technologies to map ecosystems.

Perform a literature search to identify previous research using remote sensing for land cover mapping and determine specific examples where the land cover classes represent ecosystems. Identify research that specifically links the remote sensing data to vegetation chemical, functional, or structural traits. Determine what statistical metrics have been calculated by others for use in classification algorithms. Finally, understand the current state of research and future desires for using remote sensing for biodiversity studies.

Objective 2: Generate a vegetation community map from remote sensing data. Validate the map with in-situ ground truth data.

The primary output product of this research is a land cover map that uses regional vegetation communities as the land cover classes. A classification and regression tree (CART) algorithm (Breiman et al., 1984; Therneau et al., 2015) is used to generate the vegetation community map. A series of metrics are calculated from the remote sensing data which characterize the spectral and structural variations across the landscape of the study site. The metrics serve as the input feature layers to the classification tree. Training data points for each of the vegetation communities are hand selected using photo-interpretation of high resolution imagery in combination with GIS data of the study site. The classification tree is generated using the training data and then validated using independent in-situ ground truth data by comparing the vegetation community class predicted by the classification tree against the actual vegetation community as determined by species lists and site photos from field sampling. Vegetation community maps are presented in Chapters V-VII along with corresponding accuracy

assessments using both the training data (with k-folds cross-validation) and the independent ground truth data.

Objective 3: Show that integrated hyperspectral (spectral, chemical, species composition and function) and lidar (vertical structure) data improves the accuracy of the landscape characterization algorithm compared to data from a single instrument.

Classification trees are generated using the hyperspectral instrument data alone (Chapter V) and lidar instrument data alone (Chapter VI). Based on the accuracy assessments, misclassification can occur in areas on the ground that have similar spectral or structural characteristics. If areas that are spectrally similar have different structure, or vice-versa, then combining the instrument data from the two sensors will improve the classification map. A classification tree generated using the integrated hyperspectral and lidar data is discussed in Chapter VII and shows an improvement in classification accuracy for the training data.

Objective 4: Start to relate derived remote sensing features to structure, chemistry, and function.

The input feature layers for the classification tree are statistical metrics calculated on the remote sensing data. At first glance, many of these features are specific to the remote sensing data themselves and do not immediately appear to be associated with ecologically meaningful features such as structure, chemistry, and function. For example, an input feature of the mean reflectance at a specific wavelength is specific to the hyperspectral data. If this wavelength happens to correspond to a wavelength of light that is

absorbed by chlorophyll, for instance, then this mean reflectance is now a proxy for canopy chlorophyll, which is a relationship of reflectance to canopy chemistry. Where possible, potential relationships are identified in Chapters V-VII between calculated remote sensing features (that are using by the classification tree) and physical structure, chemistry, or function.

Objective 5: Leverage the collection of pixels within a ground plot.

Just as vegetation communities are assemblages of species, a study area on the ground can be observed by an assemblage of remote sensing pixels or points. In this research, hyperspectral data are collected with a 1 m instantaneous field of view (IFOV) plus 1 m spacing pixels and the lidar has 0.8 m IFOV spots with a spacing of approximately 0.5 m. If analysis is performed at a 20 m ground plot scale, then an assemblage of 400 hyperspectral pixels and 1600 lidar pulses can be used to calculate assemblage input features, for example, fraction of vegetation pixels from hyperspectral data or fraction of ground points from lidar. While previous research has used pixel statistics such as standard deviation and range, little has been done in terms of thinking of groups of pixels as an assemblage of spectra or lidar points as an assemblage of vertical structure. Some of the metrics calculated as input feature layers for the classification trees in Chapters V-VII attempt to utilize the remote sensing data in this assemblage framework.

Objective 6: Analyze the key spectral and structural differences between vegetation communities as measured in the remote sensing data.

The classification tree algorithm generates splitting decisions based on the ability to correctly classify the training data class as efficiently as possible. The input feature layers used by the classification tree for the splitting decisions provide insight into similarities or differences in the data between different vegetation communities. Part of the results discussion in Chapters V-VII will address the differences in the spectra and structure of the different vegetation communities and areas of misclassification due to similarities in the remote sensing data.

Objective 7: Relate the classification results to a local scale biodiversity indicator.

The Grand Challenges in Environmental Sciences (NRC, 2001) call for new tools of rapid assessment of biodiversity. A local scale biodiversity indicator can easily be derived from the remote sensing data and ground observations of biodiversity. In this research, the biodiversity indicator used is species richness. A relative species richness estimation map is generated from the vegetation community classification results and described in Chapter VIII.

Knowledge Gaps

Remote sensing technologies are being used more frequently to monitor biodiversity, however, previous research has only scratched the surface of what might be possible. Four specific themes have emerged as having limited use and where significant gains could be made:

- vegetation communities as land cover classes,
- integration of hyperspectral and lidar data,
- spectral variability hypothesis applied to different, diverse ecosystems,
- and thinking of remote sensing data as an assemblage of pixels or points.

While a significant amount of research has been performed on using remote sensing for land cover classification, very rarely do the resulting land cover classes represent ecosystems, let alone specific vegetation communities. For instance, the National Land Cover Database (Fry et al., 2011) uses only 20 classes (13 are vegetation) to describe the entire nation and only four categories to describe forests: deciduous, evergreen, mixed, or woody wetlands. These vague descriptors may be successful for tracking land cover changes with time but say nothing about the chemistry, structure, and functioning of the vegetation. More specific land cover classes are needed as some forest communities are more threatened than others and cannot be monitored if general land cover classification is used (Strittholt & Steininger, 2007). Additionally, the methods presented here will enable the detection of ecosystem disturbance or degradation in areas where no land cover change has taken place (Rose et al., 2015).

Many previous studies have used a combination or fusion of hyperspectral and lidar data. It is unclear how effectively the data from the two sensors have been merged. Typically, the data are fused at

the feature layer, or more specifically, features are calculated individually for each sensor and then the resulting statistics are combined as inputs to a machine learning algorithm. The instruments on the NEON AOP view the same area on the ground plus the hyperspectral data is co-registered to the lidar data as part of camera model generation and geolocation which should easily allow more complex data fusion techniques.

Carlson et al., 2007 demonstrate a very successful application of the spectral variability hypothesis (Palmer et al., 2000; Palmer et al., 2002) by generating maps of species richness using the variability of spectral reflectance within a defined ground plot. Links to biogeochemistry are proposed though the study uses the derivative-reflectance range in the remote sensing data to correlate against species richness and does not actually derive chemistry maps from the remote sensing. Furthermore, this study takes place in Hawaii in a forest with a diverse canopy. Individuals from the species richness data can be easily observed in the remote sensing data. It is unclear how applicable the techniques and results from this study are to other regions of the U.S. or other types of ecosystems. One general concern about biodiversity regression using spectral variability is that major sources of spectral variability in remote sensing data are the illumination geometry, shadowing, and pixel size. These non-biodiversity causes of spectral variability led to biases and trends in derived species richness maps during initial data exploration for this research, for example the results had an artificial trend in richness in the across-track direction of the pushbroom sensor data. Another major issue is that hyperspectral sensors are measuring the reflected solar radiation primarily from the top of a canopy in the case of trees. The hyperspectral data includes no information about what exists on the ground underneath the canopy. In the case of certain ecosystems, such as the Sandhill ecosystem (FNAI, 2010), the herbaceous ground layer contains the most vegetation diversity. Unfortunately, the spectral variability in the hyperspectral data would not properly detect the diversity and subsequent species richness maps would underrepresent the true number of vegetation species.

In the past, access to high-spatial resolution (on the order of 1 m sampling distance) hyperspectral and lidar data has been limited. With advances in technology used in recent airborne observatories such as the NEON AOP (Kampe et al., 2010) with free and open data policies, these types of data are becoming more prevalent. Many previous studies have used summary statistics as input features for classification algorithms, for example: mean reflectance at 550 nm or mean, stdev, min, and max of canopy height. New data sets from AOP present an opportunity to begin thinking and describing the data as a collection of pixels or points. Just as vegetation communities are assemblages of species, remote sensing data are assemblages of pixels or points. For improved vegetation community mapping, new techniques need to be developed that analyze the remote sensing data as assemblages.

Research Significance

This research plays a role in efforts to conserve and restore longleaf pine forests to the Southeast U.S. The Range-Wide Conservation Plan (America's Longleaf, 2009) calls for use of remote sensing data to identify and map longleaf pine stands across the historical range of the species. The localized maps of the Sandhill ecosystem at Ordway-Swisher Biological Station achieve that goal. Future work to extend the use of the classification algorithms to other sites would be greatly beneficial to longleaf conservation efforts. The research also satisfies the desire to map ecosystem extent enabling future efforts to create an extent biodiversity indicator as well as ecosystem integrity and condition (Strand et al., 2007b).

Three high-level research themes guided the work presented here: observation of pattern, new technology evaluation, and development of analysis tools. The outputs of this research are vegetation

community class maps and a species richness map (a local biodiversity indicator). In order to generate the class maps, a series of metrics are calculated on the hyperspectral and lidar data. The spatial maps of these metrics form observations of the spectral and structural pattern of the landscape of the study site. These maps could also be used to inform the NEON TOS ground sampling strategy in order to capture the spatial variability across the site. The NEON AOP contains cutting edge remote sensing instruments, especially the NEON imaging spectrometer. The imaging spectrometer has a 5 nm spectral resolution and a 1 m ground sample distance. These specifications allow analysis of the data in different and news ways than previous remote sensing instruments, but also pose some challenges for working with the new data. Several processing steps are required to adjust the data and prepare it to be used to calculate the spectral and structural metrics. This research also attempted a new interpretation of how to use the spectra. The full hyperspectral data is reduced to 32 key wavelengths and a full set of normalized difference indices are calculated and used as input features in the classification algorithm. Working with a 20 m x 20 m plot scale of 1 m pixels, the remote sensing data can be thought of as an assemblage of pixels. New analysis tools are developed in this research to take this concept and apply it in the form of new metrics (for example fraction of green vegetation pixels) that serve as the input features in the classification algorithm. Other metrics used in this research have a potential to be directly related to physical spectral (or chemical) or structural characteristics of the landscape that are ecologically meaningful. Future research could improve upon these relationships and allow analysis of structure and function across the site. Ordway-Swisher Biological Station has a large amount of variability in the vegetation communities located at the site which serves as an excellent study area to develop analysis tools to relate observed remote sensing metrics to biophysical traits.

Once NEON becomes fully operational, the combination of the airborne remote sensing with ground sampling of foliar chemistry, vegetation diversity, and vegetation structure will allow advances in analysis of spatial patterns, scaling approaches from ground point to regional scales, as well as process

studies that can understand the relationships of cause and effect of biological change. The 2014 remote sensing data used in this research serves as baseline for long-term change studies.

With respect to the knowledge gaps identified in the previous section, this research is significant in that it uses specific vegetation communities as the land cover classes, allowing a more detailed understanding of the species composition that exists in each region of the classification map. Hyperspectral and lidar data are integrated at the feature layer, a practice that is common in the literature. The AOP data are co-registered enabling more advanced fusion techniques. Future research will include combining the data at earlier stages of processing, for instance using canopy height to stratify the hyperspectral pixels into sub-groups within the 20 m x 20 m plot, and then calculating metrics on each subgroup (for example separately calculating the mean reflectance of canopy vegetation versus ground vegetation). The spectral variability hypothesis was successfully applied to a closed-canopy system (Carlson et al., 2007) with a high degree of woody species diversity. A direct implementation of that method to the data collected at the Ordway-Swisher Biological Station is not successful because of open canopy savannah type ecosystems and the fact that the most vegetation diversity is in the herbaceous ground layer which may not be observable in the remote sensing data. For this research, vegetation communities are mapped and used as a proxy for species richness. Future work will include more pre-processing to remove the effects of illumination that may enable the spectral variability hypothesis to be applied at this site and more analysis of spectral and structural variability is needed to correlate the remote sensing data directly to species richness and diversity. As mentioned in the previous paragraph, many of the metrics calculated from the remote sensing data in this research consider the idea of assemblages of pixels, which advances the analysis tools and classification algorithm accuracies achieved over summary statistics.

This research also addresses the NRC Grand Challenge (NRC, 2001) of biological diversity and ecosystem functioning with some aspects of land-use dynamics. Several relevant action items are specifically called out in the report.

- Use remote sensing to assess diversity and develop remote sensing techniques to examine ecosystem characteristics

NEON AOP hyperspectral and lidar remote sensing data are used in this research. The AOP flew over the entire study site of the Ordway-Swisher Biological Station. Spectral metrics are calculated from the hyperspectral data which represent the spectral differences and variability across the site. Structural metrics are calculated from the lidar data which represent the structural differences and variability across the site. The metrics calculated from the two instruments allow analysis of the chemical, structure, and function of the various ecosystems located at the study site. Combined with NEON TOS ground data of species richness, diversity across the study site can be mapped with the remote sensing data.

- Improve accuracy and reduce uncertainty of vegetation classes from remote sensing data

A land cover classification algorithm is run using the metrics calculated from the remote sensing data. For this research, the land cover classes are vegetation communities as specified by the Florida Natural Areas Inventory (FNAI, 2010). These specific vegetation community classifications are a great advancement over generic vegetation classes such as the National Land Cover Database (Fry et al., 2011) for describing species composition. Additionally, utilizing the high-spatial resolution AOP data will reduce the confusion and misclassification due to the mixed-pixel signals of coarse spatial resolution remote sensing data commonly used for land cover mapping. Finally, the integration of hyperspectral and lidar data increases the accuracy of the vegetation classes by utilizing both spectral and structural information. The accuracy of the classification maps generated in this research are assessed on the training data using k-folds cross-validation and also with independent ground truth validation points. A discussion of the areas of confusion is provided, identifying areas of future research to improve the accuracy of the results. The

accuracy assessment framework can support future studies such as adding or removing metrics to simulate data from other sensors to determine the effects on accuracy of vegetation mapping.

- Use imagery analysis to detect variables or develop proxy measures

The calculated metrics from the remote sensing data represent various chemical and structural variables of the vegetation across the landscape. In this research, vegetation community classes are mapped using these variables. The resulting communities can then serve as a proxy for biodiversity by creating a relative species richness index using the ground data and applying that index to the vegetation community classes.

- Develop tools for rapid assessment of biodiversity

One output of this research is a local biodiversity indicator in the form of a species rightness map. The map highlights areas of high species richness that could be used by land managers to focus conservation efforts (Turner et al., 2003) and could inform field biologists which areas to target for specific research questions.

- Determine the relationship between ecosystem structure and functioning and diversity

Ecosystem structure can be characterized using the NEON AOP lidar data and ecosystem functioning (on a high level) by the hyperspectral data. Combining the airborne remote sensing with the NEON TOS field sampling of species richness and abundance establishes initial data sets for future use in process studies that might be able to determine a relationship between ecosystem structure and functioning with diversity.

- Link land change theory to space based imagery to move beyond local level analysis

While satellite remote sensing data is not used in this research, the airborne remote sensing data serves as an important bridge to scale from limited ground points or plots, to the local site-scale as mapped by the airborne data, to that of regional, continental, and global scale possible with satellite imagery.

Hyperspectral and lidar remote sensing sensors designed for vegetation studies from space are currently being designed and until launched, the application of vegetation community mapping at the global scale may be limited. Analysis of regional generic land cover classes and their effect on local biodiversity is possible with current satellites but not part of this research.

- Perform spatially mapping of land use and land cover with analysis of land changes with time

Spatial maps of vegetation communities are generated which are specialized land cover classes. NEON AOP will fly over each NEON site once per year around peak greenness for a 30 year period. After several years of data collection, land cover changes with time can be analyzed.

CHAPTER III

VEGETATION COMMUNITIES OF THE ORDWAY-SWISHER BIOLOGICAL STATION

Ordway-Swisher Biological Station

The study area for this research is the Ordway-Swisher Biological Station (OSBS), located in Melrose, FL approximately 22 miles East of Gainesville, FL as shown in Figure 12:

Figure 12: Location of the Ordway-Swisher Biological Station approximately 22 miles east of Gainesville, FL. Figure credit: OSBS

OSBS is operated by the University of Florida Institute of Food and Agricultural Science through the Office of the Dean for Research (OSBS, 2015a). The site covers an area of 9300 acres and consists of uplands and wetlands including 55 lakes and ponds, a hardwood swamp, and three wet prairies (OSBS, 2015a; OSBS, 2015b). The Ordway-Swisher Biological Station's mission is to "further the study and conservation of unique ecosystems through research, education, and management" (OSBS, 2014, p. 2). OSBS is also the NEON Domain 3 Southeast Core Terrestrial site with the instrument tower located at 29.68927 latitude, -81.99343 longitude (NEON, 2015k).

Archaeological evidence shows that Native Americans had a long-term seasonal occupancy at OSBS (Eisenberg & Franz, 1995). Europeans began agricultural activities (mixed farming and cotton) in the region around 1848, but this was limited due to soil depletion (Eisenberg & Franz, 1995). After the civil war, longleaf pine was exploited across the southeast for extraction of resin, logging for timber, and extraction of tree roots (Eisenberg & Franz, 1995). In 1926, Carl Swisher buys land and uses his 25000 acre private property for hunting, fishing, and limited agriculture (Eisenberg & Franz, 1995; OSBS, 2011; OSBS, 2015a). After his death, the Swisher-Wilcox family donates 3000 acres to the Nature Conservancy in 1979, named the Carl Swisher Memorial Sanctuary (Eisenberg & Franz, 1995; OSBS, 2011; OSBS, 2015a). The Sanctuary includes the Mill Creek drainage, Ashley Prairie, and Putnum Prairie in the North and West portions of the site (Eisenberg & Franz, 1995). In 1980, the University of Florida Foundation buys 6000 acres adjacent to the sanctuary with a grant from the Goodhill Foundation and names it the Katharine Ordway Preserve (OSBS, 2011; OSBS, 2015a). The tracts include the uplands and Sandhill lakes in the South and East portions of the site (Eisenberg & Franz, 1995). The Nature Conservancy and University of Florida jointly manage the properties (Eisenberg & Franz, 1995) until 1983 when University of Florida begins managing the Swisher Sanctuary (OSBS, 2011). In 2006, University of Florida and the Nature Conservancy rename the property as the Ordway-Swisher Biological Station (OSBS, 2011; OSBS, 2015a). In 2008 the Nature Conservancy donates the Sanctuary to University of

Florida (OSBS, 2011). Smaller areas have been added to the site since 1980 (Eisenberg & Franz, 1995) including 341 acres and 6 buildings between 2011 and 2014 (OSBS, 2013 & OSBS, 2014). Facilities at the site have been upgraded or added to support teaching, lodging for researchers, and maintenance and fabrication (OSBS, 2011; OSBS, 2013; OSBS, 2014).

Carl Swisher historically burned the pine uplands on an irregular schedule to reduce fuel loads, but in 1983, a three year rotational burn schedule was initiated on selected tracts (120-200 acres in size) of land (Eisenberg & Franz, 1995). The first 10 years of fire management at the site showed that the rotational burn favored longleaf pines over turkey oak and improved the condition of the uplands (Eisenberg & Franz, 1995). More recently, 27 prescribed burns were conducted in 2013 covering 2206 acres of land (OSBS, 2013) and 17 prescribed burns were conducted in 2014 covering 1763 acres of land (OSBS, 2014). A drought in the 1980s caused many of the Sandhill lakes to dry out leaving only 9 basins that still contained water in 1992 (Eisenberg & Franz, 1995). Rapid succession of lake vegetation was observed but the xeric adapted species in the uplands showed resilience to the deficits in rainfall (Eisenberg & Franz, 1995).

Florida Natural Areas Inventory

The Florida Natural Areas Inventory (FNAI) is a nonprofit organization administered by Florida State University (FNAI, 2015). The inventory was founded in 1981 and collects, interprets, and communicates information for conservation of Florida's biological diversity. In 1990, FNAI released a "Guide to the Natural Communities of Florida" (FNAI, 1990). The guide details a hierarchical classification system of 81 different natural biological associations (communities) found in Florida. Here

a natural community is defined as “a distinct and reoccurring assemblage of populations of plants, animals, fungi, and microorganisms natural associated with each other and their physical environment” (FNAI, 1990, p. 1). The guide contains 7 broad categories based on hydrology and vegetation type. In 2007, FNAI began an update to the guide with funding from the Florida Department of Environmental Protection, Division of State Lands (FNAI, 2010). The purpose of the update was to provide further clarification on distinctions between similar communities and more description of variations within each community. The hierarchical organization was modified a bit to emphasize similarities in species composition, structure, and landscape position (FNAI, 2010). Even though the focus of the guide is on natural communities, a list of altered land cover types due to human impact was added.

OSBS is in the process of updating their vegetation community map using the 2010 FNAI descriptions (S. Coates, personal communication, July 17, 2014). For this research, the 1990 FNAI descriptions are used along with the corresponding OSBS GIS data from 2004 (OSBS, 2015c; OSBS, 2015d) as shown in Figure 13:

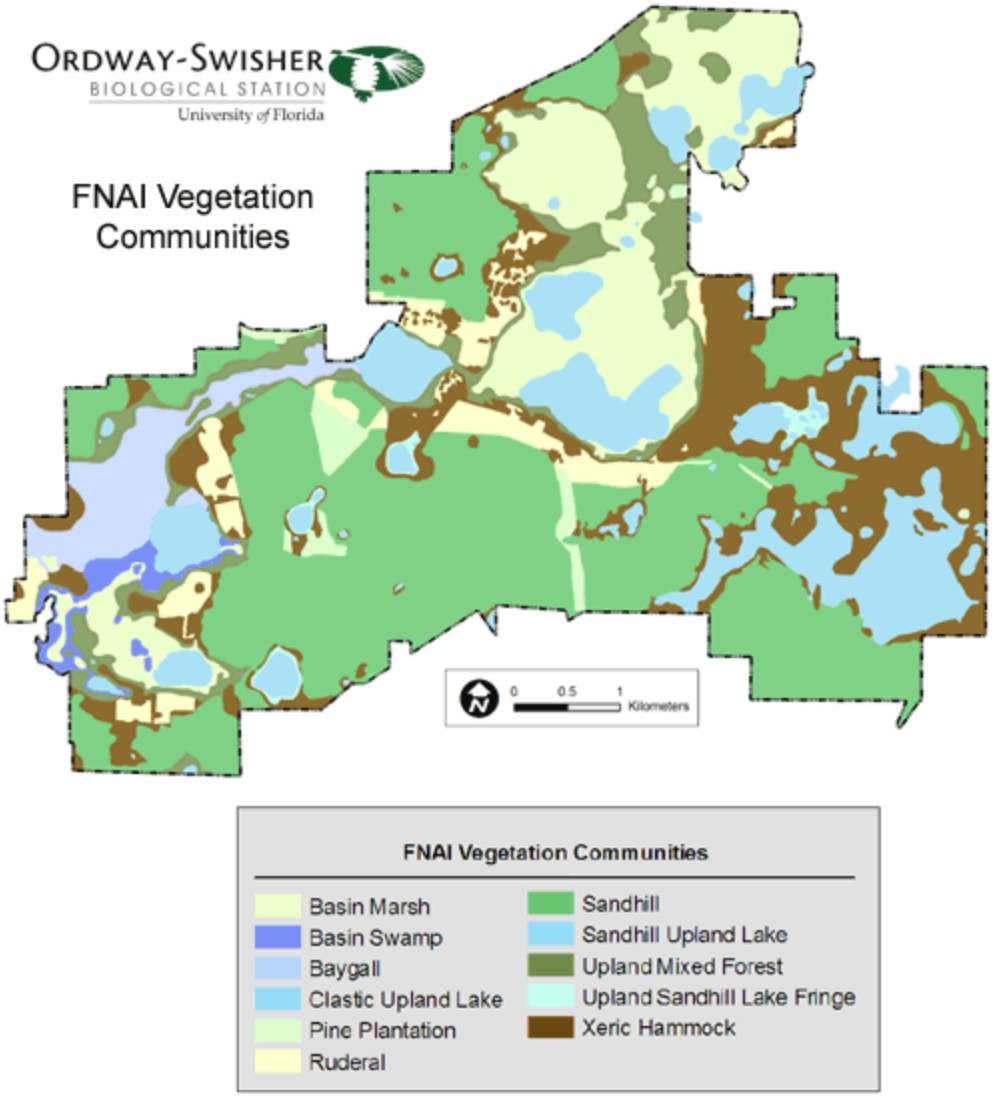

Figure 13: FNAI vegetation communities at OSBS. Figure credit: OSBS

The GIS layers were created using photo-interpretation of high resolution imagery combined with site surveys (S. Coates, personal communication, July 17, 2014). Remote sensing data and a decision tree classification algorithm will be used to generate similar maps with results shown in Chapters V-VII. Eleven vegetation communities were identified at the site including: Basin Marsh, Basin Swamp, Baygall,

Clastic Upland Lake, Pine Plantation, Ruderal, Sandhill, Sandhill Upland Lake, Upland Mixed Forest, Upland Sandhill Lake Fringe, and Xeric Hammock.

The area of each of the vegetation communities at OSBS is calculated and is shown in the pie chart in Figure 14 and listed as percent area in Table 4 along with the global and state rarity rank (FNAI, 1990).

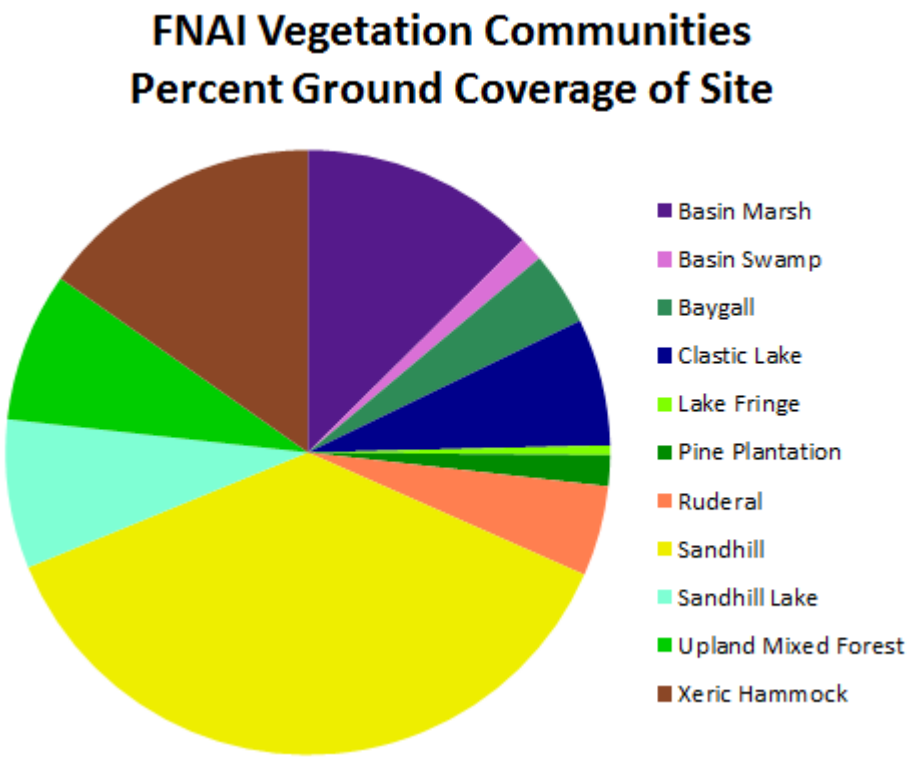

Figure 14: Pie chart showing the relative area covered by each of the vegetation communities at OSBS

Table 4: Percent of area at OSBS covered by each FNAI vegetation community along with global and state rarity rank

FNAI Community	Percent of Area at OSBS	Global and State Rarity Rank
Sandhill	37.17	G3/S2
Xeric Hammock	15.24	G3/S3
Basin Marsh	12.59	G4/S3
Upland Mixed Forest	8	G3/S3
Sandhill Upland Lake	7.97	G3/S2
Clastic Upland Lake	6.79	G3/S2
Ruderal	4.83	NA
Baygall	3.94	G4/S4
Pine Plantation	1.65	NA
Basin Swamp	1.3	G4/S3
Upland Sandhill Lake Fringe	0.52	NA

Definition of State (S) element ranks (FNAI, 1990, p. 4):

S1 = “Critically imperiled in state because of extreme rarity (5 or fewer occurrences or very little remaining area) or because of some factor(s) making it especially vulnerable to extinction;”

S2 = “Imperiled in state because of rarity (6-20 occurrences or little remaining area) or because of some factor(s) making it very vulnerable to extinction throughout it range;”

S3 = “Rare or uncommon in state (on the order of 21 to 100 occurrences);”

S4 = “Apparently secure in state, although it may be rare in some parts of its state range;”

S5 = “Demonstrably secure in state and essentially ineradicable under present conditions;”

S? = “uncertain State rank.”

The Sandhill community is the largest at OSBS covering 37.17% of the site. This is also a critically endangered ecosystem with a state rank of S2. Xeric Hammock and Basin Marsh cover the second and third largest areas at the site. The vegetation communities are defined below.

Upland Mixed Forest and Xeric Hammock belong to the FNAI category Hardwood Forested Uplands described as “mesic or xeric forest dominated mainly by hardwood trees” (FNAI, 2010, p. 4). Sandhill belongs to the category High Pine and Scrub described as “hills with mesic or xeric woodlands or shrublands; canopy, if present, open and consisting of pine or a mixture of pine and deciduous hardwoods” (FNAI, 2010, p. 4). Basin Marsh belongs to the category Freshwater Non-forested Wetlands described as “herbaceous or shrubby palustrine communities in floodplains or depressions; canopy trees, if present, very sparse and often stunted” (FNAI, 2010, p. 6), specifically the Marshes group described as “long hydroperiod; dominated by grasses, sedges, broadleaf emergents, floating aquatics, or shrubs” (FNAI, 2010, p. 6). Basin Swamp belongs to the category Freshwater Forested Wetlands described as “floodplains or depressions dominated by hydrophytic trees” (FNAI, 2010, p. 7), specifically the Cypress/Tupelo group described as “dominated entirely by cypress or tupelo, or these species important in the canopy; long hydroperiod” (FNAI, 2010, p. 7). Baygall is also in the Freshwater Forested Wetlands category but in the Hardwood group described as “dominated by a mix of hydrophytic hardwood trees; cypress or tupelo may be occasional or infrequent in the canopy; short hydroperiod” (FNAI, 2010, p. 8). Clastic Upland Lakes and Sandhill Upland Lakes belong to the category Ponds and Lakes (LACUSTRINE) described as “non-flowing wetlands of natural depressions lacking persistent emergent vegetation except around the perimeter” (FNAI, 2010, p. 9). Finally, Ruderal and Pine Plantation are altered landcover types where humans have cleared the forests in these areas and used them for

agriculture, grazing, or plantations. Each of the vegetation communities are described in detail in the following sections.

Basin Marsh

A Basin Marsh is a depression that is regularly inundated with rain water, forming herbaceous wetlands. Basin Marshes are large landscape features and the four marshes at OSBS cover the third largest area of the vegetation communities at the site (OSBS, 2015c; OSBS, 2015d). Vegetation species can be divided into submerged, floating-leaved, emergent, and grassy zones. The marshes are dominated by herbs and shrubs. Trees are sparse so there is little or no canopy. The characteristic species include: white waterlily (*Nymphaea odorata*), maidencane (*Panicum hemitomon),* sawgrass (*Cladium jamaicense*), bulltongue arrowhead (*Sagittaria lancifolia*), pickerelweed (*Pontederia cordata*), and sand cordgrass (*Spartina bakeri*) (FNAI, 1990; FNAI, 2010). Figure 15 shows site photos of two areas of Basin Marsh. As seen in the pictures, the marshes are dominated by the herbaceous ground layer and there is little or no canopy.

Figure 15: Two areas of Basin Marsh. Photos courtesy of NEON

Basin Swamp

A Basin Swamp is a type of wetland with an extended hydroperiod (can have standing water for most of the year) that contains hydrophytic trees and shrubs. This type of swamp is highly variable in size, shape, and species composition with mixed species canopies common. They occur in large landscape depressions where the soils are acidic, nutrient poor peats lying on top of a clay lens or other impervious layer. The characteristic species include: pond cypress (*Taxodium ascendens*) and swamp tupelo (*Nyssa sylvatica* var. *biflora*). Higher vegetation diversity is expected near the perimeter of a Basin Swamp where the soils are more aerated (FNAI, 1990; FNAI, 2010). Figure 16 shows site photos of a Basin Swamp canopy (upper left), understory (upper right), and ground cover (bottom). The mostly closed canopy contains a mix of species, however, the open understory in this example is dominated by pond cypress. The ground was not flooded at the time these pictures were taken.

Figure 16: Basin Swamp canopy (upper left), understory (upper right), and ground cover (bottom). Photos courtesy of NEON

Baygall

A Baygall is an evergreen forested wetland, situated in a depression, and dominated by bay species in both the canopy and understory. The canopy can be open or dense and may contain pine trees that can withstand a hydric soil. The canopy and understory typically do not form distinct strata and appear as a dense tall thicket. Vines may be abundant but herbs are absent or few. Baygalls form on wet acidic peat soils and are often drained by streams. Seepage from surrounding communities along with rainfall keep the peat saturated but rarely flooded. Baygall can be a succession from Basin Swamp. The characteristic species include: Loblolly bay (*Gordonia lasianthus*), sweetbay (*Magnolia virginiana*), and/or swamp bay (*Persea palustris*). Wetter Baygalls may contain pond cypress and swamp tupelo making them appear similar to Basin Swamp (FNAI, 1990; FNAI, 2010). Figure 17 shows site photos of two areas of Baygall. As seen in the pictures, the understory is very thick and the bottom image appears to be an area with an open canopy (with canopy trees in the top image and the background of the bottom image).

Figure 17: Two areas of Baygall. Photos courtesy of NEON

Sandhill and Clastic Upland Lakes

Sandhill Lakes are shallow rounded depressions in sandy upland communities. These lakes are typically permanent water bodies however their levels can fluctuate significantly or completely dry out in times of severe drought. Sandhill Lakes are lentic water bodies that lack significant surface inflows or outflows. They primarily get their water through lateral ground water seepage from the neighboring well-drained soils. A narrow band of vegetation may exist along the shore consisting of hydrophytic grasses and herbs or a dense shrub thicket. In cases of gradually sloping shorelines, broader bands of emergent vegetation with submerged aquatic plants may exist (FNAI, 1990; FNAI, 2010). A Sandhill Lake is shown in the site photo in the upper right of Figure 18. Emergent vegetation can be seen sticking up out of the water near the shore.

A Clastic Upland Lake is an irregular-shaped depression or basin occurring in uplands on clay substrates. These lakes range in depth from shallow to relatively deep. Clastic Lakes are lentic water bodies with surface inflows but no significant outflows where water is dissipated through evaporation and transpiration. A large variation of vegetation can be found at the water's edge. Shallow areas are densely vegetated by shallow bands of emergent, floating, and submersed aquatics (FNAI, 1990; FNAI, 2010). A Clastic Lake is shown in the site photo in the upper left of Figure 18. Emergent and floating vegetation can be seen sticking up out of the water near the shore. The aerial photo in the bottom of Figure 18 was taken by the AOP flight operators and shows significant floating vegetation on some of the lakes.

Figure 18: A Clastic Lake (upper left), a Sandhill Lake (upper right), and an arial view of lakes with a significant amount of floating vegetation. Photos courtesy of NEON

Pine Plantation

A Pine Plantation is an area altered by silviculture, in this case planted pine trees. The ground cover could have been damaged in the past by mechanical site preparation, or historical and ongoing damage due to shading of the ground layer and deep pine needle duff, preventing the growth of herbs. Pine Plantations are typically dominated by dense, even-aged loblolly pine (*Pinus taeda*), sand pine (*Pinus clausa*), or slash pine (*P. elliottii*) (FNAI, 1990; FNAI, 2010). Figure 19 shows site photos of two areas of Pine Plantation at the University of Florida Austin Cary Forest (SFRC, 2015). As seen in the pictures, the plantations contain dense but regularly spaced slash pine trees. Active logging may remove rows of trees or leave large bare openings after tree harvest as seen in the bottom picture.

Figure 19: Two areas of Pine Plantation at the Austin Cary Forest. Photos courtesy of NEON

Ruderal

At OSBS, a Ruderal vegetation community refers to the altered land cover type of an abandoned field or pasture. These are areas of old fields, formerly used for agriculture or grazing. These fields show early succession and are dominated by weeds but may also have successional oak trees (FNAI, 1990; FNAI, 2010). Figure 20 shows site photos of four Ruderal areas that used to be fields and pastures. As seen in the pictures, the Ruderal areas have been cleared of trees and are dominated by grasses, herbs, and shrubs. Some areas have patches of exposed sand.

Figure 20: Four rederal areas (abandoned fields and patures). Photos courtesy of NEON

Sandhill

Sandhill is a variation of a longleaf pine (*Pinus palustris*) ecosystem with widely spaced pines, a sparse mid-story of deciduous oaks, and dense groundcover of grasses, herbs, and low shrubs. Sandhill occurs on rolling topography with deep well drained sands (often yellow in color). The herbaceous ground cover contains the greatest vegetation diversity. The characteristic species include: longleaf pine (*Pinus palustris*), turkey oak (*Quercus laevis*), and wiregrass (*Aristida stricta var. beyrichiana*). Sandhill requires frequent (1-3 year interval) low intensity growing season fires to maintain an open canopy structure. The open canopy provides plenty of sunlight to the ground which is needed for growth of longleaf pine juveniles, shrubs, and grasses. Without regular fires, the mid-layer of oaks will form a closed canopy preventing light from reaching herbs and juvenile longleaf pines and the ecosystem will change into a Xeric Hammock community. Sandhill ecosystems are considered critically endangered and are given a state rarity rank of S2 (FNAI, 1990; FNAI, 2010). See the final section in this chapter for more information about the importance of the Sandhill community. Figure 21 shows site photos of a Sandhill canopy (upper left), understory (upper right), and ground cover (bottom). The canopy is open with widely spaced pines and the open understory contains wiregrass and longleaf pine seedlings. Charring (due to active fire management) can be seen on the trunks of the longleaf pine trees.

Figure 21: Sandhill canopy (upper left), understory (upper right), and ground cover (bottom). Photos courtesy of NEON

Upland Mixed Forest

The GIS map shown in Figure 13 uses the vegetation community Upland Mixed Forest. Since the creation of this map back in 2004, site managers have taken a closer look at the vegetation species at the site in these areas and have recommended that the map be updated and these areas be labeled as Mesic Hammock (S. Coates, personal communication, July 17, 2014). The following sentences use elements from the Mesic Hammock description in the 2010 FNAI report (FNAI, 2010).

Upland Mixed Forest is a well-developed evergreen hardwood forest with a closed canopy dominated by live oak (*Quercus virginiana*) with cabbage palm (*Sabal palmetto*). Southern magnolia (*Magnolia grandiflora*) and pignut hickory (*Carya glabra*) are characteristic in the subcanopy. Pine trees may form a sparse divergent layer and both shrubs and herbs can be sparse or patchy. At OSBS, upland Mixed Forest forms in the ecotone between wetland and upland communities which has minimal or no fires due to its landscape position (FNAI, 2010). Figure 22 shows site photos of an upland Mixed Forest canopy (upper left), understory (upper right), and ground cover (bottom). The mostly closed canopy contains a mix of species and the understory in this example is dominated by saw palmetto (*Serenoa repens*).

Figure 22: Mixed Forest canopy (upper left), understory (upper right), and ground cover (bottom). Photos courtesy of NEON

Upland Sandhill Lake Fringe

Upland Sandhill Lake Fringe is the area surrounding a Sandhill Lake. Depending on seasonal water levels, the fringe area may contain emergent vegetation around the perimeter of the lake or may be bare soil from the lake bottom (FNAI, 1990; FNAI, 2010). Figure 23 shows site photos of four Lake Fringe areas. As seen in the pictures, the Lake Fringes are dominated by grasses, herbs, and shrubs and are absent of canopy trees. Some areas have patches of exposed sand.

Figure 23: Four Lake Fringe areas. Photos courtesy of NEON

Xeric Hammock

Xeric Hammock is an evergreen oak forest on well drained sandy soils. It has a closed canopy and the canopy height is lower than other forest communities. The understory is open but contains shrubs and an herb layer that is sparse or absent. Spanish moss (*Tillandsia usneoides*) and ballmoss (*T. recurvata*) are often found hanging in the trees. The characteristic set of species include: sand live oak (*Quercus geminata*) and saw palmetto (*Serenoa repens*). In the case of succession from a Sandhill community, turkey oak (*Q. laevis*) and bluejack oak (*Q. incana*) may be dominant. Xeric Hammocks occur when fire exclusion (requires 7-16 years without fire) allows the oak trees to grow enough to close the canopy and reduce sunlight reaching the ground, therefore decreasing the growth of herbs and juvenile longleaf pines (*Pinus palustris*) (FNAI, 1990; FNAI, 2010). Figure 24 shows site photos of a xeric Hammock canopy (upper left), understory (upper right), and ground cover (bottom). The mostly closed canopy is dominated by oak species and the understory is very open with a lot of leaf litter on the ground.

Figure 24: Xeric Hammock canopy (upper left), understory (upper right), and ground cover (bottom). Photos courtesy of NEON

Implications of Species and Structure for Utilizing Remote Sensing Data

The imaging spectrometer is a passive hyperspectral instrument that measures reflected solar radiation in the form of at-sensor radiance. The measured pixel integrated column radiance values are dominated by species in top of the canopy, except in cases of canopy gaps or other significant light transmission at the 1 m spatial scale. Vegetation species may have unique spectral reflectance curves, or at least similar functional types will have unique spectra verses other functional types (Ollinger, 2011; Ustin & Gamon, 2010). In the case of areas of active fire management (Sandhill, Ruderal, Lake Fringe, and Basin Marsh), spectral signatures of burnt materials can be observed in the hyperspectral data. Lake water pixels and bare soil will also exhibit unique spectral signatures. One complication is that lakes can have floating vegetation that cover a large surface area so mapping lakes may not be as straightforward as mapping water spectra. Finally, vegetation communities are assemblages of species. Analysis of a 20 m x 20 m plot scale will include assemblages of spectra, which could be leveraged to improve classification of the vegetation communities in the hyperspectral data.

The lidar is an active instrument that emits laser pulses at a wavelength of 1064 nm. The pulse of light travels through the air and interacts with the vegetation and/or ground where some of the photons are scattered back to the receiver (after traveling back through the air). The receiver measures the return signal as a function of time, enabling it to resolve the backscattered radiation as a function of distance (using the speed of light). Each received laser shot measures the vertical structure of the scene (since the lidar instrument is located in an aircraft looking down at the ground). When analyzed at a 20 m x 20 m plot scale, the collection of received laser pulses additionally captures the 2D horizontal structure, giving a 3D picture of the vegetation structure. Physical vegetation parameters such as tree height, tree spacing, canopy closure, and understory structure can be derived from the lidar data (Lefsky et al., 2002). The top

of the canopy can be identified in the received laser pulses as the first detected signal. A canopy height model can be generated from the first returns. Similarly, the ground can be identified in the received laser pulses as the last detected signal (except where vegetation is so thick that the laser photons cannot penetrate to the ground). A bare-Earth digital terrain model can be generated from the ground returns. Topographic analysis such as slope and aspect can also be calculated.

Each vegetation community contains specific characteristic species that together form a specific vegetation structure. The NEON AOP remote sensing instruments observe species as measured spectra in the hyperspectral data and can directly measure the 3D structure of the vegetation in the lidar data. Table 5 lists the relevant species and structural parameters for each FNAI vegetation community that are measured by the remote sensing instruments and that can be leveraged by the classification algorithms used in this research.

Table 5: Relevant Species and Structural Parameters for Each FNAI Vegetation Community

FNAI	Dominant Species	Canopy	Understory	Soil & Moisture	Topography
Basin Marsh	Herbs and shrubs: white waterlily, maidencane, sawgrass, bulltongue arrowhead,	No canopy	Submerged, floating level, emergent, and grassy zones	Herbaceous wetlands from rainwater	Large landscape depression occurring around lakes

	pickerelweed, and sand cordgrass				
Basin Swamp	Hydrophytic trees: pond cypress and swamp tupelo	Mostly closed	Hydrophytic shrubs, especially on perimeter	Acidic nutrient poor peats on top of an impervious layer, wetland with extended hydroperiod	Basin in large landscape depression
Baygall	Loblolly bay, sweetbay, and swamp bay	Open to dense canopy	Mixed over and understory, no distinct strata, dense tall thicket	Wet acidic peat soils, saturated but not flooded, drained by a stream	At the base of a slope or in a depression
Clastic Upland Lake	NA	None	Shallow areas are densely vegetated by shallow bands of emergent, floating, and submersed aquatics	Clay substrate	Irregular-shaped depression or basin, surface inflows but no significant outflows

Pine Plantation	Loblolly pine, sand pine, or slash pine	Dense and even-aged	Depends on mechanical site preparation, damage due to shading of the ground layer, and deep pine needle duff	Same as Sandhill at OSBS	Same as Sandhill at OSBS
Ruderal	NA	Open canopy with sparse trees	dominated by weeds but may also have successional oak trees	Same as Sandhill at OSBS	Same as Sandhill at OSBS
Sandhill	Longleaf pine, turkey oak, and wiregrass	Open canopy of widely spaced pines	Sparse mid-story of deciduous oaks, and dense groundcover of grasses, herbs, and low shrubs	Deep well drained sands (often yellow in color)	Rolling topography
Sandhill Upland Lake	NA	None	Hydrophytic grasses and herbs may	Sand substrate, levels can fluctuate	Shallow rounded depressions,

			exist along the shore in a narrow band	significantly or completely dry out in times of severe drought	lack significant surface inflows or outflows
Upland Mixed Forest	Live oak, cabbage palm, southern magnolia, and pignut hickory	Closed canopy	Dense or open shrubby understory, sparse or patchy herb layer	Sands mixed with organic matter, may have a thick layer of leaf litter, soils have high moisture by heavy shading and accumulation of litter but rarely inundated	Ecotones between wetlands and upland communities
Upland Sandhill Lake Fringe	NA	None	May contain emergent vegetation around the perimeter	Visible bare soil from the lake bottom	Edge of depression containing a lake

Xeric Hammock	Sand live oak, saw palmetto, turkey oak, bluejack oak, spanish moss, and ballmoss	Closed, lower height canopy	Open understory of shrubs and sparse herbs	Well-drained sandy soils	Same as Sandhill at OSBS

Sandhill requires frequent (1-3 year interval) low intensity growing season fires to maintain an open canopy structure (FNAI, 2010; Oswalt et al., 2012). The open canopy provides plenty of sunlight to the ground which is needed for growth of longleaf pine juveniles, shrubs, and grasses. Without regular fires, the mid-layer of oaks will form a closed canopy preventing light from reaching herbs and juvenile longleaf pines and the ecosystem will change into a Xeric Hammock community. The herbaceous ground cover contains the greatest vegetation diversity in Sandhill, but remote sensing may not be able to directly observe this diversity. Instead, analysis of mid and understory structure may provide an indirect insight into the diversity at specific ground locations.

The FNAI community descriptions and property table suggest some relationships between FNAI classes (due to succession) or with topography and hydrology. Future work could incorporate GIS analysis techniques (ex: distance from lake as a feature layer) or incorporate object based image analysis techniques (ex: compare data to spatial neighbors and develop rules to modify the output class under different circumstances) to improve the vegetation classification using contextual information. Additionally, many of the canopy tree species are evergreens, but some like pond cypress are deciduous. Remote sensing data collections in the winter time could be used to help identify specific canopy species. Finally, a diverse herbaceous ground layer should not be as flat as open ground in lidar data. Current

discrete return data may not be able to detect this difference. Waveform lidar analysis could be utilized to further characterize ground vegetation and distinguish subtle differences in surface cover.

Importance of the Sandhill Community

The Sandhill vegetation community is one of several variants of the longleaf pine (*Pinus palustris Mill.*) ecosystem (FNAI, 2010). Longleaf pine ecosystems once covered up to 92 million acres across the Southeast U.S. (Frost, 2006; Oswalt et al., 2012) with an extent from Texas to Virginia and central Florida to the piedmont of Alabama and Georgia (Oswalt et al., 2012). The ecosystem became established between 4000-8000 years ago as the continental glacier retreated (Oswalt et al., 2012). Longleaf pine ecosystems contain a canopy of widely spaced longleaf pines with an herbaceous ground layer of grasses and forbs (Oswalt et al., 2012). Frequent, low-intensity fires are required to maintain the diversity of the ground layer, so longleaf pines flourished in the pre-Columbian era as Native Americans conducted periodic burns which maintained and enhanced the lands (Oswalt et al., 2012).

Longleaf pine ecosystems are some of the most diverse in North America with an extremely diverse ground layer containing up to 40 or more vascular plant species in a 1 m^2 area and more than 170 species in a 1000 m^2 (America's Longleaf, 2009; Oswalt et al., 2012). These ecosystems contain around 900 endemic plant species, 30 species of reptiles and amphibians that are specialists, and 29 species that are federally listed as threatened or endangered (America's Longleaf, 2009). Sandhill specifically provides an important habitat for many rare, threatened, or endangered species (FNAI, 2010; FWS, 2015a):

- gopher frog (*Rana capito*)
- gopher tortoise (*Gopherus polyphemus*)
- eastern indigo snake (*Drymarchon couperi*)
- Florida pine snake (*Pituophis melanoleucus mugitus*)
- short-tailed snake (*Stilosoma extenuatum*)
- Eastern diamondback rattlesnake (*Crotalus adamanteus*)
- red-cockaded woodpecker (*Picoides borealis*)
- southeastern American kestrel (*Falco sparverius paulus*)
- Florida mouse (*Podomys floridanus*)
- southeastern squirrels and Sherman's fox squirrels (*Sciurus niger niger* and *S. niger shermani*, respectively)
- Florida deepdigger scarab beetle (*Peltotrupes profundus*)
- Ocala deepdigger scarab beetle (*Peltotrupes youngi*)
- north peninsular mycotrupes beetle (*Mycotrupes gaigei*)
- Skelley's june beetle (*Phyllophaga skelleyi*)
- pygmy anomala scarab beetle (*Anomala exigua*)
- McCrone's burrowing wolf spider (*Geolycosa xera*)
- and several species of melanoplus grasshoppers including pygmy sandhill grasshopper (*Melanoplus pygmaeus*) and Tequesta grasshopper (*Melanoplus tequestae*).

The gopher tortoise is listed as a threatened species by the U.S. Fish and Wildlife Service in certain regions of the Southeast (FWS, 2015b). Their populations have declined by 80% over the past 100 years consistent with habitat loss and human predation (Smith, 1995). Tortoises prefer clear, unshaded areas for their nesting sites making Sandhill, old fields, and sand live oak hammocks prime habitat

(Smith, 1995). Previous research has determined that 61% of gopher tortoise burrows are occupied at any time (Smith, 1995) plus more than 60 species of vertebrates and more than 300 species of invertebrates use the burrows as shelters, making the gopher tortoise a keystone species in Sandhill ecosystems (FNAI, 2010; Smith, 1995).

Three centuries of land use change along with fire exclusion have significantly reduced the size of the longleaf pine ecosystem (Oswalt et al., 2012). The major causes of the reduced extent and condition in longleaf pine ecosystems include (America's Longleaf, 2009; Oswalt et al., 2012):

- Logging of longleaf pine and unsustainable harvest
- Extraction of naval stores
- Conversion to agriculture, industrial pine plantations, and urban development
- Landscape fragmentation
- Invasive species
- Interruption of natural fire regimes

Oswalt et al., 2012 provide an interesting historical account of the loss of longleaf pine ecosystems due to human land use, especially due to logging in the late 1800s and early 1900s. Florida continued to see a decline of 90% in longleaf pine forests between 1936 and 1995 due to conversion to pine plantations, urban development, and agriculture (FNAI, 2010). Across the Southeast, between 1987 and 1995, 92000 acres of longleaf pine forest was converted to other land cover types including 7400 acres a year to urbanization and 3700 acres a year to agriculture (Oswalt et al., 2012). The 92 million acres of longleaf pine ecosystems had been reduced to around 2.6 million acres (less than 3% of the

natural area) by 2005 (Oswalt et al., 2012). Decadal surveys of longleaf pine were conducted over the last 50 years as part of the Forest Inventory and Analysis (FIA) program by the United States Department of Agriculture and US Forest Service. These surveys found longleaf pine in eight states but found that the ecosystems were highly fragmented and in poor condition (Oswalt et al., 2012).

Several programs are underway to conserve and improve existing longleaf pine areas (Oswalt et al., 2012). One major program is run by the Longleaf Alliance to manage and restoring longleaf pine forests for ecological and economic benefits (America's Longleaf, 2015). Working groups and workshops developed a conservation plan with a 15 year goal to increase the acreage of longleaf pine forests from 3.4 to 8.0 million acres (America's Longleaf, 2009). The plan calls for "maintaining existing longleaf ecosystems in good condition, improving areas identified as longleaf pine forests but missing significant components of the understory and/or fire regimes, and restoring longleaf pine to suitable sites currently in other forest types" (America's Longleaf, 2009, p. ii). In line with the recommendations of the range wide conservation plan, Florida created a statewide forest action plan (FDACS, 2010). Longleaf pine ecosystems are called out as a specific issue. The strategy notes that one of the biggest problems was proper fire management which was only occurring on 0.2% of longleaf areas in Florida. Correspondingly, one goal is to increase fire management so that fire return intervals in longleaf pine ecosystems are every 1-5 years. Both the range wide plan and Florida strategy identify the need to inventory and assess longleaf pine ecosystems (America's Longleaf, 2009; FDACS, 2010). FNAI and the Florida Forest Service have created a "Longleaf Pine Ecosystem Geodatabase" which serves as a GIS repository of the distribution and condition of longleaf pine forests in Florida (LLPEG, 2014a; LLPEG, 2014b). To date, 2.2 million acres of longleaf pine ecosystems have been verified in Florida (LLPEG, 2014a; LLPEG, 2014b) with about 40% on private lands. This research maps all eleven vegetation communities at OSBS but due to its importance, the classification results of the Sandhill ecosystem are emphasized.

CHAPTER IV

REMOTE SENSING AND GROUND DATA

National Ecological Observatory Network Airborne Observation Platform Domain 03 Flight Campaign 2014

The Airborne Observation Platform (AOP) flew the NEON Domain 03 Southeast region in 2014 as part of a campaign to train flight operators and perform nominal flight collections over NEON sites. All three Domain 03 sites were flown in May 2014 including Ordway-Swisher Biological Station (OSBS), Disney Preserve in Florida, and the Jones Ecological Research Center in Georgia (NEON, 2015k). Two flight plans were generated to cover the Ordway-Swisher Biological Station including the surrounding watershed (J. Musinsky, personal communication, May 7, 2014). The flight plans were prioritized to ensure that the site was collected first. If extra flying time existed, then the rest of the watershed area would be flown. Flight plan 1 included 1 East-West cross-strip (used to improve the relative geolocation of the data) and 26 North-South survey lines ranging in length from 10 to 16 km. Flight plan 2 included 1 cross-strip and 15 survey lines ranging in length from 4 to 7 km. The sum of the flight plans gives 2 cross-strips and 41 survey lines. The plan flight lines are shown in Figure 25 with the priority 1 lines in yellow and priority 2 lines in blue.

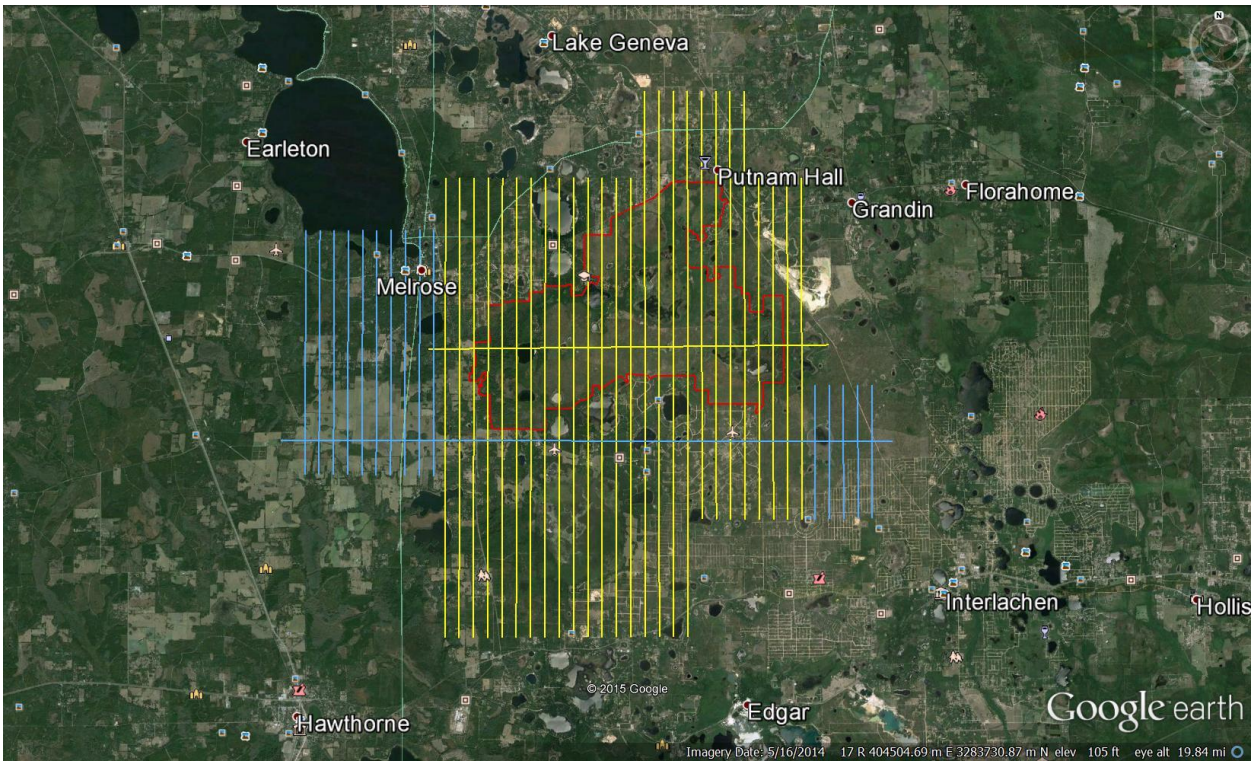

Figure 25: Flight plans for the Ordway-Swisher Biological Station. The yellow lines are the priority 1 area covering the site and the blue lines are the priority 2 area to cover the site watershed

The flight parameters and instrument configurations are listed in Table 6:

Table 6: Flight Parameters and Instrument Configuration for the AOP Domain 03 Campaign

Parameter	Value
Nominal Flight Altitude	1000 m (3280.84 ft) Above Ground Level
Airspeed	50 m/s (97.2 knots)
Flight Line Spacing	420 m (30% overlap between flight lines)
Spectrometer Line Rate	100 frames per sec
Camera Shot Overlap	50 %
Lidar Pulse Repetition Frequency	100 kHz
Lidar Mirror Scan Half Angle	± 18.5 deg
Lidar Mirror Scan Frequency	50 Hz
Lidar Beam Divergence	Wide 0.8 mrad

With the flight parameters listed above, the pixel resolution of the spectrometer is approximately 1.0 m at nadir with a cross-track spacing of 1.0 m and an along-track spacing of 0.5 m (the spectrometer is over-sampled in the along-track direction). Using a 70 mm focal length lens, the high-resolution camera takes pictures with a pixel resolution of 8.5 cm at nadir. The lidar instrument achieves a nominal 3.82 points per meter squared with a 0.8 m diameter spot and a spot spacing of about 0.524 m in the across-track direction and 0.5 m in the along-track direction.

The AOP flew the Ordway-Swisher Biological Station on two separate days: May 7 and May 19, 2014. On May 7, 2014, the AOP began the survey in the center of the site working their way West until the last flight line in the priority 1 area. The aircraft then returned to the center of the site and continued East. The aircraft landed to re-fuel but then ran into electronics issues. The next flight took place on May 19 after the issues had been resolved. The remainder of the priority 1 flight lines were surveyed as well as

all of the priority 2 watershed lines to the far West and East of the site. Figure 26 shows the lidar ground swaths of the data collected with the May 7 swaths in green and the May 19 swaths in orange.

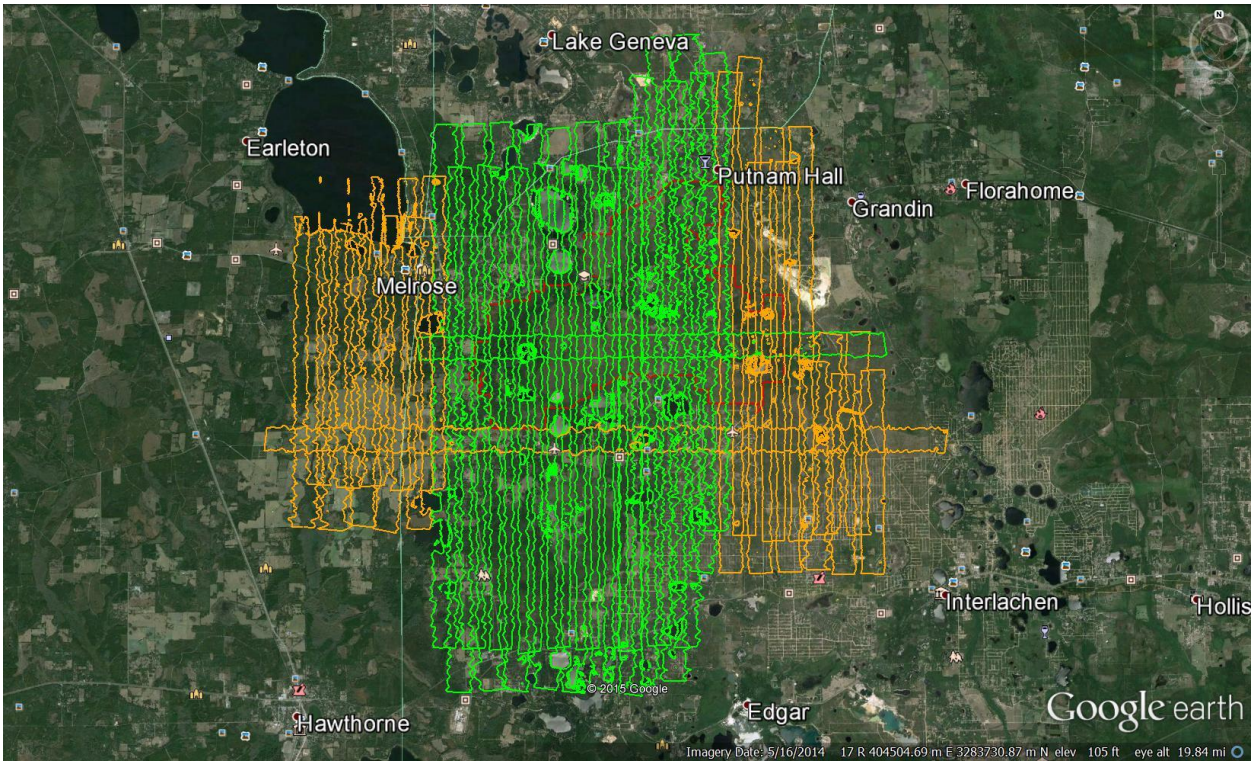

Figure 26: Ground footprints of the lidar data collected on May 7, 2014 (green polygons) and May 19, 2014 (orange polygons)

Several pieces of ground support equipment were deployed in the field in coordination with the flights including a GPS base station, a CIMEL sun photometer, and radiometric calibration tarps. The black 3% and gray 48% reflectance tarps are seen from the air in Figure 27 and are used to improve the values in the spectrometer surface reflectance product.

Figure 27: Radiometric calibration tarps as viewed from the aircraft on May 7 (left) and May 19, 2014 (right). Photos courtesy of NEON AOP

Airborne Observation Platform Data Processing

The raw remote sensing data are processed into Level 1 calibrated products for the spectrometer, lidar, and high-resolution camera. The first step is to convert the raw position and orientation data from the GPS/IMU into a smoothed best estimate of trajectory (SBET) file. The SBETs are generated using the Applanix POSPacMMS software (Applanix, 2015) including ground GPS base station data plus CORS network station data (NOAA, 2015) to improve the accuracy of the GPS positions. The output of the SBET file is the aircraft position (latitude, longitude, and elevation) and aircraft attitude (roll, pitch, and heading) as a function of GPS time. The trajectory file is essential to the geolocation processing chains for all three instruments.

Imaging Spectrometer Data Processing

The imaging spectrometer data are processed to a Level 1 directional surface reflectance product. The processing is comprised of three major steps: radiance conversion, ortho-rectification, and atmospheric correction. Radiance conversion transforms the raw imaging spectrometer digital number (DN) values into calibrated at-sensor radiance ($\mu W / cm^2 - nm - ster$). The imaging spectrometer is calibrated in the laboratory at the beginning of the flight season using an integrating sphere (Kampe et al., 2010.) to derive absolute radiometric calibration coefficients. A pixel non-uniformity correction table is generated to account for relative pixel-to-pixel brightness variations. Bad pixels are also identified and their values are replaced during the radiance conversion step. More information about the radiance conversion processing is found in an algorithm theoretical basis document (Gallery & Leisso, 2014).

The radiance spectrometer data are in their native 2D sensor format, i.e. 600 across-track pixels by 424 spectral bands by number of collected along-track data lines. The pixels are geolocated to the Earth's surface during the ortho-rectification processing step. Each pixel is raytraced using the position and orientation from the trajectory file, a camera distortion model (generated using imagery from a calibration flight at the beginning of the flight season which co-registers the spectrometer to the lidar instrument), and a surface model derived from the lidar data. The raytracing determines the corresponding ground location for each pixel in a WGS84 UTM projection with a Geoid height. The pixels are rearranged onto a fixed output 1 m resolution grid using a nearest neighbor resampling. Figure 28 shows the ortho-rectification of a portion of a single flight line.

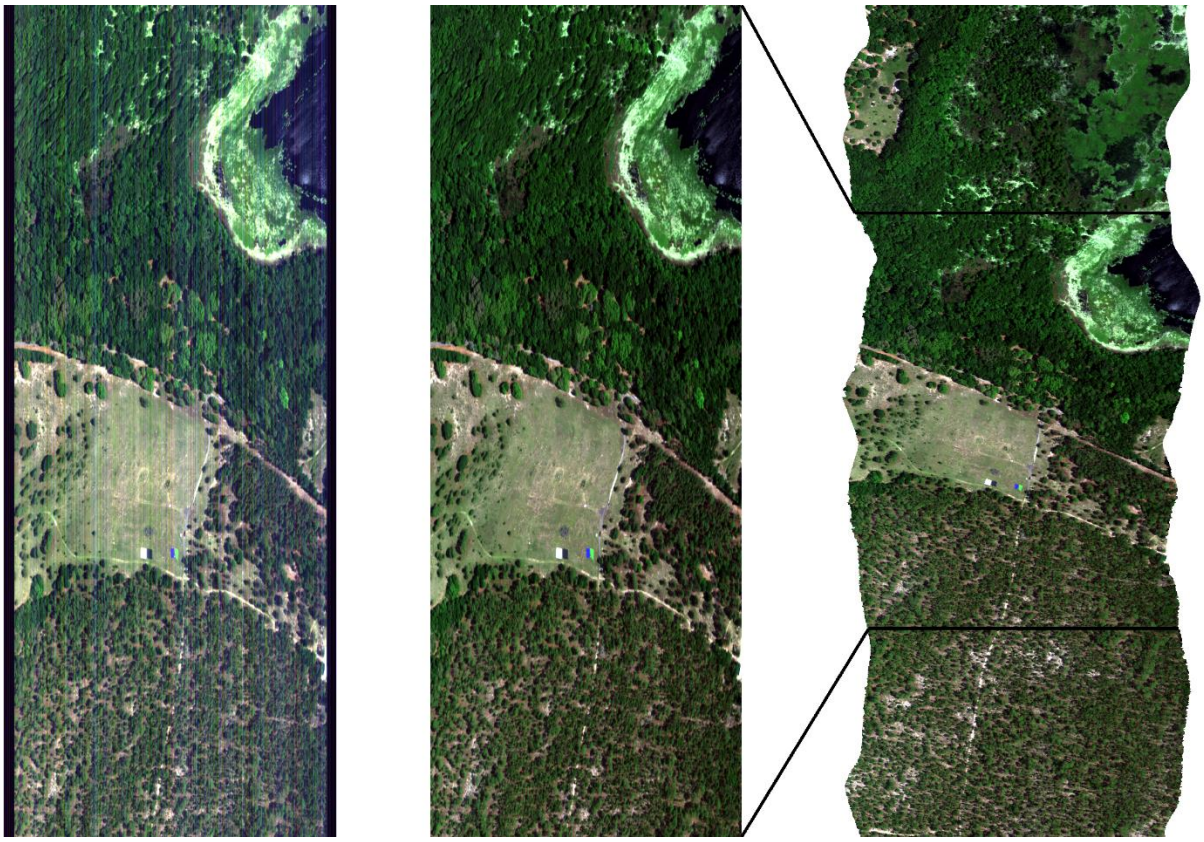

Figure 28: Radiometric correction and ortho-rectification of a portion of a single flight line of imaging spectrometer data. The left image is the raw data in the native 2D sensor rectangular format collected by the spectrometer where vertical radiometric streaks are visible. The center image has been converted to radiance, removing the streaks, but showing wavy lines due to aircraft turbulence and is oversampled by a factor of two in the along-track direction. The flight line is ortho-rectified onto a fixed 1 m output grid with a WGS84 UTM projection using nearest neighbor resampling as shown on the right

The left side of the figure is a color image in the native sensor format. The image is square and waves/curves can be seen in the trees and roads due to aircraft turbulence and the optical distortion. The image is also oversampled by a factor of two with an across-track (columns) pixel sampling of 1 m but an along-track (rows) pixel sampling of 0.5 m. The right side of the figure shows the output ortho image which is half as long as the original image. The roads are now straight and the edges of the image are wavy showing the projection of the ground swath due to turbulence. More information about the geolocation processing is found in an algorithm theoretical basis document (Kampe, 2014).

Atmospheric correction is performed using the ATCOR 4 Software (ReSe, 2015) to convert at-sensor radiance to directional surface reflectance. ATCOR attempts to compensate for absorption and scattering in the atmosphere and solves for the surface reflectance (surface leaving radiance divided by the downwelling solar spectral irradiance). ATCOR uses the hyperspectral data to solve for column water vapor absorption on a pixel-by-pixel basis but solves for atmospheric absorption and scattering over a larger portion of the scene. The software also estimates the molecular path radiance and aerosol scattering using dark targets (the contribution of scattered light by the atmosphere into the radiance signal). Finally, the ATCOR software solves for the adjacency effect of nearby ground targets whose radiance is scattered by the atmosphere into the instantaneous field of view of a given pixel. Figure 29 shows spectral at-sensor radiance curves on the left for several natural targets such as sand, grass and shrubs in a pasture, the canopy of a pine tree, the canopy of an oak tree, and water.

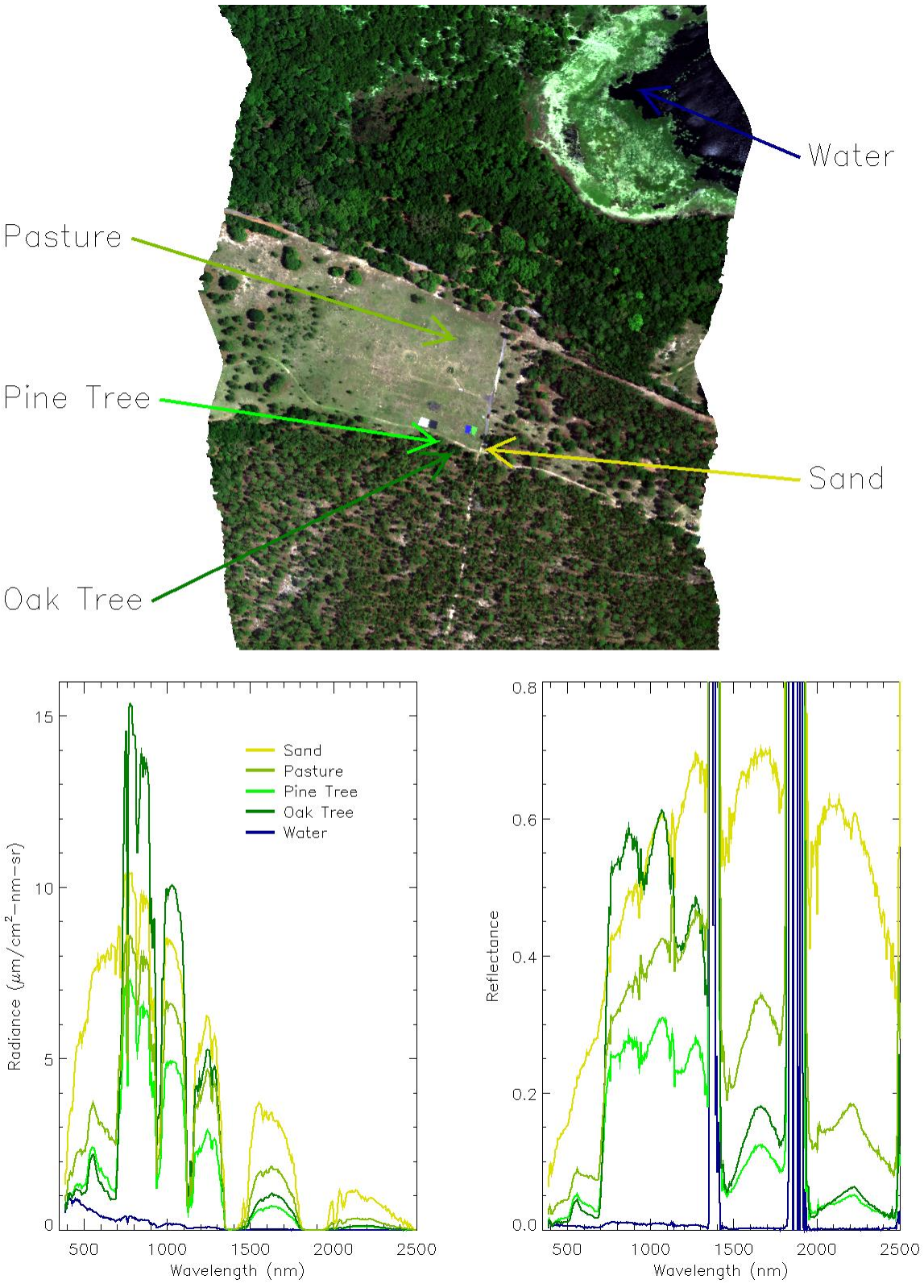

Figure 29: At-sensor radiance (bottom left) and directional surface reflectance (bottom right) curves for several natural targets whose spatial locations are shown in the top of the figure

The shape of these curves is dominated by the shape of the solar spectral irradiance and atmospheric absorption bands can be seen at several wavelengths. The plot on the right side of the figure shows the corresponding directional surface reflectance curves. Here the pasture, pine tree, and oak tree spectra show the classic vegetation "red edge" where their reflectance values increase significantly around 700 nm. More information about the atmospheric correction is found in an algorithm theoretical basis document (Karpowicz & Kampe, 2015).

All 26 of the North-South survey flight lines in the priority 1 flight plan are used in the research. Imaging spectrometer data from these flight lines are mosaicked together into a seamless data set as described further in Chapter V and shown in Figure 30. A few of the longer lines in the center of the image appear brighter because they were collected later in the afternoon at higher sun angles and the shadows are smaller.

Figure 30: Mosaic of 26 flight lines of imaging spectrometer data

Discrete Return Lidar Data Processing

The discrete return lidar data are processed to a Level 1 point cloud (a series of geolocated points that can be displayed in 3D) product as well as Level 3 gridded surface model products. The processing begins by converting the raw discrete return lidar range file into a geolocated point for each return. The lidar instrument emits a laser pulse and then records the received energy as a function of time. Figure 31 illustrates a single laser pulse with the outgoing pulse shape on the left and a return waveform from vegetation on the right.

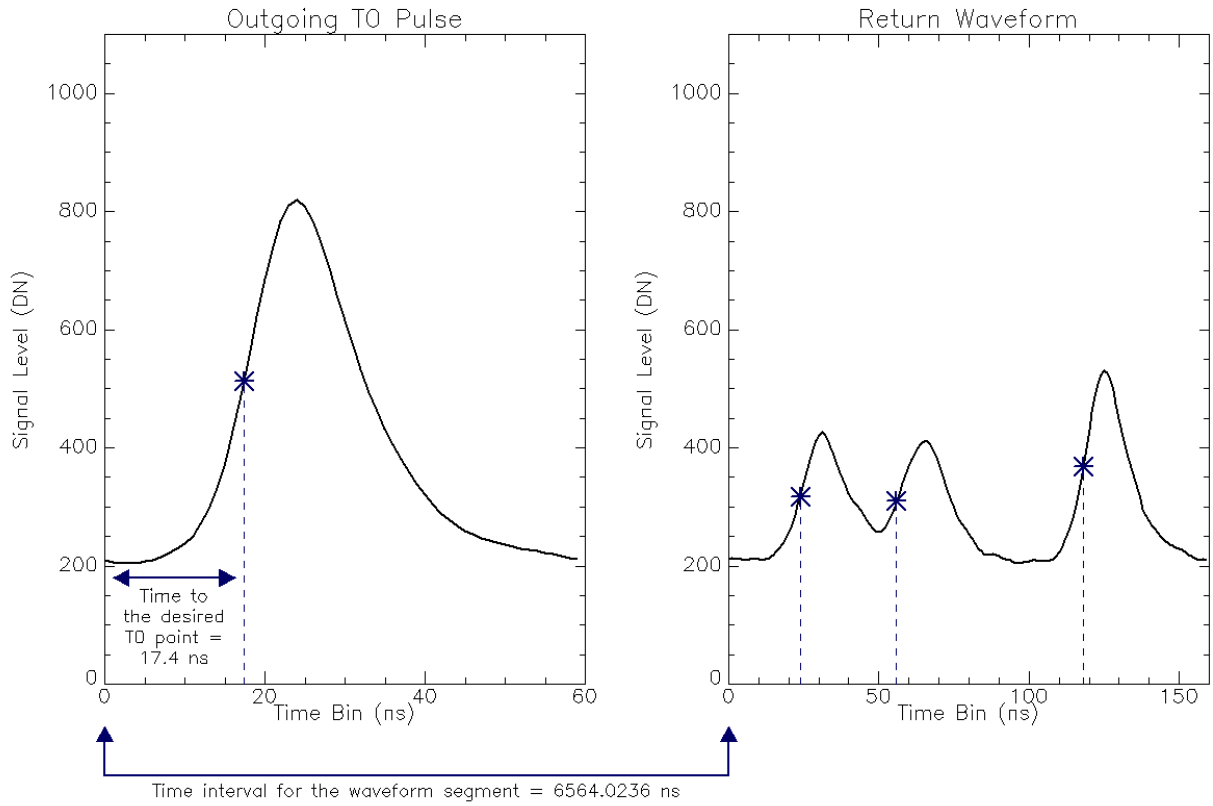

Figure 31: Discrete lidar ranging example using a simple 50% leading edge target detection algorithm

The system records the time interval between the outgoing pulse and the received signal. The figure shows a simple object detection algorithm that marks the 50% leading edge of any peaks in the signal. The range to a given target can be calculated by using the time of flight between the leading edge of the outgoing pulse and the leading edge of a given detected target (Stilla & Jutzi, 2009) as well as taking into account the speed of light and the index of refraction of air according to:

$$R = \frac{c}{n} \cdot \frac{\tau}{2},$$

Where R is the range in m, c is the speed of light (m/s), and τ is the time interval between the outgoing pulse and the received signal, and n is the index of refraction of air which varies with temperature, pressure, and humidity.

The Optech Gemini analyzes the received signals in real-time to detect targets using a proprietary algorithm. Instrument and timing information is stored in the raw range file which is processed by NEON using the Optech LMS software. LMS calculates the range for each detected return and then uses the mirror angle data as well as the SBET file to calculate the geometric position of the return on the Earth's surface. Special test flights are conducted by NEON at the beginning of the flight season to calibrate the boresight angles for the lidar instrument and the LMS software uses the cross-strip and overlapping areas to adjust the flight lines to be self-consistent. More information about the lidar processing is found in an algorithm theoretical basis document (Krause & Goulden, 2015). The output of LMS is a Level 1 geolocated point cloud data product where each return has a set of X, Y, and Z coordinates (Easting, Northing, and Geoid height) and a corresponding peak signal intensity value. Figure 32 illustrates discrete return lidar points for a couple trees located in a field as a 2D perspective view rendering of the 3D data.

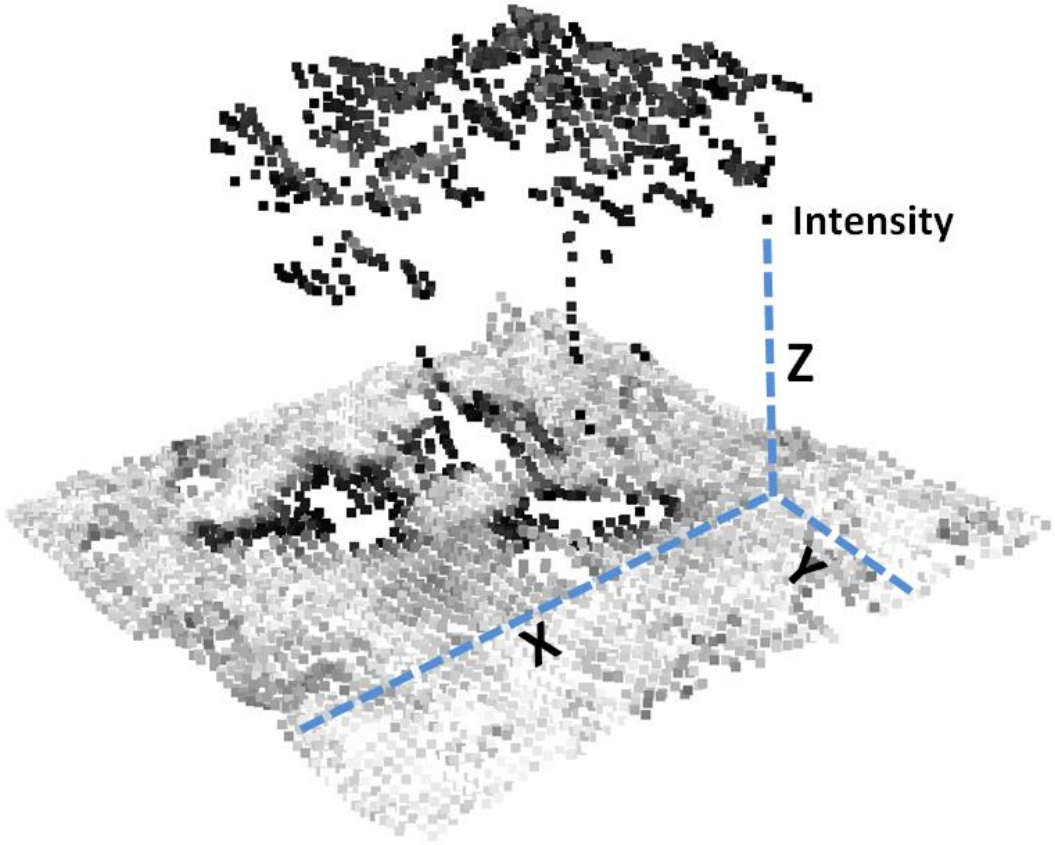

Figure 32: Discrete return lidar point cloud showing a couple trees in a field. Each return has a set of X, Y, and Z coordinates as well as a corresponding intensity value

Figure 33 shows discrete return lidar points for a single flight line at OSBS colorized by height with low elevations in blue and high elevations in red.

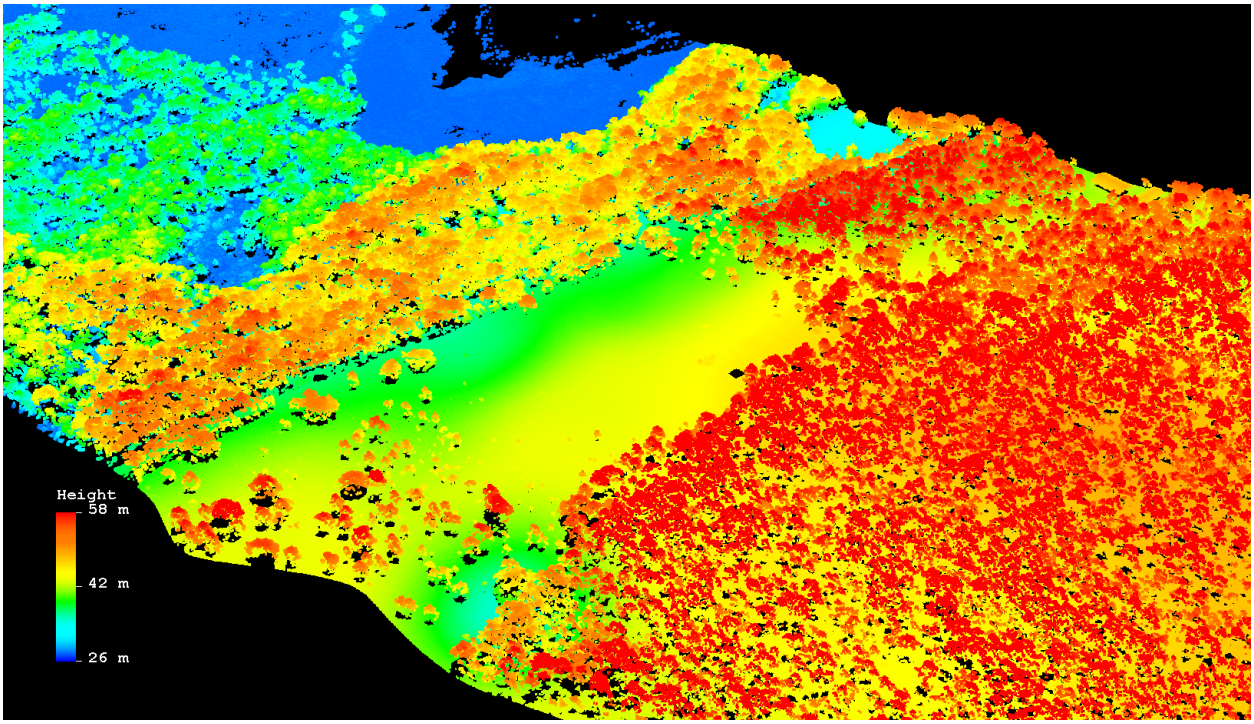

Figure 33: Discrete return lidar points for a single flight line colorized by height (blue for low elevations and red for high)

All 26 of the North-South survey flight lines in the priority 1 flight plan are used in the research. Discrete return lidar data from these flight lines are mosaicked together into a seamless data set as shown in Figure 34:

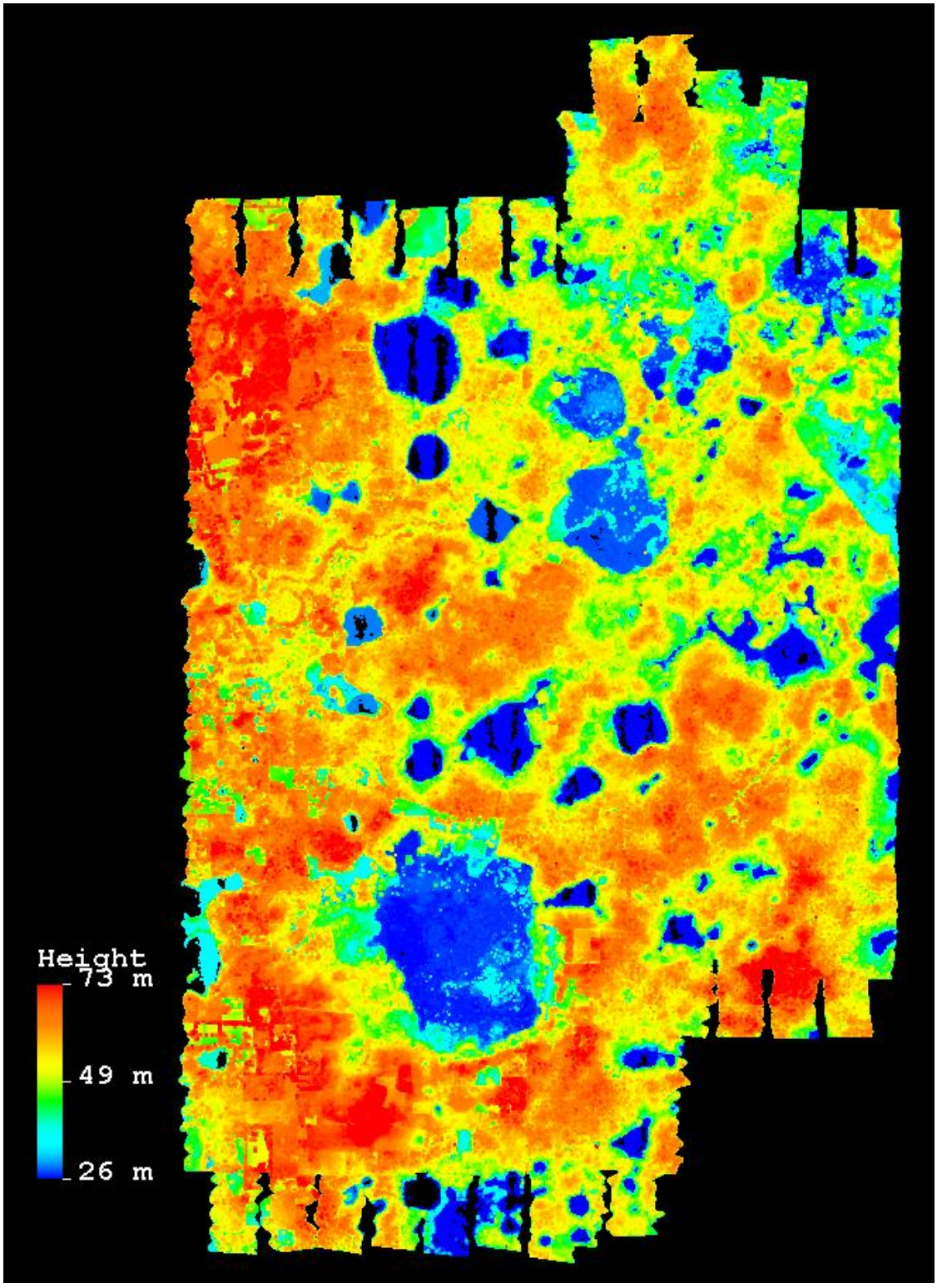

Figure 34: Mosaic of 26 flight lines of discrete return lidar data

Level 3 surface model products are derived from the discrete return point clouds using LASTools (Rapidlasso, 2015). First, noisy data are removed from the point clouds using the “lasnoise” routine. Next, ground points are classified using the “lasground” routine and buildings and high vegetation points are classified using the “lasclassify” routine. The ground points are gridded into a digital terrain model (DTM) using the “las2dem” routine. The first return points are gridded into a digital surface model (DSM) using the “las2DEM” routine. Finally, the height above ground is calculated for each vegetation point by subtracting the DTM using the “lasheight” routine. The height above ground vegetation points are gridded into a canopy height model (CHM) using the “lasgrid” routine. All three of the surface model products are created on a fixed 1 m resolution output grid. Figure 35 shows a 1 km x 1 km area with the DSM in the upper left, the DTM in the upper right, and the CHM in the lower left.

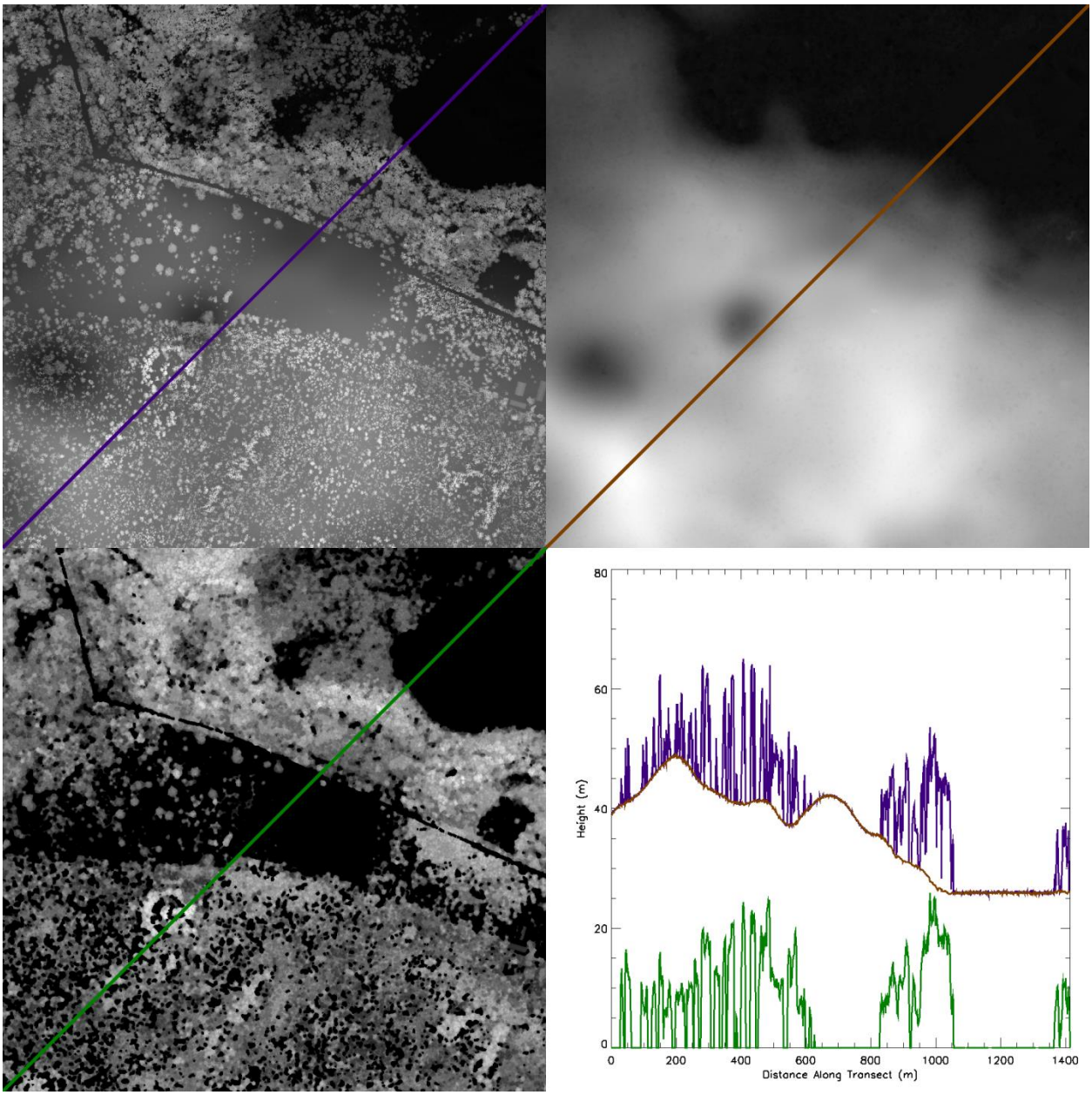

Figure 35: Gridded Digital Surface Model (DSM) in the upper left, Bare-Earth Digital Terrain Model (DTM) in the upper right, and Canopy Height Model (CHM) in the lower left. The plot in the lower right shows the heights along a diagonal transect with the DSM in purple, DTM in brown, and CHM in green

The height of each model is plotted for a diagonal transect starting at the lower left of each data set and moving across to the upper right. The DSM (purple curve) and DTM (brown curve) are in absolute GEOID height. Surfaces such as trees stand above the bare ground but both curves match in the middle

where there is an open pasture and near the right of the plot where there is a lake. The DTM is subtracted from the DSM to derive the CHM which shows the height of the vegetation (green curve).

National Ecological Observatory Network Ground Validation Data

NEON conducted ground sampling at Ordway-Swisher Biological Station in 2010, 2011, 2013 and 2014. The 2010 Pathfinder Campaign combined airborne remote sensing with terrestrial observations of vegetation. At the time, the AOP instruments hadn't been purchased yet so NASA JPL was contracted to collect imagine spectrometer data with the AVIRIS Classic (JPL, 2015b) instrument and the National Center for Airborne Laser Mapping (NCALM, 2015) was contracted to collect lidar data with an Optech Gemini, flown separately than AVIRIS. The terrestrial observations performed by NEON personnel and contractors focused on the airshed region surrounding the proposed location of the NEON tower in the Sandhill ecosystem. Leaf area index measurements were made along 8 transects on the ground and vegetation structure data (canopy height, canopy width, and trunk diameter at breast height DBH) were collected for a small area along one transect. In addition, vegetation species lists were created for 26 ground locations distributed across the site in different ecosystems. Species were identified in a 7.31 m radius circle (168 m^2) at each location. In 2011, a more detailed site characterization effort was conducted by NEON. Species lists were created for 47 ground locations distributed across the site and photographs were taken. An example of a species list from one location is given:

Plot: os343c

No doubt of identification

aevi	aeschynomene viscidula, sticky jointvetch
agju2	ageratina jucunda, hammock snakeroot
ammu2	amphicarpum muehlenbergianum, muhlenberg maidencane
anfl	andropogon floridanus, florida bluestem
angy2	andropogon gyrans, elliott`s bluestem
ante2	andropogon ternarius, splitbeard bluestem
anvi2	andropogon virginicus, broomsedge bluestem
arco7	aristida condensata, piedmont threeawn
armo3	aristida mohrii, mohr`s threeawn
arpu8	aristida purpurascens, arrowfeather threeawn
arst5	aristida stricta, pineland threeawn
arse3	aristolochia serpentaria, virginia snakeroot
asin12	asimina incana, woolly pawpaw
buci	bulbostylis ciliatifolia, capillary hairsedge
chni2	chamaecrista nictitans, partridge pea
chfl2	chapmannia floridana, florida alicia
clma4	clitoria mariana, atlantic pigeonwings
crro5	crotalaria rotundifolia, rabbitbells
crar2	croton argyranthemus, healing croton
cycr6	cyperus croceus, baldwin`s flatsedge
cyre5	cyperus retrorsus, pine barren flatsedge
dapi2	dalea pinnata, summer farewell
diac	dichanthelium aciculare, needleleaf rosette grass
dico6	digitaria cognata, fall witchgrass
difi	digitaria filiformis, slender crabgrass
divi5	diospyros virginiana, common persimmon
erto2	eriogonum tomentosum, dogtongue buckwheat
erar13	eryngium aromaticum, fragrant eryngo
gare2	galactia regularis, eastern milkpea
gavo	galactia volubilis, downy milkpea
hopr	houstonia procumbens, roundleaf bluet
lehi2	lespedeza hirta, hairy lespedeza
lipa8	liatris pauciflora, fewflower blazing star
lite6	liatris tenuifolia, shortleaf blazing star
limi5	licania michauxii, gopher apple
ophu	opuntia humifusa, devil`s-tongue
pase5	paspalum setaceum, thin paspalum
phgr12	phoebanthus grandiflorus, florida false sunflower
pipa2	pinus palustris, longleaf pine
pigr4	pityopsis graminifolia, narrowleaf silkgrass
ptaq	pteridium aquilinum, western brackenfern
ptpy2	pterocaulon pycnostachyum, wand blackroot
quge2	quercus geminata, sand live oak
quin	quercus incana, bluejack oak
qula2	quercus laevis, turkey oak
rhci3	rhynchosia cinerea, brownhair snoutbean
rhmi10	rhynchosia michauxii, michaux`s snoutbean
rhre	rhynchosia reniformis, dollarleaf

ruca4	ruellia caroliniensis, carolina wild petunia
scmu4	scutellaria multiglandulosa, small`s skullcap
seto7	sericocarpus tortifolius, dixie whiettop aster
smau	smilax auriculata, earleaf greenbrier
sood	solidago odora, anisescented goldenrod
sose5	sorghastrum secundum, lopsided indiangrass
spju	sporobolus junceus, pineywoods dropseed
stsy	stillingia sylvatica, queen`s-delight
stab	stylisma abdita, showy dawnflower
stpa8	stylisma patens, coastalplain dawnflower
stbi2	stylosanthes biflora, sidebeak pencilflower
tefl	tephrosia florida, florida hoarypea
tevi	tephrosia virginiana, virginia tephrosia
trro4	tradescantia roseolens, longleaf spiderwort
	Uncertain of genus
spju	sporobolus junceus, pineywoods dropseed

This ground location is within a Sandhill ecosystem based on the indicator species longleaf pine (Pinus palustris), turkey oak (Quercus laevis), and wiregrass (Aristida stricta var. beyrichiana). A photo of the location also confirms the classification as Sandhill as shown in Figure 36:

Figure 36: Photo of NEON 2011 ground site os343c, a Sandhill ecosystem (photo courtesy of NEON)

Charring seen on the trunks of the pine trees is a reminder that OSBS actively burns the Sandhill regions periodically to preserve those ecosystems.

FNAI communities are assigned to the 2010 and 2011 ground validation locations using the species lists, site photos, and photo-interpretation of USGS NAIP imagery. Table 7 lists the number of ground validation locations in each of the 10 FNAI vegetation communities and Figure 37 shows the locations on a map.

Table 7: Number of Ground Validation Locations for Each FNAI Vegetation Community

	Number of Validation Locations		
FNAI	NEON 2010	NEON 2011	Total
Basin Marsh	4	3	7
Basin Swamp	1	2	3
Baygall	0	0	0
Lake	0	0	0
Lake Fringe	3	1	4
Pine Plantation	0	0	0
Ruderal	0	2	2
Sandhill	10	20	30
Upland Mixed Forest	3	9	12
Xeric Hammock	5	10	15
Total	26	47	73

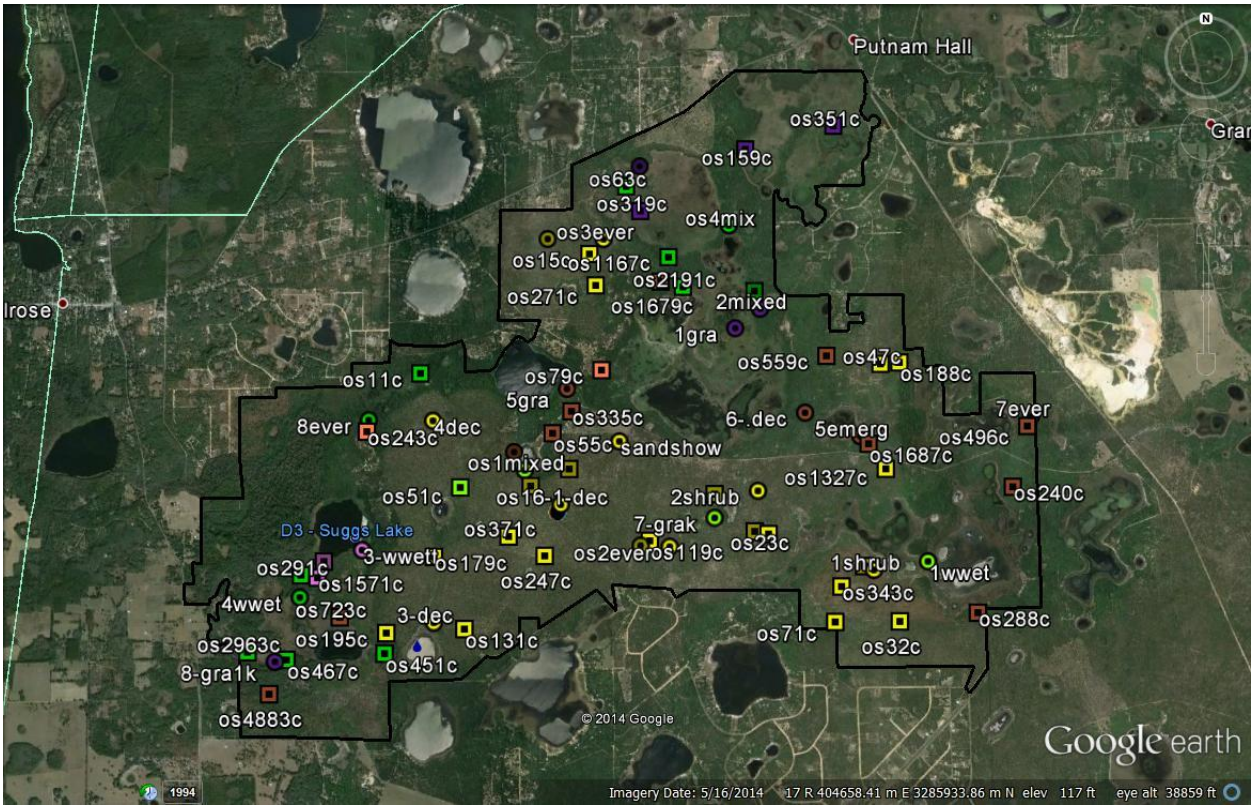

Figure 37: Map of the 2010 and 2011 NEON ground validation locations where species lists were generated

In total 30 Sandhill validation locations were identified, 10 from 2010 and 20 from 2011. These ground locations are used in the research as independent validation points to assess the accuracy of the classification models in the following chapters.

The total number of vegetation species at each ground location is tallied and plotted in Figure 38 sorted by FNAI vegetation community.

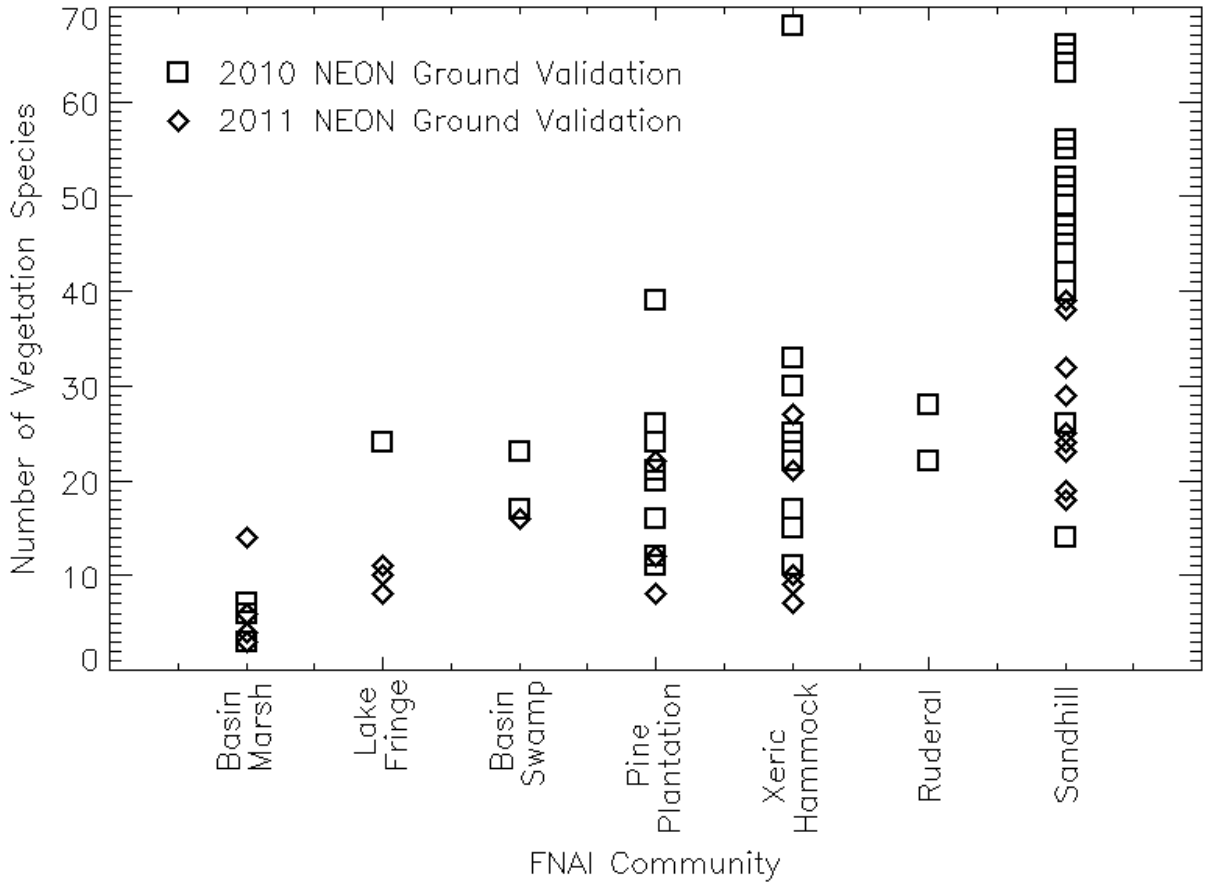

Figure 38: Number of vegetation species at each of the 73 ground validation locations

A relationship exists between species richness and FNAI community, i.e. some communities support a larger diversity of vegetation species. The data suggest that a map of the FNAI communities could serve as a proxy of vegetation diversity.

NEON also measured the spectral reflectance of several vegetation species in 2010 and 2014, performed site selection of the distributed plots (for use in NEON operations for the 30 year observatory period) in 2013, and collected foliar samples for biogeochemical analysis in 2014.

Classification Model Training Data

Several popular machine learning classification algorithms exist but decision trees are used in this research as they are easy to interpret and understand the relationship between inputs and outputs. Classification and Regression Trees (CART) is the specific implementation of the algorithm used (Breiman et al., 1984). CART creates a decision tree through recursive splitting of the data, where each split has two branches. The decision rules are similar to a parallelepiped classifier in that only a single feature is considered in a split and the decision boundaries are orthogonal to the feature axes (Tso & Mather, 2009). The algorithm builds the tree by testing each possible input feature and all possible split values for each feature. A Gini impurity index is calculated for each hypothetical split to analyze how well the output training data point classes are separated (Breiman et al., 1984; Therneau et al., 2015). Essentially, the split that separates the data classes into the two most dissimilar groups is chosen for the decision tree node. If the training point classes could be split perfectly into two groups, the Gini impurity index would equal 0 and this would be the best split. If the training classes are evenly mixed after the split, the Gini impurity index would equal 0.5. The lower this index, the better the splitting decision. Interpretation of the resulting decision tree structure allows understanding of the relationships between objects (data input features representing the objects) at varying scales of observation (Tso & Mather, 2009).

Classification and regression trees (CART) are a type of supervised classification algorithm. CART models such as package “rpart” in R require training data where the output classes are explicitly labeled (Therneau et al., 2015). A set of 1000 training locations are hand selected through photo interpretation of USGS NAIP imagery using the OSBS vegetation community GIS vectors for reference.

100 locations are selected for each of the 10 FNAI vegetation communities. Since these points are hand-picked, they are spatially distributed across the site as best as possible. Features are calculated from the hyperspectral and lidar data for each of the 1000 locations and used to train the classification tree. A 20 m x 20 m spatial plot is chosen to calculate the feature metrics as that size is consistent with the NEON TOS distributed vegetation sampling ground plot size (Barnett, 2014). The classification model will then be applied to predict the FNAI vegetation community at other locations across the site (also using calculated features from a 20 m x 20 m plot), as discussed further in the following chapters. A map of the training points is shown in Figure 39:

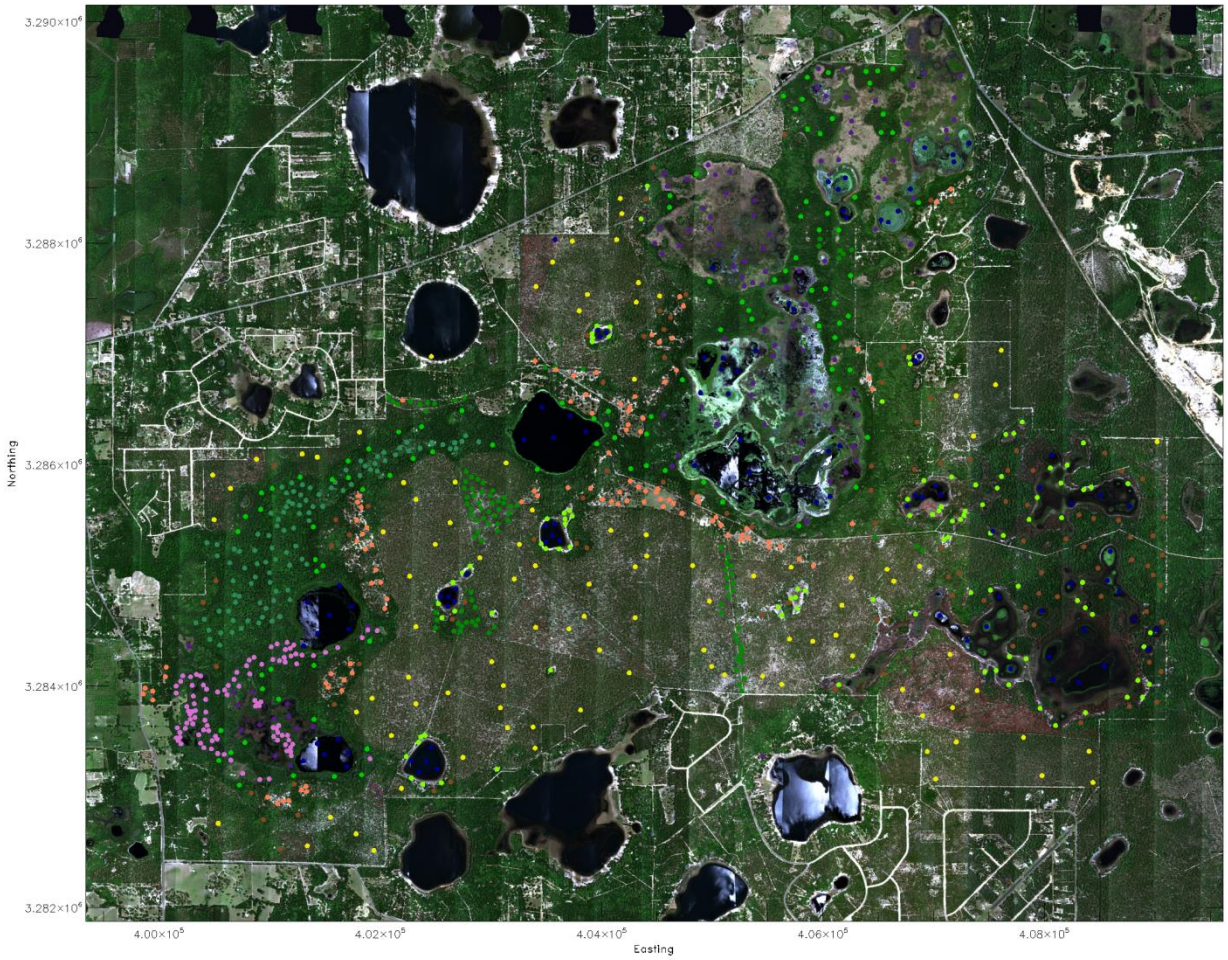

Figure 39: Map of the hand-picked locations used in training of the classification algorithm

CHAPTER V

VEGETATION CLASSIFICATION USING SPECTRAL VARIABILITY AS MEASURED BY HYPERSPECTRAL REMOTE SENSING

The NEON Imaging Spectrometer (NIS) measures the reflected solar radiation in 424 channels over the 380 nm-2500 nm spectral range with a spectral resolution of about 5-6 nm. The instrument can detect medium-to broad spectral features due to absorption and scattering of light by vegetation and other ground targets. In this chapter, the hyperspectral data provided by the Airborne Observation Platform (Kampe et al., 2010; Kampe et al., 2011; NEON, 2015h) of the National Ecological Observatory Network (NEON, 2015a; NRC 2003; Schimel et al., 2011) are prepared with some data processing before a series of hyperspectral features are calculated. The features for the ground training locations are used to generate a decision tree classification algorithm with the goal of producing a classification map for the Ordway-Swisher Biological Station (OSBS, 2015a; OSBS, 2015c; OSBS, 2015d) site using the Florida Natural Areas Inventory (FNAI, 2010) vegetation communities as the output classes. Classification accuracy is assessed using k-folds cross validation and the independent ground truth data collected by NEON.

Objectives

Objectives

- Investigate how hyperspectral data can be used to observe vegetation species diversity through spectral (chemical or functional) differences.
- Generate a vegetation community map using spectral differences in the data. This will serve as a proxy for vegetation species diversity.

Science questions

- How do vegetation communities differ spectrally?
- How is spectral diversity related to species diversity?
- What spectral differences can we observe in the remote sensing data?
- How is the measured variability in the remote sensing data related to biodiversity or vegetation community type?

Hyperspectral Data Preparation

The NEON hyperspectral products require some data preparation before feature layers are calculated and the classification algorithm is run. As part of the NEON standard processing, an atmospheric correction is applied and the data are processed to surface reflectance. Each flight line is also geolocated onto the Earth's surface and an ortho is generated with a pixel spacing of 1 m and a North-up WGS-84 UTM projection on a standard grid. First, the data from all of the flight lines need to be combined into a single mosaic for the entire site. The flight lines are planned to have an overlap of 30% but in reality the actual overlap can be smaller or larger due to the aircraft position and attitude or turbulence. In the case where multiple pixels from more than one flight line fall within the same grid cell of the mosaic, the pixel with the smallest off-nadir angle is used. A portion of the mosaic is shown in Figure 40:

Figure 40: A true color image of a portion of the hyperspectral data showing several flight lines stitched together to form a mosaic for the entire site. Ground calibration tarps can be seen in the field

Here four flight lines are stitched together. The seam lines can be seen due to solar illumination, brdf, and shadow differences between adjacent flight lines and in some cases flight line edge artifacts from the atmospheric correction algorithm show up as slight brightness variations. Some of the features calculated in the next section will cancel out these illumination geometries through band ratios but hopefully enough training locations have been selected across the mosaic to account for the variations within each vegetation community.

Next the quality of the surface reflectance curves is evaluated. A series of endmembers are chosen for a single flight line. These endmembers are unique objects on the ground with a size larger than 1 m such that the spectra from a single pixel within the object will represent a spectrally pure pixel, for

example: water, sand, or vegetation from a single tree. The delivered reflectance curves versus wavelength for the endmembers are plotted in Figure 41:

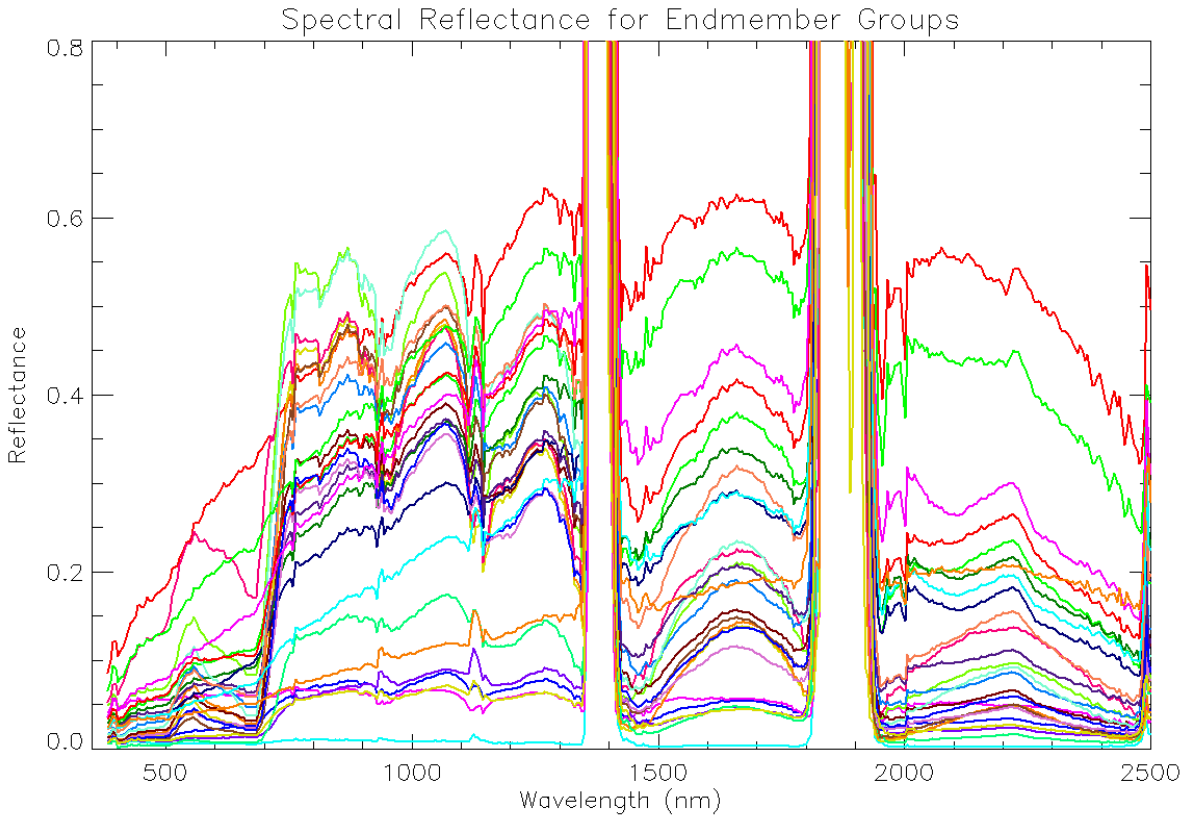

Figure 41: Delivered reflectance versus wavelength for endmembers pixels. Atmospheric artifacts and other noise can be seen in the data

The reflectance curves appear to have noise plus some bias at wavelengths with atmospheric absorption. Several potential error sources include: stray light within the spectrometer, spectral or radiometric calibration errors, and issues with the atmospheric correction algorithm not using the best atmospheric parameters.

Black (3%) and white (48%) calibration tarps were deployed during the data collection and can be seen in the grass field in Figure 40. The reflectance of the tarps is measured in the field with an ASD spectrometer. The ground-truth reflectance can be compared to the data product reflectance as measured by the airborne NEON Imaging Spectrometer and is shown in Figure 42:

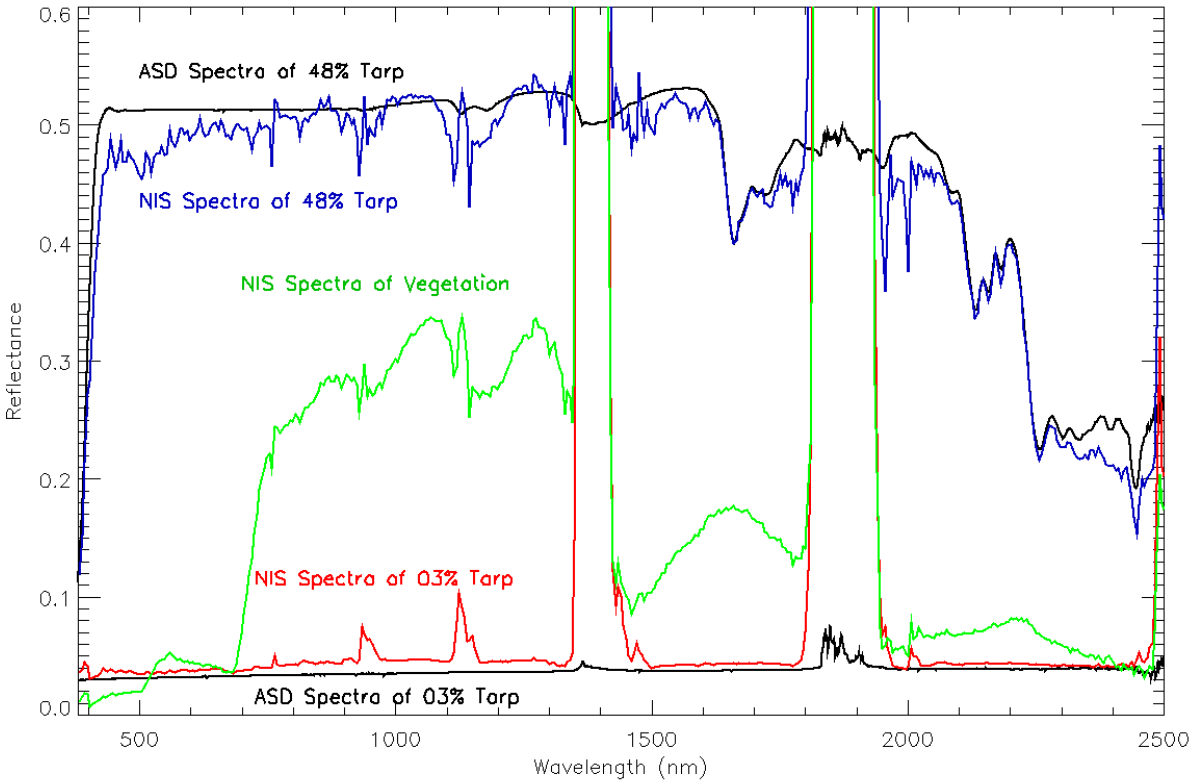

Figure 42: The as-delivered reflectance of calibration tarps (03% tarp in red and the 48% tarp in blue) and a vegetation pixel (green) measured by the airborne NEON Imaging Spectrometer. Noise and atmospheric artifacts are seen in the NIS curves. The black curves are ground truth reflectance measured by an ASD spectrometer

The 3% tarp (red curve), the 48% tarp (blue), and a vegetation pixel (green) curves are noisy and have atmospheric artifacts compared to the ground truth ASD curves (black) which are very smooth.

An empirical line correction is performed to improve the quality of the airborne reflectance spectra. Gain and offset values are calculated by adjusting the airborne spectra of the tarps so that they perfectly match the ground truth ASD measured reflectance, i.e. adjust the blue and red curves so they match their respective black curve. Tarps were deployed in the field for each of the two flight collection dates, and so two sets of empirical line coefficients are calculated to account for the atmospheric differences between May 7 and May 19, 2014. The gain and offset adjustment coefficients are shown in Figure 43 for both dates:

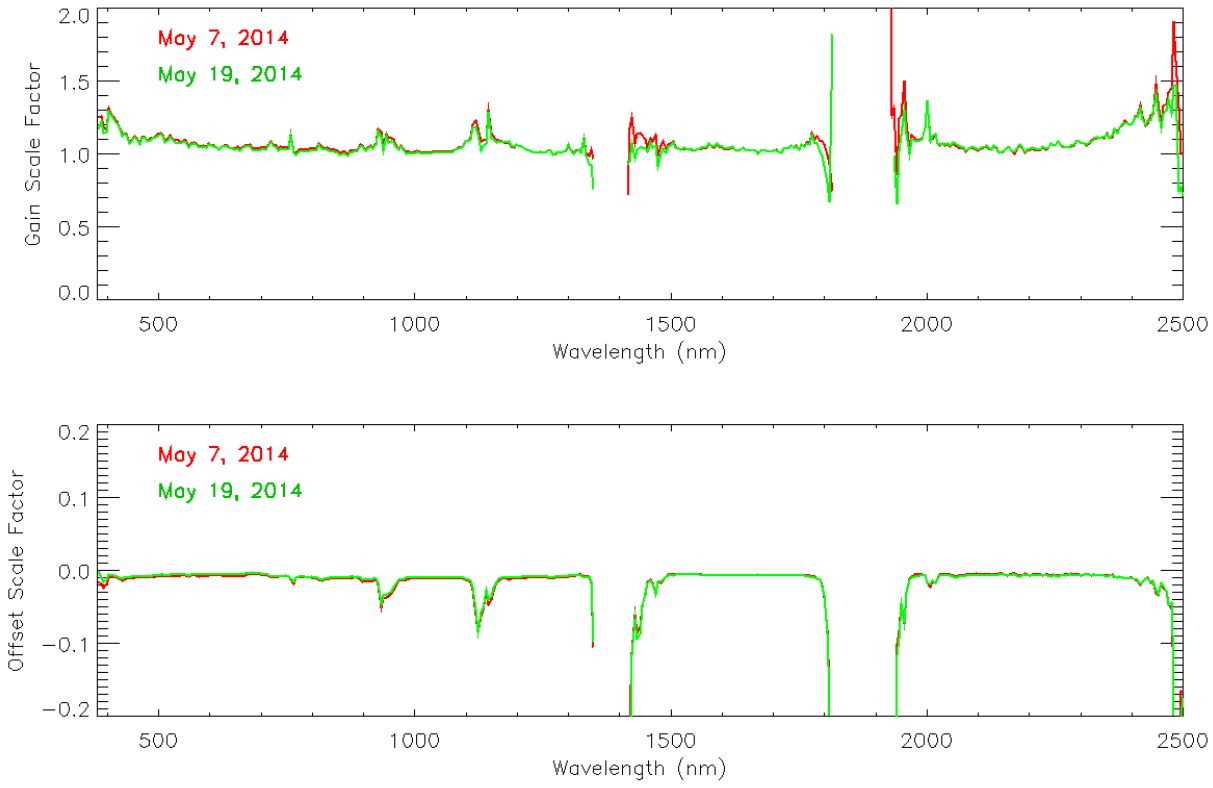

Figure 43: Gain and offset coefficients used in the empirical line adjustment to improve the quality of the NEON Imaging Spectrometer reflectance products collected on May 7 and May 19, 2014 at Ordway-Swisher Biological Station

Flight line reflectance values are adjusted using the scale factors from their respective dates.

Once adjusted, the tarps in the hyperspectral data will perfectly match the ground data. Figure 44 illustrates the improvement with the atmospheric spikes significantly reduced.

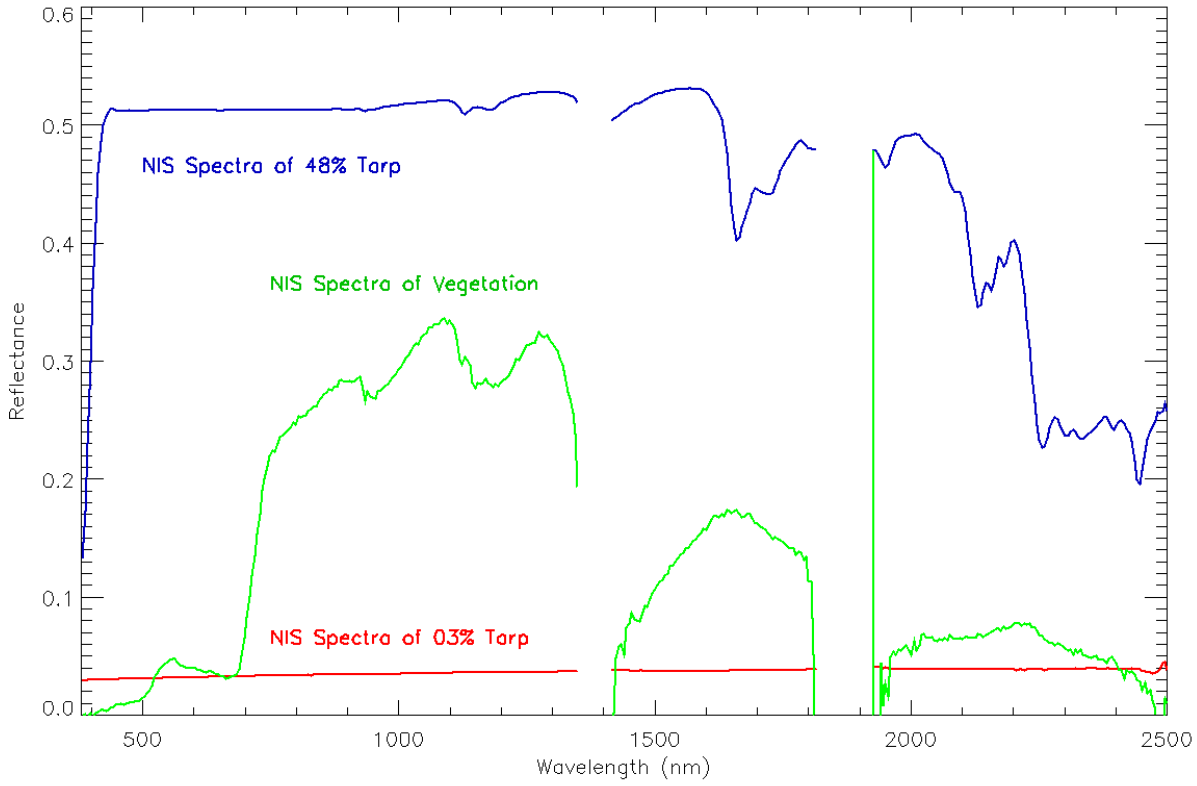

Figure 44: Adjusted reflectance of calibration tarps (03% tarp in red and the 48% tarp in blue) and a vegetation pixel (green) measured by the airborne NEON Imaging Spectrometer

The adjusted endmember reflectance curves are shown in Figure 45:

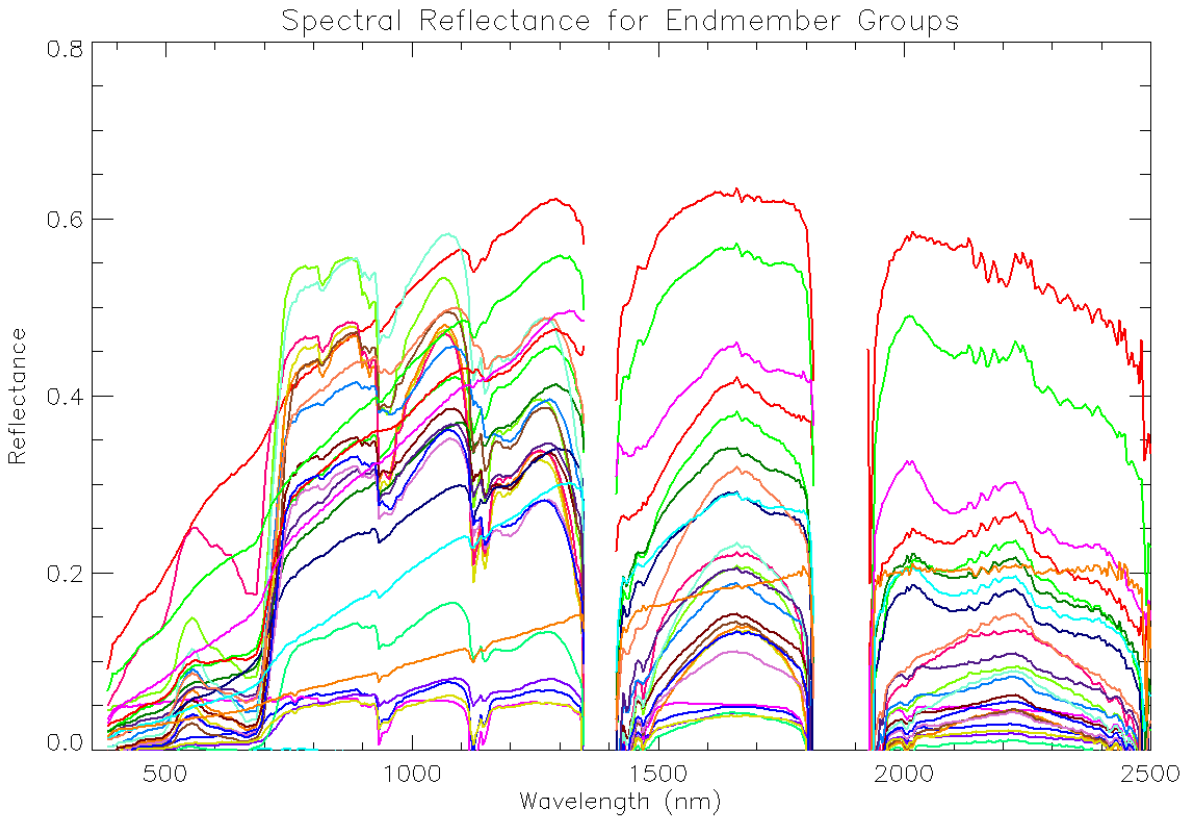

Figure 45: Reflectance versus wavelength for endmembers pixels, adjusted using scale factors generated by using ground calibration tarps

The curves are considerably smoother than those in Figure 41 except at wavelengths longer than 2100 nm where scene targets are brighter than the 48% reflectance tarp leading to extrapolation of the scale factors beyond their useful range.

Finally, 424 spectral bands is a large number of inputs considering that the reflectance is highly correlated for neighboring bands. The desire is to represent the overall shapes of the reflectance curves, but using a subset of wavelengths. Based on this ideal, 32 wavelengths have been hand selected. Those

wavelengths are 467.8, 502.9, 527.9, 553.0, 578.0, 603.1, 643.2, 668.2, 708.3, 743.4, 783.4, 878.6, 993.8, 1064.0, 1104.0, 1199.2, 1249.3, 1284.4, 1314.4, 1514.8, 1544.9, 1585.0, 1615.0, 1650.1, 1685.1, 1730.2, 2035.8, 2105.9, 2161.1, 2216.2, 2281.3, and 2326.4 nm. The chosen wavelengths are shown in Figure 46 as the vertical dashed lines.

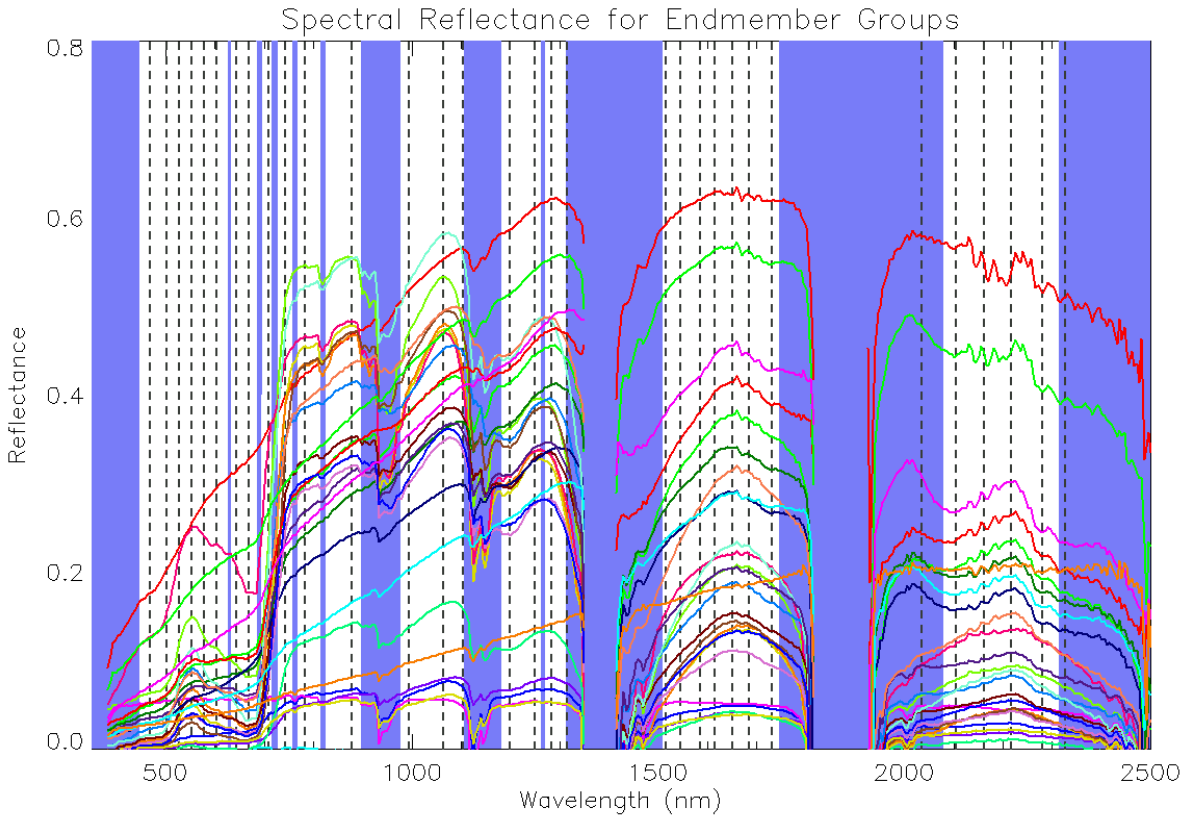

Figure 46: Selection of 32 wavelengths for use in feature calculation, show as the vertical dashed lines on top of the endmember spectra. The blue regions are wavelengths with higher atmospheric absorption

Wavelengths with higher absorption by the atmosphere are also colored blue. While a few bands were selected near the edges of the blue atmospheric keep out zones, most are within atmospheric transmissions windows.

The full hyperspectral mosaic image containing 424 bands is reduced to 32 bands. In order to clean up noisy data a bit more, each output band is the average of three consecutive hyperspectral bands (effectively the average reflectance over 15 nm) centered on each of the 32 selected wavelengths. The first derivative of reflectance is also calculated for each of the 32 wavelengths using the slope calculated from the two adjacent hyperspectral bands on either side of a given output wavelength (± 5 nm).

A 20 m x 20 m spatial area is subset, centered on each of the classification training locations and the mean reflectance is calculated for each of the 32 selected wavelengths. Figure 47 shows the mean reflectance for each of the 10 FNAI vegetation communities versus the 32 selected wavelengths.

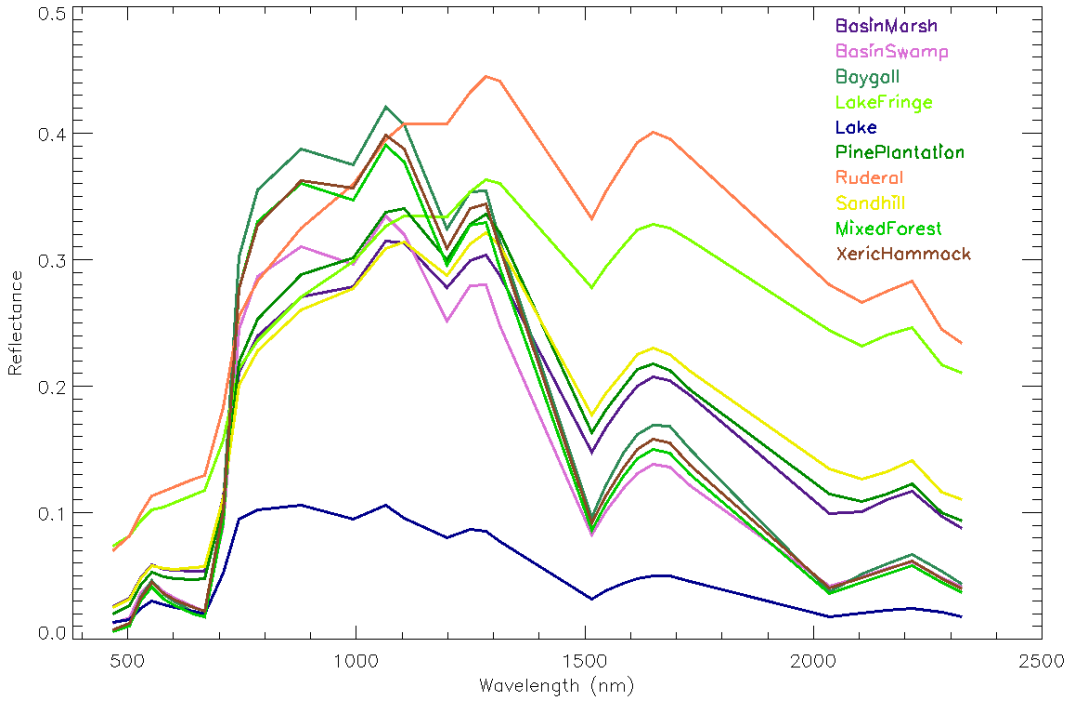

Figure 47: The mean reflectance of the classification training data for each of the 10 FNAI vegetation communities versus the 32 selected wavelengths

In a general sense, the orange Ruderal and chartreuse Lake Fringe reflectance is higher for most wavelengths and the dark blue Lake reflectance is lower for most wavelengths. The relative shapes from wavelength to wavelength are different for the different vegetation communities, which could be taken advantage of by an appropriate classification feature layer such as a band ratio or normalized index. However, several areas of confusion can be seen. For instance the green Mixed Forest and brown Xeric Hammock curves are similar to each other at all wavelengths (these can be identified at 900 nm in the figure as they have a higher reflectance than all other curves except aqua Baygall).

A considerable amount of spectral variability exists with a given vegetation community. The variation is a combination of vegetation variability across the site and the illumination and reflectance

differences between different flight lines in the airborne data. Figure 48 shows the plot level mean reflectance versus the selected wavelengths for all 100 of the Sandhill training locations.

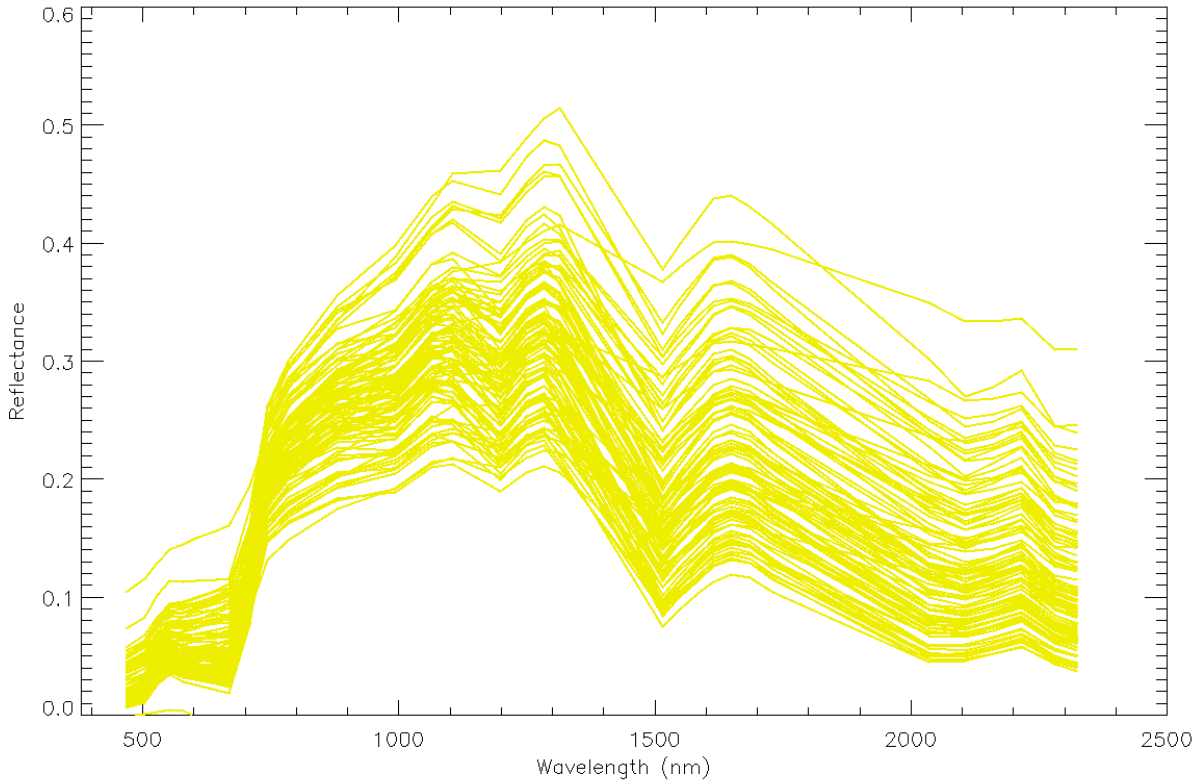

Figure 48: Plot level mean reflectance versus the 32 selected wavelengths for all 100 of the Sandhill training locations

While most of the curves share a similar shape, they vary considerably in absolute reflectance in the near-infrared and shortwave-infrared portions of the spectrum. Also, a few of the curves appear to be outliers with a high reflectance between 500-700 nm and a relatively flat reflectance at wavelengths longer than 1300 nm. Comparing these outlier Sandhill curves to Figure 47, they seem to match quite well with the shapes of the Lake Fringe and Ruderal curves. It is likely that the decision tree will have a difficult correctly classifying these as Sandhill.

The first derivative of reflectance is also calculated for each pixel and the mean value is calculated for a 20 m x 20 m spatial area centered on each of the classification training locations for each of the 32 selected wavelengths. The first derivative captures the slope or shape of the reflectance curves at various wavelengths. Any variations in absolute reflectance are ignored and the relative slope differences can be compared. Figure 49 shows the mean of the first derivative of reflectance for each of the 10 FNAI vegetation communities versus the 32 selected wavelengths.

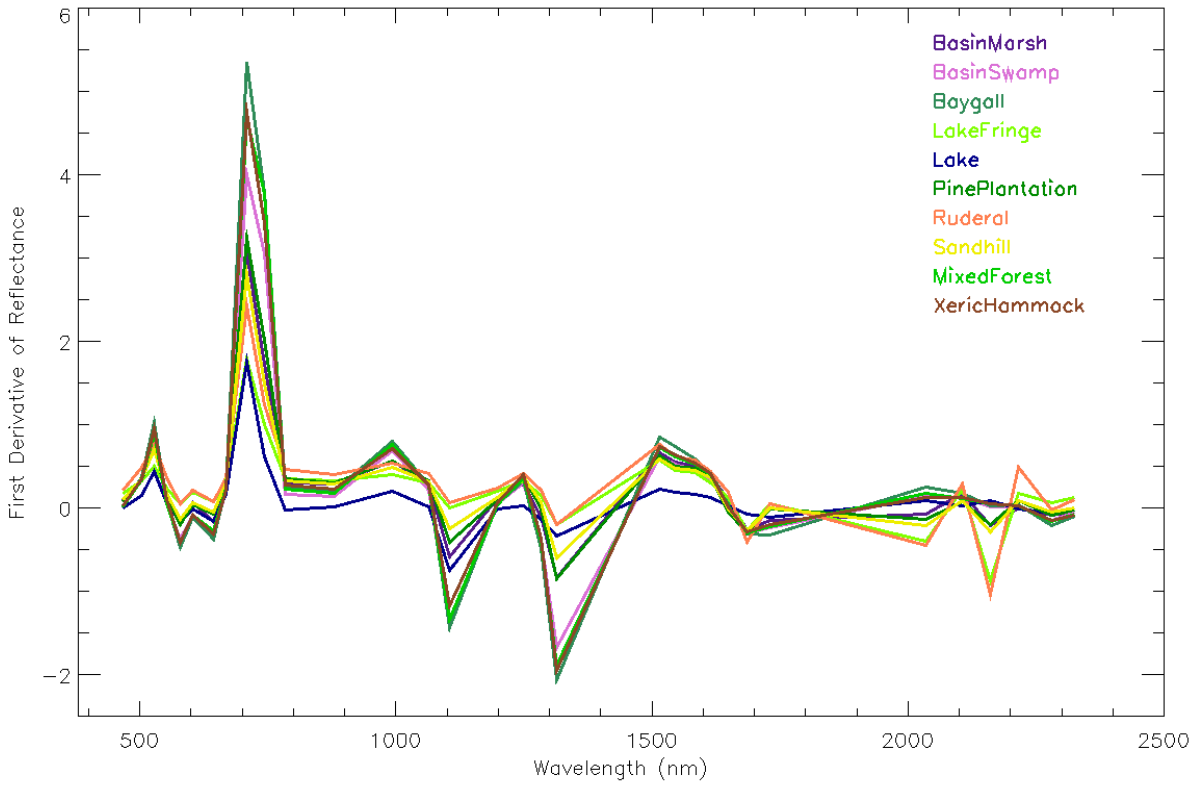

Figure 49: The mean of the first derivative of reflectance of the classification training data for each of the 10 FNAI vegetation communities versus the 32 selected wavelengths

The largest variation occurs around 700 nm at the "red edge" where the reflectance rises significantly for vegetation due to absorption by chlorophyll around 650 nm (low reflectance) and multiple scattering above 750 nm (high reflectance). Communities with more vegetation cover such as Baygall, Mixed Forest, and Xeric Hammock will have a higher slope at this wavelength than communities with open ground and sand such as Lake Fringe and Ruderal, allowing the classification algorithm to take advantage of this feature to separate these classes. Also, at several wavelengths, the dark blue Lake curve has a lower value that stands apart from the other groups making it easier to classify (such as using 1064 nm).

Calculation of Hyperspectral Features

Before the decision tree classification model can be generated, a series of input features must be calculated, in this case hyperspectral features. A 20 m x 20 m data subset centered on a ground plot location is extracted for the feature calculation. A major benefit of this airborne remote sensing data is the high spatial resolution. Working at a plot scale, collections of pixels can be analyzed and related to the vegetation communities which are collections of species. As mentioned previously, reflectance and first derivative of reflectance is calculated on the 1 m full resolution mosaic for each of the selected 32 wavelengths. A basic endmember analysis is also performed on the 1 m data. A suite of feature layers are created:

- The mean and standard deviation of the reflectance for each of the 32 wavelengths is calculated for the 20 m x 20 m plot.
- The mean and standard deviation of the first derivative of reflectance for each of the 32 wavelengths is also calculated for the 20 m x 20 m plot.

- Vegetation pixels are identified using the endmember analysis and the mean and standard deviation of the reflectance for each of the 32 wavelengths is calculated for the vegetation pixels within the 20 m x 20 m plot.
- Vegetation pixels are identified using the endmember analysis and the mean and standard deviation of the first derivative of reflectance for each of the 32 wavelengths is also calculated for the vegetation pixels within the 20 m x 20 m plot.
- Normalized difference indices (NDIs) are calculated for each band pair combination on the mean reflectance values for each of the 32 wavelengths according to:

$$NDI_{\lambda 1-\lambda 2} = \frac{\mu(\rho_{\lambda 1}) - \mu(\rho_{\lambda 2})}{\mu(\rho_{\lambda 1}) + \mu(\rho_{\lambda 2})}.$$

- Normalized difference indices (NDIs) are also calculated for each band pair combination on the mean reflectance values of the vegetation pixels for each of the 32 wavelengths.
- Triangle metrics, similar to a band-depth formula, are calculated on the mean reflectance values using 503 nm, 553 nm, and 668 nm for a green peak and 553 nm, 668 nm, and 743 nm for a red depth according to:

$$Green\ peak = \mu(\rho_{553}) - \frac{\mu(\rho_{503}) + \mu(\rho_{668})}{2}$$

Or

$$Red\ depth = \mu(\rho_{668}) - \frac{\mu(\rho_{553}) + \mu(\rho_{743})}{2}.$$

- A second variation of this triangle metric applies a coordinate rotation using the edge wavelengths and calculating the peak or depth from the center wavelength normal to the line that connects the edge wavelengths.

- Triangle metrics are calculated on the mean reflectance values of the vegetation pixels using 503 nm, 553 nm, and 668 nm for a green peak and 553 nm, 668 nm, and 743 nm for a red depth.
- The basic endmember analysis is used to calculate the fraction of pixels in the 20 m x 20 m plot that were classified as each endmember.

In total, a suite of 1283 hyperspectral features or metrics are calculated for each ground location. A summary of the hyperspectral feature categories is listed in Table 8 and a full list of the hyperspectral features can be found in Appendix A.1. These metrics serve as the input feature layer to the classification model.

Table 8: Summary of Hyperspectral Feature Categories

Category	Examples
Summary Statistics	Mean of reflectance at 743 nm, standard deviation of reflectance at 528 nm, mean of the first derivative of reflectance at 2106 nm
Normalized Difference Indices (NDIs)	NDI 578 nm and 643 nm, NDI 1515 nm and 2216 nm for vegetation pixels
Triangle Metrics	Green peak, Green peak rotated, Red depth, Red depth rotated
Endmember Analysis	Fraction of sand road, fraction of dark water, fraction of Sandhill vegetation type 1, etc.

The features are meant to be proxies of the 2D spectral (chemical, functional, and structural) differences of the vegetation communities as measured by the NEON Imaging Spectrometer instrument.

The imaging spectrometer measures reflected solar radiation, which interacts with the atmosphere and canopy or surface showing influences of scattering and absorption at different wavelengths. Figure 50 shows a false color composite of three features where the red, green, and blue display channels are “vegNDI1515i2216”, “NDI578i643”, and “muDeriv2106” respectively.

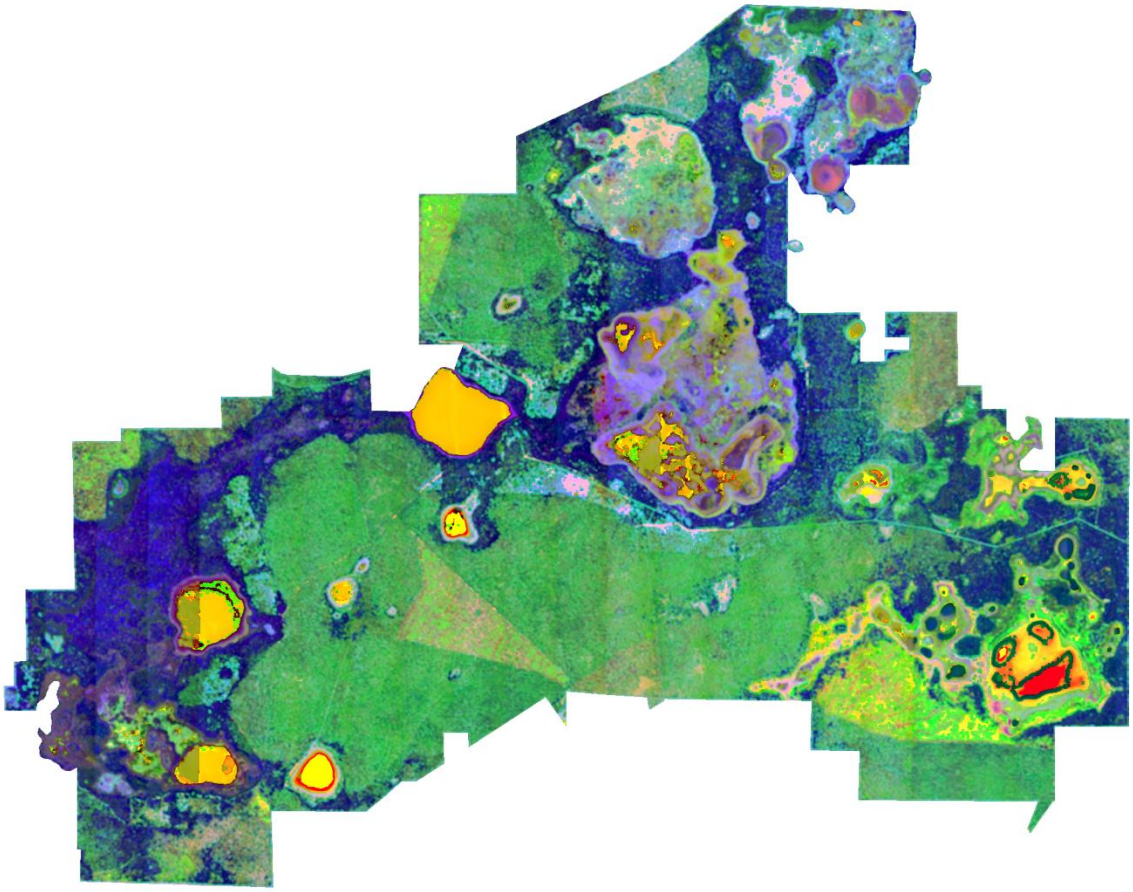

Figure 50: False color composite of three hyperspectral features where the red, green, and blue display channels are “vegNDI1515i2216”, “NDI578i643”, and “muDeriv2106” respectively

In the figure, different FNAI vegetation communities display as different colors, illustrating the potential for a decision tree to classify the vegetation communities using the feature data.

Basin Marsh and burnt Sandhill areas have higher values in the "vegNDI1515i2216" feature, which is displayed in the red channel of the false color image. The green display channel is the hyperspectral feature "NDI578i643". The normalized difference index between wavelengths 578 nm and 643 nm is correlated to the chlorophyll content in a given ground area. Chlorophyll absorbs light at 643 nm so ground locations dominated by vegetation with high chlorophyll content will have a negative "NDI578i643" value. Other natural targets such as soil and sand have an increasing reflectance curve between 578 nm and 643 nm so these will have a positive "NDI578i643" if a ground location is dominated by soil. The 20 m x 20 m spatial sampling size contains soil/sand pixels for Basin Marsh, Sandhill, and Ruderal areas which have higher "NDI578i643" values. Other community types are greener and have higher chlorophyll content so they have lower "NDI578i643" values. The blue channel is the feature "muDeriv2106" in which Basin Marsh, Lake Fringe, Ruderal, and some Baygall areas have higher values. Based on the values in the display channels, Basin Marsh shows up as hues of lavender or chartreuse, Lakes are yellow or red, Baygall, Mixed Forest, and Xeric Hammock are blue and dark purple, Ruderal is pink, and Sandhill is green where burnt areas of active fire management show up as a lighter color than the rest of the Sandhill.

Decision Tree Training

A classification decision tree is generated using the "rpart" package (Breiman et al., 1984; Therneau et al., 2015) in R (R Core Team, 2015). The variable "minbucket" is set to a value of 1 allowing

terminal nodes to contain a single training location. Package “rpart” also has a complexity parameter variable “cp” that determines if the tree should generate more splits along a branch. A split must decrease the lack of fit by a factor of cp or else it is not attempted. The decision tree algorithm will iterate through different threshold values of each input feature, and determine the best data partitioning according to the fitting function of classified outputs (i.e. it picks the split that most correctly classifies the outputs). As described in the previous sections, the lidar data are subset for each of the training ground locations, features are calculated, and those features are brought into package “rpart”, shown in Figure 51:

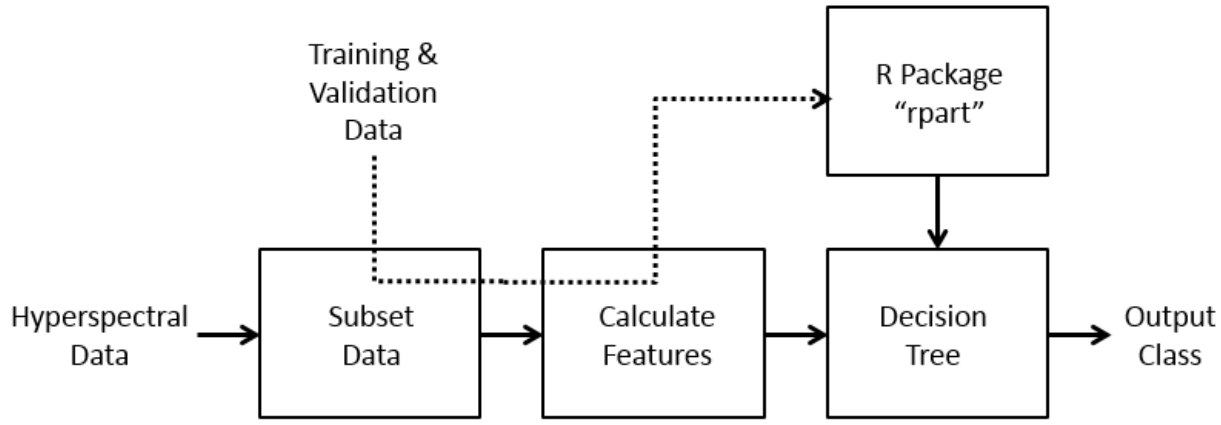

Figure 51: Hyperspectral decision tree classification algorithm flowchart

K-folds cross validation is employed to determine the optimal value of the complexity parameter. A random number generator is used to separate the training data into 10 groups. For each fold, data from 9 of the groups is used to train the classification model and the data from the 10th group is used as independent validation. The process is performed a total of 10 times, each time generating a new decision tree and using a different group as the validation set. The complexity parameter value is varied to get an

idea of the classification accuracy versus complexity parameter. If a very small complexity value is used, the classification model will over-fit the training data. The model will achieve a 100% classification accuracy for the training data but will have lower accuracy values for the validation data. By iterating the complexity parameter and performing the k-folds cross validation 10 times, the complexity parameter value that gives the overall best classification accuracy can be chosen. The results from the k-folds cross validation are shown in Figure 52:

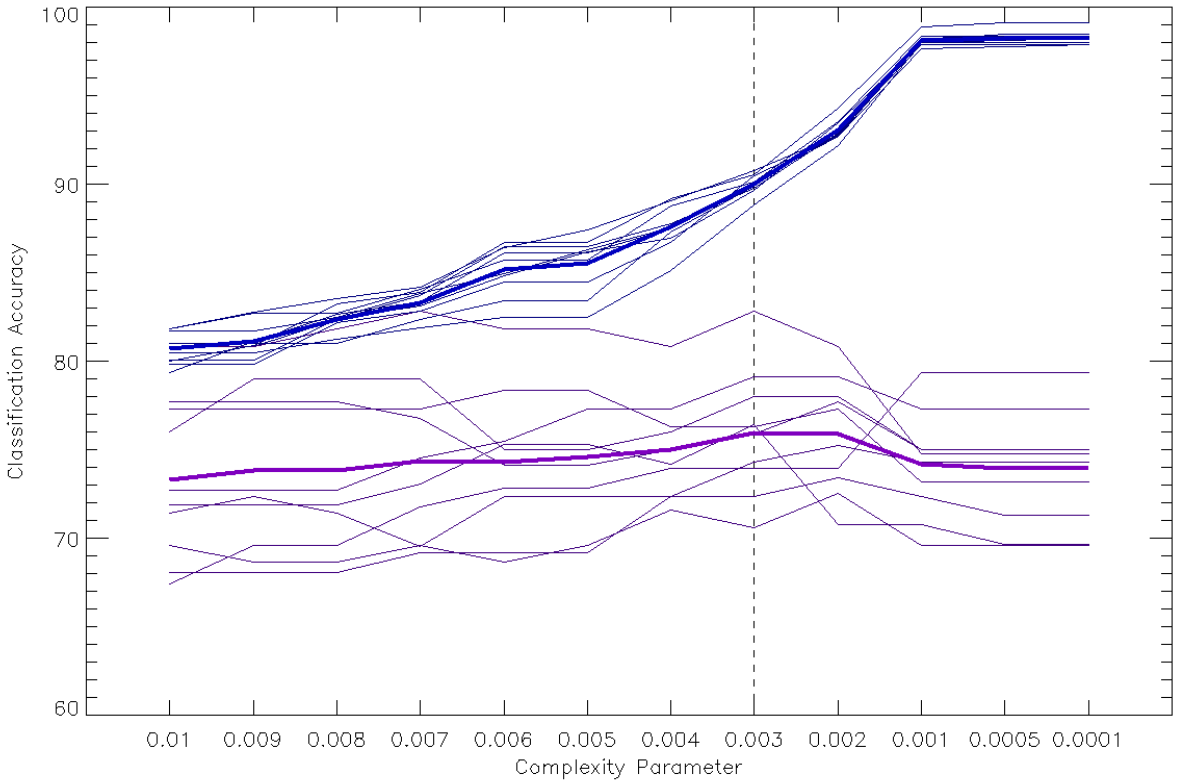

Figure 52: K-folds cross validation to determine the optimal value of the complexity parameter for the hyperspectral decision tree

As seen in the figure, as the value of the complexity parameter decreases, the classification accuracy of the training data increases toward 100%. The classification accuracy of the validation data

increases to a point, but then starts to decrease as the training data is over fit. Based on the k-fold cross validation results of the spectrometer data, a complexity parameter value of 0.003 produced a model with the best validation results. A classification accuracy of 90.0% ± 0.54 is achieved for the training data and an accuracy of 76.0% ± 3.50 is achieved for the validation data.

Classification Model

A full decision tree model is generated using the combined 1283 hyperspectral features, all 1000 of the training points, and a complexity parameter value of 0.003. The final model only uses 33 of the total hyperspectral features, using 27 of 32 wavelengths. 18 metrics use single wavelengths and 14 metrics are Normalized Difference Indices of wavelength pairs. The model also reports which variables or features are important based on the statistics of each split. The variable importance is listed in Table 9:

Table 9: Variable Importance in the Hyperspectral Data Classification Decision Tree

Variable	Variable Importance	Used in Decision Tree
muDeriv1249	3	Yes
muDeriv1064	2	Yes
muDeriv783	2	Yes
NDI578i643	2	Yes
NDI528i668	2	No
NDI553i668	2	No
NDI578i668	2	No

NDI603i643	2	No
NDI553i643	2	No
muDeriv1650	2	No
muDeriv879	2	No
vegNDI1515i2216	2	Yes
muR1314	2	No
NDI1515i2216	2	No
muDeriv2106	2	Yes
vegNDI1515i2161	2	No
NDI1515i2161	2	No
muR1249	2	No
vegNDI1545i2216	2	No
vegNDI1515i2106	2	No
muDeriv528	2	Yes
NDI1199i1249	1	No
NDI1585i1730	1	Yes
muDeriv994	1	No
NDI1650i1685	1	Yes
vegNDI1650i1685	1	No
vegNDI1585i1730	1	Yes
muDeriv2216	1	Yes
FracSandhillVeg2	1	No
NDI1615i1730	1	No
NDI2161i2281	1	No
vegNDI2161i2281	1	No

muDeriv503	1	No
muDeriv468	1	No
muR1545	1	No
stdDeriv2216	1	No
vegNDI2216i2281	1	No
NDI2216i2281	1	No
GreenPeakRot	1	No
NDI1249i1284	1	No
vegNDI1249i1284	1	Yes
stdDeriv2106	1	No
stdDeriv2161	1	No
muR1585	1	No
NDI783i879	1	No
muVegDeriv1650	1	No
vegNDI1615i1685	1	No
NDI503i603	1	No
vegNDI783i879	1	Yes
NDI783i1064	1	No
vegNDI1199i1314	1	No
vegNDI783i994	1	No
NDI1199i1314	1	No
vegNDI2106i2281	1	No
muVegDeriv668	1	Yes
NDI783i994	1	No
NDI2106i2281	1	No

The table lists 57 features that are significant in a relative sense, i.e. they appear at the top of the split decision list more often than other features. Many of these are not used in the decision tree because they were beat out by other features that could achieve a higher classification for a single split. Looking at the top variables in a spectral context: “muDeriv1249”, “muDeriv1064”, “muDeriv783”, “muDeriv1650”, “muDeriv879”, “muDeriv2106”, and “muDeriv528” capture the slope of the reflectance curves in several different wavelength regions; “NDI578i643”, “NDI528i668”, “NDI553i668”, “NDI578i668”, “NDI603i643”, “NDI553i643” are inversely related to high chlorophyll or green areas; and “muR1314” and “muR1249” take advantage of absolute reflectance at individual wavelengths. More specifically, “muDeriv1064” can detect Lake areas and “muDeriv2106” can separate Pine Plantation and Sandhill from Basin Marsh, Lake Fringe, and Ruderal solely based on the slope of the reflectance. Eighteen additional features are used in the decision tree but were not identified as important by the algorithm: “vegNDI743i994”, “stdR528”, “NDI743i1585”, “muDeriv2036”, “stdR668”, “muR743”, “NDI503i2161”, “NDI2281i2326”, “vegNDI1199i1284”, “muVegDeriv1730”, “FracSandhillGrass”, “NDI994i1104”, “stdVegDeriv643”, “muDeriv708”, “stdDeriv2326”, “muVegDeriv503”, “vegNDI1104i1284”, and “stdDeriv603”. These features achieved the best classification for their individual split but did not appear enough times throughout the tree to be listed as important.

A version of the decision tree generated with a complexity parameter of 0.01 is shown in Figure 53:

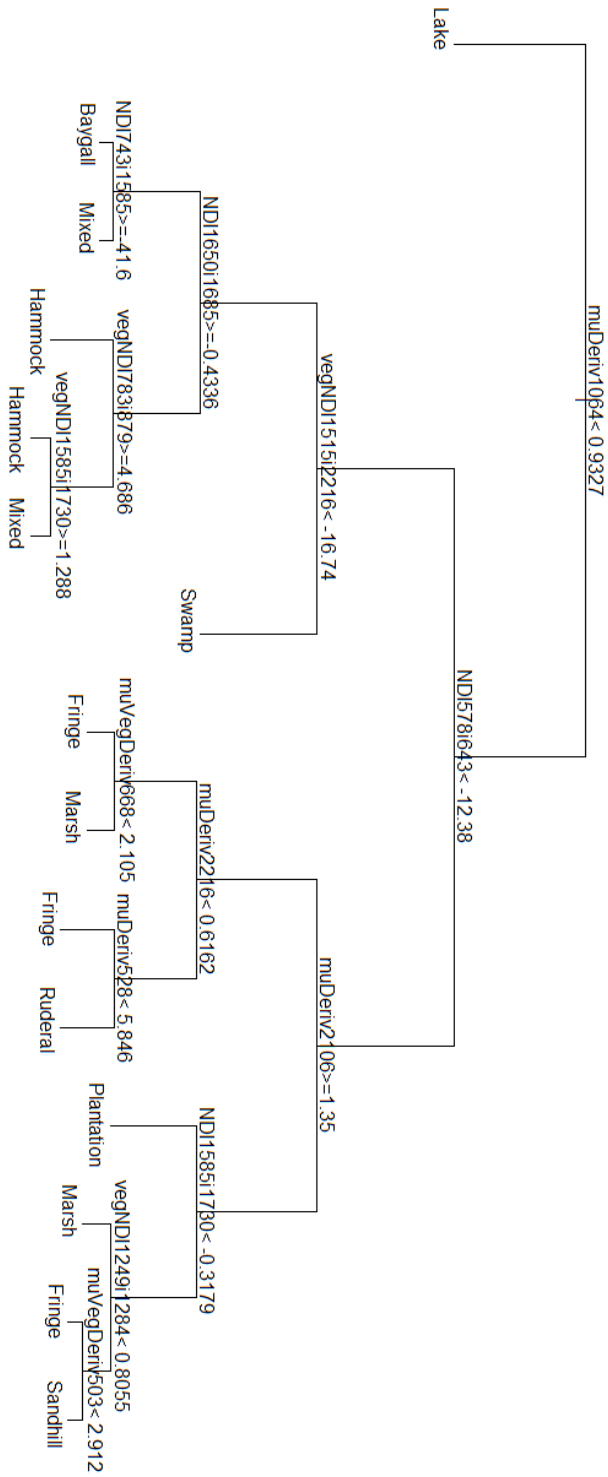

Figure 53: Hyperspectral decision tree generated with a complexity parameter of 0.01

The higher cp value allows a cleaner view of the top split points of the tree. The first split is on the feature "muDeriv1064" which separates Lakes from the rest of the vegetation communities. The split decision is shown in further detail in Figure 54 displaying a scatter plot of "muDeriv1064" for each FNAI vegetation community.

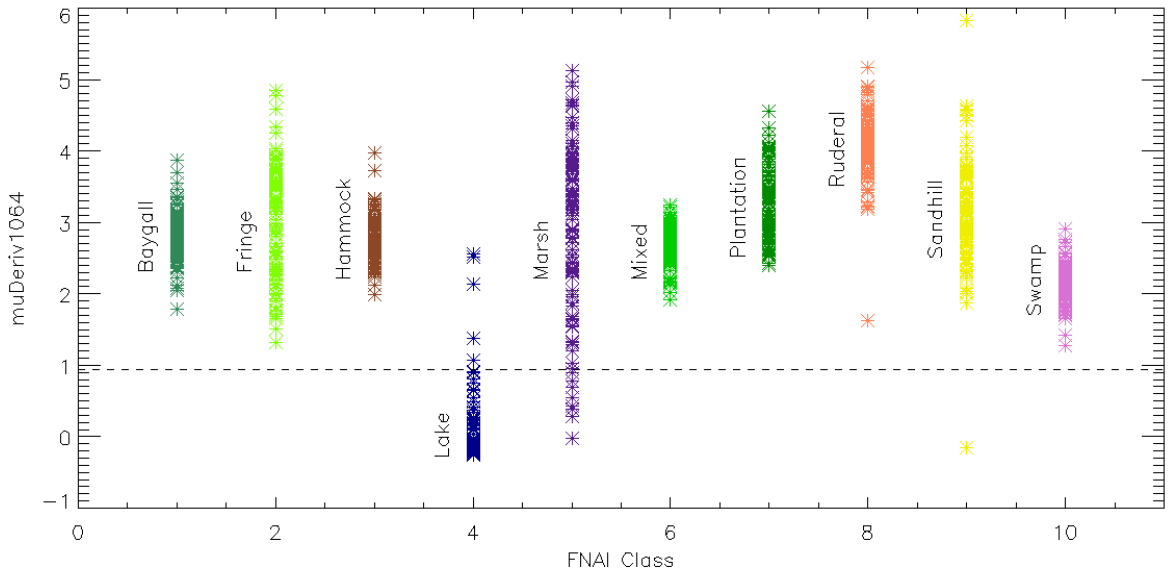

Figure 54: Mean of the first derivative of reflectance at 1064nm as the first split of the decision tree, separating Lake areas from the other vegetation communities

While this split isolates most of the Lake locations below the dashed line, a few Lake points remain above the line and a few Basin Marsh points fall below it.

The second split (Figure 55) is on “NDI578i643” separating the closed canopy (areas of mostly green pixels) Baygall, Basin Swamp, Mixed Forest, and Xeric Hammock areas from the open canopy or areas with significant ground signal: Basin Marsh, Lake Fringe, Pine Plantation, Ruderal, and Sandhill.

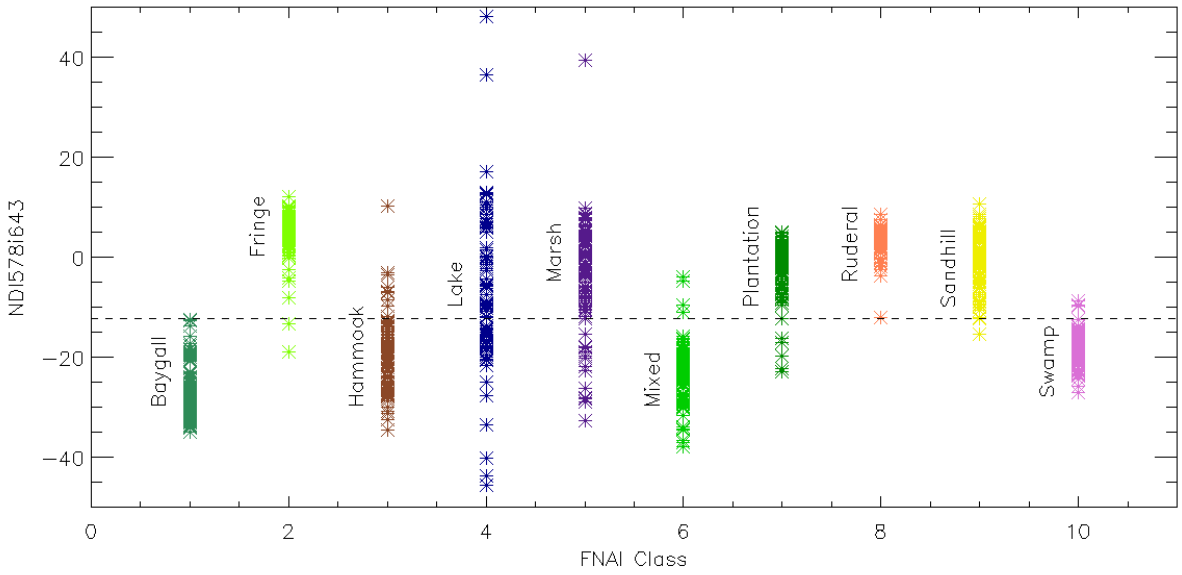

Figure 55: “NDI578i643” as the second split of the decision tree, separating the closed canopy (areas of mostly green pixels) Baygall, Basin Swamp, Mixed Forest, and Xeric Hammock areas from open the canopy or areas with significant ground signal: Basin Marsh, Lake Fringe, Pine Plantation, Ruderal, and Sandhill

These two hyperspectral features are able to separate the data for the entire site establishing the best starting route for classification accuracy.

The third split on the left side of the tree uses “vegNDI1515i2216” to separate Swamp areas from the rest of the closed canopy classes. The fourth split on the left side of tree uses “NDI1650i1685” to

separate Baygall from Xeric Hammock and most of the Mixed Forest. The left side of Figure 56 shows "vegNDI1515i2216" plotted against "NDI1650i1685".

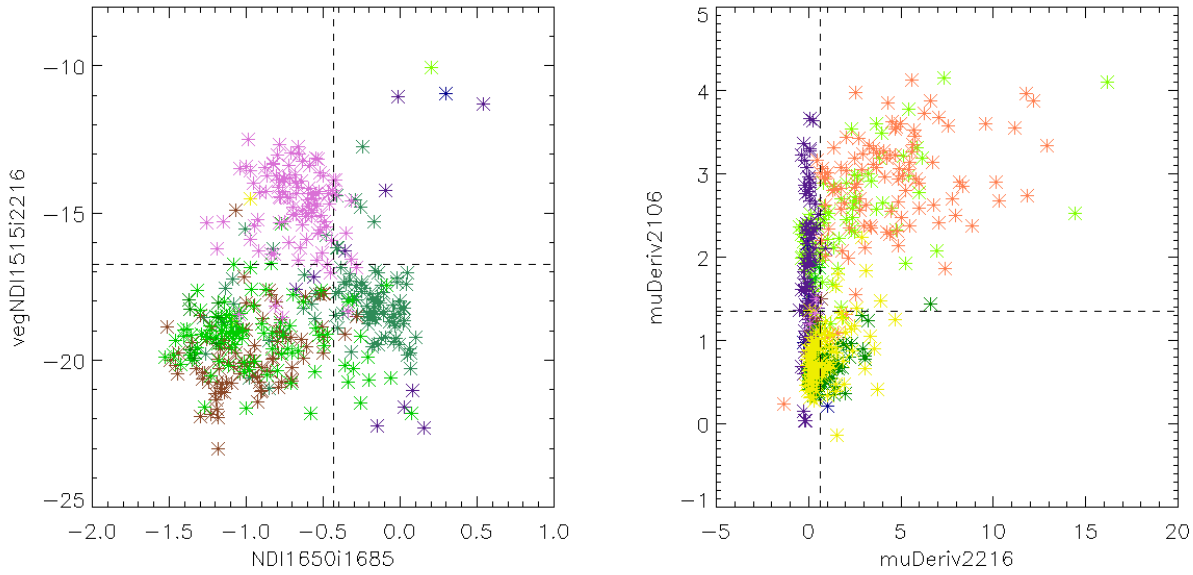

Figure 56: Scatter plots showing how a few of the decision tree splits separate the data. The left side is the normalized difference index for vegetation pixels using the reflectance at wavelengths of 1515 nm and 2216 nm versus the normalized difference index using the reflectance at wavelengths of 1650 nm and 1685 nm. The right side is the mean of the first derivative of reflectance at 2106 nm versus the mean of the first derivative of reflectance at 2216 nm

The pink Basin Swamp points are easily separated into the top half of the plot as they have higher values in the "vegNDI1515i2216" feature. The brown Xeric Hammock points are separated in the lower left quadrant as they have low "vegNDI1515i2216" and low "NDI1650i1685" values whereas the aqua Baygall points are in the lower right quadrant with low "vegNDI1515i2216" and high "NDI1650i1685" values. While the green Mixed Forest points are separated in the bottom of the plot from the Basin Swamp points and mostly in the lower left quadrant, they cover both the Xeric Hammock and Baygall

points. The final model requires more splits using other data features in order to separate these classes better.

The third split on the right side of the tree uses “muDeriv2106” to separate the bare-Earth Basin Marsh, Lake Fringe, and Ruderal areas from the open canopy Pine Plantation and Sandhill areas. One of the fourth splits on the right side of the tree uses “muDeriv2216” to further separate the Basin Marsh from the Lake Fringe and Ruderal areas. The right side of Figure 56 shows “muDeriv2106” plotted against “muDeriv2216”. The dark green Pine Plantation and yellow Sandhill points are in the bottom half of the plot with “muDeriv2106” values less than 1.35. The purple Basin Marsh points are separated in the upper left quadrant with high values of “muDeriv2106” and low values of “muDeriv2216”. The chartreuse Lake Fringe and orange Ruderal points are separated in the upper right quadrant with high values of “muDeriv2106” and high values of “muDeriv2216”. More splits are required to separate Pine Plantation from Sandhill and Lake Fringe from Ruderal.

The final hyperspectral based decision tree classification model using a complexity parameter of 0.003 is listed in full in Appendix B.1 and shown in Figure 57:

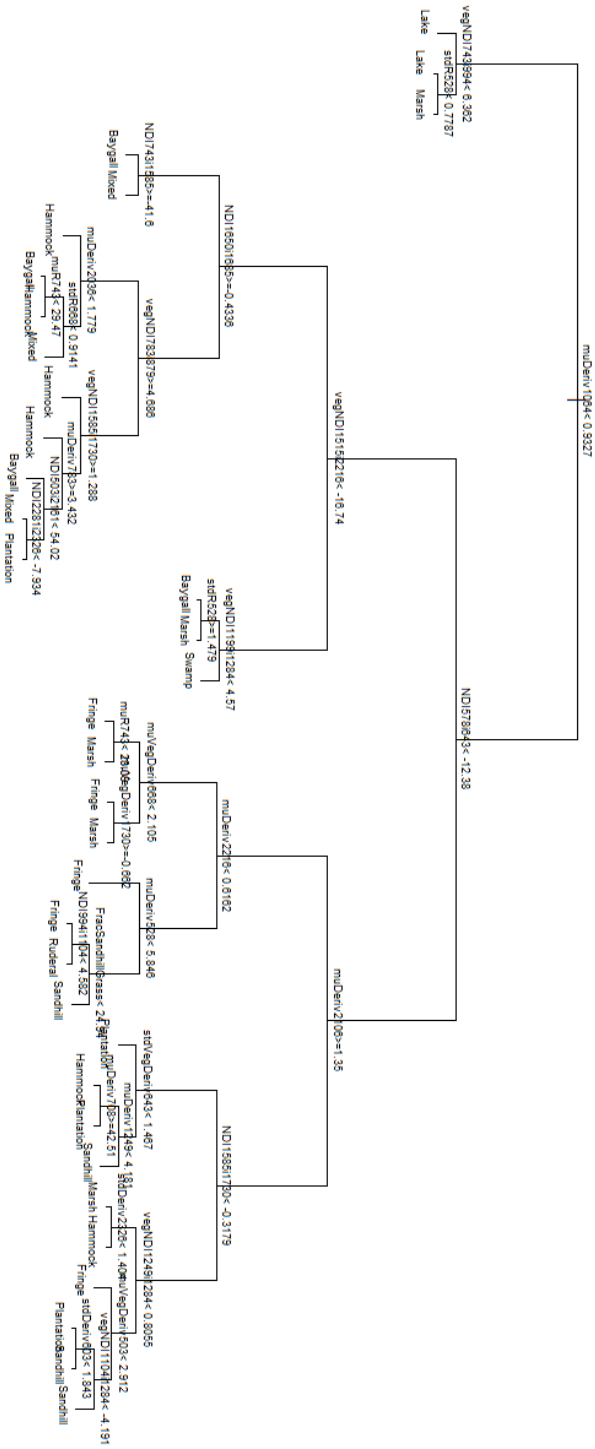

Figure 57: Final hyperspectral decision tree generated with a complexity parameter of 0.003

The Basin Swamp and Ruderal classes each have one terminal node and the Lake class has two terminal nodes suggesting that hyperspectral data from ground locations of these classes were most homogeneous. Several other classes such as Basin Marsh, Lake Fringe, and Xeric Hammock have 5 or 6 terminal nodes highlighting the heterogeneity of these classes. Based on terminal node pairs, the decision tree is trying to separate similarities between: Basin Marsh versus Lake; Lake Fringe versus Basin Marsh and Ruderal; Mixed Forest versus Baygall and Xeric Hammock; Basin Swamp versus Basin Marsh and Baygall; and Sandhill versus Lake Fringe, Pine Plantation, and Ruderal.

Results

The final decision tree is generated using all of the training points. Evaluating the training points through the decision tree, an overall classification accuracy of 89.8% is achieved for all vegetation communities. The classification accuracy for each FNAI vegetation community is listed in the confusion matrix in Table 10:

Table 10: Hyperspectral Decision Tree, Training Data Confusion Matrix

	FNAI Vegetation Community	Reference Sites											
		Baygall	Lake Fringe	Xeric Hammock	Lake	Basin Marsh	Mixed Forest	Pine Plantation	Ruderal	Sandhill	Basin Swamp	Total	User's Accuracy
Map	Baygall	89	0	1	0	2	2	0	0	0	2	96	93%
	Lake Fringe	0	95	2	1	2	0	0	7	2	1	110	86%
	Xeric Hammock	2	0	89	0	1	7	1	0	0	3	103	86%
	Lake	0	0	0	95	2	0	0	0	1	0	98	97%
	Basin Marsh	0	1	0	3	88	2	0	1	0	1	96	92%
	Mixed Forest	7	1	5	0	2	85	3	0	0	3	106	80%
	Pine Plantation	0	2	1	0	1	2	91	2	6	0	105	87%
	Ruderal	0	1	1	0	1	0	0	87	1	0	91	96%
	Sandhill	0	0	0	1	0	0	5	3	89	0	98	91%
	Basin Swamp	2	0	1	0	1	2	0	0	1	90	97	93%
	Total	100	100	100	100	100	100	100	100	100	100	898/1000	
	Producer's Accuracy	89%	95%	89%	95%	88%	85%	91%	87%	89%	90%		90%

Lake and Lake Fringe achieved the highest classification accuracies at 95%. Sandhill achieved a producer's accuracy of 89% (89% of Sandhill training locations are correctly classified) and a user's accuracy of 91% (91% of the locations labeled as Sandhill actually are Sandhill). Sandhill is primarily misclassified as Pine Plantation and occasionally misclassified as Lake Fringe. Pine Plantation is primarily misclassified as Sandhill. The confusion likely stems from the fact that the Sandhill ecosystem was disturbed to plant the Pine Plantations and both ecosystems contain pines, although the pines are

more regularly spaced in the plantations and might contain Slash Pine versus Longleaf Pine. The confusion of the rest of the output classes is described in Table 11:

Table 11: Confusion of Output Classes by the Decision Tree

Output Class	Confused As
Baygall	Mixed Forest and a little as Basin Swamp and Xeric Hammock
Lake Fringe	A little as Pine Plantation
Xeric Hammock	Mixed Forest and a little as Lake Fringe
Lake	Basin Marsh
Basin Marsh	A little as Baygall, Lake, Lake Fringe and Mixed Forest
Mixed Forest	Xeric Hammock and a little as Basin Marsh, Basin Swamp, Baygall, and Pine Plantation
Pine Plantation	Sandhill and a little as Mixed Forest
Ruderal	Lake Fringe and a little as Pine Plantation an Sandhill
Sandhill	Pine Plantation and a little as Lake Fringe
Basin Swamp	A little as Baygall, Mixed Forest, and Xeric Hammock

Hyperspectral data features are also calculated for a 20 m x 20 m plot covering the independent ground validation locations collected by NEON and are run through the decision tree. An overall accuracy

of 66% is achieved for all vegetation communities with Sandhill achieving a producer's accuracy of 80% and a user's accuracy of 100%. For this hyperspectral model, Sandhill achieved the highest classification accuracy of all of the vegetation communities. The classification accuracy for each ground validation location is listed in the confusion matrix in Table 12:

Table 12: Hyperspectral Decision Tree, NEON Ground Validation Data Confusion Matrix

	FNAI	Reference Sites											
	Vegetation Community	Baygall	Lake Fringe	Xeric Hammock	Lake	Basin Marsh	Mixed Forest	Pine Plantation	Ruderal	Sandhill	Basin Swamp	Total	User's Accuracy
Map	Baygall	-		1			1					2	0%
	Lake Fringe		3			2			1			6	50%
	Xeric Hammock			8			4			2	2	16	50%
	Lake				-							-	NA
	Basin Marsh			1		5						6	83%
	Mixed Forest			1			7				1	9	78%
	Pine Plantation			3				-		4		7	0%
	Ruderal		1						1			2	50%
	Sandhill									24		24	100%
	Basin Swamp			1							0	1	0%
	Total	-	4	15	-	7	12	-	2	30	3	48/73	
	Producer's Accuracy	NA	75%	53%	NA	71%	58%	NA	50%	80%	0%		66%

Finally, hyperspectral data features are calculated for a 20 m x 20 m plot on a 5 m posting covering the entire site. Data from each 5 m posting is run through the decision tree to generate a classification map. The output class map is shown in Figure 58 and has been cropped by the OSBS site boundary.

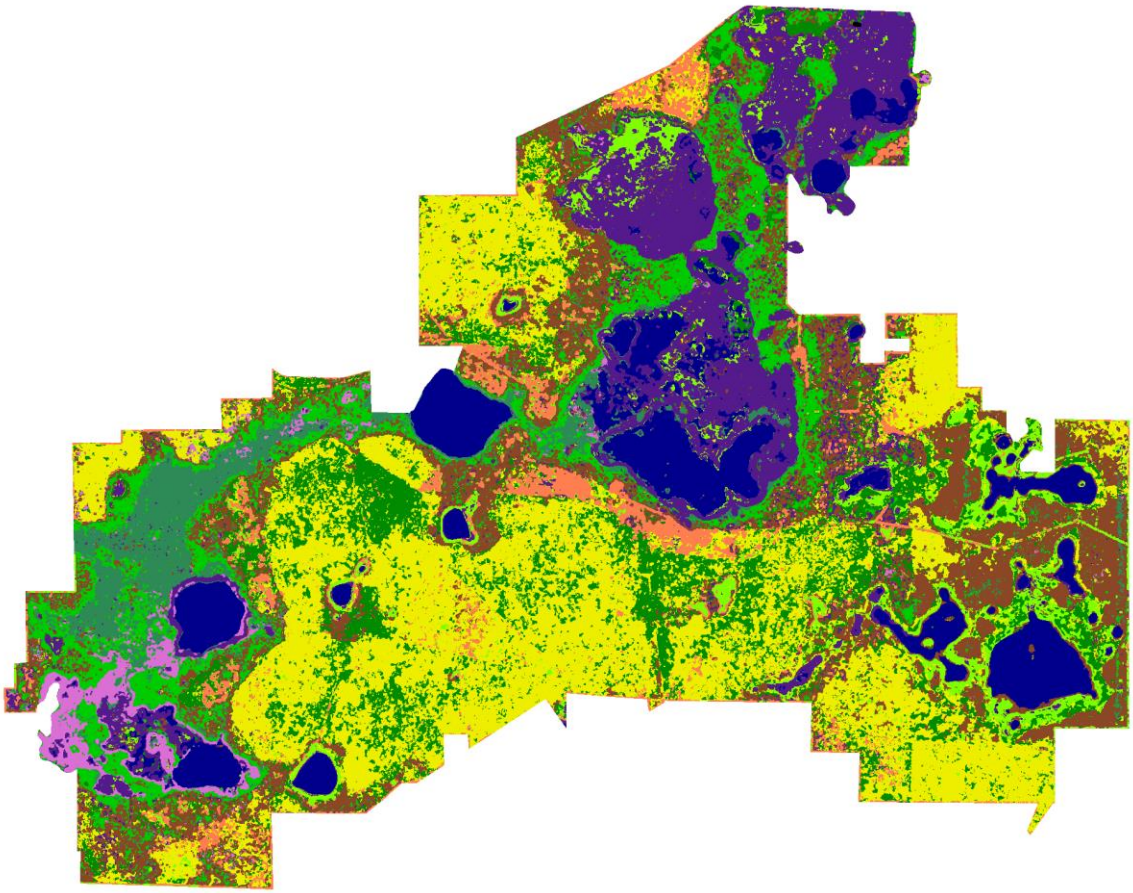

Figure 58: FNAI vegetation community classification map generated using hyperspectral data

Figure 59: FNAI classmap legend

At a first glance, the classification map closely resembles the OSBS GIS vegetation community map shown in Chapter III. As the classification accuracy states, the dark blue Lakes visually appear correct. Small Lake pixels in the purple Basin Marsh region in the North center of the image are probably areas of standing water. The classification confusion discussed earlier can also been seen in this map: dark green Pine Plantation pixels are littered throughout the chartreuse Lake Fringe area in the Southeast; brown Xerick Hammock is misclassified as green Mixed Forest in the Southwest and purple Basin Marsh or aqua Baygall in the East; senescent vegetation (showing up as bright areas in the color imagery) in purple Basin Marsh region in the North center are mislabeled as chartreuse Lake Fringe and Basin Marsh areas with trees are mislabeled as aqua Baygall; some of the green Mixed Forest in the North right of the map (between the Basin Marshes) is mislabeled as brown Xeric Hammock; and the yellow Sandhill regions are speckled with small dark green Pine Plantation and orange Ruderal pixels.

CHAPTER VI

VEGETATION CLASSIFICATION USING STRUCTURAL VARIABILITY AS MEASURED BY DISCRETE RETURN LIDAR

Discrete return lidar data measure the vertical structure of the vegetation and ground topography as collected by a pulsed laser instrument operating at a wavelength of 1064 nm. In this chapter, the lidar data provided by the Airborne Observation Platform (Kampe et al., 2010; Kampe et al., 2011; NEON, 2015h) of the National Ecological Observatory Network (NEON, 2015a; NRC, 2003; Schimel et al., 2011) are prepared with some data processing before a series of lidar features are calculated. The features for the ground training locations are used to generate a decision tree classification algorithm with the goal of producing a classification map for the Ordway-Swisher Biological Station (OSBS, 2015a; OSBS, 2015c; OSBS, 2015d) site using the Florida Natural Areas Inventory (FNAI, 2010) vegetation communities as the output classes. Classification accuracy is assessed using k-folds cross validation and the independent ground truth data collected by NEON.

Objectives

Objectives

- Investigate how discrete return lidar data can be used to observe vegetation species diversity through structural differences.
- Generate a vegetation community map using structural differences in the data. This will serve as a proxy for vegetation species diversity.

Science questions

- How do vegetation communities differ structurally?
- How is structural diversity related to species diversity?
- What structural differences can we observe in the remote sensing data?
- How is the measured variability in the remote sensing data related to biodiversity or vegetation community type?

Lidar Data Preparation

The NEON lidar products require some data preparation before feature layers are calculated and the classification algorithm is run. The digital surface model (DSM), the bare-Earth digital terrain model (DTM), and the canopy height model (CHM) provided by NEON contain ignore values of -9999 in areas where no discrete return lidar points exist. These null or void areas exist within the interior of a given flight line such as a lake, or in the overlap between adjacent mosaicked flight lines caused by a gap in data collection. The void areas in the DSM and DTM are filled in using an iterative median filter. Void areas in the CHM are set to a value of 0. The elevation models are also buffered in extent to make sure that they have valid values for every discrete return lidar point plus an extra area, referred to as the backfill area. A coarse resolution (~10 m gsd) digital elevation model is downloaded from the USGS, resampled to a UTM projection grid in ENVI. The backfill areas of the DSM and DEM are replaced with the USGS DEM values using nearest neighbor resampling. Once again the CHM backfill areas are set to 0. The slope of the DTM is calculated in ENVI using a fine resolution 3 x 3 pixel kernel and a coarse resolution 21 x 21 pixel kernel. Also surface roughness or rugosity is calculated for a moving 3 x 3 pixel kernel using an algorithm that calculates the total surface area of a set of triangles within the kernel.

The discrete return lidar points with an absolute height less than -75.0 m and greater than +75.0 m are removed to filter out noisy points. In order to analyze vegetation height and structure across multiple sites, it is desirable to remove topography from the data. The bare-Earth digital terrain model (DTM) is subtracted from each lidar point using a nearest-neighbor lookup to convert the discrete returns from absolute height to height above ground (relative height in meters) as shown in Figure 60:

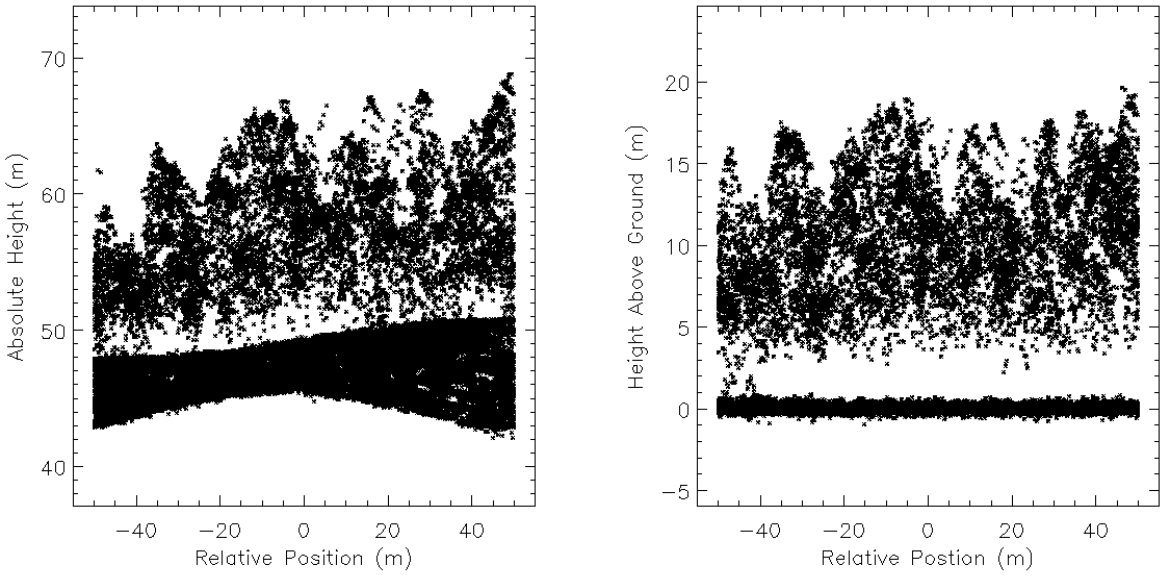

Figure 60: Subtraction of digital terrain model to convert absolute height into height above ground

The data are subset into 20 m x 20 m boxes centered on each training plot location. Figure 61 shows an example of the discrete lidar data for a Basin Marsh training plot location.

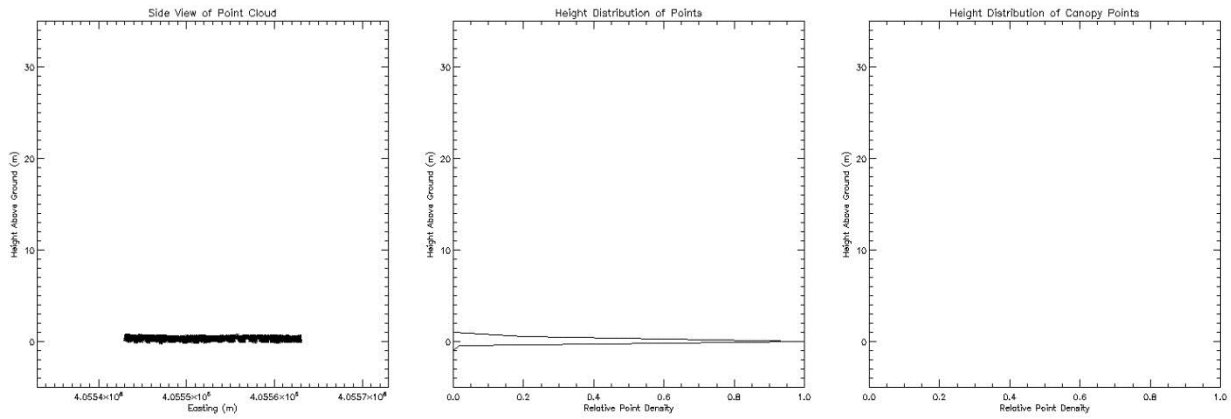

Figure 61: Example of the discrete lidar data for a Basin Marsh training plot location. The left is a side view of the lidar point cloud subset, the center is a normalized histogram of point heights, and the right is a normalized histogram of canopy height points

The left is a side view of the lidar point cloud subset, the center is a normalized histogram of point heights, and the right is a normalized histogram of canopy height points. In the case of Basin Marsh, there is little or no canopy so most or all of the lidar points fall on the ground. Figure 62 shows an example of the discrete lidar data for a Sandhill training plot location.

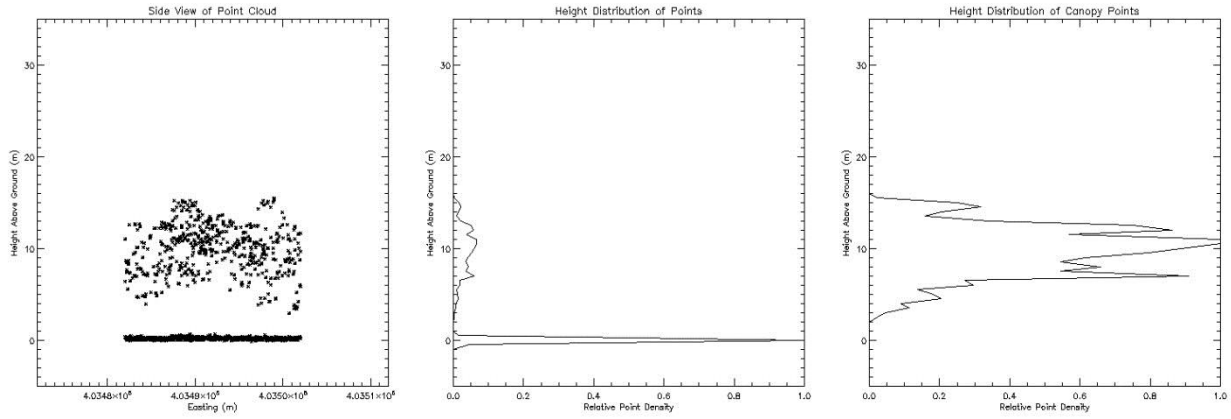

Figure 62: Example of the discrete lidar data for a Sandhill training plot location. The left is a side view of the lidar point cloud subset, the center is a normalized histogram of point heights, and the right is a normalized histogram of canopy height points

The left and center figures show that Sandhill has a fairly open canopy so a significant number of ground points are still recorded. Also, Sandhill has an open understory so there are very few points in height between 1.0 – 2.0 m. The canopy vertical height distribution seems to be fairly well balanced around the mean canopy height (see the right side of the figure). Figure 63 shows an example of the discrete lidar data for a Xeric Hammock training plot location.

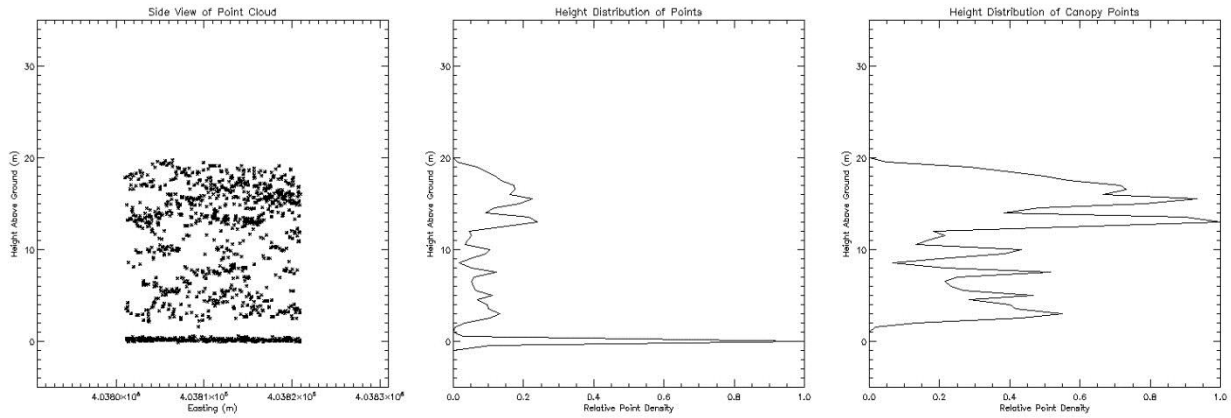

Figure 63: Example of the discrete lidar data for a Sandhill training plot location. The left is a side view of the lidar point cloud subset, the center is a normalized histogram of point heights, and the right is a normalized histogram of canopy height points

The left and center figures show that the canopy is starting to close in Xeric Hammock and less ground points are recorded. Also, Xeric Hammock has an open understory so there are very few points in height between 1.0 – 2.0 m. The canopy vertical height distribution seems to be top heavy where more laser points are returned from higher up in the canopy (see the right side of the figure). Finally, Figure 64 shows an example of the discrete lidar data for a Basin Swamp training plot location.

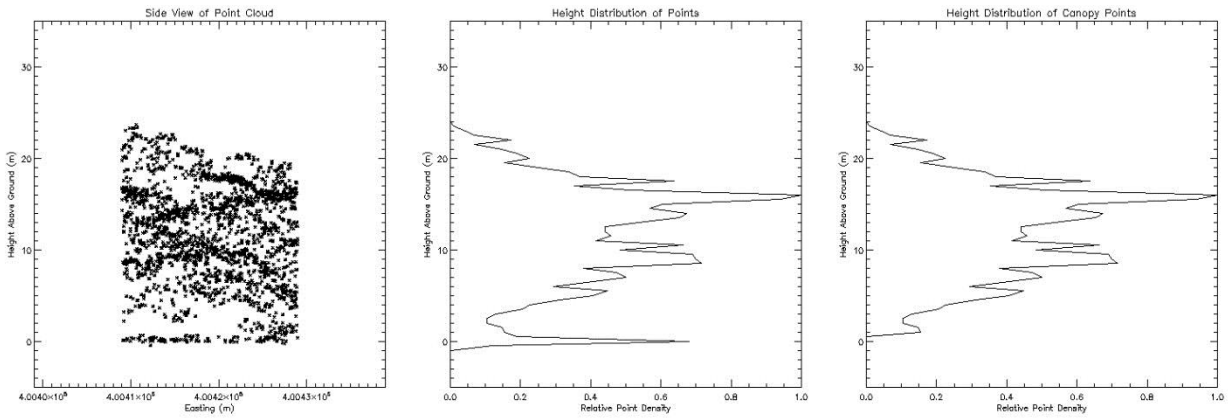

Figure 64: Example of the discrete lidar data for a Basin Marsh training plot location. The left is a side view of the lidar point cloud subset, the center is a normalized histogram of point heights, and the right is a normalized histogram of canopy height points

Unlike, Sandhill and Xeric Hammock, Basin Swamp has some understory so there are lots of points in height between 1.0 – 2.0 m. The left and center figures show a very thick and closed canopy where there are fewer ground points. The fraction of ground points is no longer the histogram bin with the highest count.

Calculation of Lidar Features

Before the decision tree classification model can be generated, a series of input features must be calculated, in this case lidar features. Most of the features are calculated on the discrete return lidar point clouds, adjusted as height above ground. A 20 m x 20 m data subset centered on a ground plot location is extracted for the feature calculation. The point clouds are categorized into several groups: all points, canopy points, and ground points. The data is classified into ground or canopy categories according to:

- Ground point if height above ground < 1.37 m
- Canopy point if height above ground ≥ 1.37 m

Also, other categories are calculated:

- Below ground aoi0 if height above ground < 0.0 m
- Below ground aoi1 if height above ground < -0.25 m
- Below ground aoi2 if height above ground < -0.5 m
- Shrub point if height above ground > 0.3 m and < 1.37 m
- First return if return number = 1
- One_one if return number = 1 and total number of returns in the laser pulse = 1
- Two_two if return number = 2 and total number of returns in the laser pulse = 2
- Three_three if return number = 3 and total number of returns in the laser pulse = 3
- Four_four if return number = 4 and total number of returns in the laser pulse = 4

Next height histograms are calculated for both ground and canopy groups using a vertical bin size of 0.5 m. Percentile heights are also calculated. A series of features are derived from the percentiles such as specific percentile height range (ex. 90th minus 50th percentile), percentile height range fraction (ex. percentile height range divided by the 100th minus 0 percentile height value), and normalized difference height index (ex. [(P90 – P50) – (P50 – P10)] / [(P90 – P50) + (P50 – P10)]). Point distribution calculations are also performed:

- Median absolute deviation or MAD = Median(abs[Height – median(height)])
- Mean absolute deviation or AAD = Mean(abs[Height – mean(height)])
- Vertical distribution ratio or VDR = (P100 – P50) / P100
- Canopy Relief Ratio or CRR = (P50-P0) / (P100 – P0)
- Canopy density = fraction of first returns above 3 m / total number of first returns

Summary statistics are calculated for all points, ground points, canopy points, and first returns. Additionally, the fraction of 1st return ground points and the 95th percentile height of the first returns are calculated. Finally, several features are derived from the 2D gridded canopy height model and bare-Earth digital elevation/terrain model including summary statistics, fraction of ground pixels in the canopy height model, rugosity (integrated triangular surface area as a measure of surface roughness), coarse resolution slope of the terrain model, and total of the fine resolution slope (another measure of roughness). Fractions are multiplied by 100 to help balance out size of the numbers, though this step isn't required for the "rpart" algorithm. Intensity values were not used as part of the feature set in order to focus the features on structural differences.

In total, a suite of 89 lidar features or metrics are calculated for each ground location. A summary of the lidar feature categories is listed in Table 13 and a full list of the lidar features can be found in Appendix A.2. These metrics serve as the input feature layer to the classification model.

Table 13: Summary of Lidar Feature Categories

Category	Examples
Summary Statistics	Mean, standard deviation, min, and max
Point Classification	Ground point, below ground, shrub, canopy, high canopy, first return, 1st of 1, 2nd of 2, etc.
Percentile Height	0, 5th, 10th, 15th, 20th, 25th, 30th, 35th, 40th, 45th, 50th, 55th, 60th, 65th, 70th, 75th, 80th, 85th, 90th, 95th, and 100th percentiles
Percentile Height Range (including interquartile range)	90th minus 50th percentile, 75th minus 50th percentile, 100th minus 0 percentile height, etc.
Percentile Height Range Fraction	Percentile height range divided by the 100th minus 0 percentile height value
Normalized Difference Height Index	[(P90 – P50) – (P50 – P10)] / [(P90 – P50) + (P50 – P10)]
Point Distribution	Median absolute deviation, mean absolute deviation, canopy vertical distribution ratio, and canopy relief ratio
Elevation Model	Bare-Earth slope, canopy height model rugosity

The features are meant to be proxies of the horizontal 2D and vertical 3D structure of the different vegetation communities as measured by the discrete return lidar instrument. The lidar point clouds represent not only the 3D structure of the canopies and ground, based on their reflectance of light at 1064 nm, but also exhibit optical phenomena such as light penetration and multiple scattering. Figure 65 shows a false color composite of three features where the red, green, and blue display channels are “CanopyDensity”, “Frac22”, and “CHMrugosity” respectively.

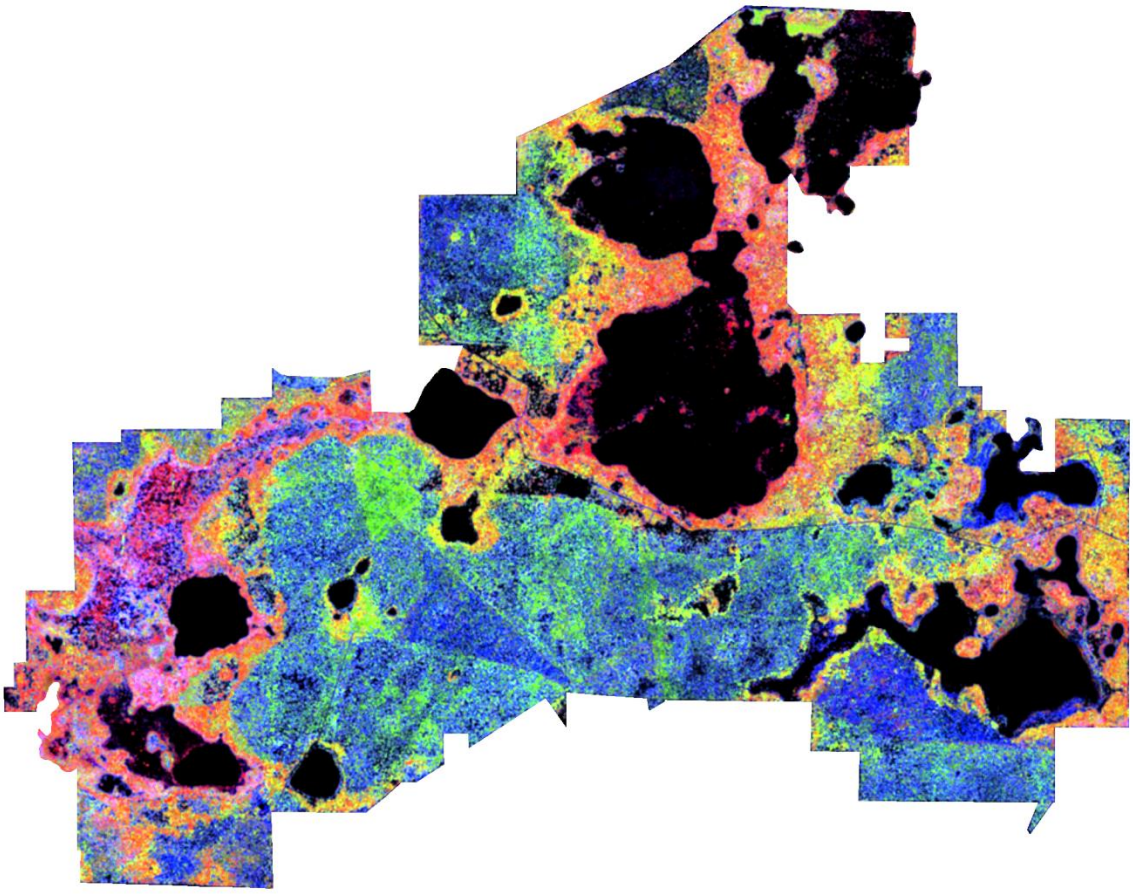

Figure 65: False color composite of three lidar features where the red, green, and blue display channels are “CanopyDensity”, “Frac22”, and “CHMrugosity” respectively

In the figure, different FNAI vegetation communities display as different colors, illustrating the potential for a decision tree to classify the vegetation communities using the feature data.

The canopy density feature did not end up being used by the lidar decision tree but is used by the integrated hyperspectral and lidar decision tree in Chapter VII. Baygall, Mixed Forest, and Xeric

Hammock have higher canopy density values because they have thicker canopies and a larger fraction of the lidar first returns are from heights 3 m or higher above the ground. Basin Swamp and Mixed Forest areas typically show up as pink and red hues in the false color image. Xeric Hammock and Pine Plantation have higher values in the second of two feature layer suggesting the light penetration within the canopy for these vegetation communities is higher and causing the Pine Plantation areas to show up as green and the Xeric Hammock areas to show up as yellow or orange (high in both the red and green display channels) in the false color image. The "Frac22" feature is used in both the lidar and integrated decision trees. The surface roughness of the canopy height model "CHMrugosity" is the third feature in the false color image where Baygall, Basin Swamp, and Sandhill high higher values or their canopy tops have a higher roughness. Sandhill areas show up as blue hues and Baygall areas show up as blue or magenta hues in the false color image. Rugosity shows up on the lidar variable importance list given later in this chapter but is not used by any of the final decision trees. The Basin Marsh, Lake, and Ruderal areas of little or no canopy have low values in all three features used in the false color image and show up as black.

Decision Tree Training

A classification decision tree is generated using the "rpart" package (Breiman et al., 1984; Therneau et al., 2015) in R (R Core Team, 2015). The variable "minbucket" is set to a value of 1 allowing terminal nodes to contain a single training location. Package "rpart" also has a complexity parameter variable "cp" that determines if the tree should generate more splits along a branch. A split must decrease the lack of fit by a factor of cp or else it is not attempted. The decision tree algorithm will iterate through different threshold values of each input feature, and determine the best data partitioning according to the fitting function of classified outputs (i.e. it picks the split that most correctly classifies the outputs). As

described in the previous sections, the lidar data are subset for each of the training ground locations, features are calculated, and those features are brought into package “rpart”, shown in Figure 66:

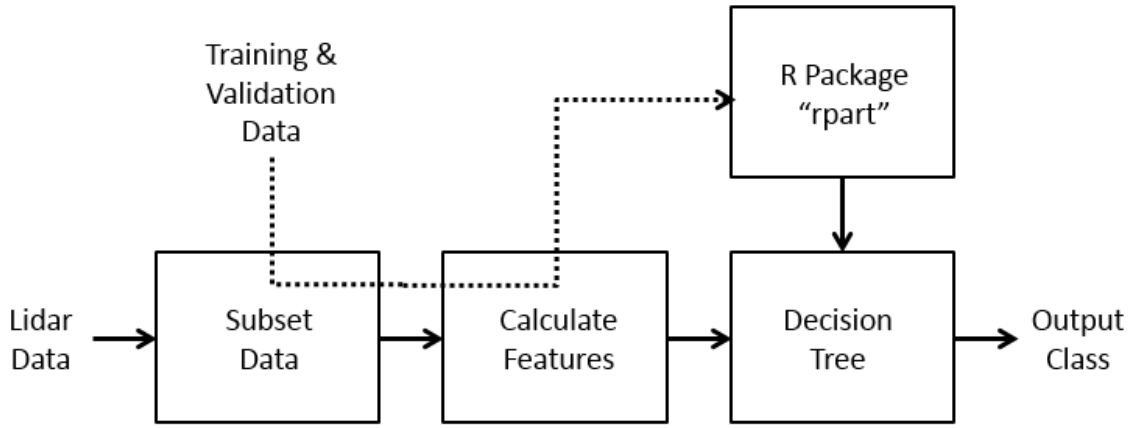

Figure 66: Lidar decision tree classification algorithm flowchart

K-folds cross validation is employed to determine the optimal value of the complexity parameter. A random number generator is used to separate the training data into 10 groups. For each fold, data from 9 of the groups is used to train the classification model and the data from the 10th group is used as independent validation. The process is performed a total of 10 times, each time generating a new decision tree and using a different group as the validation set. The complexity parameter value is varied to get an idea of the classification accuracy versus complexity parameter. If a very small complexity value is used, the classification model will over-fit the training data. The model will achieve a 100% classification accuracy for the training data but will have lower accuracy values for the validation data. By iterating the complexity parameter and performing the k-folds cross validation 10 times, the complexity parameter

value that gives the overall best classification accuracy can be chosen. The results from the k-folds cross validation are shown in Figure 67:

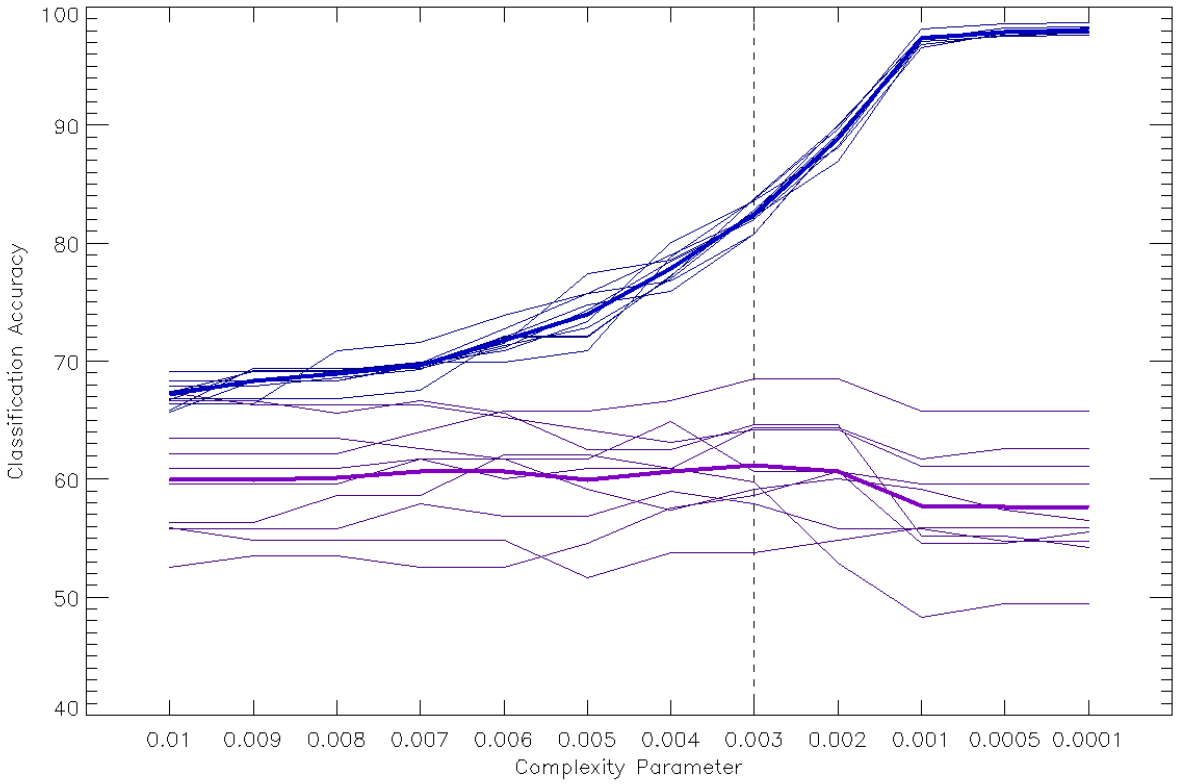

Figure 67: K-folds cross validation to determine the optimal value of the complexity parameter for the lidar decision tree

As seen in the figure, as the value of the complexity parameter decreases, the classification accuracy of the training data increases toward 100%. The classification accuracy of the validation data increases to a point, but then starts to decrease as the training data is over fit. Based on the k-fold cross validation results for the lidar data, a complexity parameter value of 0.003 produced a model with the best validation results. A classification accuracy of 82.4% ± 1.08 is achieved for the training data and an accuracy of 61.1% ± 4.25 is achieved for the validation data.

Classification Model

A full decision tree model is generated using the 89 lidar features, all 1000 of the training points, and a complexity parameter value of 0.003. The final model only uses 37 of the lidar features. The model also reports which variables or features are important based on the statistics of each split. The variable importance is listed in Table 14:

Table 14: Variable Importance in the Lidar Data Classification Decision Tree

Variable	Variable Importance	Used in Decision Tree
FracGround	6	Yes
CanopyDensity	5	No
mu1st	4	Yes
muAll	4	No
muCHM	4	Yes
grP75mP25	4	No
FracGroundCHM	3	Yes
grP90mP10	3	Yes
stdGround	3	Yes
grP90	3	No
grAAD	3	No
grP90mP50	3	Yes
Frac1stGround	3	Yes

caP75	2	Yes
caP90	2	Yes
caP25	2	Yes
caP50	2	Yes
caP10	2	Yes
DEMslopeK21	2	Yes
FracLT0	2	No
grP75mP50	2	No
caP75mP25	1	No
maxCHM	1	No
stdAll	1	No
std1st	1	Yes
Frac11	1	No
Frac22	1	Yes
stdCanopy	1	Yes
grMAD	1	Yes
caP90mP10	1	No
Frac1st	1	Yes
p951st	1	Yes
AAD	1	Yes
grP50mP10	1	Yes
caP100mP75	1	Yes
grP100mP75	1	Yes
caP0	1	No
grP50mP25	1	Yes

totDEMslopeK3	1	Yes
FracShrub	1	Yes
caP100mP50	1	Yes
grP100mP75Frac	1	Yes
caP100	1	No
muCanopy	1	No
grP75	1	Yes
caP100mP0	1	Yes
caP50mP25	1	Yes
CHMrugosity	1	No
grP100	1	No
caP50mP0	1	Yes
muGround	1	No
caP50mP10	1	No

The table lists 52 features that are significant in a relative sense, i.e. they appear at the top of the split decision list more often than other features. Many of these are not used in the decision tree because they were beat out by other features that could achieve a higher classification for a single split. Looking at the top six important variables in a structural context: “FracGround” and “CanopyDensity” represent the portion of the lidar returns that come from the 3D canopy or ground, “mu1st” and “muCHM” are the mean height of the top of the canopy, “muAll” is the overall mean height of the lidar returns, and “grP75mP25” is the height difference of the 25th to 75th percentile ground points representing surface roughness or existence of shrubs/understory. Four additional features are used in the decision tree but were not identified as important by the algorithm: “grP75mP25Frac”, “stdCHM”, “caP100mP75Frac”,

“caP50mP25Frac”. These features achieved the best classification for their individual split but did not appear enough times throughout the tree to be listed as important.

A version of the decision tree generated with a complexity parameter of 0.01 is shown in Figure 68:

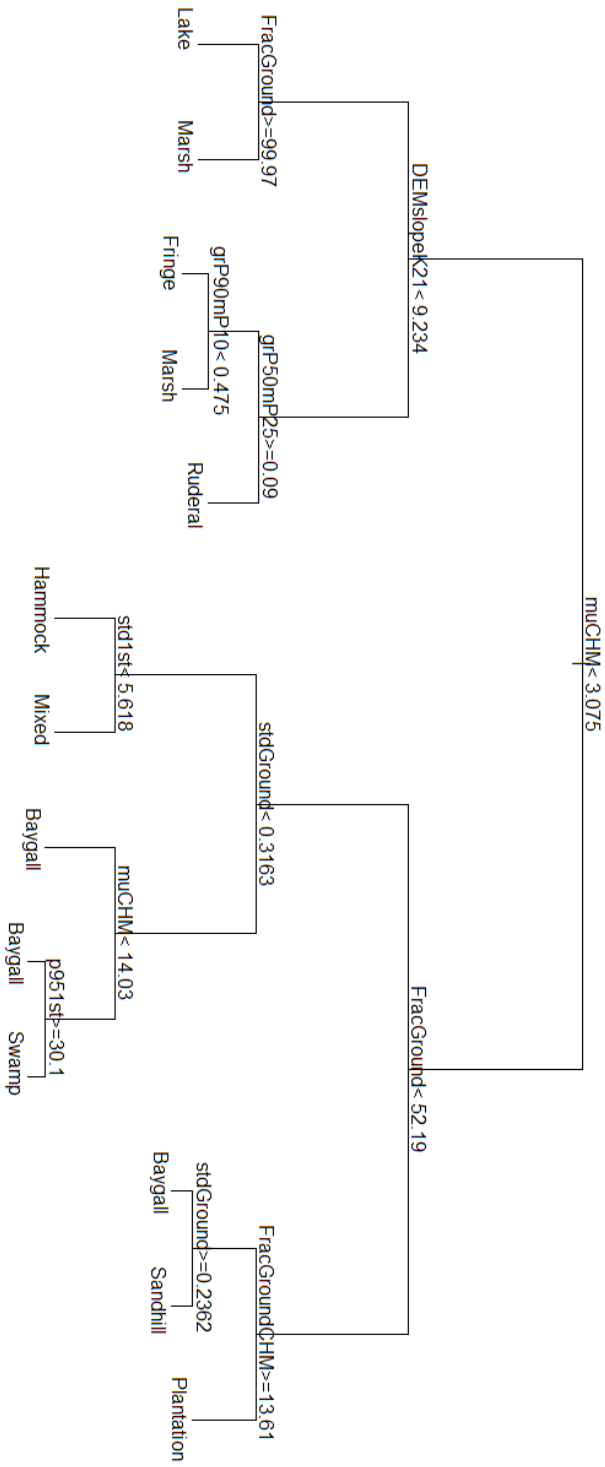

Figure 68: Lidar decision tree generated with a complexity parameter of 0.01

The higher cp value allows a cleaner view of the top split points of the tree. The first split is on the feature "muCHM" which uses the mean height of the gridded canopy height model to separate areas with and without canopies (such as the open areas of Basin Marsh, Lake, Lake Fringe, and Ruderal). The split decision is shown in further detail in Figure 69 showing a scatter plot of "muCHM" for each FNAI vegetation community.

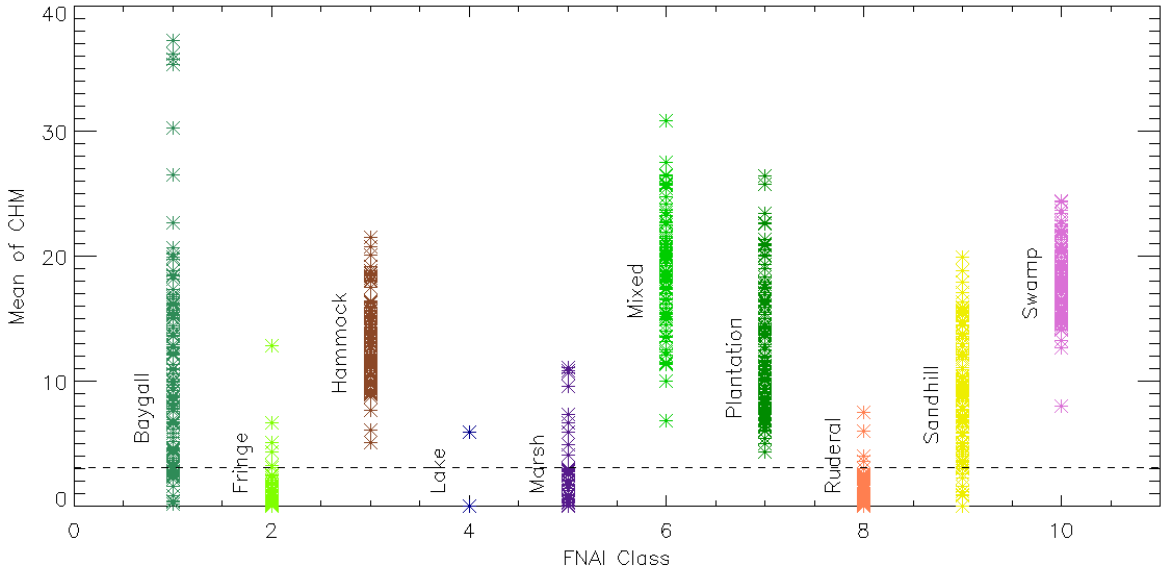

Figure 69: Mean of the Canopy Height Model as the first split of the decision tree, separating closed from open canopy areas

Most of the Basin Marsh, Lake, Lake Fringe, and Ruderal points are below the dashed line, however, a few ground locations with some trees are misclassified. Also, open canopy areas in Baygall and Sandhill are grouped with the other open canopy classes.

The second split on the left side of the tree uses the feature “DEMsplopek21” to separate flat Lakes and Basin Marsh from sloped ground areas of Lake Fringe, Basin Marsh, and Ruderal. The left side of Figure 70 shows the log of “DEMsplopek21” plotted against “grP50mP25”.

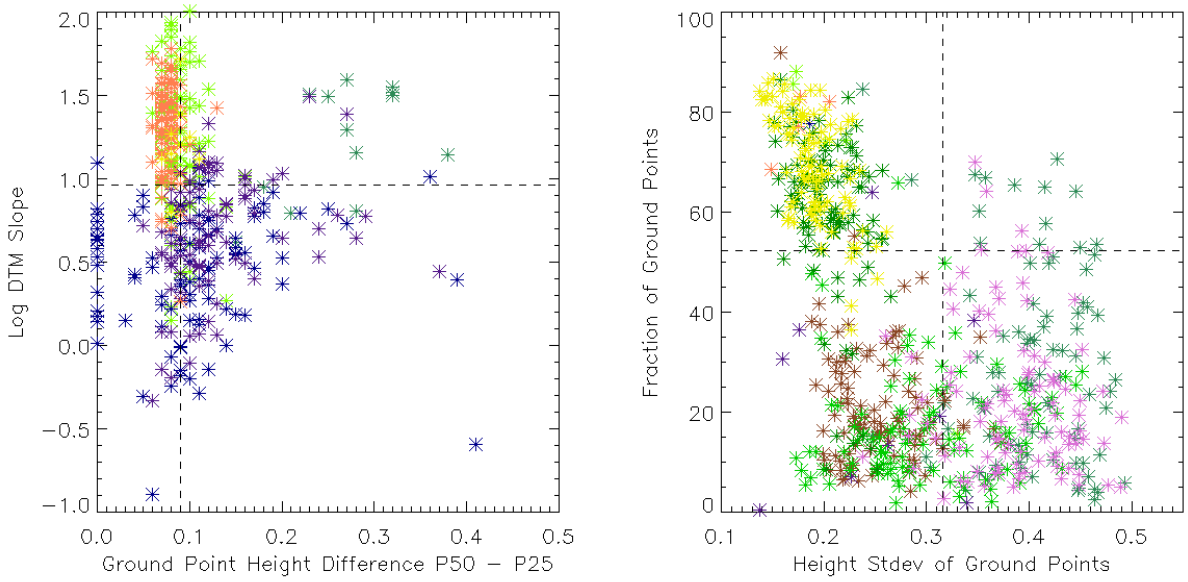

Figure 70: Scatter plots showing how a few of the decision tree splits separate the data. The left side is the log of the coarse resolution bare-Earth slope versus the height difference of the 50th minus 25th percentile height of ground points. The right side is fraction of ground points versus the standard deviation of height of the ground points

As previously described, the dark blue Lake and purple Basin Marsh points tend to fall in the bottom half of the plot with chartreuse Lake Fringe and orange Ruderal points separated in the top half. The “grP50mP25” decision attempts to pull the Lake Fringe points to the right of the Ruderal points but there is still some overlap between these classes. The final model requires more splits using other data features in order to separate these classes better.

The second split on the right side of the tree uses “FracGround” to separate closed canopy Basin Swamp, Baygall, Xeric Hammock, and Mixed Forest areas from partially open canopy Baygall, Pine Plantation, and Sandhill areas. The right side of Figure 70 shows “FracGround” plotted against “stdGround”. Brown Xeric Hammock and green Mixed Forest are separated in the lower left quadrant of the plot with low fraction of ground points and a low ground height standard deviation, pink Basin Swamp and aqua Baygall are separated in the lower right quadrant with a low fraction of ground points and a higher ground height standard deviation, and yellow Sandhill and dark green Pine Plantation are separated in the upper left quadrant with a high fraction of ground points and a low ground height standard deviation.

The final lidar based decision tree classification model using a complexity parameter of 0.003 is listed in full in Appendix B.2 and shown in Figure 71:

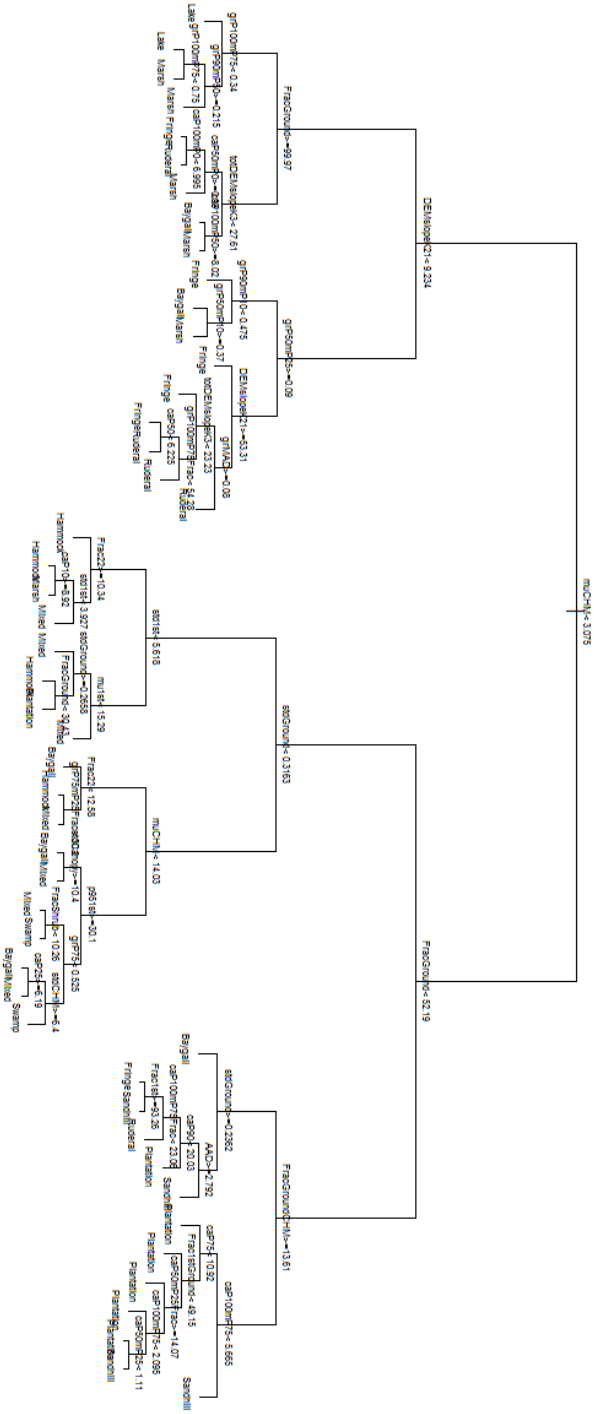

Figure 71: Final lidar decision tree generated with a complexity parameter of 0.003

The Basin Swamp and Lake classes each have two terminal nodes suggesting that lidar data from ground locations of these classes were most homogeneous. Several other classes such as Basin Marsh, Baygall, Lake Fringe, Mixed Forest, and Pine Plantation have 6 or 7 terminal nodes highlighting the heterogeneity of these classes. Based on terminal node pairs, the decision tree is trying to separate similarities between Basin Marsh versus Lake, Lake Fringe versus Ruderal, Pine Plantation versus Sandhill, and Mixed Forest versus Baygall, Basin Swamp, and Xeric Hammock.

Results

The final decision tree is generated using all of the training points. Evaluating the training points through the decision tree, an overall classification accuracy of 82.3% is achieved for all vegetation communities. The classification accuracy for each FNAI vegetation community is listed in the confusion matrix in Table 15:

Table 15: Lidar Decision Tree, Training Data Confusion Matrix

	FNAI Vegetation Community	Reference Sites											
		Baygall	Lake Fringe	Xeric Hammock	Lake	Basin Marsh	Mixed Forest	Pine Plantation	Ruderal	Sandhill	Basin Swamp	Total	User's Accuracy
Map	Baygall	78	0	0	1	3	4	0	1	0	4	91	86%
	Lake Fringe	0	72	0	2	2	0	0	5	5	0	86	84%
	Xeric Hammock	1	1	91	0	0	7	2	0	0	0	102	89%
	Lake	0	0	0	87	12	0	0	0	0	0	99	88%
	Basin Marsh	1	10	0	8	80	0	0	2	0	0	101	79%
	Mixed Forest	9	0	6	0	0	77	5	0	0	13	110	70%
	Pine Plantation	1	0	2	0	2	4	84	0	9	2	104	81%
	Ruderal	0	16	0	1	1	0	0	92	4	0	114	81%
	Sandhill	1	1	1	1	0	0	9	0	82	1	96	85%
	Basin Swamp	9	0	0	0	0	8	0	0	0	80	97	82%
	Total	100	100	100	100	100	100	100	100	100	100	823/1000	
	Producer's Accuracy	78%	72%	91%	87%	80%	77%	84%	92%	82%	80%		82%

The producer's accuracy is lower than for that of the hyperspectral based model suggesting the decision tree had a more difficult time correctly separating classes due to a higher similarity in the lidar data for different classes. Xeric Hammock and Ruderal achieved the highest classification accuracies at 91% and 92% respectively. Sandhill achieved a producer's accuracy of 82% (82% of Sandhill training locations are correctly classified) and a user's accuracy of 85% (85% of the locations labeled as Sandhill actually are Sandhill). Sandhill is primarily misclassified as Pine Plantation and occasionally misclassified as Lake

Fringe and Ruderal. Pine Plantation is primarily misclassified as Sandhill. The confusion likely stems from the fact that the Sandhill ecosystem was disturbed to plant the Pine Plantations and both ecosystems contain pines, although the pines are more regularly spaced in the plantations and might contain Slash Pine versus Longleaf Pine. The confusion of the rest of the output classes is described in Table 16:

Table 16: Confusion of Output Classes by the Decision Tree

Output Class	Confused As
Baygall	Mixed Forest and Basin Swamp
Lake Fringe	Basin Marsh and Ruderal
Xeric Hammock	A little as Mixed Forest
Lake	A little as Basin Marsh
Basin Marsh	Lake and a little as Baygall
Mixed Forest	A little as Xeric Hammock, Basin Swamp, Baygall, and Pine Plantation
Pine Plantation	Sandhill and a little as Mixed Forest
Ruderal	A little as Lake Fringe
Sandhill	Pine Plantation and a little as Lake Fringe and Ruderal
Basin Swamp	Mixed Forest and a little as Baygall

Lidar data features are also calculated for a 20 m x 20 m plot covering the independent ground validation locations collected by NEON and are run through the decision tree. An overall accuracy of 64% is achieved for all vegetation communities with Sandhill achieving a producer's accuracy of 70%

and a user's accuracy of 91%. Xeric Hammock has the highest classification accuracy at 73%. The classification accuracy for each ground validation location is listed in the confusion matrix in Table 17:

Table 17: Lidar Decision Tree, NEON Ground Validation Data Confusion Matrix

	FNAI	Reference Sites											
	Vegetation Community	Baygall	Lake Fringe	Xeric Hammock	Lake	Basin Marsh	Mixed Forest	Pine Plantation	Ruderal	Sandhill	Basin Swamp	Total	User's Accuracy
Map	Baygall	-					1			1		2	0%
	Lake Fringe		2			2				2		6	33%
	Xeric Hammock			11			3				1	15	73%
	Lake		1		-	1						2	0%
	Basin Marsh		1			4						5	80%
	Mixed Forest			1			8				2	11	73%
	Pine Plantation			2				-		3		5	0%
	Ruderal								1	3		4	25%
	Sandhill			1					1	21		23	91%
	Basin Swamp										0	-	NA
	Total	-	4	15	-	7	12	-	2	30	3	47/73	
	Producer's Accuracy	NA	50%	73%	NA	57%	67%	NA	50%	70%	0%		64%

Finally, lidar data features are calculated for a 20 m x 20 m plot on a 5 m posting covering the entire site. Data from each 5 m posting is run through the decision tree to generate a classification map. The output class map is shown in Figure 72 and has been cropped by the OSBS site boundary.

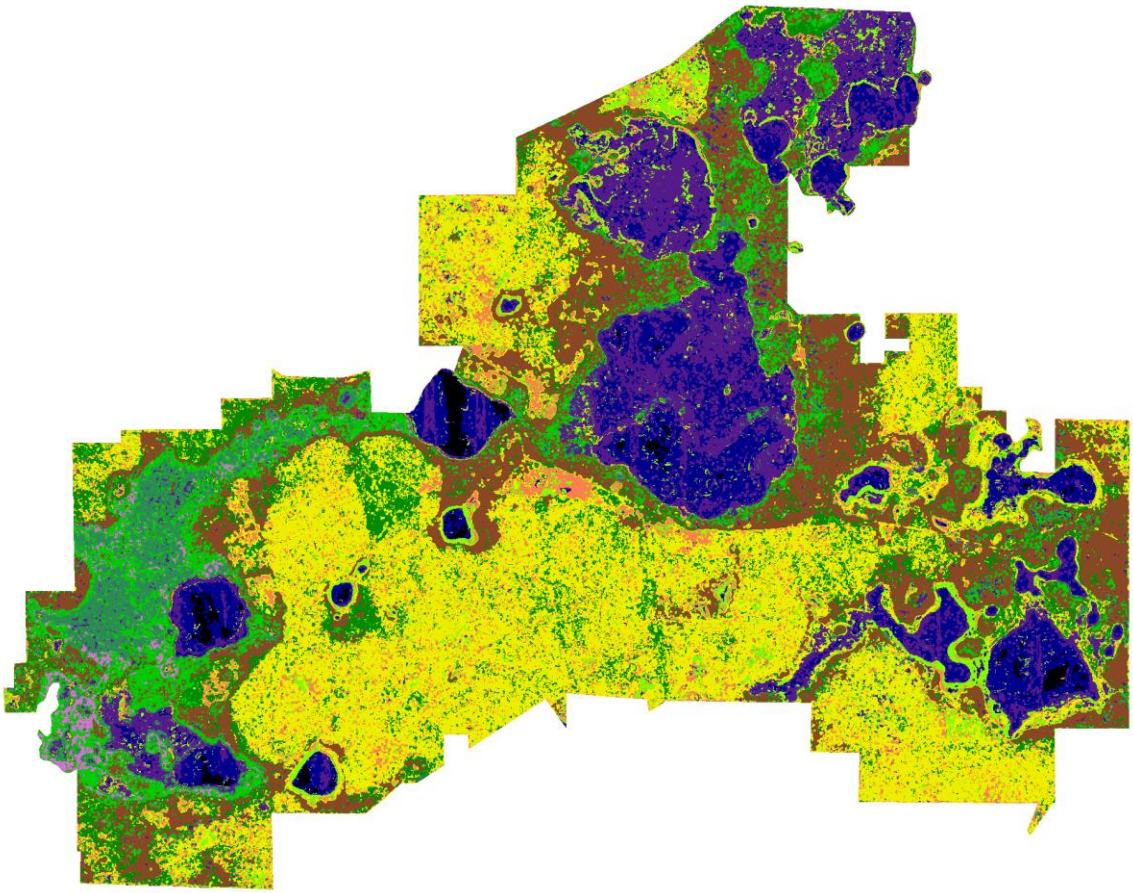

Figure 72: FNAI vegetation community classification map generated using lidar data

Figure 73: FNAI classmap legend

The black areas in the lakes represent regions of the lakes where no lidar exists. The lidar instrument might not detect a return over dark water at off-nadir angles due to the specular reflection of the water surface sending photons away from the receiver. Lidar data from adjacent flight lines may not be perfectly registered due to drifts and errors in GPS location, attitude of the aircraft, and sensor mirror and ranging. The lidar pre-processing attempts to block adjust all of the lidar lines to maximize their co-registration. Even after this process, some vertical mis-registration is seen in the flight line overlap areas leading to higher standard deviation values, especially in the calculation of ground features. These errors show up as vertical bands of Basin Marsh that can be seen in the lakes of the classification map. The areas of confusion discussed earlier can also been seen in the classification map: purple Basin Marsh and dark blue Lake pixels are intermixed and a lot of extra lake areas are displayed in the Basin Marsh in the top center of the map; aqua Baygall, pink Basin Swamp, and green Mixed Forest pixels are intermixed in the left portion of the map; and the yellow Sandhill regions are speckled with small dark green Pine Plantation and orange Ruderal pixels.

CHAPTER VII

INTEGRATED REMOTE SENSING OBSERVATIONS FOR IMPROVED VEGETATION MAPPING

In this chapter, the hyperspectral and lidar data provided by the Airborne Observation Platform (Kampe et al., 2010; Kampe et al., 2011; NEON, 2015h) of the National Ecological Observatory Network (NEON, 2015a; NRC, 2003; Schimel et al., 2011) are integrated at the feature level to take full advantage of the cutting edge AOP instrument payload. The features for the ground training locations are used to generate a decision tree classification algorithm with the goal of producing a classification map for the Ordway-Swisher Biological Station (OSBS, 2015a; OSBS, 2015c; OSBS, 2015d) site using the Florida Natural Areas Inventory (FNAI, 2010) vegetation communities as the output classes. Classification accuracy is assessed using k-folds cross validation and the independent ground truth data collected by NEON.

Objectives

Objectives

- Investigate how integrated hyperspectral and discrete return lidar data can be used to observe vegetation species diversity through both spectral (chemical or functional) and structural differences.
- Generate a vegetation community map using integrated spectral and structural differences in the data. This will serve as a proxy for vegetation species diversity.

- Show that integrated hyperspectral (spectral, chemical, species composition and function) and lidar (vertical structure) data improves the accuracy of the landscape characterization algorithm compared to data from a single instrument.
- Relate the community map to a local scale biodiversity indicator.

Science questions

- How do vegetation communities differ spectrally and structurally?
- How is combined spectral and structural diversity related to species diversity?
- What spectral and structural differences can we observe in the remote sensing data?
- How is the measured variability in the remote sensing data related to biodiversity or vegetation community type?

Integrated Hyperspectral and Lidar Features

The hyperspectral features from Chapter V are combined with the lidar features from Chapter VI to form a set of integrated features, representing both spectral and structural information. The integrated data contains a total number of 1372 features. As with the previous chapters, the data are subset for each of the training ground locations, features are calculated, and those features are brought into package "rpart" for decision tree generation, shown in Figure 74:

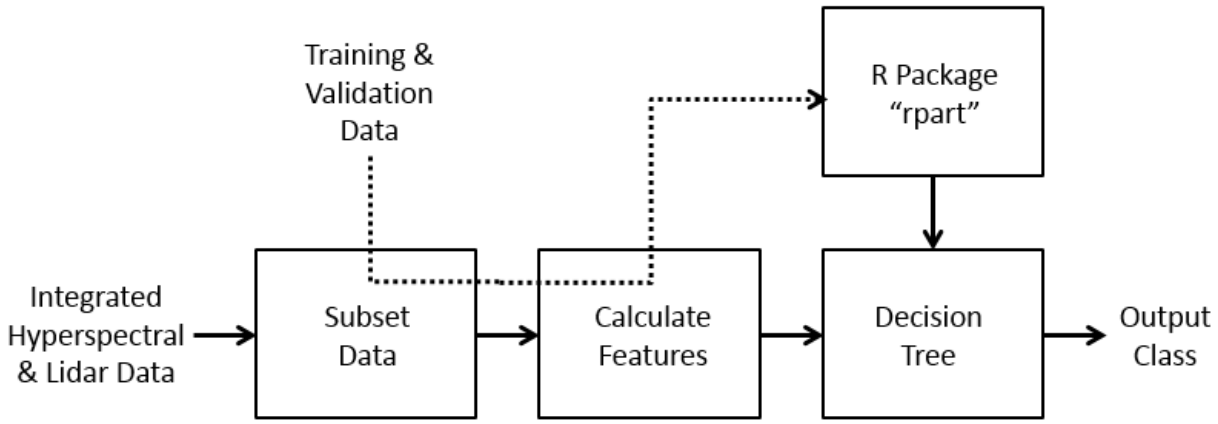

Figure 74: Integrated hyperspectral and lidar decision tree classification algorithm flowchart

Figure 75 shows a false color composite of three integrated features.

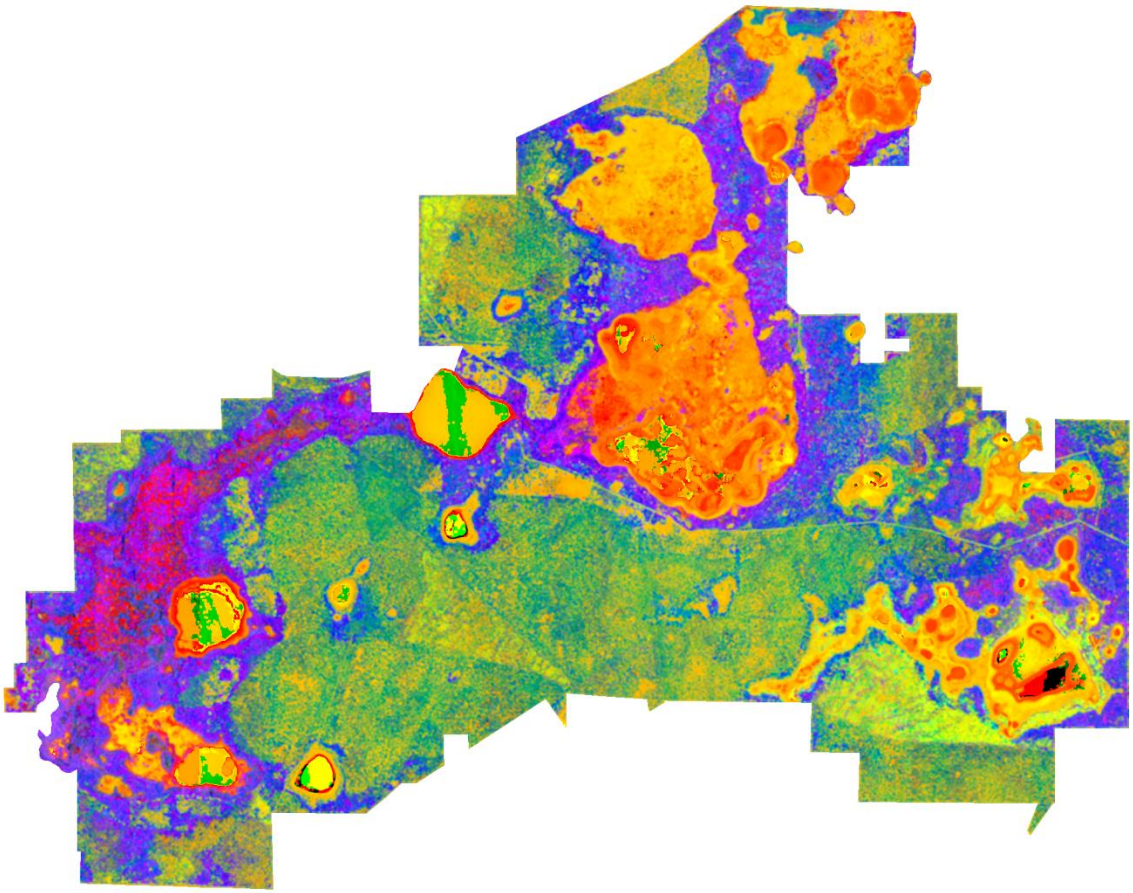

Figure 75: False color composite of three features where the red, green, and blue display channels are lidar “Frac1st”, hyperspectral “NDI578i643”, and lidar “CanopyDensity” respectively

The red display channel is the lidar feature “Frac1st” or fraction of first returns. Areas of bare ground such as Basin Marsh, Lake, Lake Fringe, and Ruderal which contain a significant number of single, first returns have very high “Frac1st” values. Also the open canopy areas in the Baygall and Sandhill communities will have an above average fraction of first returns. The green display channel is the hyperspectral feature “NDI578i643”. The normalized difference index between wavelengths 578 nm and 643 nm is correlated to the chlorophyll content in a given ground area. Chlorophyll absorbs light at 643 nm so ground locations dominated by vegetation with high chlorophyll content will have a negative

"NDI578i643" value. Other natural targets such as soil and sand have an increasing reflectance curve between 578 nm and 643 nm so these will have a positive "NDI578i643" if a ground location is dominated by soil. The 20 m x 20 m spatial sampling size contains soil/sand pixels for Basin Marsh, Sandhill, and Ruderal areas which have higher "NDI578i643" values. Other community types are greener and have higher chlorophyll content so they have lower "NDI578i643" values. The blue display channel is the lidar feature "CanopyDensity". Baygall, Mixed Forest, and Xeric Hammock have higher canopy density values because they have thicker canopies and a larger fraction of the lidar first returns are from heights 3 m or higher above the ground. Table 18 describes the display color of each vegetation community in the false color image as well as the relative value within each of the three features.

Table 18: Display Color and Relative Value of Each Vegetation Community in the False Color Feature Display

FNAI	False color	Frac1st	NDI578i643	CanopyDensity
Baygall	Purple, red, or maroon with patches of blue	Low to High	Low	Low to High
Lake Fringe	Red or orange	High	Low to Medium	Zero
Xeric Hammock	Blue	Low	Low	High
Lake	Yellow or green (no lidar points)	High	Medium to High	Zero
Basin Marsh	Yellow or orange	High	Medium to High	Zero to Low
Mixed Forest	Purple	Medium to High	Low	High
Pine Plantation	Aqua	Low to Medium	Medium to High	High
Ruderal	Orange	High	Medium	Zero to Low
Sandhill	Green with intermixed pixels of blue, yellow, and orange	Medium to High	Medium to High	Low to High
Basin Swamp	Purple	Medium to High	Low	High

In Figure 75, different FNAI vegetation communities display as different colors, illustrating the potential for a decision tree to classify the vegetation communities using the feature data. The three features shown

and listed here can quantitatively separate vegetation communities and all are used in the final integrated classification model.

Decision Tree Training

A classification decision tree is generated using the "rpart" package (Breiman et al., 1984; Therneau et al., 2015) in R (R Core Team, 2015). The variable "minbucket" is set to a value of 1 allowing terminal nodes to contain a single training location. Package "rpart" also has a complexity parameter variable "cp" that determines if the tree should generate more splits along a branch. A split must decrease the lack of fit by a factor of cp or else it is not attempted. The decision tree algorithm will iterate through different threshold values of each input feature, and determine the best data partitioning according to the fitting function of classified outputs (i.e. it picks the split that most correctly classifies the outputs).

K-folds cross validation is employed to determine the optimal value of the complexity parameter. A random number generator is used to separate the training data into 10 groups. For each fold, data from 9 of the groups is used to train the classification model and the data from the 10th group is used as independent validation. The process is performed a total of 10 times, each time generating a new decision tree and using a different group as the validation set. The complexity parameter value is varied to get an idea of the classification accuracy versus complexity parameter. If a very small complexity value is used, the classification model will over-fit the training data. The model will achieve a 100% classification accuracy for the training data but will have lower accuracy values for the validation data. By iterating the complexity parameter and performing the k-folds cross validation 10 times, the complexity parameter

value that gives the overall best classification accuracy can be chosen. The results from the k-folds cross validation are shown in Figure 76:

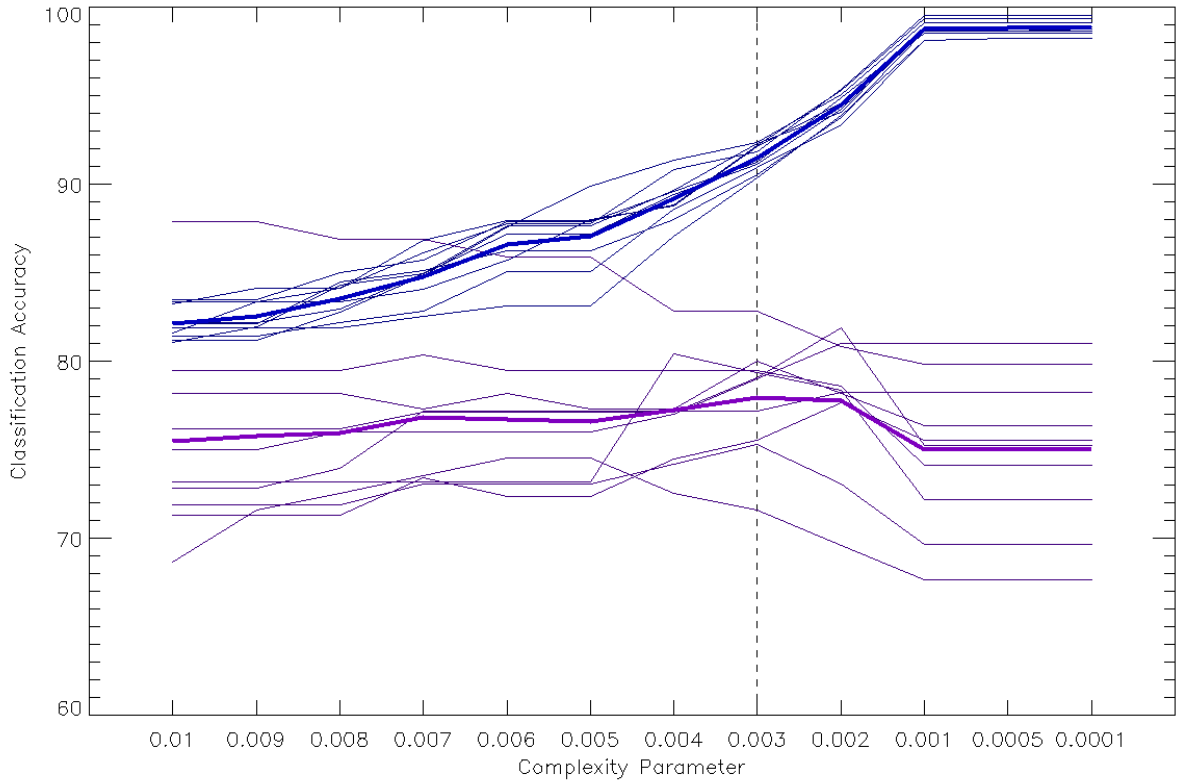

Figure 76: K-folds cross validation to determine the optimal value of the complexity parameter for the integrated hyperspectral and lidar decision tree

As seen in the figure, as the value of the complexity parameter decreases, the classification accuracy of the training data increases toward 100%. The classification accuracy of the validation data increases to a point, but then starts to decrease as the training data is over fit. Based on the k-fold cross validation results for the lidar data, a complexity parameter value of 0.003 produced a model with the best validation results. A classification accuracy of 91.5% ± 0.75 is achieved for the training data and an accuracy of 77.9% ± 3.14 is achieved for the validation data.

Classification Model

A full decision tree model is generated using the combined 1372 hyperspectral and lidar features, all 1000 of the training points, and a complexity parameter value of 0.003. The final model only uses 34 of the total features, 21 hyperspectral (using 17 of 32 wavelengths) and 13 lidar. The model also reports which variables or features are important based on the statistics of each split. The variable importance is listed in Table 19:

Table 19: Variable Importance in the Integrated Hyperspectral and Lidar Data Classification Decision Tree

Variable	Variable Importance	Used in Decision Tree
CanopyDensity	2	Yes
muCHM	2	No
Frac11	2	No
Frac22	2	Yes
muDeriv1064	2	Yes
NDI578i643	2	Yes
NDI528i668	2	No
NDI553i668	2	No
NDI578i668	2	No
Frac1st	2	Yes
NDI603i643	2	No

NDI553i643	2	No
muDeriv1249	2	No
FracGroundCHM	2	No
muDeriv783	2	No
muDeriv879	2	No
muR1314	2	Yes
vegNDI1515i2216	2	Yes
muR1249	2	No
NDI1515i2216	2	No
vegNDI1515i2161	2	No
NDI1515i2161	2	No
vegNDI1545i2216	2	No
vegNDI1515i2106	2	No
stdDeriv2216	2	No
muDeriv528	1	Yes
muDeriv994	1	No
stdDeriv2326	1	Yes
grP90mP10	1	No
stdGround	1	Yes
stdDeriv2106	1	No
grP90	1	No
grAAD	1	No
grP90mP50	1	No
NDI1585i1730	1	Yes
GreenPeakRot	1	No

grP75mP25	1	No
stdDeriv2161	1	No
stdDeriv2281	1	No
muDeriv503	1	No
muDeriv708	1	No
caP100mP0	1	Yes
muR2216	1	No
caP100	1	No
maxCHM	1	No
muDeriv1515	1	No
NDI1650i1685	1	Yes
muDeriv1650	1	No
vegNDI1585i1730	1	No
p951st	1	No
caP90	1	No
stdCanopy	1	No
muVegDeriv1650	1	No
vegNDI1615i1685	1	No
muDeriv2216	1	Yes
muDeriv668	1	Yes

The table lists 56 features that are significant in a relative sense, i.e. they appear at the top of the split decision list more often than other features. Many of these are not used in the decision tree because they were beat out by other features that could achieve a higher classification for a single split. Looking at the top variables in a structural context: “Frac11” and “FracGroundCHM” represent hard-target lidar

returns from the ground; “muCHM” is the average canopy height; “CanopyDensity” is the proportion of the lidar returns from the canopy; and finally “Frac22” and “Frac1st” represent multiple scattering and light penetration within the canopy. Looking at the top variables in a spectral context: “muDeriv1064” can detect Lake areas and “NDI578i643”, “NDI528i668”, “NDI553i668”, and “NDI578i668” are inversely related to high chlorophyll or green areas. Nineteen additional features are used in the decision tree but were not identified as important by the algorithm: “grP100mP0”, “FracMarshGreen”, “muVegDeriv2036”, “caP90mP50”, “muDeriv1685”, “caP50mP10”, “NDI1545i2106”, “FracGround”, “mu1st”, “stdR528”, “muVegDeriv668”, “muVegDeriv578”, “DEMslopeK21”, “stdDeriv643”, “caP100mP75Frac”, “stdVegDeriv643”, “caP100mP50”, “FracMixedDark”, and “vegNDI1104i1249”. These features achieved the best classification for their individual split but did not appear enough times throughout the tree to be listed as important.

A version of the decision tree generated with a complexity parameter of 0.01 is shown in Figure 77:

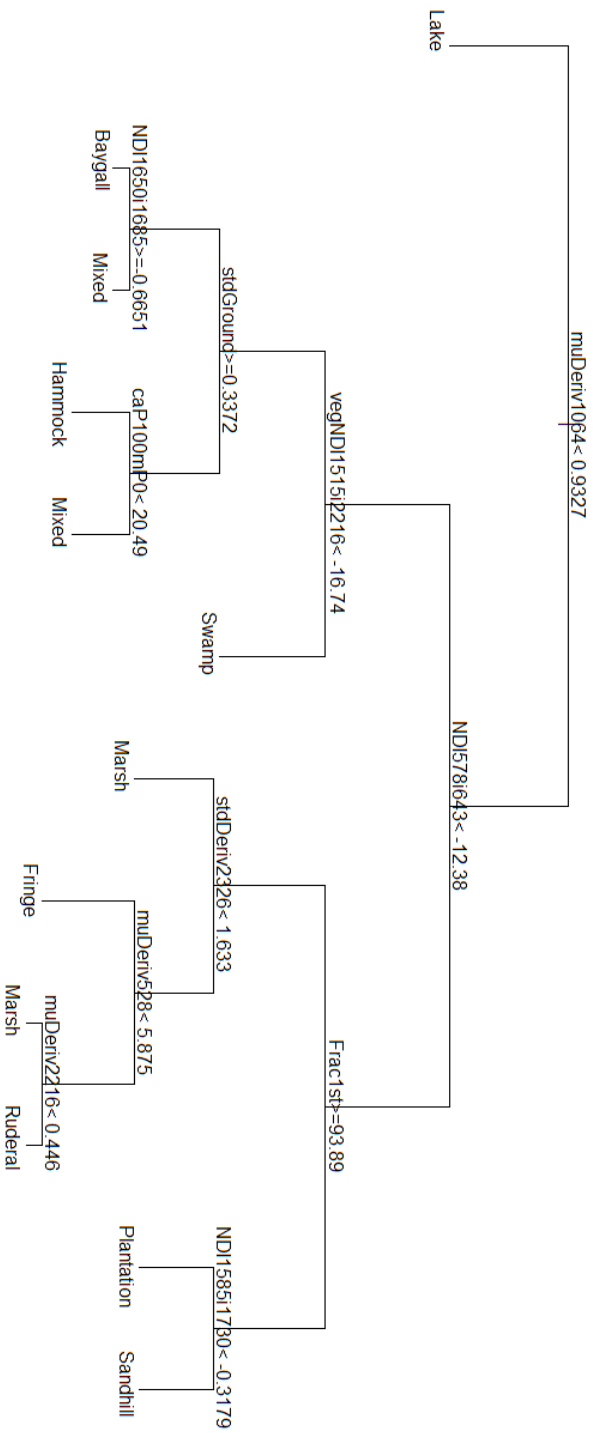

Figure 77: Integrated hyperspectral and lidar decision tree generated with a complexity parameter of 0.01

The higher cp value allows a cleaner view of the top split points of the tree. Just as in the hyperspectral decision tree, the first split is on the feature “muDeriv1064” which separates Lakes and the second split is on “NDI578i643” separating the closed canopy (areas of mostly green pixels) Baygall, Basin Swamp, Mixed Forest, and Xeric Hammock areas from open the canopy or areas with significant ground signal: Basin Marsh, Lake Fringe, Pine Plantation, Ruderal, and Sandhill. These two hyperspectral features are able to separate the data for the entire site establishing the best starting route for classification accuracy.

The third split on the left side of the tree uses “vegNDI1515i2216” to separate Swamp areas from the rest of the closed canopy classes. The fourth split on the left side of tree uses “stdGround” to separate Baygall from Xeric Hammock taking advantage of the open understory of the Xeric Hammock that would have a lower standard deviation in the ground points. The left side of Figure 78 shows “vegNDI1515i2216” plotted against “stdGround”.

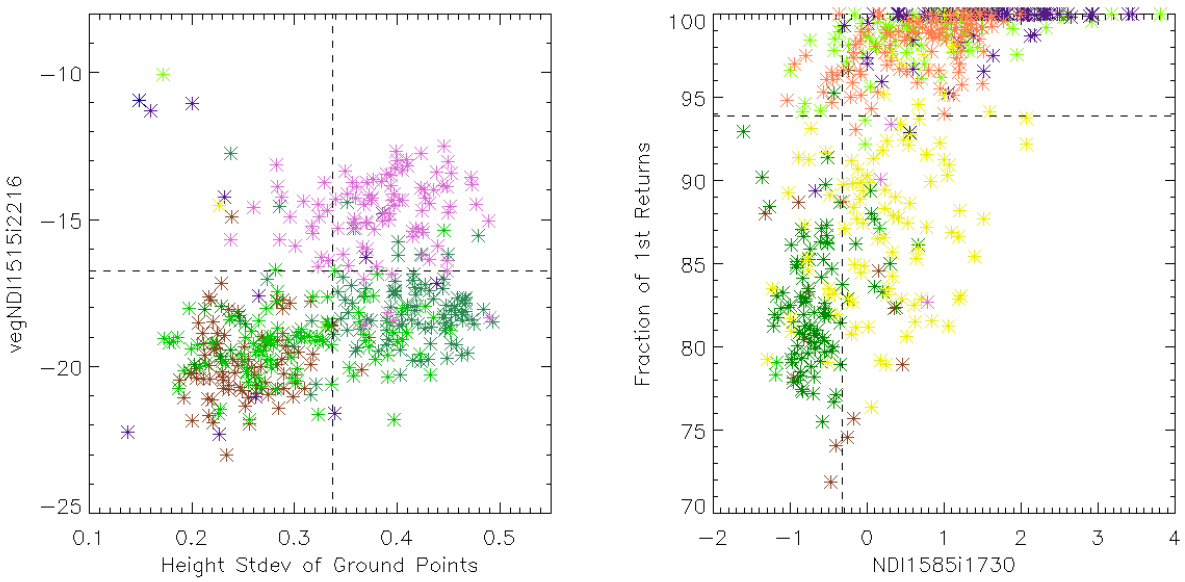

Figure 78: Scatter plots showing how a few of the decision tree splits separate the data. The left side is the normalized difference index for vegetation pixels using the reflectance at wavelengths of 1515 nm and 2216 nm versus the standard deviation of height of the ground points. The right side is the fraction of first returns versus the normalized difference index using the reflectance at wavelengths of 1585 nm and 1730 nm

The pink Basin Swamp points are easily separated into the top half of the plot as they have higher values in the "vegNDI1515i2216" feature. The brown Xeric Hammock points are separated in the lower left quadrant as they have low "vegNDI1515i2216" and low "stdGround" values whereas the aqua Baygall points are in the lower right quadrant with low "vegNDI1515i2216" and high "stdGround" values. While the green Mixed Forest points are separated in the bottom of the plot from the Basin Swamp points, they cover both the Xeric Hammock and Baygall points. The final model requires more splits using other data features in order to separate these classes better.

The third split on the right side of the tree uses "Frac1st" to separate the wide open Basin Marsh, Lake Fringe, and Ruderal areas from the open canopy Pine Plantation and Sandhill areas. One of the

fourth splits on the right side of the tree uses “NDI1585i1730” to further separate the Pine Plantation from Sandhill. The right side of Figure 78 shows “Frac1st” plotted against “NDI1585i1730”. The dark green Pine Plantation points are in the lower left quadrant with fraction of first returns below 93.9% and “NDI1585i1730” less than -0.3, the yellow Sandhill points are in the lower right quadrant with fraction of first returns below 93.9% and “NDI1585i1730” greater than -0.3, and the rest of the points are in the top of the plot. The purple Basin Marsh, chartreuse Lake Fringe, and orange Ruderal points require more splits using other data features in order to separate these classes better.

The final integrated hyperspectral and lidar based decision tree classification model using a complexity parameter of 0.003 is listed in full in Appendix B.3 and shown in Figure 79:

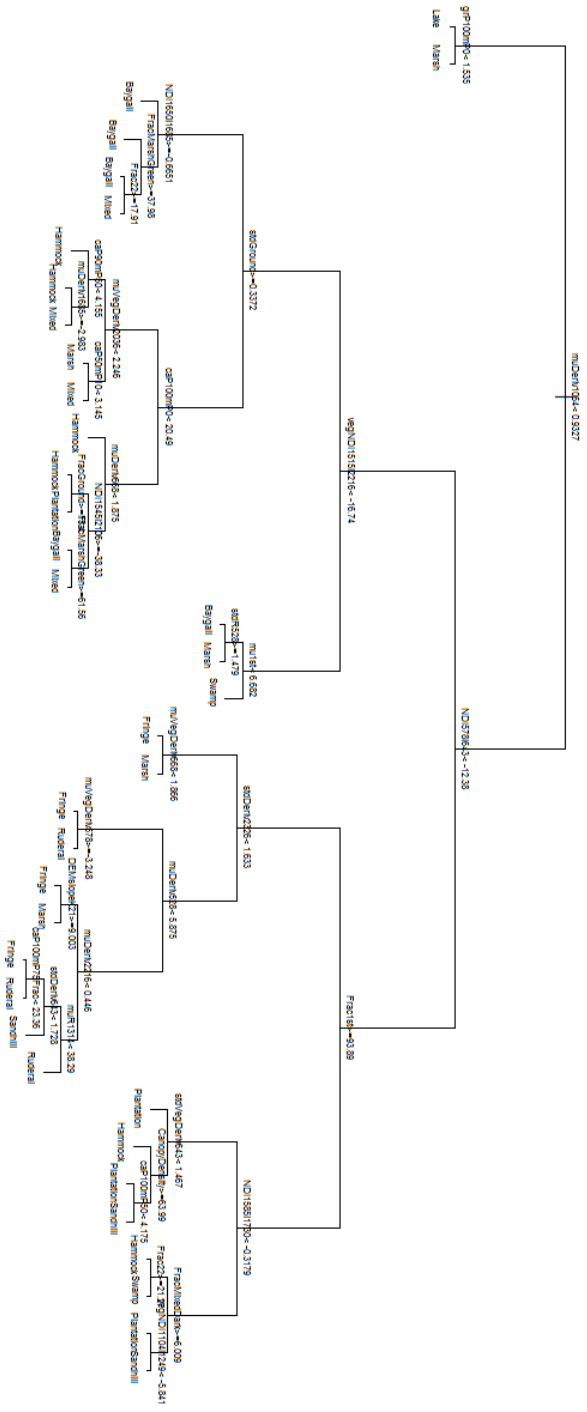

Figure 79: Final integrated hyperspectral and lidar decision tree generated with a complexity parameter of 0.003

The Lake class has a single terminal node and Basin Swamp has 2 terminal nodes suggesting that the integrated data from ground locations of these classes were most homogeneous. Several other classes such as Basin Marsh, Baygall, and Xeric Hammock have 5 or 6 terminal nodes highlighting the heterogeneity of these classes. Based on terminal node pairs, the decision tree is trying to separate similarities between: Basin Marsh versus Lake; Lake Fringe versus Basin Marsh and Ruderal; Mixed Forest versus Baygall and Xeric Hammock; Basin Swamp versus Basin Marsh, Baygall, and Xeric Hammock; and Sandhill versus Lake Fringe, Pine Plantation, and Ruderal.

Results

The final decision tree is generated using all of the training points. Evaluating the training points through the decision tree, an overall classification accuracy of 91.9% is achieved for all vegetation communities. The classification accuracy for each FNAI vegetation community is listed in the confusion matrix in Table 20:

Table 20: Integrated Hyperspectral and Lidar Decision Tree, Training Data Confusion Matrix

	FNAI Vegetation Community	Reference Sites: Baygall	Lake Fringe	Xeric Hammock	Lake	Basin Marsh	Mixed Forest	Pine Plantation	Ruderal	Sandhill	Basin Swamp	Total	User's Accuracy
Map	Baygall	95	0	2	0	2	5	0	0	0	5	109	87%
	Lake Fringe	0	92	1	1	1	0	0	5	3	0	103	89%
	Xeric Hammock	2	1	93	0	0	0	1	0	0	0	97	96%
	Lake	0	0	0	93	2	0	0	0	1	0	96	97%
	Basin Marsh	1	2	1	5	94	0	0	0	0	0	104	90%
	Mixed Forest	1	0	0	0	0	89	3	0	0	1	94	95%
	Pine Plantation	0	0	1	1	1	2	87	1	5	1	98	89%
	Ruderal	0	3	0	0	0	0	1	93	0	0	97	96%
	Sandhill	0	0	1	0	0	1	6	1	90	0	99	91%
	Basin Swamp	1	2	1	0	0	3	2	0	1	93	103	90%
	Total	100	100	100	100	100	100	100	100	100	100	919/1000	
	Producer's Accuracy	95%	92%	93%	93%	94%	89%	87%	93%	90%	93%		92%

The producer's accuracy is higher than for that of the hyperspectral based model for most of the output classes except Lake, Lake Fringe, and Pine Plantation suggesting the decision tree has a higher success at correctly separating classes. Furthermore all of producer's accuracies are at or above 90% except Mixed Forest (89%) and Pine Plantation (87%). Basin Marsh and Baygall achieved the highest classification accuracies at 94% and 95% respectively. Sandhill achieved a producer's accuracy of 90% (90% of Sandhill training locations are correctly classified) and a user's accuracy of 91% (91% of the locations

labeled as Sandhill actually are Sandhill). Sandhill is primarily misclassified as Pine Plantation and occasionally misclassified as Lake Fringe. Pine Plantation is primarily misclassified as Sandhill. The confusion likely stems from the fact that the Sandhill ecosystem was disturbed to plant the Pine Plantations and both ecosystems contain pines, although the pines are more regularly spaced in the plantations and might contain Slash Pine versus Longleaf Pine. The confusion of the rest of the output classes is described in Table 21:

Table 21: Confusion of Output Classes by the Decision Tree

Output Class	Confused As
Baygall	A little as Xeric Hammock
Lake Fringe	A little as Basin Marsh, Basin Swamp, and Ruderal
Xeric Hammock	A little as Baygall
Lake	Basin Marsh
Basin Marsh	A little as Baygall and Lake
Mixed Forest	Baygall and a little as Basin Swamp and Pine Plantation
Pine Plantation	Sandhill and a little as Mixed Forest and Basin Swamp
Ruderal	Lake Fringe
Sandhill	Pine Plantation and a little as Lake Fringe
Basin Swamp	Baygall

Hyperspectral and lidar data features are also calculated for a 20 m x 20 m plot covering the independent ground validation locations collected by NEON and are run through the decision tree. An overall accuracy of 66% is achieved for all vegetation communities with Sandhill achieving a producer's accuracy of 67% and a user's accuracy of 87%, which is worse than the hyperspectral model from Chapter V. Mixed Forest has the highest classification accuracy at 92%. The classification accuracy for each ground validation location is listed in the confusion matrix in Table 22:

Table 22: Integrated Hyperspectral and Lidar Decision Tree, NEON Ground Validation Data Confusion Matrix

	FNAI	Reference Sites											
	Vegetation Community	Baygall	Lake Fringe	Xeric Hammock	Lake	Basin Marsh	Mixed Forest	Pine Plantation	Ruderal	Sandhill	Basin Swamp	Total	User's Accuracy
Map	Baygall	-									1	1	0%
	Lake Fringe		2			2				2		6	33%
	Xeric Hammock			10			1			2		13	77%
	Lake				-							-	NA
	Basin Marsh					4						4	100%
	Mixed Forest			1			11				2	14	79%
	Pine Plantation			1				-		5		6	0%
	Ruderal		2			1			1			4	25%
	Sandhill			2					1	20		23	87%
	Basin Swamp			1						1	0	2	0%
	Total	-	4	15	-	7	12	-	2	30	3	48/73	
	Producer's Accuracy	NA	50%	67%	NA	57%	92%	NA	50%	67%	0%		66%

Finally, lidar data features are calculated for a 20 m x 20 m plot on a 5 m posting covering the entire site. Data from each 5 m posting is run through the decision tree to generate a classification map. The output class map is shown in Figure 80 and has been cropped by the OSBS site boundary.

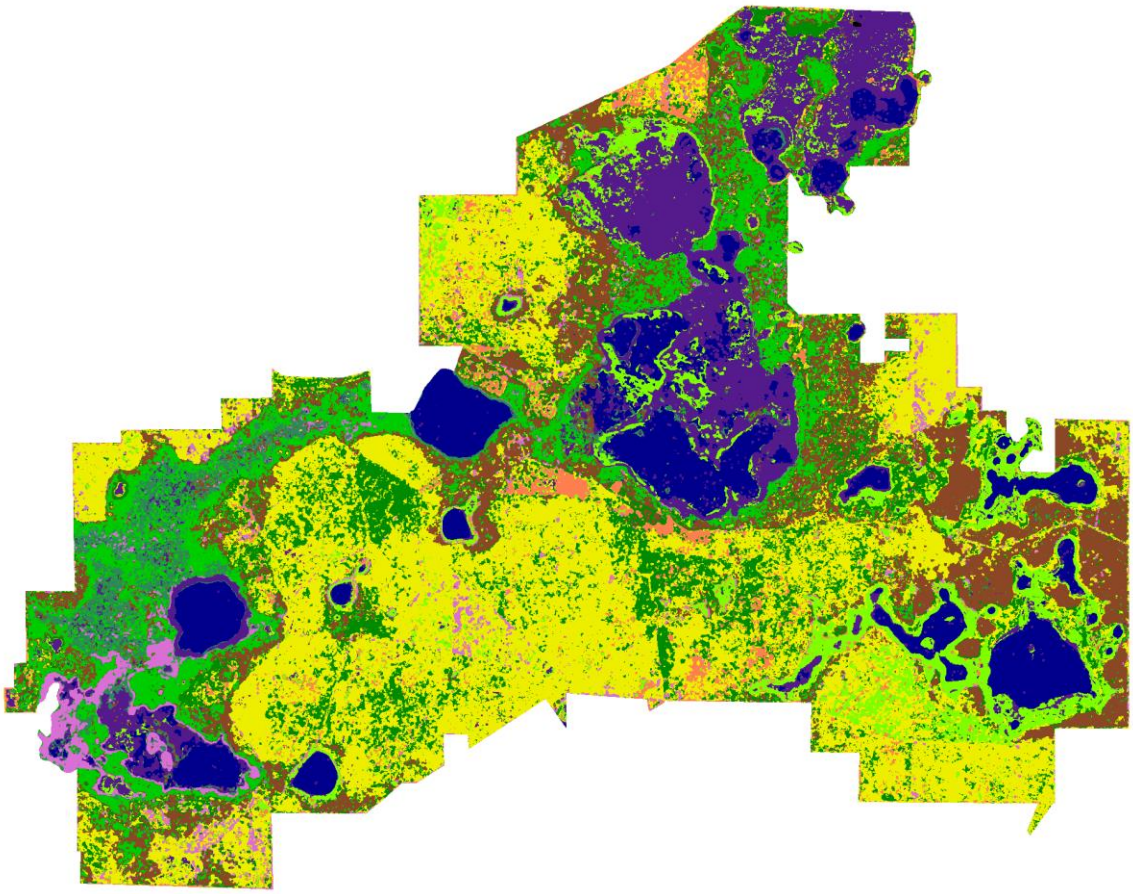

Figure 80: FNAI vegetation community classification map generated using the integrated hyperspectral and lidar data

Figure 81: FNAI classmap legend

The areas of confusion discussed earlier can also been seen in the classification map: purple Basin Marsh pixels dot the lakes in areas that have a significant amount of floating vegetation; pixels in the aqua Baygall area on the left of the map are misclassified as brown Mixed Forest, pink Basin Swamp, and purple Basin Marsh; yellow Sandhill regions are speckled with small dark green Pine Plantation and in the case of recently burned management units, Sandhill is misclassified as Lake Fringe (this confusion may be due to the fact that prescribed burning recently took place in both Sandhill and Lake Fringe units in the Southeast portion of the site). Unfortunately, some of the thick vegetation areas in the Sandhill are misclassified as pink Basin Swamp.

CHAPTER VIII

CONCLUSIONS

Local Scale Biodiversity Indicator

The Grand Challenges in Environmental Sciences (NRC, 2001) call for new tools of rapid assessment of biodiversity. A local scale biodiversity indicator can easily be derived from the remote sensing data and the ground observations of biodiversity. In this research, the biodiversity indicator used is vegetation species richness. The NEON ground diversity data (species number derived from species lists) from Figure 38 in Chapter IV shows that a relationship exists between species richness and FNAI community, i.e. some communities support a larger diversity of vegetation species. A relative ranking is generated using the species number data and is listed in Table 23:

Table 23: Species Richness Relative Rank

FNAI Vegetation Community	Species Richness Relative Rank
Upland Clastic Lake	0
Upland Sandhill Lake	0
Basin Marsh	2
Upland Sandhill Lake Fringe	3
Basin Swamp	4
Upland Mixed Forest	5
Baygall	6
Xeric Hammock	7
Ruderal	8
Pine Plantation	9
Sandhill	10

A map of the FNAI vegetation communities serves as a proxy of vegetation species richness by applying the relative rank to the identified vegetation community in the classification map for each pixel location. The resulting relative vegetation species richness map generated from the hyperspectral classification map is shown in Figure 82:

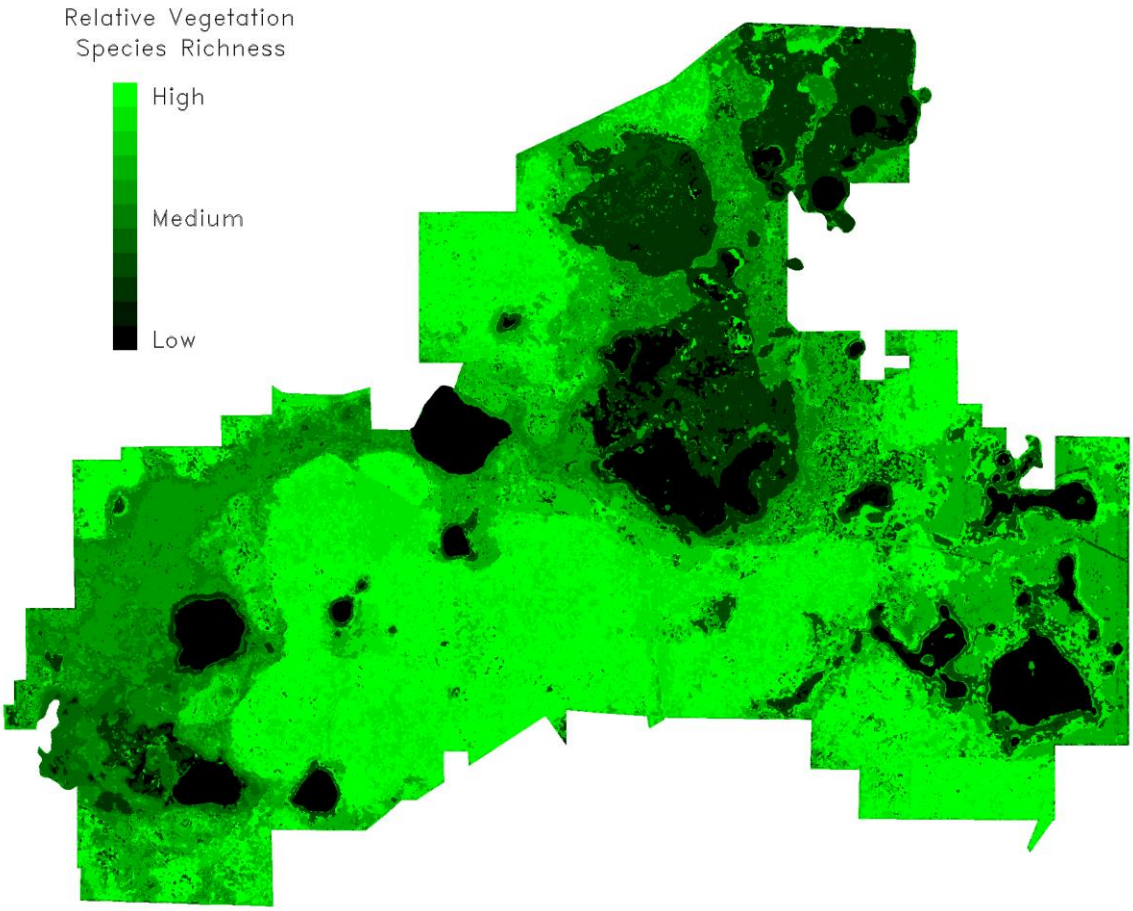

Figure 82: Relative vegetation species richness as a local scale biodiversity indicator

Here the relative rank is displayed as shades of green where darker areas have lower vegetation species richness and saturated green areas have higher richness. Since the brightness scale on the map is relative to the vegetation communities at the study site, it is categorized as a local scale biodiversity indicator. The species richness map is a type of alpha diversity metric. A total of 329530 pixels are labeled as Sandhill with a relative rank of 10. The spatial resolution of the map is 5 m so this comes to a total area of 9.81 km^2 or 2424.91 acres. 29.35% of the pixels are labeled Sandhill which is a bit lower

than the 37% estimated from the GIS data in Chapter III. This underestimation of Sandhill area is consistent with the 76.0% classification accuracy measured during the k-folds cross validation.

Discussion of Results

Project Objectives

- Objective 1: Investigate remote sensing technologies to map ecosystems.

The literature on biodiversity, remote sensing of biodiversity, and ecosystem mapping was reviewed. Remote sensing primarily observes vegetation from above with hyperspectral data influenced by chemical, structural, and functional traits of the vegetation and lidar data capable of measuring the 3D physical structure of vegetation. Remote sensing data are commonly used to map generalized land cover with a few studies focusing on mapping specific ecosystems. Another common practice is to identify and map specific species. Some biogeochemistry studies looked at mapping vegetation traits such as nitrogen, carbon, and water thickness in order to assess the spatial variability in the vegetation and then use proxies to determine the species richness. The prior literature and early data prototypes led this research in the direction of mapping vegetation communities using both hyperspectral and lidar remote sensing data.

- Objective 2: Generate a vegetation community map from remote sensing data. Validate the map with in-situ ground truth data.

A classification and regression tree (CART) algorithm (Breiman et al., 1984; Therneau et al., 2015) is used to generate the vegetation community map. A series of metrics are calculated for a 20 m x 20 m plot size from the remote sensing data which characterize the spectral and structural variations across the landscape of the study site. The metrics serve as the input feature layers to the classification tree. Training data points for each of the vegetation communities are hand selected using photo-interpretation of high resolution imagery in combination with GIS data of the study site. The classification tree is generated using the training data and then validated using independent in-situ ground truth data by comparing the vegetation community class predicted by the classification tree against the actual vegetation community as determined by species lists and site photos from field sampling. Vegetation community classification maps are presented in Chapters V-VII.

Using k-folds cross validation, a complexity parameter value of 0.003 is chosen as both the best hyperspectral model and lidar model. The hyperspectral decision tree produces a classification accuracy of 90.0% ± 0.54 for the training data and an accuracy of 76.0% ± 3.50 for the validation data. Sandhill achieved a producer's accuracy of 89% (89% of Sandhill training locations are correctly classified) and a user's accuracy of 91% (91% of the locations labeled as Sandhill actually are Sandhill). Assessing the independent ground truth points, an overall accuracy of 66% is achieved for all vegetation communities with Sandhill having a producer's accuracy of 80% and a user's accuracy of 100%. The lidar decision tree produces a classification accuracy of 82.4% ± 1.08 for the training data and an accuracy of 61.1% ± 4.25 for the validation data. Sandhill achieved a producer's accuracy of 82% and a user's accuracy of 85%. Assessing the independent ground truth points, an overall accuracy of 64% is achieved for all vegetation

communities with Sandhill having a producer's accuracy of 70% and a user's accuracy of 91%. The accuracy numbers for both the hyperspectral and lidar data models are listed in Table 24 in the next section discussing Objective 3.

The lidar data alone didn't perform as well as the hyperspectral data suggesting that the decision tree had a more difficult time correctly separating classes due to a higher similarity in the lidar data (vegetation structure similarity) for different classes. The lidar classification accuracy might be improved by using the intensity values as an input feature to the classification model. Intensity might be able to distinguish areas that have similar structure, though intensity values are dependent on the range between the instrument and the ground and a correction would need to be applied to properly scale the intensities before they could be used. The hyperspectral model showed a higher accuracy over the lidar model by +7.6% for the k-folds training data, +14.9% for the k-folds validation data, and +2.0% for the independent ground validation. For both models, Sandhill is primarily misclassified as Pine Plantation and occasionally misclassified as Lake Fringe and Ruderal. Pine Plantation is primarily misclassified as Sandhill. The confusion likely stems from the fact that the Sandhill ecosystem was disturbed to plant the Pine Plantations and both ecosystems contain pines, although the pines are more regularly spaced in the plantations and might contain Slash Pine versus Longleaf Pine. Sandhill and Xeric Hammock achieved the highest classification accuracies of all of the vegetation communities for the hyperspectral and lidar data models respectively for the independent ground truth data.

- Objective 3: Show that integrated hyperspectral (spectral, chemical, species composition and function) and lidar (horizontal and vertical structure) data improves the accuracy of the landscape characterization algorithm compared to data from a single instrument.

Using k-folds cross validation, a complexity parameter value of 0.003 is also chosen as the best integrated hyperspectral plus lidar data model. The integrated decision tree produces a classification accuracy of 91.5% ± 0.75 for the training data and an accuracy of 77.9% ± 3.14 for the validation data. Sandhill achieved a producer's accuracy of 90% (90% of Sandhill training locations are correctly classified) and a user's accuracy of 91% (91% of the locations labeled as Sandhill actually are Sandhill). Assessing the independent ground truth points, an overall accuracy of 66% is achieved for all vegetation communities with Sandhill having a producer's accuracy of 67% and a user's accuracy of 87%. The accuracy numbers for the integrated hyperspectral and lidar data model are compared against the individual sensor models and listed in Table 24:

Table 24: Classification Accuracies for Hyperspectral, Lidar, and Integrated Decision Trees of Vegetation Community at OSBS

	Lidar	Hyperspectral	Integrated
K-Folds Training Accuracy	82.4% ± 1.08	90.0% ± 0.54	91.5% ± 0.75
K-Folds Validation Accuracy	61.1% ± 4.25	76.0% ± 3.50	77.9% ± 3.14
Independent Ground Validation	64.0%	66.0%	66.0%

Overall the integrated model achieves better accuracy where the producer's accuracy is higher than for that of the hyperspectral based model for most of the output classes except Lake, Lake Fringe, and Pine Plantation suggesting the decision tree has a higher success at correctly separating classes. The integrated model showed a higher accuracy over the hyperspectral model by +1.5% for the k-folds training data, +1.9% for the k-folds validation data. The accuracy for the independent ground validation points is the same for the integrated and hyperspectral models. Furthermore all of producer's accuracies are at or above 90% except Mixed Forest and Pine Plantation. However, the Sandhill independent validation points were worse for the integrated model versus those of the hyperspectral and lidar only models. Sandhill is primarily misclassified as Pine Plantation and occasionally misclassified as Lake Fringe, Xeric Hammock, and Basin Swamp (some pixels in burn areas).

- Objective 4: Start to relate derived remote sensing features to structure, chemistry, and function.

The hyperspectral features are meant to be proxies of the 2D spectral (chemical, functional, and structural) differences of the vegetation communities as measured by the NEON Imaging Spectrometer instrument. The subset of 32 wavelengths capture the spectral shapes of the ground objects and are influenced by absorption and scattering. Feature layers using visible wavelengths shorter than 680 nm are highly correlated to chlorophyll absorption. Several of the normalized difference indices are influenced by liquid water absorption in leaves. Others features (depending on the wavelengths used) are tied to lignin and cellulose. Higher reflectance in the near-infrared wavelengths may indicate multiple scattering due to increased leaf area index and is related to the structure of the leaves and canopies. Finally, the spectra themselves, or the assemblage of certain endmembers in a 20 m x 20 m ground plot may be indicative or certain vegetation functional groups and can separate closed canopy from open canopy areas based on

fraction of green pixels. Vegetation chemistry and structure (leaf area index for example) hyperspectral products were not available from NEON at the time of the research but could create a tighter link between reflectance and the underlying vegetation traits.

The lidar features are meant to be proxies of the horizontal 2D and vertical 3D structure of the different vegetation communities as measured by the discrete return lidar instrument. The lidar point clouds represent not only the 3D structure of the canopies and ground, based on their reflectance of light at 1064 nm, but also exhibit optical phenomena such as light penetration and multiple scattering. Open verses closed canopy areas can be differentiated based on the fraction of discrete returns classified as canopy or ground based on their relative height above a bare-Earth terrain model. Mean and max height in the discrete points or the canopy height model can detect differences in species, stand age, and functional groups. Analysis of multiple returns in a canopy is directly related to leaf area and clumping and may provide some insight into the absorption, transmission, and reflection properties of leaves. The vertical distribution of discrete returns (a pseudo waveform) can also capture vertical profiles of leaf area and canopy stratification (overstory versus understory layers). Finally, statistics calculated on discrete returns near the ground can provide insight into how flat the ground is and whether or not tall grasses or shrubs exist.

- Objective 5: Leverage the collection of pixels within a ground plot.

Analysis of 1 m ground sample distance hyperspectral pixels in a 20 m x 20 m ground plot includes 400 pixels that are used for the metric/feature calculations. The objective strives to move beyond basic summary statistics such as mean, standard deviation, min, max, and range, and utilize the variability

of the 400 pixels in new ways. While standard deviation and range have been shown to be correlated to biodiversity, grouping pixels into endmembers allows the data to be analyzed as a distribution of pixel reflectance values which may share a stronger link to vegetation communities. This research performs a basic pixel classification on the hyperspectral data and then calculates features as "fraction of" in order to capture pixel distributions. The fraction feature list is: FracSandRoad, FracLakeFringe, FracWaterShallow, FracWaterDark, FracWaterGlint, FracWaterDarkVeg, FracWaterMedVeg, FracWaterLightVeg, FracWaterWhiteVeg, FracShadowGround, FracShadowCanopy, FracSandhillVeg1, FracSandhillVeg2, FracSandhillGrass, FracSandhillBurnt, FracTreesBright, FracTreesDark, FracMixedDark, FracTreesGray, FracRuderalGreen, FracRuderalMixed, FracRuderalBright, FracMarshGreen, FracMarshPink, FracMarshLight, FracMarshDark, and FracBurnScar. This research subsets vegetation endmember pixels and calculates features on them separately, which is enabled by having the group of 1 m pixels. The fraction feature list is fairly generic but future versions could improve the vegetation classes to directly represent actual species or possibly use unsupervised training techniques to group pixels into distinct clusters of spectra. Foliar chemistry maps (such as chlorophyll) could be generated and the distribution of values within the ground plot could be analyzed as another advanced assemblage technique.

Analysis of 3.82 lidar points per meter^2 in a 20 m x 20 m ground plot includes 1528 pixels that are used for the metric/feature calculations. Each of these points can be classified as either a ground point, below ground, shrub, canopy, high canopy, first return, 1st of 1, 2nd of 2, 3rd of 3, and 4th of 4. Similar to the hyperspectral data, "fraction of" metrics can be calculated on the lidar data. Other features common in the literature include combining the discrete return points into a vertical distribution (pseudo waveform) of height above ground. Percentile heights, percentile height ranges (including interquartile range), and point distribution statistics are calculated. Future work might include analyzing full-waveform lidar for species identification or possibly using the lidar heights in an integrated fashion with the hyperspectral

data to calculate features for low vegetation pixels separately than high vegetation pixels for example. See concept of assemblages of pixels or points in “Future Work” Section of this chapter for more information about future work on this objective.

- Objective 6: Analyze the key spectral and structural differences between vegetation communities as measured in the remote sensing data.

Each split within the decision tree separates the data into two groups. Ideally, the intra-variability within a single group should be low and the inter-variability between groups should be high (i.e. the two groups should be different in terms of the remote sensing data). The decision tree forms a dendrogram diagram where output classes whose branches are closer together on the graph are more similar to each other in terms of their remote sensing data (i.e. data from these classes are grouped together during the earlier splits).

The first hyperspectral split is on the feature “muDeriv1064” which separates Lakes from the rest of the vegetation communities based on water spectra having decreasing reflectance versus wavelength in the near-infrared region. While this split isolates most of the Lake locations, a few Lake points are excluded and a few Basin Marsh points are included. This may be a case where the excluded Lake points have significant floating vegetation and the included Basin Marsh points may have standing water. The second split is on “NDI578i643” separating the closed canopy (areas of mostly green pixels) Baygall, Basin Swamp, Mixed Forest, and Xeric Hammock areas from open the canopy or areas with significant ground signal (sand and grasses): Basin Marsh, Lake Fringe, Pine Plantation, Ruderal, and Sandhill. Further splits attempt to separate Swamp areas from the rest of the closed canopy classes; separate

Baygall from Xeric Hammock and most of the Mixed Forest; separate the bare-Earth Basin Marsh, Lake Fringe, and Ruderal areas from the open canopy Pine Plantation and Sandhill areas; and separate the Basin Marsh from the Lake Fringe and Ruderal areas. The Basin Swamp and Ruderal classes each have one terminal node and the Lake class has two terminal nodes suggesting that hyperspectral data from ground locations of these classes were most homogeneous. Several other classes such as Basin Marsh, Lake Fringe, and Xeric Hammock have 5 or 6 terminal nodes highlighting the heterogeneity of these classes. Based on terminal node pairs, the decision tree is trying to separate similarities between: Basin Marsh versus Lake; Lake Fringe versus Basin Marsh and Ruderal; Mixed Forest versus Baygall and Xeric Hammock; Basin Swamp versus Basin Marsh and Baygall; and Sandhill versus Lake Fringe, Pine Plantation, and Ruderal.

The first lidar split is on the feature "muCHM" which uses the mean height of the gridded canopy height model to separate areas with and without canopies (such as the open areas of Basin Marsh, Lake, Lake Fringe, and Ruderal). Most of the Basin Marsh, Lake, Lake Fringe, and Ruderal points are grouped together after the split, however, a few ground locations with some trees are misclassified. Also, open canopy areas in Baygall and Sandhill are grouped with the other open canopy classes. The second split on the left side of the tree uses the feature "DEMsplopek21" to separate flat Lakes and Basin Marsh from sloped ground areas of Lake Fringe, Basin Marsh, and Ruderal. Lake and Basin Marsh points are grouped together but separated from Lake Fringe and Ruderal points. A further decision attempts to separate the Lake Fringe points from the Ruderal points but there is still some overlap between these classes. The final model requires more splits using other data features in order to separate these classes better. The second split on the right side of the tree uses "FracGround" to separate closed canopy Basin Swamp, Baygall, Xeric Hammock, and Mixed Forest areas from partially open canopy Baygall, Pine Plantation, and Sandhill areas. With splits using "FracGround" and "stdGround", Xeric Hammock and Mixed Forest are separated with low fraction of ground points and a low ground height standard deviation, Basin Swamp

and Baygall are separated with a low fraction of ground points and a higher ground height standard deviation, and Sandhill and Pine Plantation are separated with a high fraction of ground points and a low ground height standard deviation. The Basin Swamp and Lake classes each have two terminal nodes suggesting that lidar data from ground locations of these classes were most homogeneous. Several other classes such as Basin Marsh, Baygall, Lake Fringe, Mixed Forest, and Pine Plantation have 6 or 7 terminal nodes highlighting the heterogeneity of these classes. Based on terminal node pairs, the decision tree is trying to separate similarities between Basin Marsh versus Lake, Lake Fringe versus Ruderal, Pine Plantation versus Sandhill, and Mixed Forest versus Baygall, Basin Swamp, and Xeric Hammock.

The integrated model first tries to separate out Lake points similar to the hyperspectral model. Next, the closed canopy (areas of mostly green pixels) Baygall, Basin Swamp, Mixed Forest, and Xeric Hammock areas are separated from open the canopy or areas with significant ground signal: Basin Marsh, Lake Fringe, Pine Plantation, Ruderal, and Sandhill. The Lake areas are distinct in the hyperspectral data making them easy to separate but they have a similar structure to other flat areas without trees in the lidar data (which is why they aren't separated in the lidar model). A common theme in all three decision tree models is the separation of open canopy from closed canopy vegetation communities. Fraction of canopy manifests as greenness in the hyperspectral data and 2D canopy structure in the lidar data.

- Objective 7: Relate the classification results to a local scale biodiversity indicator.

A local scale biodiversity indicator is presented in the first section of this chapter. The biodiversity indicator is species richness. Using the relationship between FNAI community and species richness, a relative rank of diversity is created. A map of the FNAI vegetation communities serves as a

proxy of vegetation species richness by applying the relative rank to the identified vegetation community in the classification map for each pixel location. Saturated green areas have higher richness, with the maximum richness given to pixels classified as Sandhill. A future version of this result could take advantage of spatial variability in the remote sensing data along with some regression against species richness to display a finer gradation in the richness relative rank and spatial variability in richness. The map can be used to inform which areas of the site should be prioritized for conservation and serves as a baseline for long-term change analysis with future AOP remote sensing data of OSBS.

General discussion

The integrated remote sensing decision tree has the highest classification accuracy, although the hyperspectral only model was very close. This suggests that hyperspectral sensors capture structural information in the reflectance curves either directly due to scattering or indirectly due to endmembers or individual species as observed in the pixel data. The accuracy results may also suggest that the hyperspectral and lidar data integration was not as successful as originally planned. Integrating at the feature layer does not provide a significant improvement in overall classification accuracy. Future work includes fusing earlier in the processing chain such as separating high and low vegetation and calculating individual features/metrics for each of those classes.

Several error sources exist which might have prevented the achievement of higher accuracy numbers. The decision tree training points are hand selected using photo-interpretation of high resolution imagery and GIS layers provided by the study site. Confusion about boundaries between vegetation communities and errors in the GIS layers can cause training points to be selected in the wrong community

type. Human error could also cause the ground coordinates of training points to be recorded incorrectly. The independent ground validation points could also be mislabeled with the wrong vegetation community. While species lists are helpful, vegetation communities tend to have dominant canopy species that are primarily observed by the remote sensing data. Vegetation abundance, especially in the canopy, would be extremely valuable to confirm that the correct community class is assigned to the validation data. Considering that only 73 independent ground validation points were available at the time of the research, a few misclassified points causes a significant reduction in accuracy. Finally, noise or other artifacts in the remote sensing data could cause misclassification errors.

The success of vegetation community mapping using remote sensing relies on detectability of community differences in the sensor data. The hyperspectral instrument measures reflected solar energy from the top of the canopy. This sensor cannot penetrate all the way through the canopy to the ground unless the canopy is open. In the case of the Sandhill community, the herbaceous ground layer has the most diversity. This is an open canopy ecosystem, but even 1 m spatial pixels cannot measure individual herbs and grasses. Lidar is able to directly measure the 3D structure of the canopy and ground, but also has limits in the vertical resolution of what size objects can be individually detected.

Vegetation communities that only have one or two terminal nodes in the classification model have homogeneous remote sensing data and can be considered endemic in terms of the chemical, functional, and structural traits as observed by remote sensing. Other heterogeneous vegetation communities that follow multiple branches in the decision tree are more difficult to correctly classify as these tend to look similar to other vegetation communities. For instance, the decision trees misclassify Sandhill as Pine Plantation and vice-versa. The confusion likely stems from the fact that the Sandhill ecosystem was disturbed to plant the Pine Plantations and both ecosystems contain pines, although the

pines are more regularly spaced in the plantations and might contain Slash Pine versus Longleaf Pine. Ground locations that flow out of terminal nodes whose branches are grouped near each other at the later splits in the decision tree are more similar to each other in terms of their remote sensing data, and likely have similar chemical, functional, and structural traits.

Biodiversity is much more than just trees and plants. Overall, the airborne remote sensing data have enough resolution to map habitats which can serve as a proxy for species diversity including other taxa (such as fauna) that may not be directly detected by remote sensing. More work is required to characterize biodiversity in finer detail than just vegetation community. Variability in the remote sensing data can be correlated to vegetation species diversity as long as illumination geometry and other sensor artifacts are corrected for. Other biodiversity indicators besides species richness could be applied such as diversity measures like the Shannon Index or assessment of ecosystem condition and fragmentation. This research contains vegetation community maps collected at a single point in time and serves as a baseline for analysis of ecosystem change with time, a very important metric in many biodiversity indicators. The research is also only performed at a single study site. The next section addresses extending this research to other sites and space remote sensing missions. This research provides a solid framework for classification of vegetation community. During the course of the work, many ideas were born or evolved but require further investigation. The final section of this chapter contains an extensive list of possible follow on or derivative work that is complimentary to the research presented here.

Extendibility to Other Sites and Space Remote Sensing Missions

In this research, vegetation community classification maps are only generated for a single site (Ordway-Swisher Biological Station) using a state/regional based classification system (FNAI). The desire exists to apply the techniques developed here across multiple sites and ecoregions. In 2014, NEON AOP also collected data over two other sites in the NEON Southeast Domain 03 including Florida (Disney Preserve) and Georgia (Jones Ecological Research Center). Jones has longleaf pine but not necessarily in a Sandhill ecosystem. One issue is that detailed vegetation community maps of these two sites were not available at the time of the research, nor were independent ground truth points. The classification trees could be applied to these data sets, and would inappropriately predict one of the ten FNAI vegetation community classes from the OSBS data. Ground truth data at these sites would need to be collected to determine what new vegetation communities exist (new compared to OSBS) and select training locations to build a new decision tree. Independent ground truth data would also be required to perform an accuracy assessment.

Similarity/dissimilarity analysis tools could be developed to compare data from new areas and determine if the areas are similar to previous ones in terms of the spectra and structure as measured by the remote sensing instruments. The current decision tree could be applied to the new data where appropriate or in the case of dissimilar data, new training points and new output classes could be added to the decision tree. A generalized framework for extension to other sites would include the calculation of all of the metrics used here in this research, but would require site specific vegetation community output classes and additional training points for classification tree model generation. NEON AOP also has data available for various sites in the Northeast Domain 01 (2014), Appalachians and Cumberland Plateau Domain 07

(2015), Ozarks Complex Domain 08 (2015), Central Plains Domain 10 (2013), and Pacific Southwest Domain 17 (2013) covering a variety of different ecosystems.

Satellite remote sensing is a cost effective means to achieve global coverage and monitor change with time. The benefits of global and seasonal coverage are obtained by sacrificing spatial resolution, where most satellite based instruments with high spectral resolution have a pixel size of 30-60 m. Here, high spatial resolution pixels do not exist so the idea of an assemblage cannot be leveraged. The coarse pixels may or may not be able to capture the variability of vegetation chemistry, structure, and functioning, however, they will measure the mean values. In other cases, some commercial imaging satellites have high spatial resolution but only a limited number of spectral bands with coarse spectral resolution. Different biodiversity indicators may need to be used for satellite remote sensing versus airborne remote sensing due to what level of spatial and spectral variability is observable.

Two upcoming missions are proposing instruments to collect hyperspectral and lidar data from space (HyspIRI and GEDI). The Hyperspectral Infrared Imager (HyspIRI) is a proposed decadal study mission to collect hyperspectral imagery from low earth orbit (JPL, 2015c). The mission includes a visible to shortwave infrared (VSWIR) imaging spectrometer with a spectral range of 380-2500 nm and a 10 nm fwhm and also includes a thermal infrared (TIR) multispectral imager covering the 3-12 μm spectral range. The VSWIR instrument has a spatial resolution of 60 m at nadir and a revisit time of 19 days. A prime example of a space based lidar instrument to measure vegetation is the Global Ecosystem Dynamics Investigation (GEDI) Lidar (NASA, 2015). GEDI is an Earth System Science Pathfinder mission that proposes to place a lidar instrument on the International Space Station. The GEDI instrument will have three laser beams that are optically divided into 14 tracks on the ground (spaced 500 m apart) and covering a total swath width of 6.5 km. Each beam divergence is designed to illuminate a 25 m spot

on the ground. The return signal as a function of time is captured as a waveform where canopy height can be determined to an accuracy of 1 m. GEDI is due to be completed in 2018. One major limitation is that the two instruments will be on different platforms and will not have coincident ground coverage. The following paragraphs address how this research could be applied and/or adapted for use on data collected from space as well as looking at issues that arise from the coarse spatial resolution of the proposed space instruments. In addition, the airborne data may be used as validation for algorithms developed for satellite remote sensing.

The NEON AOP high spatial resolution airborne imaging spectrometer collects 1 m ground sample distance pixels at a nominal flight altitude. Analysis of a 20 m x 20 m ground plot includes 400 pixels that are used for the metric/feature calculations. NASA spaceborne multispectral and hyperspectral instruments typically have pixels with a ground sample distance of 30 m or larger (HyspIRI will be 60 m). The at-sensor radiance values of all ground objects that fall within that pixel are averaged together into a single radiance value. Any spatially explicit information about where the radiance came from is lost. A major complication is that the scaling of radiances is linear (assuming spatial response uniformity within a pixel for this example) but feature layers such as vegetation indices scale in a non-linear fashion depending on the mix of surface objects within the pixel (Wessman & Bateson, 2004).

In terms of the hyperspectral feature categories, some are already compatible with a 60 m pixel, some calculations could be modified, and a few are lost due to the large pixel size that cannot detect ground objects smaller than that scale. The radiance of a 60 m pixel equals the mean of radiance of 1 m pixels within a 60 m cell so this summary statistic exists but stdev cannot be calculated without the high resolution data. Mean and stdev of the first derivative of reflectance is lost but the first derivative of mean reflectance could be calculated as an alternative feature (but these two are not equal). For this research,

the normalized vegetation indices and triangle metrics are calculated on the spatial mean (the entire 20 m x 20 m ground plot) of reflectance and are applicable to larger resolution instruments. Finally, endmember analysis techniques (Roberts et al., 1998) could be applied to calculate "fraction of" metrics (like fraction of vegetation) but actual vegetation spectra cannot be separated for use in feature calculations. About half of the splitting features used early in the decision tree are not possible with the coarse pixel size. These might be easily replaced by other features that split the data in the same manner as the features used. Effectively, the coarse resolution of spaceborne hyperspectral systems makes it more difficult to analyze the data as an assemblage of points (depending on how well endmember analysis works), but does not prevent the use of decision trees to map vegetation communities.

Commercial imaging satellites can collect high spatial resolution data but at a few wide spectral bands. WorldView-2 has 8 multispectral bands covering 400-1040 nm with a nadir pixel ground sample distance of 1.85 m and pan-sharpened products can be generated at 0.5 m (DigitalGlobe, 2015a). WorldView-3 has increased spectral coverage by including the same 8 bands as WorldView-2 and adding 8 additional bands in the shortwave-infrared (SWIR) covering 1195-2365 nm (DigitalGlobe, 2015a). The multispectral bands have a ground resolution of 1.24 m at nadir and the SWIR bands have a resolution of 3.7 m. With these instruments, the spectral bandwidth is greater than 30 nm which is significantly wider than the NEON imaging spectrometer. A series of input features could be generated taking advantage of the high spatial resolution pixels at the 20 m x 20 m ground plot scale, albeit using fewer spectral bands (8 or 16 instead of 32). The accuracy of classification models derived from WorldView data cannot be assumed as the wider band widths may reduce the effectiveness of certain spectral bands to discriminate the spectral differences between vegetation communities.

The NEON AOP high spatial resolution airborne lidar collects approximately 3.82 points per meter^2 at a nominal flight altitude with a 0.8 m spot on a 0.5 m point spacing. This comes out to 1528 return laser pulses that are analyzed as part of a 20 m x 20 m ground plot. In the case of spaceborne lidar instruments, the divergence angle creates a 25 m (ex: GEDI) or larger spot. Assuming the ability to save full waveform data with a fine vertical resolution (say 1 m bins), the returned laser energy from all ground objects located within the spot is grouped together into a single return waveform. Any spatially explicit information about where the return energy came from is lost. Another issue is that the waveform shape will be smeared out vertically if any terrain exists or the data are collected at an off-nadir angle. This will cause return energy from bare-Earth and ground vegetation or shrubs to blend together and making those difficult to detect as individual structural layers.

In terms of the lidar feature categories, calculations could be modified to analyze the single return waveform: summary statistics can be calculated on waveform energy directly (for example, stdev of ground could be the full-width at half maximum of the potion of the waveform identified as ground energy); point classification can be performed based on analyzing the fraction of waveform energy at various heights; percentile height, percentile height range, percentile height range fraction, and normalized height difference index can be performed based on analyzing the fraction of waveform energy at versus height; and point distribution calculations could be modified to be applied to the waveform vertical energy distribution.

One issue is that the spaceborne lidar data cannot be evaluated to describe the top of the surface (the first returns in the discrete lidar data) since the waveform energy is not known in a spatially explicit manner. Another issue is that a high spatial resolution bare-Earth terrain map cannot be generated. Slope could be calculated at a more coarse resolution but the calculation of the variability of the canopy height

model at high resolution and calculation of surface roughness (such as rugosity) are not possible. A final issue is that spaceborne lidar instruments typically cannot maintain a contiguous point spacing on the ground leaving large gaps on the Earth's surface where vegetation structure isn't measured resulting in maps that either have gaps or require interpolation. A few of the splitting features used early in the decision tree are not possible with the coarse spot size such as std1st, p951st, and FracGroundCHM. These might be easily replaced by other features that split the data in the same manner as the features used. Effectively, the coarse resolution of spaceborne lidar systems makes it more difficult to analyze the data as an assemblage of points, but does not prevent the use of decision trees to map vegetation communities.

Future Work

The methods utilized in this research achieve a 91.5% classification accuracy for all k-folds training points, a 77.9% accuracy for all k-folds validation points, and 66% accuracy for all of the independent ground validation points. A 90% classification accuracy is achieved for Sandhill training points and a 67% accuracy for Sandhill independent ground validation points. While these results are promising, more could be done to improve the overall classification accuracy and reduce uncertainty in the output vegetation classes. During the course of the research, many ideas were sparked to improve or extend the methods used and some advanced concepts were conceived that will require further research and development. Future research tasks cover the following themes: improvements to current methodology (improved classification algorithm training and validation and new or improved features); investigation of scale; extendibility (extendibility to other sites and extendibility to space remote sensing missions); concept of assemblages of pixels or points; alternative classification techniques; advanced

analysis (vegetation diversity models and similarity/dissimilarity analysis); tool development; direct remote sensing of vegetation diversity; and process studies and long-term change.

Improved classification algorithm training and validation:

- Get the updated vegetation GIS layers from Ordway-Swisher Biological Station and double check that training points are labeled with the correct vegetation community
- Identify and fix any errors in ground training or validation coordinates
- Use more training points and make sure to adequately sample burn areas (especially in the Sandhill vegetation community)
- Acquire more independent ground truth validation points, preferably from the NEON vegetation structure ground plots that will have canopy species lists and abundances species to determine the dominant species with respect to vegetation community

New or improved data features:

- Perform a more detailed study to quantitatively determine the optimum number and wavelength of spectral bands needed to capture the spectral shape of the ground targets of interest
- Generate new features that describe or characterize the unique spectral shapes and curvatures of endmembers or vegetation communities
- Revisit the endmember analysis on hyperspectral pixels. What are the most important high level classes (sand, shadow, water, veg, etc.)?
- Calculate normalized difference indices on the 1m hyperspectral pixels, then calculate plot summary statistics

- Derive chemical maps from the hyperspectral data and use to calculate new input features (for example: chlorophyll, nitrogen, effective water thickness, lignin, cellulose, etc.)
- Apply a Fourier Transform to the lidar canopy height model data to see if there is a detectable spatial frequency in the case of Pine Plantations with regularly spaced trees
- Clumping index using range to 1st return (M. van Leeuwen & J. van Aardt, personal communication, August 27, 2014)
- Use lidar intensity values
- Conduct a future site survey using multi-wavelength lidar systems
- Investigate the misclassified cases of similar vegetation communities. Look to see if there are differences in the data that could be captured with new features (for instance Sandhill versus Pine Plantation)
- Perform data fusion earlier, for instance using lidar data to identify categories of vegetation in the hyperspectral data and then calculate features on those pixels separately (example: canopy vegetation versus ground vegetation)

Investigation of scale:

- More testing of spatial scale. Repeat classification with 5 m, 10 m, 30 m, 40 m, and 60 m ground plots to see if results improve
- Analysis could be strengthened with similarity/dissimilarity tools described in the advanced analysis section below

Extendibility to other sites:

- Validate current algorithm in other portions of Florida with Sandhill communities

- Include data from other sites with different FNAI vegetation communities to train the classification tree to identify more FNAI classes
- Generate classification trees for other regions of the U.S. using the current input features but with region specific vegetation classes
- Develop an analysis capability that could determine the validity of the current output classes of the classification tree on new data (i.e. the analysis would show if the remote sensing data at a new location was representative of any of the training data, and if not, it would produce a low certainty quality flag). This could be strengthened with similarity/dissimilarity tools described in the advanced analysis section below

Extendibility to space remote sensing missions:

- Run current classification tree using appropriate spatial scales and the input features that could be calculated on data from current or near-future space remote sensing missions
- Investigate endmember analysis and spectral unmixing techniques

Concept of assemblages of pixels or points:

- Refine the concept of assemblages of remote sensing data pixels or points
- Investigate the best way to use the collection of 1 m pixels in a 20 m x 20 m plot scale
- What metrics should come out of this?
- Cluster pixels into endmembers. Analyze distributions of the endmembers to relate to vegetation communities

- Use histogram bins of a parameter (such as reflectance, normalized difference indices, chlorophyll levels, or other chemical traits) and/or percentile metrics (ex. 50^{th} percentile of chlorophyll)

Alternative classification techniques:

- Investigate hierarchical classification techniques
- Try other machine learning classification algorithms: random forests, support vector machines, neural networks, etc.
- Incorporate GIS analysis techniques for contextual analysis (ex: distance from lake as a feature layer)
- Incorporate object based image analysis techniques for contextual analysis (ex: compare data to spatial neighbors and develop rules to modify the output class under different circumstances)

Vegetation diversity models:

- Apply biodiversity theory to remote sensing data pixel diversity
- Develop a remote sensing pixel diversity index (spectral and structural)
- Investigate spatial heterogeneity and relate to a remote sensing diversity index
- Characterize the intra- and inter-specific diversity in the remote sensing data for different vegetation communities
- Build a regression model of variability in remote sensing vegetation parameters vs ground truth data of species richness

Similarity/dissimilarity analysis:

- Investigate similarity/dissimilarity calculations
- Investigate spectral and structural data clustering
- For each input feature layer, look at the inter- and intraspecific variability across FNAI communities. Identify features that show the most promise for separating data
- Investigate biotic similarity (Gotelli & Chao, 2013) as an alternative classification technique (ex. Horn overlap (Horn, 1966) or Morisita-Horn (Morisita, 1959) similarity measure)

Tool development:

- Develop a tool for rapid assessment of biodiversity
- Improve species richness map based on spectral variability hypothesis (Palmer et al., 2000; Palmer et al., 2002) and diversity regression models
- Investigate other local scale biodiversity indicators (besides species richness) such as ecosystem extent, fragmentation, condition, etc.
- Develop tools to inform ground sampling strategy
- Develop tools to identify areas of interest which could be unusual habitats, such as places where biodiversity is higher or where different or rare species may be found

Direct remote sensing of vegetation diversity:

- Identify which species dominate the canopies or visible portions of the ground
- Investigate remote sensing algorithms to classify and map individual canopy species
- Use data from TOS structure plots as ground validation of species maps

- Examine the data closer to identify differences between similar species and communities (can remote sensing data observe spectral and structural differences between similar species such as sand live oak vs live oak or a longleaf pine vs loblolly or slash pine?)

Process studies and long-term change:

- Investigate the relationship between ecosystem structure and functioning and diversity. Combine NEON AOP with TOS data for advanced analysis
- Monitor changes in vegetation community maps with time as NEON AOP re-flies sites (the 2014 data serves as a baseline for long-term change)
- Spatially map land use and land cover and analyze LULC change with time
- Link land change theory to spaced based imagery for regional or continental analysis

BIBLIOGRAPHY

AIBS (American Institute of Biological Sciences) (2004). Biodiversity, species composition, and ecosystem functioning: Report from a NEON science workshop. Washington, DC: AIBS.

America's Longleaf (2009). Range-Wide Conservation Plan for Longleaf Pine

America's Longleaf (2015). Retrieved October 1, 2015, from http://www.americaslongleaf.org/

Applanix (2015). Retrieved October 1, 2015, from http://www.applanix.com/products/airborne/pospac-mms.html

Asner, G.P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment, 64*, 234-253.

Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Jones, M.O., Martin, R.E., Boardman, J., & Field, C.B. (2007). Carnegie Airborne Observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging (wLiDAR) for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing, 1*(013536), 1-21. Doi: 10.1117/1.2794018

Asner, G.P., Hughes, R.F., Vitousek, P.M., Knapp, D.E., Kennedy-Bowdoin, T., Boardman, J., Martin, R.E., Eastwood, M., & Green, R.O. (2008). Invasive plants transform the three-dimensional structure of rain forests. *Proceedings of the National Academy of Sciences, 105*(11), 4519-4523. Doi: 10.1073/pnas.0710811105

Asner, G.P. & Martin, R.E. (2009). Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment, 2009*, 7(5), 269–276. Doi: 10.1890/070152

Asner, G.P., Martin, R.E., Ford, A.J., Metcalfe, D.J., & Liddell, M.J. (2009). Leaf chemical and spectral diversity in Australian tropical forests. *Ecological Applications, 19*(1), 236–253.

Asner, G.P., Knapp, D.E., Boardman, J., Green, R.O., Kennedy-Bowdoin, T., Eastwood, M., Martin, R.E., Anderson, C., & Field, C.B. (2012). Carnegie Airborne Observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sensing of Environment, 124*, 454-465. Doi: 10.1016/j.rse.2012.06.012

Asner, G.P. (2013). Biological diversity mapping comes of age. *Remote Sensing, 5*, 374-376. Doi: 10.3390/rs5010374

Asner, G.P. (2015). Organismic remote sensing for tropical forest ecology and conservation. *Annals of the Missouri Botanical Garden, 100*(3), 127-140. Doi: 10.3417/2012016

Asner, G.P., Martin, R.E., Anderson, C.B., & Knapp, D.E. (2015). Quantifying forest canopy traits: Imaging spectroscopy versus field survey. *Remote Sensing of Environment 158*, 15–27. http://dx.doi.org/10.1016/j.rse.2014.11.011

Barnett, D. (2014). TOS science design for plant diversity. *NEON Doc. # 000912 Rev A.*

Barnett, D., Duffy, P., Schimel, D., Azuaje, E., Krauss, R., Thorpe, A., Gudex-Cross, D., McCorkel, J. (2015). TOS science design for spatial sampling design. *NEON Doc. # 000913 Rev A.*

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees*. Belmont, CA: Chapman and Hall/CRC.

Carlson, K.M., Asner, G.P., Hughes, R.F., Ostertag, R., & Martin, R.E. (2007). Hyperspectral remote sensing of canopy biodiversity in Hawaiian lowland rainforests. *Ecosystems, 10*, 536–549. Doi: 10.1007/s10021-007-9041-z

CBD (Convention on Biological Diversity) (2003). Monitoring and indicators: Designing national-level monitoring programmes and indicators. Ninth meeting. Montreal, November 10–14. Item 5.3 of the provisional agenda. Subsidiary Body on Scientific, Technical and Technological Advice, Secretariat of the Convention on Biological Diversity, Montreal, Québec, Canada. UNEP/CBD/SBSTTA/9/10. Retrieved September 17, 2015, from http://www.cbd.int/doc/meetings/sbstta/sbstta-09/official/sbstta-09-10-en.pdf

CBD (2015a). Retrieved September 26, 2015, from https://www.cbd.int/sp/targets/

CBD (2015b). Retrieved September 26, 2015, from https://www.cbd.int/convention/

CBD (2015c). Retrieved September 26, 2015, from https://www.cbd.int/2011-2020/

Colwell, R.K. (2009). Biodiversity: concepts, patterns, and measurement. In S.A. Levin (Ed.), *The Princeton guide to ecology* (pp. 257-263). Princeton, NJ: Princeton University Press.

Comer, P., Faber-Langendoen, D., Evans, R., Gawler, S., Josse, C., Kittel, G., Menard, S., Pyne, M., Reid, M., Schulz, K., Snow, K., & Teague, J. (2003). Ecological systems of the United States: A working classification of U.S. terrestrial systems. Arlington, Virginia: NatureServe.

Cook, D.B. & Corp, L.A. (2012). G-LIHT: Goddard's LiDAR, Hyperspectral, and Thermal Airborne Imager. Retrieved September 17, 2015, from ftp://fusionftp.gsfc.nasa.gov/multimedia/docs/G-LiHT_White_Paper.docx

Curran, P.J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment, 30,* 271-278.

CVS (Carolina Vegetation Survey) (2015). Retrieved September 17, 2015, from http://cvs.bio.unc.edu/

Dahlin, K.M., Asner, G.P., & Field, C.B. (2013). Environmental and community controls on plant canopy chemistry in a Mediterranean-type ecosystem. *Proceedings of the National Academy of Sciences, 110*(17), 6895-6900. Doi: 10.1073/pnas.1215513110

DigitalGlobe (2015a). WorldView-2 data sheet. Retrieved November 22, 2015, from https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/98/DG_WorldView2_DS_PROD.pdf

DigitalGlobe (2015b). WorldView-3 data sheet. Retrieved November 22, 2015, from https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/95/DG_WorldView3_DS_forWeb_0.pdf

Eisenberg J.F. & Franz, R. (1995). Forward. In J.F. Eisenberg & R. Franz (Eds.), *Bulletin of the Florida Museum of Natural History Volume 38, Pts I. & II., Nos. 1-9, Natural history of the Katharine Ordway Preserve-Swisher Memorial Sanctuary, Putnam County, Florida* (pp. i-v). Gainesville, FL: University of Florida.

Elmendorf S. and Jones, K. (2015). TOS science design for plant phenology. *NEON Doc. # 000907 Rev. A.*

Fancy, S. G., Gross, J.E., & Carter, S.L. (2009). Monitoring the condition of natural resources in US national parks. *Environmental Monitoring and Assessment, 151*, 161-174. Doi: 10.1007/s10661-008-0257-y

Fauth, J.E., Bernardo, J., Camara, M., Resetarits, Jr., W.J., Van Buskirk, J. & McCollum, S.A. (1996). Simplifying the jargon of community ecology: a conceptual approach. *American Naturalist, 147*(2), 282-286.

FDACS (Florida Department of Agriculture and Consumer Services) (2010). Forest Resources – 2010 Florida's Statewide Strategies.

FGDC (Federal Geographic Data Committee) (2008). National Vegetation Classification Standard, Version 2. Vegetation Subcommittee, Federal Geographic Data Committee. Retrieved September 29, 2015, from http://www.fgdc.gov/standards/projects/FGDC-standards-projects/vegetation/NVCS_V2_FINAL_2008-02.pdf

FNAI (Florida Natural Areas Inventory) (1990). Guide to the natural communities of Florida. Florida Natural Areas Inventory and Florida Department of Natural Resources, Tallahassee, FL.

FNAI (2010). Guide to the natural communities of Florida: 2010 edition. Florida Natural Areas Inventory, Tallahassee, FL.

FNAI (2015). About FNAI. Retrieved August 9, 2015, from http://www.fnai.org/about.cfm

Frost, C. (2006). History and future of the longleaf pine ecosystem. In: S. Jose, E.J. Jokela, and D.L. Miller (Eds.), *The longleaf ecosystem: ecology, silviculture and restoration* (pp. 9-48). New York: Springer Science.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J. (2011). Completion of the 2006 National Land Cover Database for the Conterminous United States, *PE&RS, 77*(9), 858-864.

FWS (Fish and Wildlife Service) (2015a). South Florida multi-species recovery plan: High pine. Retrieved September 25, 2015, from, http://www.fws.gov/verobeach/MSRPPDFs/HighPine.pdf

FWS (2015b). Gopher Tortoise fact sheet. Retrieved September 25, 2015, from http://www.fws.gov/northflorida/gophertortoise/gopher_tortoise_fact_sheet.html

Gallery, W. & Leisso, N. (2014). NEON algorithm theoretical basis document: NEON Imaging Spectrometer (NIS) level 1B calibrated radiance. *NEON Doc. # 001210 Rev A.*

Gamon, J.A., Peñulas, J., & Field, C.B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sensing of Environment, 41, 35-44.

Gamon, J.A., & Surfus, J.S. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist, 143*, 105-117.

GEOBON (2015). Retrieved September 29, 2015, from http://geobon.org/

Gotelli, N.J. (2008). *A primer of ecology* (4th ed.). Massachusetts: Sinauer Associates, Inc.

Gotelli, N.J. & Anne, C. (2013). Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In: S.A. Levin (Ed.), *Encyclopedia of biodiversity*, (2nd ed.), Volume 5 (pp. 195-211). Waltham, MA: Academic Press.

Hargrove, W.W., & Hoffman, F.M. (1999). Using multivariate clustering to characterize ecoregion borders. *Computing in Science & Engineering, 1*(4), 18-25.

Hargrove, W.W., & Hoffman, F.M. (2005). Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental Management, 34*, S39-S60. Doi: 10.1007/s00267-003-1084-0

Harper, J.L. & Hawksworth, D.L. (1995). Preface. In D.L. Hawksworth (Ed.), *Biodiveristy: measurement and estimation* (pp. 5-12). London: Chapman & Hall.

Hinckley, E.S. (2014). TOS science design for terrestrial biogeochemistry. NEON Doc. # 000906 Rev A.

Horn, H.S. (1966). Measurement of overlap in comparative ecological studies. *American Naturalist, 100*, 419–424.

Horning, N., Robinson, J.A., Sterling, E.J., Turner, W., & Spector, S. (2010). *Remote sensing for ecology and conservation: A handbook of techniques*. New York, NY: Oxford University Press.

JPL (2015a). Retrieved September 13, 2015, from http://aso.jpl.nasa.gov/

JPL (2015b). Retrieved September 13, 2015, from http://aviris.jpl.nasa.gov/

Kampe, T.U., Johnson, B.R., Kuester, M., & Keller, M. (2010). NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *Journal of Applied Remote Sensing, 4*(043510), 1-24. DOI: 10.1117/1.3361375

Kampe, T.U., McCorkel, J.M., Hamlin, L., Green, R.O., Krause, K.S., and Johnson, B.J. (2011). Progress in the development of airborne remote sensing instrumentation for the National Ecological Observatory Network. *Proc. SPIE, 8156*(81560A), 1-12. Doi: 10.1117/12.892572

Kampe, T. (2014). NEON imaging spectrometer geolocation processing algorithm theoretical basis document. *NEON Doc. # 001290 Rev A.*

Karpowicz, B. & Kampe, T. (2015). NEON imaging spectrometer radiance to reflectance algorithm theoretical document. *NEON Doc. # 001288 Rev A.*

Kao, R.H., Gibson, C.M., Gallery, R.E., Meier, C.L., Barnett, D.T., Docherty, K.M., Blevins, K.K., Travers, P.D., Azuaje, E., Springer, Y.P., Thibault, K.M., McKenzie, V.J., Keller, M.A., Luciana, F., Hinckley, E.S., Parnell, J., Schimel, D. (2012). NEON terrestrial field observations: designing continental-scale, standardized sampling. Ecosphere, 3(12), article 115, 1-17. Doi: 10.1890/ES12-00196.1

Keller, M., Schimel, D.S., Hargrove, W.W., & Hoffman, F.M. (2008). A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment, 6*, 282-284.

Kokaly, R.F., Asner, G.P., Ollinger, S.V., Martin, M.E., & Wessman, C.A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment, 113*, S78–S91. Doi: 10.1016/j.rse.2008.10.018

Krause, K. & Goulden, T. (2015). NEON L0-to-L1 discrete return lidar algorithm theoretical basis document. *NEON Doc. # 001292 Rev A.*

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar remote sensing for ecosystem studies. *BioScience, 52*(1), 19-30.

Loucks, C. & Leimgruber, P. (2007). Trends in species populations. In H. Strand, R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, & W. Turner (Eds.), *Sourcebook on remote sensing and biodiversity indicators: Technical series no. 32* (pp. 121-130). Montreal: Secretariat of the Convention on Biological Diversity.

LPEGDB (Longleaf Pine Ecosystem Geodatabase) (2014a). Longleaf Pine Ecosystem Geodatabase v.1 Final Report June 2014

LPEGDB (2014b). Version 2 Summary Report Supplement to the Longleaf Pine Ecosystem Geodatabase v.1 Final Report October 2014

Maestre, F.T., Quero, J.L., Gotelli, N.J., Escudero, A., Ochoa, V., Delgado-Baquerizo, M., Garcia-Gomez, M., Bowker, M.A., Soliveres, S., Escolar, C., Garcia-Palacios, P., Berdugo, M., Valencia, E., Gozalo, B., Gallardo, A., Aguilera, L., Arredondo, T., Blones, J., Boeken, B., Bran, D., Conceicao, A.A., Cabrera, O., Chaieb, M., Derak, M., Eldridge, D.J., Espinosa, C.I., Florentino, A., Gaitan, J., Gatica, M.G., Ghiloufi, W., Gomez-Gonzalez, S., Gutierrez, J.R., Hernandez, R.M., Huang, X.W., Huber-Sannwald, E., Jankju, M., Miriti, M., Monerris, J., Mau, R.L., Morici, E., Naseri, K., Ospina, A., Polo, V., Prina, A., Pucheta, E., Ramirez-Collantes, D.A., Romao, R., Tighe, M., Torres-Diaz, C., Val, J., Veiga, J.P., Wang, D.L., & Zaady, E. (2012). Plant species richness and ecosystem multifunctionality in global drylands. *Science, 335*, 214-218. Doi: 10.1126/science.1215442

Magurran, A.E. (2004). *Measuring biological diversity*. Malaysia: Blackwell Publishing.

Martin, M.E., Plourde, L.C., Ollinger, S.V., Smith, M.-L., & McNeil, B.E. (2008). A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. *Remote Sensing of Environment, 112*, 3511-3519. Doi: 10.1016/j.rse.2008.04.008

Meier, C. & Jones, K. (2014). TOS science design for plant biomass, productivity, and leaf area index. NEON Doc. # 000914 Rev A.

Morisita, M. (1959). Measuring of interspecific association and similarity between communities. *Memoires of the Faculty of Science, Kyushu University, Series E (Biology), 3*, 65–80.

MRLC (Multi-Resolution Land Characteristics Consortium) (2015). Retrieved September 29, 2015, from, http://www.mrlc.gov/nlcd2011.php

NCALM (National Center for Airborne Laser Mapping) (2015). Retrieved October 1, 2015, from http://ncalm.cive.uh.edu/

NEON (National Ecological Observatory Network) (2015a). Retrieved September 12, 2015, from http://www.neoninc.org

NEON (2015b). Retrieved September 12, 2015, from http://www.neoninc.org/science-design/collection-methods

NEON (2015c). Retrieved September 12, 2015, from http://www.neoninc.org/science-design/collection-methods/flux-tower-measurements

NEON (2015d). Retrieved September 12, 2015, from http://www.neoninc.org/science-design/collection-methods/soil-sensors-measurements

NEON (2015e). Retrieved September 12, 2015, from http://www.neoninc.org/science-design/collection-methods/terrestrial-organismal-sampling

NEON (2015f). Retrieved September 12, 2015, from http://www.neoninc.org/science-design/collection-methods/aquatic-sensor-measurements

NEON (2015g). Retrieved September 12, 2015, from http://www.neoninc.org/science-design/collection-methods/aquatic-observations

NEON (2015h). Retrieved September 12, 2015, from http://www.neoninc.org/science-design/collection-methods/airborne-remote-sensing

NEON (2015i). Retrieved September 12, 2015, from http://www.neoninc.org/data-resources/data-themes

NEON (2015j). Retrieved September 12, 2015, from http://www.neoninc.org/sites/default/files/basic-page-files/NEON_DOC_002652.pdf

NEON (2015k). Retrieved August 9, 2015, from http://www.neoninc.org/science-design/field-sites/list

NOAA (2015). Retrieved September 26, 2015, from http://geodesy.noaa.gov/CORS/

Norse, E.A. & Ecological Society of America & Wilderness Society (U.S.) (1986). *Conserving biological diversity in our national forests*. Washington, DC: The Wilderness Society.

NRC (National Research Council) (2001). *Grand Challenges in Environmental Sciences.* Washington, DC: National Academies Press.

NRC (2003). *NEON: Addressing the nation's environmental challenges.* Washington, DC: National Academies Press.

NSF (National Science Foundation) (1998). Final Report: Biodiversity Monitoring Workshop, held September 10-11 at the Blandy Experimental Farm, University of Virginia, VA: National Science Foundation.

NSF (1999a). Report of the Second Workshop on the Biodiversity Observation Network, held January 14-16 at the National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA. Arlington, VA: National Science Foundation.

NSF (1999b). Report of the Third Workshop on the Biodiversity Observation Network, held May 6-7 at the California Academy of Sciences, San Francisco, CA, VA: National Science Foundation.

NSF (1999c). Report of the Fourth Workshop on the Biodiversity Observation Network (BON), held August 27-29 at the National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA. Arlington, VA: National Science Foundation.

NSF (2000a). Report on First Workshop on the National Ecological Observatory Network, held January 10-12 at Archbold Biological Station, Lake Placid, FL. Arlington, VA: National Science Foundation.

NSF (2000b). Report to National Science Foundation from the Second Workshop on the Development of a National Ecological Observatory Network (NEON), held on March 9-13 at San Diego Supercomputer Center, La Jolla, CA. Arlington, VA: National Science Foundation.

NSF (2000c). Report to National Science Foundation from the Third Workshop on the Development of a National Ecological Observatory Network (NEON), held on May 3-4, 2000 at Sante Fe Institute, Santa Fe, NM. Arlington, VA: National Science Foundation.

NSF (2002a). Report to the National Science Foundation from the Fourth Workshop on the Development of a National Ecological Observatory Network (NEON): Standard Measurements and Infrastructure Needs, held on June 4-5, 2002 at the Millenium Hotel, Boulder, CO. Arlington, VA: National Science Foundation.

NSF (2002b). Final Report on NEON-V: CRIPTON Workshop, held on June 14-16, 2002 at Field Museum of Natural History, Chicago, IL. Arlington, VA: National Science Foundation.

NSF (2002c). Report to the National Science Foundation from the Sixth Workshop of a National Ecological Observatory Network (NEON): Information Management, held on September 16-18 at the National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA. Arlington, VA: National Science Foundation.

Ollinger, S.V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist, 189*, 375–394. doi: 10.1111/j.1469-8137.2010.03536.x

Optech (2010a). Gemini Summary Specification Sheet Version 100908

Optech (2010b). ALTM + DiMAC Summary Specification Sheet Version 100909

OSBS (Ordway-Swisher Biological Station) (2011). The Sandhill: The Newsletter of the Ordway-Swisher Biological Station. Ordway-Swisher Biological Station, University of Florida IFAS, Issue 1. Summer 2011.

OSBS (2013). 2013 UF/IFAS Ordway-Swisher Biological Station Annual Report. Retrieved August 9, 2015, from http://ordway-swisher.ufl.edu/forms/13_OSBS_AR.pdf

OSBS (2014). 2014 UF/IFAS Ordway-Swisher Biological Station Annual Report. Retrieved August 9, 2015, from http://ordway-swisher.ufl.edu/forms/14_OSBS_AR.pdf

OSBS (2015a). About the Station. Retrieved August 9, 2015, from http://ordway-swisher.ufl.edu/about.htm

OSBS, (2015b). Water Resources. Retrieved August 9, 2015, from http://ordway-swisher.ufl.edu/water.htm

OSBS (2015c). Geospatial Information. Retrieved August 9, 2015, from http://ordway-swisher.ufl.edu/geo.htm

OSBS (2015d). Plant Communities. Retrieved August 9, 2015, from http://ordway-swisher.ufl.edu/communities/index.htm

Oswalt, C.M., Cooper, J.A., Brockway, D.G., Brooks, H.W., Walker, J.L., Connor, K.F., Oswalt, S.N., & Conner, R.C. (2012). History and current condition of Longleaf Pine in the Southern United States. *Gen. Tech. Rep. SRS–166*. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station.

Palmer, M.W., Wohlgemuth, T., Earls, P., Arévalo, J.R., & Thompson, S.D. (2000). Opportunities for long-term ecological research at the Tallgrass Prairie Preserve, Oklahoma. In: K. Lajtha & K. Vanderbilt (Eds.), *Cooperation in Long Term Ecological Research in Central and Eastern Europe: Proceedings of ILTER Regional Workshop* (pp. 123–128). Budapest, Hungary.

Palmer, M.W., Earls, P., Hoagland, B.W., White, P.S., & Wohlgemuth, T. (2002). Quantitative tools for perfecting species lists. *Environmetrics, 13*, 121–137. Doi: 10.1002/env.516

Peet, R.K., Wentworth, T.R., & White, P.S. (1998). A flexible, multipurpose method for recording vegetation composition and structure. Castanea, 63(3), 262-274.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., & Wegmann, M. (2013). Essential biodiversity variables, *Science, 339*, 277-278. Doi: 10.1126/science.1229931

Rapidlasso (2015). LAStools. Retrieved October 1, 2015, from http://rapidlasso.com/lastools/

R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved May 2, 2015, from URL http://www.R-project.org/

ReSe (2015). Retrieved October 1, 2015, from http://www.atcor.com/software/atcor/index.html

Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R.O. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment, 65*, 267–279.

Rocchini, D. (2015). Earth observation for ecosystems monitoring in space and time: A special issue in Remote Sensing. *Remote Sensing, 7*, 8102-8106. Doi: 10.3390/rs70608102

Rose, R.A., Byler, D., Eastman, J.R., Fleishman, E., Geller, G., Goetz, S., Guild, L., Hamilton, H., Hansen, M., Headley, R., Hewson, J., Horning, N., Kaplin, B.A., Laporte, N., Leidner, A., Leimgruber, P., Morisette, J., Musinsky, J., Pintea, L., Prados, A., Radeloff, V.C., Rowen, M., Saatchi, S., Schill, S., Tabor, K., Turner, W., Vodacek, A., Vogelmann, J., Wegmann, M., Wilkie, D., & Cara Wilson, C. (2015). Ten ways remote sensing can contribute to conservation. *Conservation Biology, 29*(2), 350:359. Doi: 10.1111/cobi.12397

Schaepman, M.E., Itten, K.I, & the APEX Team (2003). APEX: Current status of the airborne dispersive pushbroom imaging spectrometer. *Proc. SPIE, 5234*, 202-210. Doi: 10.1117/12.513745

Schaepman, M.E., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., Schneider, F.D., Laurent, V., Popp, C., Seidel, F.C., Lenhard, K., Gege, P., Küchler, C., Brazile, J., Kohler, P., De Vos, L., Meuleman, K., Meynart, R., Schläpfer, D., Kneubühler, M., & Itten, K.I. (2015). Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX). *Remote Sensing of Environment, 158*, 207-219. Doi: 10.1016/j.rse.2014.11.014

Schimel, D., Keller, M., Berukoff, S., Kao, R., Loescher, H., Powell, H., Kampe, T., Moore, D., Gram, W., Barnett, D., Gallery, R., Gibson, C., Goodman, K., Meier, C., Parker, S., Pitelka, L., Springer, Y., Thibault, K., & Utz, R. (2011). *The National Ecological Observatory Network 2011 science strategy: Enabling continental-scale ecological forecasting.* Retrieved September 13, 2015, from http://www.neoninc.org/sites/default/files/basic-page-files/NEON_Strategy_2011u2_0.pdf

Schimel, D.S., Asner, G.P., & Moorcroft, P. (2013). Observing changing ecological diversity in the Anthropocene, *Frontiers in Ecology and the Environment, 11*(3), 129-137. Doi: 10.1890/120111

SFRC (School of Forest Resources and Conservation) (2015). Austin Cary Forest. Retrieved October 1, 2015, from http://sfrc.ufl.edu/forest/about/austincary/

Smith, L.L. (1995). Nesting ecology, female home range and activity, and population size-class structure of the Gopher Tortoise, *Gopherus Polyphemus*, on the Katharine Ordway Preserve, Putnum County, Florida. In J.F. Eisenberg & R. Franz (Eds.), *Bulletin of the Florida Museum of Natural History Volume 38, Pts I. & II., Nos. 1-9, Natural history of the Katharine Ordway Preserve-Swisher Memorial Sanctuary, Putnam County, Florida* (pp. 97-126). Gainesville, FL: University of Florida.

Steininger, M. and Horning, N. (2007). The basics of remote sensing. In H. Strand, R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, & W. Turner (Eds.), *Sourcebook on remote sensing and biodiversity indicators: Technical series no. 32* (pp. 23-33). Montreal: Secretariat of the Convention on Biological Diversity.

Stilla, U. & Jutzi, B. (2009). Waveform analysis for small-footprint pulsed laser systems. In J. Shan & C.K. Toth (Eds.), *Topographic laser ranging and scanning: Principles and processing* (pp. 215-234). Boca Raton, FL: CRC Press Taylor & Francis Group.

Strand, H., Höft, R., Strittholt, J., Miles, L., Horning, N., Fosnight, E., & Turner, W. (Eds.). (2007a). *Sourcebook on Remote Sensing and Biodiversity Indicators: Technical Series no. 32*. Montreal: Secretariat of the Convention on Biological Diversity.

Strand, H., Fosnight, E., Herkenrath, P., & Höft, R. (2007b). Remote sensing and monitoring for the Convention on Biological Diversity. In H. Strand, R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, & W. Turner (Eds.), *Sourcebook on remote sensing and biodiversity indicators: Technical series no. 32* (pp. 9-21). Montreal: Secretariat of the Convention on Biological Diversity.

Strand H., Leimgruber, P., & Mueller, T. (2007c). Trends in selected biomes, habitats, and ecosystems: Dry and sub-humid lands. In H. Strand, R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, & W. Turner (Eds.), *Sourcebook on remote sensing and biodiversity indicators: Technical series no. 32* (pp. 65-84). Montreal: Secretariat of the Convention on Biological Diversity.

Strittholt, J. (2007). Coverage of protected areas. In H. Strand, R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, & W. Turner (Eds.), *Sourcebook on remote sensing and biodiversity indicators: Technical series no. 32* (pp. 131-141). Montreal: Secretariat of the Convention on Biological Diversity.

Strittholt J. & and Steininger, M. (2007). Trends in selected biomes, habitats, and ecosystems: Forests. In H. Strand, R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, & W. Turner (Eds.), *Sourcebook on remote sensing and biodiversity indicators: Technical series no. 32* (pp. 35-63). Montreal: Secretariat of the Convention on Biological Diversity.

Theobald, D. M., Stevens, D.L. Jr., White, D., Urquhart, N.S., Olsen, A.R., & Norman. J.B. (2007). Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environmental Management, 40,* 134-146. DOI 10.1007/s00267-005-0199-x

Therneau, T., Atkinson, B., & Ripley, B. (2015). rpart: Recursive Partitioning and Regression Trees. R Package version 4.1-9. Retrieved May 4, 2015, from https://cran.r-project.org/web/packages/rpart/index.html

Townsend, P.A., Foster, J.R., Chastain, Jr., R.A., & Currie, W.S. (2003). Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS. *IEEE Transactions on Geoscience and Remote Sensing, 41*(6), 1347-1354. Doi: 10.1109/TGRS.2003.813205

Townsend, A.R., Asner, G.P., & Cleveland, C.C. (2008). The biogeochemical heterogeneity of tropical forests. *Trends in Ecology and Evolution, 23*(8), 424-431. Doi: 10.1016/j.tree.2008.04.009

Tso, B. & Mather, P.M. (2009). *Classification methods for remotely sensed data: Second Edition*. Boca Raton, FL: CRC Press Taylor & Francis Group.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. TRENDS in Ecology and Evolution, 18(6), 306-314. Doi: 10.1016/S0169-5347(03)00070-3

Turner, W. (2014). Sensing biodiversity. Science, 346, 301-302. Doi: 10.1126/science.1256014

Underwood, E. & Ustin, S. (2007). Trends in invasive alien species. In H. Strand, R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, & W. Turner (Eds.), *Sourcebook on remote sensing and biodiversity indicators: Technical series no. 32* (pp. 163-179). Montreal: Secretariat of the Convention on Biological Diversity.

UNEP (2007). *Global Environmental Outlook GEO4: environment for development*. Malta: Progress Press LTD.

Ustin, S.L., Roberts, D.A., Gamon, J.A. Asner, G.P., & Green, R.O. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *BioScience, 54*(6), 523-534. http://www.bioone.org/doi/full/10.1641/0006-3568%282004%29054%5B0523%3AUISTSE%5D2.0.CO%3B2

Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment, 113*, S67–S77. Doi: 10.1016/j.rse.2008.10.019

Ustin, S.L. & Gamon, J.A. (2010). Remote sensing of plant functional types. *New Phytologist, 186*, 795–816. Doi: 10.1111/j.1469-8137.2010.03284.x

Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., Schlesinger, W.H., and Tilman, D.G. (1997). Human alteration of the global nitrogen cycle: sources and consequences. *Ecological Applications, 7*(3), 737–750.

Wessman, C.A. (1990). Evaluation of canopy biochemistry. In R.J. Hobbs & H.A. Mooney (Eds.), *Remote sensing of biosphere functioning* (pp. 135-156). New York: Springer Verlag.

Wessman, C.A., Aber, J.D., Peterson, D.L., & Melillo, J.M. (1998). Remote sensing of canopy chemistry and nitrogen cycling in temperate forest ecosystems. *Nature, 335*(8), 154-156.

Wessman, C.A. & Bateson, C.A. (2004). Building up with a top-down approach: The role of remote sensing in deciphering functional and structural diversity. In J. Wu, K.B. Jones, H. Li, & O.L. Loucks (Eds.), *Scaling and uncertainty analysis in ecology* (pp. 147-163). Netherlands: Springer.

White, J.C., Gómez, C., Wulder, M.A., & Coops, N.C. (2010). Characterizing temperate forest structural and spectral diversity with Hyperion EO-1 data. *Remote Sensing of Environment, 114*, 1576–1589. Doi: 10.1016/j.rse.2010.02.012

APPENDIX

Appendix A.1 Full List of Hyperspectral Data Features

Metric Category	Variable	Description
Summary Statistics	muR468, muR503, muR528, muR553, muR578, muR603, muR643, muR668, muR708, muR743, muR783, muR879, muR994, muR1064, muR1104, muR1199, muR1249, muR1284, muR1314, muR1515, muR1545, muR1585, muR1615, muR1650, muR1685, muR1730, muR2036, muR2106, muR2161, muR2216, muR2281, muR2326	Mean reflectance of all pixels for each chosen wavelength
Summary Statistics	stdR468, stdR503, stdR528, stdR553, stdR578, stdR603, stdR643, stdR668, stdR708, stdR743, stdR783, stdR879, stdR994, stdR1064, stdR1104, stdR1199, stdR1249, stdR1284, stdR1314, stdR1515, stdR1545, stdR1585, stdR1615, stdR1650, stdR1685, stdR1730, stdR2036, stdR2106, stdR2161, stdR2216, stdR2281, stdR2326	Standard deviation of reflectance of all pixels for each chosen wavelength
Summary Statistics	muDeriv468, muDeriv503, muDeriv528, muDeriv553, muDeriv578, muDeriv603, muDeriv643, muDeriv668, muDeriv708, muDeriv743, muDeriv783, muDeriv879, muDeriv994, muDeriv1064, muDeriv1104, muDeriv1199, muDeriv1249, muDeriv1284, muDeriv1314, muDeriv1515, muDeriv1545, muDeriv1585, muDeriv1615, muDeriv1650, muDeriv1685, muDeriv1730, muDeriv2036, muDeriv2106, muDeriv2161, muDeriv2216, muDeriv2281, muDeriv2326	Mean of the first derivative of reflectance of all pixels for each chosen wavelength
Summary Statistics	stdDeriv468, stdDeriv503, stdDeriv528, stdDeriv553, stdDeriv578, stdDeriv603, stdDeriv643, stdDeriv668, stdDeriv708, stdDeriv743, stdDeriv783, stdDeriv879, stdDeriv994, stdDeriv1064, stdDeriv1104, stdDeriv1199, stdDeriv1249, stdDeriv1284, stdDeriv1314, stdDeriv1515, stdDeriv1545, stdDeriv1585, stdDeriv1615, stdDeriv1650, stdDeriv1685, stdDeriv1730, stdDeriv2036, stdDeriv2106, stdDeriv2161, stdDeriv2216, stdDeriv2281, stdDeriv2326	Standard deviation of the first derivative of reflectance of all pixels for each chosen wavelength
Summary Statistics	muVegR468, muVegR503, muVegR528, muVegR553, muVegR578, muVegR603, muVegR643, muVegR668, muVegR708, muVegR743, muVegR783, muVegR879, muVegR994, muVegR1064, muVegR1104, muVegR1199, muVegR1249, muVegR1284, muVegR1314, muVegR1515, muVegR1545, muVegR1585, muVegR1615, muVegR1650, muVegR1685, muVegR1730, muVegR2036, muVegR2106, muVegR2161, muVegR2216, muVegR2281, muVegR2326	Mean reflectance of vegetation pixels for each chosen wavelength
Summary Statistics	stdVegR468, stdVegR503, stdVegR528, stdVegR553, stdVegR578, stdVegR603, stdVegR643, stdVegR668, stdVegR708, stdVegR743, stdVegR783, stdVegR879,	Standard deviation of reflectance of vegetation pixels for

	stdVegR994, stdVegR1064, stdVegR1104, stdVegR1199, stdVegR1249, stdVegR1284, stdVegR1314, stdVegR1515, stdVegR1545, stdVegR1585, stdVegR1615, stdVegR1650, stdVegR1685, stdVegR1730, stdVegR2036, stdVegR2106, stdVegR2161, stdVegR2216, stdVegR2281, stdVegR2326	each chosen wavelength
Summary Statistics	muVegDeriv468, muVegDeriv503, muVegDeriv528, muVegDeriv553, muVegDeriv578, muVegDeriv603, muVegDeriv643, muVegDeriv668, muVegDeriv708, muVegDeriv743, muVegDeriv783, muVegDeriv879, muVegDeriv994, muVegDeriv1064, muVegDeriv1104, muVegDeriv1199, muVegDeriv1249, muVegDeriv1284, muVegDeriv1314, muVegDeriv1515, muVegDeriv1545, muVegDeriv1585, muVegDeriv1615, muVegDeriv1650, muVegDeriv1685, muVegDeriv1730, muVegDeriv2036, muVegDeriv2106, muVegDeriv2161, muVegDeriv2216, muVegDeriv2281, muVegDeriv2326	Mean of the first derivative of reflectance of vegetation pixels for each chosen wavelength
Summary Statistics	stdVegDeriv468, stdVegDeriv503, stdVegDeriv528, stdVegDeriv553, stdVegDeriv578, stdVegDeriv603, stdVegDeriv643, stdVegDeriv668, stdVegDeriv708, stdVegDeriv743, stdVegDeriv783, stdVegDeriv879, stdVegDeriv994, stdVegDeriv1064, stdVegDeriv1104, stdVegDeriv1199, stdVegDeriv1249, stdVegDeriv1284, stdVegDeriv1314, stdVegDeriv1515, stdVegDeriv1545, stdVegDeriv1585, stdVegDeriv1615, stdVegDeriv1650, stdVegDeriv1685, stdVegDeriv1730, stdVegDeriv2036, stdVegDeriv2106, stdVegDeriv2161, stdVegDeriv2216, stdVegDeriv2281, stdVegDeriv2326	Standard deviation of the first derivative of reflectance of vegetation pixels for each chosen wavelength
Normalized Difference Indices (NDIs)	NDI468i503, NDI468i528, NDI468i553, NDI468i578, NDI468i603, NDI468i643, NDI468i668, NDI468i708, NDI468i743, NDI468i783, NDI468i879, NDI468i994, NDI468i1064, NDI468i1104, NDI468i1199, NDI468i1249, NDI468i1284, NDI468i1314, NDI468i1515, NDI468i1545, NDI468i1585, NDI468i1615, NDI468i1650, NDI468i1685, NDI468i1730, NDI468i2036, NDI468i2106, NDI468i2161, NDI468i2216, NDI468i2281, NDI468i2326, NDI503i528, NDI503i553, NDI503i578, NDI503i603, NDI503i643, NDI503i668, NDI503i708, NDI503i743, NDI503i783, NDI503i879, NDI503i994, NDI503i1064, NDI503i1104, NDI503i1199, NDI503i1249, NDI503i1284, NDI503i1314, NDI503i1515, NDI503i1545, NDI503i1585, NDI503i1615, NDI503i1650, NDI503i1685, NDI503i1730, NDI503i2036, NDI503i2106, NDI503i2161, NDI503i2216, NDI503i2281, NDI503i2326, NDI528i553, NDI528i578, NDI528i603, NDI528i643, NDI528i668, NDI528i708, NDI528i743, NDI528i783, NDI528i879, NDI528i994, NDI528i1064, NDI528i1104, NDI528i1199, NDI528i1249, NDI528i1284, NDI528i1314, NDI528i1515, NDI528i1545, NDI528i1585, NDI528i1615, NDI528i1650, NDI528i1685, NDI528i1730, NDI528i2036, NDI528i2106, NDI528i2161, NDI528i2216, NDI528i2281, NDI528i2326, NDI553i578, NDI553i603, NDI553i643, NDI553i668, NDI553i708, NDI553i743,	Normalized difference index between two chosen wavelengths for all pixels

NDI553i783, NDI553i879, NDI553i994, NDI553i1064, NDI553i1104, NDI553i1199, NDI553i1249, NDI553i1284, NDI553i1314, NDI553i1515, NDI553i1545, NDI553i1585, NDI553i1615, NDI553i1650, NDI553i1685, NDI553i1730, NDI553i2036, NDI553i2106, NDI553i2161, NDI553i2216, NDI553i2281, NDI553i2326, NDI578i603, NDI578i643, NDI578i668, NDI578i708, NDI578i743, NDI578i783, NDI578i879, NDI578i994, NDI578i1064, NDI578i1104, NDI578i1199, NDI578i1249, NDI578i1284, NDI578i1314, NDI578i1515, NDI578i1545, NDI578i1585, NDI578i1615, NDI578i1650, NDI578i1685, NDI578i1730, NDI578i2036, NDI578i2106, NDI578i2161, NDI578i2216, NDI578i2281, NDI578i2326, NDI603i643, NDI603i668, NDI603i708, NDI603i743, NDI603i783, NDI603i879, NDI603i994, NDI603i1064, NDI603i1104, NDI603i1199, NDI603i1249, NDI603i1284, NDI603i1314, NDI603i1515, NDI603i1545, NDI603i1585, NDI603i1615, NDI603i1650, NDI603i1685, NDI603i1730, NDI603i2036, NDI603i2106, NDI603i2161, NDI603i2216, NDI603i2281, NDI603i2326, NDI643i668, NDI643i708, NDI643i743, NDI643i783, NDI643i879, NDI643i994, NDI643i1064, NDI643i1104, NDI643i1199, NDI643i1249, NDI643i1284, NDI643i1314, NDI643i1515, NDI643i1545, NDI643i1585, NDI643i1615, NDI643i1650, NDI643i1685, NDI643i1730, NDI643i2036, NDI643i2106, NDI643i2161, NDI643i2216, NDI643i2281, NDI643i2326, NDI668i708, NDI668i743, NDI668i783, NDI668i879, NDI668i994, NDI668i1064, NDI668i1104, NDI668i1199, NDI668i1249, NDI668i1284, NDI668i1314, NDI668i1515, NDI668i1545, NDI668i1585, NDI668i1615, NDI668i1650, NDI668i1685, NDI668i1730, NDI668i2036, NDI668i2106, NDI668i2161, NDI668i2216, NDI668i2281, NDI668i2326, NDI708i743, NDI708i783, NDI708i879, NDI708i994, NDI708i1064, NDI708i1104, NDI708i1199, NDI708i1249, NDI708i1284, NDI708i1314, NDI708i1515, NDI708i1545, NDI708i1585, NDI708i1615, NDI708i1650, NDI708i1685, NDI708i1730, NDI708i2036, NDI708i2106, NDI708i2161, NDI708i2216, NDI708i2281, NDI708i2326, NDI743i783, NDI743i879, NDI743i994, NDI743i1064, NDI743i1104, NDI743i1199, NDI743i1249, NDI743i1284, NDI743i1314, NDI743i1515, NDI743i1545, NDI743i1585, NDI743i1615, NDI743i1650, NDI743i1685, NDI743i1730, NDI743i2036, NDI743i2106, NDI743i2161, NDI743i2216, NDI743i2281, NDI743i2326, NDI783i879, NDI783i994, NDI783i1064, NDI783i1104, NDI783i1199, NDI783i1249, NDI783i1284, NDI783i1314, NDI783i1515, NDI783i1545, NDI783i1585, NDI783i1615, NDI783i1650, NDI783i1685, NDI783i1730, NDI783i2036, NDI783i2106, NDI783i2161, NDI783i2216, NDI783i2281, NDI783i2326, NDI879i994, NDI879i1064, NDI879i1104, NDI879i1199, NDI879i1249, NDI879i1284, NDI879i1314, NDI879i1515, NDI879i1545, NDI879i1585, NDI879i1615, NDI879i1650, NDI879i1685, NDI879i1730,

NDI879i2036, NDI879i2106, NDI879i2161, NDI879i2216,
NDI879i2281, NDI879i2326, NDI994i1064, NDI994i1104,
NDI994i1199, NDI994i1249, NDI994i1284, NDI994i1314,
NDI994i1515, NDI994i1545, NDI994i1585, NDI994i1615,
NDI994i1650, NDI994i1685, NDI994i1730, NDI994i2036,
NDI994i2106, NDI994i2161, NDI994i2216, NDI994i2281,
NDI994i2326, NDI1064i1104, NDI1064i1199,
NDI1064i1249, NDI1064i1284, NDI1064i1314,
NDI1064i1515, NDI1064i1545, NDI1064i1585,
NDI1064i1615, NDI1064i1650, NDI1064i1685,
NDI1064i1730, NDI1064i2036, NDI1064i2106,
NDI1064i2161, NDI1064i2216, NDI1064i2281,
NDI1064i2326, NDI1104i1199, NDI1104i1249,
NDI1104i1284, NDI1104i1314, NDI1104i1515,
NDI1104i1545, NDI1104i1585, NDI1104i1615,
NDI1104i1650, NDI1104i1685, NDI1104i1730,
NDI1104i2036, NDI1104i2106, NDI1104i2161,
NDI1104i2216, NDI1104i2281, NDI1104i2326,
NDI1199i1249, NDI1199i1284, NDI1199i1314,
NDI1199i1515, NDI1199i1545, NDI1199i1585,
NDI1199i1615, NDI1199i1650, NDI1199i1685,
NDI1199i1730, NDI1199i2036, NDI1199i2106,
NDI1199i2161, NDI1199i2216, NDI1199i2281,
NDI1199i2326, NDI1249i1284, NDI1249i1314,
NDI1249i1515, NDI1249i1545, NDI1249i1585,
NDI1249i1615, NDI1249i1650, NDI1249i1685,
NDI1249i1730, NDI1249i2036, NDI1249i2106,
NDI1249i2161, NDI1249i2216, NDI1249i2281,
NDI1249i2326, NDI1284i1314, NDI1284i1515,
NDI1284i1545, NDI1284i1585, NDI1284i1615,
NDI1284i1650, NDI1284i1685, NDI1284i1730,
NDI1284i2036, NDI1284i2106, NDI1284i2161,
NDI1284i2216, NDI1284i2281, NDI1284i2326,
NDI1314i1515, NDI1314i1545, NDI1314i1585,
NDI1314i1615, NDI1314i1650, NDI1314i1685,
NDI1314i1730, NDI1314i2036, NDI1314i2106,
NDI1314i2161, NDI1314i2216, NDI1314i2281,
NDI1314i2326, NDI1515i1545, NDI1515i1585,
NDI1515i1615, NDI1515i1650, NDI1515i1685,
NDI1515i1730, NDI1515i2036, NDI1515i2106,
NDI1515i2161, NDI1515i2216, NDI1515i2281,
NDI1515i2326, NDI1545i1585, NDI1545i1615,
NDI1545i1650, NDI1545i1685, NDI1545i1730,
NDI1545i2036, NDI1545i2106, NDI1545i2161,
NDI1545i2216, NDI1545i2281, NDI1545i2326,
NDI1585i1615, NDI1585i1650, NDI1585i1685,
NDI1585i1730, NDI1585i2036, NDI1585i2106,
NDI1585i2161, NDI1585i2216, NDI1585i2281,
NDI1585i2326, NDI1615i1650, NDI1615i1685,
NDI1615i1730, NDI1615i2036, NDI1615i2106,
NDI1615i2161, NDI1615i2216, NDI1615i2281,

	NDI1615i2326, NDI1650i1685, NDI1650i1730, NDI1650i2036, NDI1650i2106, NDI1650i2161, NDI1650i2216, NDI1650i2281, NDI1650i2326, NDI1685i1730, NDI1685i2036, NDI1685i2106, NDI1685i2161, NDI1685i2216, NDI1685i2281, NDI1685i2326, NDI1730i2036, NDI1730i2106, NDI1730i2161, NDI1730i2216, NDI1730i2281, NDI1730i2326, NDI2036i2106, NDI2036i2161, NDI2036i2216, NDI2036i2281, NDI2036i2326, NDI2106i2161, NDI2106i2216, NDI2106i2281, NDI2106i2326, NDI2161i2216, NDI2161i2281, NDI2161i2326, NDI2216i2281, NDI2216i2326, NDI2281i2326	
Normalized Difference Indices (NDIs)	vegNDI468i503, vegNDI468i528, vegNDI468i553, vegNDI468i578, vegNDI468i603, vegNDI468i643, vegNDI468i668, vegNDI468i708, vegNDI468i743, vegNDI468i783, vegNDI468i879, vegNDI468i994, vegNDI468i1064, vegNDI468i1104, vegNDI468i1199, vegNDI468i1249, vegNDI468i1284, vegNDI468i1314, vegNDI468i1515, vegNDI468i1545, vegNDI468i1585, vegNDI468i1615, vegNDI468i1650, vegNDI468i1685, vegNDI468i1730, vegNDI468i2036, vegNDI468i2106, vegNDI468i2161, vegNDI468i2216, vegNDI468i2281, vegNDI468i2326, vegNDI503i528, vegNDI503i553, vegNDI503i578, vegNDI503i603, vegNDI503i643, vegNDI503i668, vegNDI503i708, vegNDI503i743, vegNDI503i783, vegNDI503i879, vegNDI503i994, vegNDI503i1064, vegNDI503i1104, vegNDI503i1199, vegNDI503i1249, vegNDI503i1284, vegNDI503i1314, vegNDI503i1515, vegNDI503i1545, vegNDI503i1585, vegNDI503i1615, vegNDI503i1650, vegNDI503i1685, vegNDI503i1730, vegNDI503i2036, vegNDI503i2106, vegNDI503i2161, vegNDI503i2216, vegNDI503i2281, vegNDI503i2326, vegNDI528i553, vegNDI528i578, vegNDI528i603, vegNDI528i643, vegNDI528i668, vegNDI528i708, vegNDI528i743, vegNDI528i783, vegNDI528i879, vegNDI528i994, vegNDI528i1064, vegNDI528i1104, vegNDI528i1199, vegNDI528i1249, vegNDI528i1284, vegNDI528i1314, vegNDI528i1515, vegNDI528i1545, vegNDI528i1585, vegNDI528i1615, vegNDI528i1650, vegNDI528i1685, vegNDI528i1730, vegNDI528i2036, vegNDI528i2106, vegNDI528i2161, vegNDI528i2216, vegNDI528i2281, vegNDI528i2326, vegNDI553i578, vegNDI553i603, vegNDI553i643, vegNDI553i668, vegNDI553i708, vegNDI553i743, vegNDI553i783, vegNDI553i879, vegNDI553i994, vegNDI553i1064, vegNDI553i1104, vegNDI553i1199, vegNDI553i1249, vegNDI553i1284, vegNDI553i1314, vegNDI553i1515, vegNDI553i1545, vegNDI553i1585, vegNDI553i1615, vegNDI553i1650, vegNDI553i1685, vegNDI553i1730, vegNDI553i2036, vegNDI553i2106,	Normalized difference index between two chosen wavelengths for vegetation pixels

vegNDI553i2161, vegNDI553i2216, vegNDI553i2281,
vegNDI553i2326, vegNDI578i603, vegNDI578i643,
vegNDI578i668, vegNDI578i708, vegNDI578i743,
vegNDI578i783, vegNDI578i879, vegNDI578i994,
vegNDI578i1064, vegNDI578i1104, vegNDI578i1199,
vegNDI578i1249, vegNDI578i1284, vegNDI578i1314,
vegNDI578i1515, vegNDI578i1545, vegNDI578i1585,
vegNDI578i1615, vegNDI578i1650, vegNDI578i1685,
vegNDI578i1730, vegNDI578i2036, vegNDI578i2106,
vegNDI578i2161, vegNDI578i2216, vegNDI578i2281,
vegNDI578i2326, vegNDI603i643, vegNDI603i668,
vegNDI603i708, vegNDI603i743, vegNDI603i783,
vegNDI603i879, vegNDI603i994, vegNDI603i1064,
vegNDI603i1104, vegNDI603i1199, vegNDI603i1249,
vegNDI603i1284, vegNDI603i1314, vegNDI603i1515,
vegNDI603i1545, vegNDI603i1585, vegNDI603i1615,
vegNDI603i1650, vegNDI603i1685, vegNDI603i1730,
vegNDI603i2036, vegNDI603i2106, vegNDI603i2161,
vegNDI603i2216, vegNDI603i2281, vegNDI603i2326,
vegNDI643i668, vegNDI643i708, vegNDI643i743,
vegNDI643i783, vegNDI643i879, vegNDI643i994,
vegNDI643i1064, vegNDI643i1104, vegNDI643i1199,
vegNDI643i1249, vegNDI643i1284, vegNDI643i1314,
vegNDI643i1515, vegNDI643i1545, vegNDI643i1585,
vegNDI643i1615, vegNDI643i1650, vegNDI643i1685,
vegNDI643i1730, vegNDI643i2036, vegNDI643i2106,
vegNDI643i2161, vegNDI643i2216, vegNDI643i2281,
vegNDI643i2326, vegNDI668i708, vegNDI668i743,
vegNDI668i783, vegNDI668i879, vegNDI668i994,
vegNDI668i1064, vegNDI668i1104, vegNDI668i1199,
vegNDI668i1249, vegNDI668i1284, vegNDI668i1314,
vegNDI668i1515, vegNDI668i1545, vegNDI668i1585,
vegNDI668i1615, vegNDI668i1650, vegNDI668i1685,
vegNDI668i1730, vegNDI668i2036, vegNDI668i2106,
vegNDI668i2161, vegNDI668i2216, vegNDI668i2281,
vegNDI668i2326, vegNDI708i743, vegNDI708i783,
vegNDI708i879, vegNDI708i994, vegNDI708i1064,
vegNDI708i1104, vegNDI708i1199, vegNDI708i1249,
vegNDI708i1284, vegNDI708i1314, vegNDI708i1515,
vegNDI708i1545, vegNDI708i1585, vegNDI708i1615,
vegNDI708i1650, vegNDI708i1685, vegNDI708i1730,
vegNDI708i2036, vegNDI708i2106, vegNDI708i2161,
vegNDI708i2216, vegNDI708i2281, vegNDI708i2326,
vegNDI743i783, vegNDI743i879, vegNDI743i994,
vegNDI743i1064, vegNDI743i1104, vegNDI743i1199,
vegNDI743i1249, vegNDI743i1284, vegNDI743i1314,
vegNDI743i1515, vegNDI743i1545, vegNDI743i1585,
vegNDI743i1615, vegNDI743i1650, vegNDI743i1685,
vegNDI743i1730, vegNDI743i2036, vegNDI743i2106,
vegNDI743i2161, vegNDI743i2216, vegNDI743i2281,
vegNDI743i2326, vegNDI783i879, vegNDI783i994,

vegNDI783i1064, vegNDI783i1104, vegNDI783i1199,
vegNDI783i1249, vegNDI783i1284, vegNDI783i1314,
vegNDI783i1515, vegNDI783i1545, vegNDI783i1585,
vegNDI783i1615, vegNDI783i1650, vegNDI783i1685,
vegNDI783i1730, vegNDI783i2036, vegNDI783i2106,
vegNDI783i2161, vegNDI783i2216, vegNDI783i2281,
vegNDI783i2326, vegNDI879i994, vegNDI879i1064,
vegNDI879i1104, vegNDI879i1199, vegNDI879i1249,
vegNDI879i1284, vegNDI879i1314, vegNDI879i1515,
vegNDI879i1545, vegNDI879i1585, vegNDI879i1615,
vegNDI879i1650, vegNDI879i1685, vegNDI879i1730,
vegNDI879i2036, vegNDI879i2106, vegNDI879i2161,
vegNDI879i2216, vegNDI879i2281, vegNDI879i2326,
vegNDI994i1064, vegNDI994i1104, vegNDI994i1199,
vegNDI994i1249, vegNDI994i1284, vegNDI994i1314,
vegNDI994i1515, vegNDI994i1545, vegNDI994i1585,
vegNDI994i1615, vegNDI994i1650, vegNDI994i1685,
vegNDI994i1730, vegNDI994i2036, vegNDI994i2106,
vegNDI994i2161, vegNDI994i2216, vegNDI994i2281,
vegNDI994i2326, vegNDI1064i1104, vegNDI1064i1199,
vegNDI1064i1249, vegNDI1064i1284, vegNDI1064i1314,
vegNDI1064i1515, vegNDI1064i1545, vegNDI1064i1585,
vegNDI1064i1615, vegNDI1064i1650, vegNDI1064i1685,
vegNDI1064i1730, vegNDI1064i2036, vegNDI1064i2106,
vegNDI1064i2161, vegNDI1064i2216, vegNDI1064i2281,
vegNDI1064i2326, vegNDI1104i1199, vegNDI1104i1249,
vegNDI1104i1284, vegNDI1104i1314, vegNDI1104i1515,
vegNDI1104i1545, vegNDI1104i1585, vegNDI1104i1615,
vegNDI1104i1650, vegNDI1104i1685, vegNDI1104i1730,
vegNDI1104i2036, vegNDI1104i2106, vegNDI1104i2161,
vegNDI1104i2216, vegNDI1104i2281, vegNDI1104i2326,
vegNDI1199i1249, vegNDI1199i1284, vegNDI1199i1314,
vegNDI1199i1515, vegNDI1199i1545, vegNDI1199i1585,
vegNDI1199i1615, vegNDI1199i1650, vegNDI1199i1685,
vegNDI1199i1730, vegNDI1199i2036, vegNDI1199i2106,
vegNDI1199i2161, vegNDI1199i2216, vegNDI1199i2281,
vegNDI1199i2326, vegNDI1249i1284, vegNDI1249i1314,
vegNDI1249i1515, vegNDI1249i1545, vegNDI1249i1585,
vegNDI1249i1615, vegNDI1249i1650, vegNDI1249i1685,
vegNDI1249i1730, vegNDI1249i2036, vegNDI1249i2106,
vegNDI1249i2161, vegNDI1249i2216, vegNDI1249i2281,
vegNDI1249i2326, vegNDI1284i1314, vegNDI1284i1515,
vegNDI1284i1545, vegNDI1284i1585, vegNDI1284i1615,
vegNDI1284i1650, vegNDI1284i1685, vegNDI1284i1730,
vegNDI1284i2036, vegNDI1284i2106, vegNDI1284i2161,
vegNDI1284i2216, vegNDI1284i2281, vegNDI1284i2326,
vegNDI1314i1515, vegNDI1314i1545, vegNDI1314i1585,
vegNDI1314i1615, vegNDI1314i1650, vegNDI1314i1685,
vegNDI1314i1730, vegNDI1314i2036, vegNDI1314i2106,
vegNDI1314i2161, vegNDI1314i2216, vegNDI1314i2281,
vegNDI1314i2326, vegNDI1515i1545, vegNDI1515i1585,

	vegNDI1515i1615, vegNDI1515i1650, vegNDI1515i1685, vegNDI1515i1730, vegNDI1515i2036, vegNDI1515i2106, vegNDI1515i2161, vegNDI1515i2216, vegNDI1515i2281, vegNDI1515i2326, vegNDI1545i1585, vegNDI1545i1615, vegNDI1545i1650, vegNDI1545i1685, vegNDI1545i1730, vegNDI1545i2036, vegNDI1545i2106, vegNDI1545i2161, vegNDI1545i2216, vegNDI1545i2281, vegNDI1545i2326, vegNDI1585i1615, vegNDI1585i1650, vegNDI1585i1685, vegNDI1585i1730, vegNDI1585i2036, vegNDI1585i2106, vegNDI1585i2161, vegNDI1585i2216, vegNDI1585i2281, vegNDI1585i2326, vegNDI1615i1650, vegNDI1615i1685, vegNDI1615i1730, vegNDI1615i2036, vegNDI1615i2106, vegNDI1615i2161, vegNDI1615i2216, vegNDI1615i2281, vegNDI1615i2326, vegNDI1650i1685, vegNDI1650i1730, vegNDI1650i2036, vegNDI1650i2106, vegNDI1650i2161, vegNDI1650i2216, vegNDI1650i2281, vegNDI1650i2326, vegNDI1685i1730, vegNDI1685i2036, vegNDI1685i2106, vegNDI1685i2161, vegNDI1685i2216, vegNDI1685i2281, vegNDI1685i2326, vegNDI1730i2036, vegNDI1730i2106, vegNDI1730i2161, vegNDI1730i2216, vegNDI1730i2281, vegNDI1730i2326, vegNDI2036i2106, vegNDI2036i2161, vegNDI2036i2216, vegNDI2036i2281, vegNDI2036i2326, vegNDI2106i2161, vegNDI2106i2216, vegNDI2106i2281, vegNDI2106i2326, vegNDI2161i2216, vegNDI2161i2281, vegNDI2161i2326, vegNDI2216i2281, vegNDI2216i2326, vegNDI2281i2326	
Triangle Metrics	GreenPeak, GreenPeakRot, RedDepth, RedDepthRot	Three band triangle metrics for all pixels
Triangle Metrics	vegGreenPeak, vegGreenPeakRot, vegRedDepth, vegRedDepthRot	Three band triangle metrics for vegetation pixels
Endmember Analysis	FracSandRoad, FracLakeFringe, FracWaterShallow, FracWaterDark, FracWaterGlint, FracWaterDarkVeg, FracWaterMedVeg, FracWaterLightVeg, FracWaterWhiteVeg, FracShadowGround, FracShadowCanopy, FracSandhillVeg1, FracSandhillVeg2, FracSandhillGrass, FracSandhillBurnt, FracTreesBright, FracTreesDark, FracMixedDark, FracTreesGray, FracRuderalGreen, FracRuderalMixed, FracRuderalBright, FracMarshGreen, FracMarshPink, FracMarshLight, FracMarshDark, FracBurnScar	Fraction of endmember pixels

Appendix A.2 Full List of Lidar Data Features

Metric Category	Variable	Description
Point Classification	Fraction of ground points	Fraction of ground points
Percentile Height	Ground percentile height array[0]	0 percentile height of ground points
Percentile Height	Ground percentile height array[2]	10th percentile height of ground points
Percentile Height	Ground percentile height array[5]	25th percentile height of ground points
Percentile Height	Ground percentile height array[10]	50th percentile height of ground points
Percentile Height	Ground percentile height array[15]	75th percentile height of ground points
Percentile Height	Ground percentile height array[18]	90th percentile height of ground points
Percentile Height	Ground percentile height array[20]	100th percentile height of ground points
Percentile Height Range	Ground p100 p0 range	Height range between 0 and 100th percentile of ground points
Percentile Height Range	Ground p90 p10 range	Height range between 10th and 90th percentile of ground points
Percentile Height Range	Ground p75 p25 range	Height range between 25th and 75th percentile of ground points
Percentile Height Range	Ground p90 p50 range	Height range between 50th and 90th percentile of ground points
Percentile Height Range	Ground p50 p10 range	Height range between 10th and 50th percentile of ground points
Percentile Height Range	Ground p100 p50 range	Height range between 50th and 100th percentile of ground points
Percentile Height Range	Ground p50 p0 range	Height range between 0 and 50th percentile of ground points
Interquartile Height Range	Ground p100 p75 range	Height range between 75th and 100th percentile of ground points
Interquartile Height Range	Ground p75 p50 range	Height range between 50th and 75th percentile of ground points
Interquartile Height Range	Ground p50 p25 range	Height range between 25th and 50th percentile of ground points
Interquartile Height Range	Ground p25 p0 range	Height range between 0 and 25th percentile of ground points
Percentile Height Range Fraction	Ground p100 p50 fraction	Fraction of ground points between 50th and 100th percentile
Percentile Height Range Fraction	Ground p75 p25 fraction	Fraction of ground points between 25th and 75th percentile
Percentile Height Range Fraction	Ground p50 p0 fraction	Fraction of ground points between 0 and 50th percentile
Interquartile Height Range Fraction	Ground p100 p75 fraction	Fraction of ground points between 75th and 100th percentile
Interquartile Height Range Fraction	Ground p75 p50 fraction	Fraction of ground points between 50th and 75th percentile
Interquartile Height Range Fraction	Ground p50 p25 fraction	Fraction of ground points between 25th and 50th percentile

Interquartile Height Range Fraction	Ground p25 p0 fraction	Fraction of ground points between 0 and 25th percentile
Normalized Difference Height Index	Ground p90 p10 ndhi	Normalized difference height index using 10th, 50th, and 90th percentile of ground points
Normalized Difference Height Index	Ground p75 p25 ndhi	Normalized difference height index using 25th, 50th, and 75th percentile of ground points
Point Distribution	Ground median absolute deviation	Median absolute deviation of ground points
Point Distribution	Ground mean absolute deviation	Mean absolute deviation of ground points
Percentile Height	Canopy percentile height array[0]	0 percentile height of canopy points
Percentile Height	Canopy percentile height array[2]	10th percentile height of canopy points
Percentile Height	Canopy percentile height array[5]	25th percentile height of canopy points
Percentile Height	Canopy percentile height array[10]	50th percentile height of canopy points
Percentile Height	Canopy percentile height array[15]	75th percentile height of canopy points
Percentile Height	Canopy percentile height array[18]	90th percentile height of canopy points
Percentile Height	Canopy percentile height array[20]	100th percentile height of canopy points
Percentile Height Range	Canopy p100 p0 range	Height range between 0 and 100th percentile of canopy points
Percentile Height Range	Canopy p90 p10 range	Height range between 10th and 90th percentile of canopy points
Percentile Height Range	Canopy p75 p25 range	Height range between 25th and 75th percentile of canopy points
Percentile Height Range	Canopy p90 p50 range	Height range between 50th and 90th percentile of canopy points
Percentile Height Range	Canopy p50 p10 range	Height range between 10th and 50th percentile of canopy points
Percentile Height Range	Canopy p100 p50 range	Height range between 50th and 100th percentile of canopy points
Percentile Height Range	Canopy p50 p0 range	Height range between 0 and 50th percentile of canopy points
Interquartile Height Range	Canopy p100 p75 range	Height range between 75th and 100th percentile of canopy points
Interquartile Height Range	Canopy p75 p50 range	Height range between 50th and 75th percentile of canopy points
Interquartile Height Range	Canopy p50 p25 range	Height range between 25th and 50th percentile of canopy points
Interquartile Height Range	Canopy p25 p0 range	Height range between 0 and 25th percentile of canopy points
Percentile Height Range Fraction	Canopy p100 p50 fraction	Fraction of canopy points between 50th and 100th percentile
Percentile Height Range Fraction	Canopy p75 p25 fraction	Fraction of canopy points between 25th and 75th percentile
Percentile Height Range Fraction	Canopy p50 p0 fraction	Fraction of canopy points between 0 and 50th percentile
Interquartile Height Range Fraction	Canopy p100 p75 fraction	Fraction of canopy points between 75th and 100th percentile
Interquartile Height Range Fraction	Canopy p75 p50 fraction	Fraction of canopy points between 50th and 75th percentile

Interquartile Height Range Fraction	Canopy p50 p25 fraction	Fraction of canopy points between 25th and 50th percentile
Interquartile Height Range Fraction	Canopy p25 p0 fraction	Fraction of canopy points between 0 and 25th percentile
Normalized Difference Height Index	Canopy p90 p10 ndhi	Normalized difference height index using 10th, 50th, and 90th percentile of canopy points
Normalized Difference Height Index	Canopy p75 p25 ndhi	Normalized difference height index using 25th, 50th, and 75th percentile of canopy points
Summary Statistics	μ all	Mean height of all points
Summary Statistics	σ all	Standard deviation of heights for all points
Summary Statistics	μ ground	Mean height of ground points
Summary Statistics	σ ground	Standard deviation of heights for ground points
Summary Statistics	μ canopy	Mean height of canopy points
Summary Statistics	σ canopy	Standard deviation of heights for canopy points
Point Classification	Fraction of below ground aoi0 points	Fraction of below ground points
Point Classification	Fraction of below ground aoi1 points	Fraction of below ground points
Point Classification	Fraction of below ground aoi2 points	Fraction of below ground points
Point Classification	Fraction of shrub points	Fraction of shrub points
Point Distribution	Canopy vertical distribution ratio	Ratio of the top half of the canopy height range to the max canopy height
Point Distribution	Canopy median absolute deviation	Median absolute deviation of canopy points
Point Distribution	Canopy mean absolute deviation	Mean absolute deviation of canopy points
Point Distribution	Canopy relief ratio	Ratio of the bottom half of the canopy height range to the total canopy height range
Point Classification	Fraction of first return points	The fraction of first return points
Summary Statistics	μ first	Mean height of first return points
Summary Statistics	σ first	Standard deviation of heights for first return points
Point Classification	Fraction of first return ground points	The fraction of first return points that are classified as ground points
Percentile Height	First P95	95th percentile height of first return points
Point Classification	Canopy density	The fraction of high canopy first return points to total first return points
Point Classification	Fraction 1 of 1	The fraction of points that are the first return of one
Point Classification	Fraction 2 of 2	The fraction of points that are the second return of two
Point Classification	Fraction 3 of 3	The fraction of points that are the third return of three
Point Classification	Fraction 4 of 4	The fraction of points that are the fourth return of four
Elevation Model	μ CHM	Mean height of the canopy height model

Elevation Model	σ CHM	Standard deviation of height of the canopy height model
Elevation Model	Min CHM	Min height of the canopy height model
Elevation Model	Max CHM	Max height of the canopy height model
Elevation Model	Fraction of ground points CHM	Fraction of ground points in the canopy height model
Elevation Model	CHM Rugosity	Surface roughness (rugosity) of the canopy height model
Elevation Model	DEM Slope 21x21	Coarse resolution slope of the bare-Earth digital terrain model
Elevation Model	Total DEM Slope 3x3	Fine resolution slope of the bare-Earth digital terrain model

Appendix B.1 Hyperspectral Decision Tree

```
n= 1000

node), split, n, loss, yval, (yprob)
      * denotes terminal node

  1) root 1000 900 Baygall (0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1)
    2) muDeriv1064< 0.932652 105  10 Lake (0 0 0 0.9 0.086 0 0 0 0.0095 0)
      4) vegNDI743i994< 6.361945 95   3 Lake (0 0 0 0.97 0.021 0 0 0 0.011 0) *
      5) vegNDI743i994>=6.361945 10   3 Marsh (0 0 0 0.3 0.7 0 0 0 0 0)
       10) stdR528< 0.7787 3   0 Lake (0 0 0 1 0 0 0 0 0 0) *
       11) stdR528>=0.7787 7   0 Marsh (0 0 0 0 1 0 0 0 0 0) *
    3) muDeriv1064>=0.932652 895 795 Baygall (0.11 0.11 0.11 0.0056 0.1 0.11 0.11 0.11 0.11 0.11)
      6) NDI578i643< -12.37815 401 301 Baygall (0.25 0.005 0.22 0.0025 0.025 0.24 0.017 0 0.0025 0.24)
       12) vegNDI1515i2216< -16.7446 290 196 Mixed (0.31 0.0034 0.3 0 0.021 0.32 0.024 0 0 0.021)
         24) NDI1650i1685>=-0.4336285 93  20 Baygall (0.78 0 0.022 0 0.043 0.13 0 0 0 0.022)
           48) NDI743i1585>=-41.5957 80   7 Baygall (0.91 0 0.012 0 0.025 0.025 0 0 0 0.025) *
           49) NDI743i1585< -41.5957 13   3 Mixed (0 0 0.077 0 0.15 0.77 0 0 0 0) *
         25) NDI1650i1685< -0.4336285 197 113 Hammock (0.086 0.0051 0.43 0 0.01 0.42 0.036 0 0 0.02)
           50) vegNDI783i879>=4.68648 74  17 Hammock (0.054 0 0.77 0 0.014 0.14 0.014 0 0 0.014)
            100) muDeriv2036< 1.77944 61   7 Hammock (0.016 0 0.89 0 0 0.066 0.016 0 0 0.016) *
            101) muDeriv2036>=1.77944 13   7 Mixed (0.23 0 0.23 0 0.077 0.46 0 0 0 0)
              202) stdR668< 0.914061 7   4 Baygall (0.43 0 0.43 0 0.14 0 0 0 0 0)
                404) muR743< 29.46505 3   0 Baygall (1 0 0 0 0 0 0 0 0 0) *
                405) muR743>=29.46505 4   1 Hammock (0 0 0.75 0 0.25 0 0 0 0 0) *
              203) stdR668>=0.914061 6   0 Mixed (0 0 0 0 0 1 0 0 0 0) *
           51) vegNDI783i879< 4.68648 123  51 Mixed (0.11 0.0081 0.22 0 0.0081 0.59 0.049 0 0 0.024)
            102) vegNDI1585i1730>=1.28819 18   2 Hammock (0.056 0 0.89 0 0 0.056 0 0 0 0) *
            103) vegNDI1585i1730< 1.28819 105  34 Mixed (0.11 0.0095 0.1 0 0.0095 0.68 0.057 0 0 0.029)
              206) muDeriv783>=3.43161 9   2 Hammock (0 0 0.78 0 0 0.22 0 0 0 0) *
              207) muDeriv783< 3.43161 96  27 Mixed (0.12 0.01 0.042 0 0.01 0.72 0.062 0 0 0.031)
                414) NDI503i2161< 54.0168 5   0 Baygall (1 0 0 0 0 0 0 0 0 0) *
                415) NDI503i2161>=54.0168 91  22 Mixed (0.077 0.011 0.044 0 0.011 0.76 0.066 0 0 0.033)
                  830) NDI2281i2326< -7.9344 87  18 Mixed (0.08 0.011 0.046 0 0 0.79 0.034 0 0 0.034) *
                  831) NDI2281i2326>=-7.9344 4   1 Plantation (0 0 0 0 0.25 0 0.75 0 0 0) *
       13) vegNDI1515i2216>=-16.7446 111  20 Swamp (0.09 0.009 0.009 0.009 0.036 0.018 0 0 0.009 0.82)
         26) vegNDI1199i1284< 4.570385 14   6 Baygall (0.57 0.071 0 0.071 0.21 0 0 0 0 0.071)
           52) stdR528>=1.47904 8   0 Baygall (1 0 0 0 0 0 0 0 0 0) *
           53) stdR528< 1.47904 6   3 Marsh (0 0.17 0 0.17 0.5 0 0 0 0 0.17) *
         27) vegNDI1199i1284>=4.570385 97   7 Swamp (0.021 0 0.01 0 0.01 0.021 0 0 0.01 0.93) *
      7) NDI578i643>=-12.37815 494 394 Ruderal (0 0.2 0.026 0.0081 0.16 0.0081 0.19 0.2 0.2 0.0061)
       14) muDeriv2106>=1.34994 256 161 Ruderal (0 0.33 0.0039 0.0078 0.24 0.0078 0.0039 0.37 0.027 0.0039)
         28) muDeriv2216< 0.6161705 102  41 Marsh (0 0.33 0 0.02 0.6 0.02 0 0.0098 0.0098 0.0098)
           56) muVegDeriv668< 2.10495 40  11 Fringe (0 0.72 0 0.025 0.22 0 0 0 0 0.025)
            112) muR743< 23.08555 33   4 Fringe (0 0.88 0 0.03 0.061 0 0 0 0 0.03) *
            113) muR743>=23.08555 7   0 Marsh (0 0 0 0 1 0 0 0 0 0) *
           57) muVegDeriv668>=2.10495 62  10 Marsh (0 0.081 0 0.016 0.84 0.032 0 0.016 0.016 0)
            114) muVegDeriv1730>=-0.6620275 6   1 Fringe (0 0.83 0 0 0 0 0 0 0.17 0) *
            115) muVegDeriv1730< -0.6620275 56   4 Marsh (0 0 0 0.018 0.93 0.036 0 0.018 0 0) *
         29) muDeriv2216>=0.6161705 154  60 Ruderal (0 0.33 0.0065 0 0.0065 0 0.0065 0.61 0.039 0)
```

```
   58) muDeriv528< 5.846135 52   7 Fringe (0 0.87 0 0 0 0 0 0.12 0.019 0) *
   59) muDeriv528>=5.846135 102  14 Ruderal (0 0.059 0.0098 0 0.0098 0 0.0098 0.86 0.049 0)
    118) FracSandhillGrass< 24.9433 97   9 Ruderal (0 0.062 0.01 0 0.01 0 0 0.91 0.01 0)
      236) NDI994i1104< 4.58154 6   1 Fringe (0 0.83 0 0 0 0 0 0.17 0 0) *
      237) NDI994i1104>=4.58154 91   4 Ruderal (0 0.011 0.011 0 0.011 0 0 0.96 0.011 0) *
    119) FracSandhillGrass>=24.9433 5   1 Sandhill (0 0 0 0 0 0 0.2 0 0.8 0) *
 15) muDeriv2106< 1.34994 238 146 Plantation (0 0.055 0.05 0.0084 0.08 0.0084 0.39 0.021 0.38 0.0084)
   30) NDI1585i1730< -0.3178745 114  31 Plantation (0 0.018 0.053 0 0 0.018 0.73 0.018 0.17 0)
     60) stdVegDeriv643< 1.466715 84   9 Plantation (0 0.012 0.012 0 0 0.024 0.89 0.012 0.048 0) *
     61) stdVegDeriv643>=1.466715 30  15 Sandhill (0 0.033 0.17 0 0 0 0.27 0.033 0.5 0)
      122) muDeriv1249< 4.18075 16   9 Plantation (0 0.062 0.31 0 0 0 0.44 0.062 0.12 0)
        244) muDeriv708>=42.505 5   0 Hammock (0 0 1 0 0 0 0 0 0 0) *
        245) muDeriv708< 42.505 11   4 Plantation (0 0.091 0 0 0 0 0.64 0.091 0.18 0) *
      123) muDeriv1249>=4.18075 14   1 Sandhill (0 0 0 0 0 0 0.071 0 0.93 0) *
   31) NDI1585i1730>=-0.3178745 124  52 Sandhill (0 0.089 0.048 0.016 0.15 0 0.073 0.024 0.58 0.016)
     62) vegNDI1249i1284< 0.8054935 26   7 Marsh (0 0 0.15 0.038 0.73 0 0 0 0 0.077)
      124) stdDeriv2326< 1.40391 20   1 Marsh (0 0 0 0.05 0.95 0 0 0 0 0) *
      125) stdDeriv2326>=1.40391 6   2 Hammock (0 0 0.67 0 0 0 0 0 0 0.33) *
     63) vegNDI1249i1284>=0.8054935 98  26 Sandhill (0 0.11 0.02 0.01 0 0 0.092 0.031 0.73 0)
      126) muVegDeriv503< 2.911595 13   2 Fringe (0 0.85 0.15 0 0 0 0 0 0 0) *
      127) muVegDeriv503>=2.911595 85  13 Sandhill (0 0 0 0.012 0 0 0.11 0.035 0.85 0)
        254) vegNDI1104i1284< -4.19105 12   6 Plantation (0 0 0 0.083 0 0 0.5 0.17 0.25 0)
          508) stdDeriv603< 1.84285 6   0 Plantation (0 0 0 0 0 0 1 0 0 0) *
          509) stdDeriv603>=1.84285 6   3 Sandhill (0 0 0 0.17 0 0 0 0.33 0.5 0) *
        255) vegNDI1104i1284>=-4.19105 73   4 Sandhill (0 0 0 0 0 0 0 0.041 0.014 0.95 0) *
```

Appendix B.2 Lidar Decision Tree

```
n= 1000

node), split, n, loss, yval, (yprob)
      * denotes terminal node

  1) root 1000 900 Baygall (0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1)
    2) muCHM< 3.075415 403 304 Lake (0.032 0.24 0 0.25 0.23 0 0 0.24 0.022 0)
      4) DEMslopeK21< 9.23421 196 102 Lake (0.02 0.097 0 0.48 0.37 0 0 0.031 0.0051 0)
        8) FracGround>=99.96925 138  44 Lake (0 0.043 0 0.68 0.28 0 0 0 0 0)
         16) grP100mP75< 0.3400005 31   0 Lake (0 0 0 1 0 0 0 0 0 0) *
         17) grP100mP75>=0.3400005 107  44 Lake (0 0.056 0 0.59 0.36 0 0 0 0 0)
           34) grP90mP50>=0.2149995 79  21 Lake (0 0.013 0 0.73 0.25 0 0 0 0 0)
             68) grP100mP75< 0.7499995 68  12 Lake (0 0 0 0.82 0.18 0 0 0 0 0) *
             69) grP100mP75>=0.7499995 11   3 Marsh (0 0.091 0 0.18 0.73 0 0 0 0 0) *
           35) grP90mP50< 0.2149995 28  10 Marsh (0 0.18 0 0.18 0.64 0 0 0 0 0) *
        9) FracGround< 99.96925 58  24 Marsh (0.069 0.22 0 0 0.59 0 0 0.1 0.017 0)
         18) totDEMslopeK3< 27.6088 21  10 Fringe (0 0.52 0 0 0.19 0 0 0.24 0.048 0)
           36) caP50mP0>=0.93 18   7 Fringe (0 0.61 0 0 0.056 0 0 0.28 0.056 0)
             72) caP100mP0< 6.995 9   0 Fringe (0 1 0 0 0 0 0 0 0 0) *
             73) caP100mP0>=6.995 9   4 Ruderal (0 0.22 0 0 0.11 0 0 0.56 0.11 0) *
           37) caP50mP0< 0.93 3   0 Marsh (0 0 0 0 1 0 0 0 0 0) *
         19) totDEMslopeK3>=27.6088 37   7 Marsh (0.11 0.054 0 0 0.81 0 0 0.027 0 0)
           38) caP100mP50>=8.02 4   1 Baygall (0.75 0 0 0 0 0 0 0.25 0 0) *
           39) caP100mP50< 8.02 33   3 Marsh (0.03 0.061 0 0 0.91 0 0 0 0 0) *
      5) DEMslopeK21>=9.23421 207 117 Ruderal (0.043 0.37 0 0.024 0.092 0 0 0.43 0.039 0)
        10) grP50mP25>=0.0899993 74  41 Fringe (0.12 0.45 0 0.054 0.26 0 0 0.068 0.054 0)
         20) grP90mP10< 0.475 42  11 Fringe (0 0.74 0 0.048 0.048 0 0 0.071 0.095 0) *
         21) grP90mP10>=0.475 32  15 Marsh (0.28 0.062 0 0.062 0.53 0 0 0.062 0 0)
           42) grP50mP10>=0.370001 10   1 Baygall (0.9 0 0 0.1 0 0 0 0 0 0) *
           43) grP50mP10< 0.370001 22   5 Marsh (0 0.091 0 0.045 0.77 0 0 0.091 0 0) *
        11) grP50mP25< 0.0899993 133  48 Ruderal (0 0.32 0 0.0075 0 0 0 0.64 0.03 0)
         22) DEMslopeK21>=53.3116 8   1 Fringe (0 0.88 0 0 0 0 0 0.12 0 0) *
         23) DEMslopeK21< 53.3116 125  41 Ruderal (0 0.29 0 0.008 0 0 0 0.67 0.032 0)
           46) grMAD>=0.07999985 75  34 Ruderal (0 0.4 0 0 0 0 0 0.55 0.053 0)
             92) totDEMslopeK3< 23.22615 9   0 Fringe (0 1 0 0 0 0 0 0 0 0) *
             93) totDEMslopeK3>=23.22615 66  25 Ruderal (0 0.32 0 0 0 0 0 0.62 0.061 0)
              186) grP100mP75Frac< 54.27695 32  15 Fringe (0 0.53 0 0 0 0 0 0.38 0.094 0)
                372) caP50< 6.225 15   2 Fringe (0 0.87 0 0 0 0 0 0.067 0.067 0) *
                373) caP50>=6.225 17   6 Ruderal (0 0.24 0 0 0 0 0 0.65 0.12 0) *
              187) grP100mP75Frac>=54.27695 34   5 Ruderal (0 0.12 0 0 0 0 0 0.85 0.029 0) *
           47) grMAD< 0.07999985 50   7 Ruderal (0 0.12 0 0.02 0 0 0 0.86 0 0) *
    3) muCHM>=3.075415 597 497 Hammock (0.15 0.0084 0.17 0.0017 0.015 0.17 0.17 0.0067 0.15 0.17)
      6) FracGround< 52.19205 395 296 Mixed (0.18 0.0025 0.25 0 0.02 0.25 0.041 0 0.01 0.24)
       12) stdGround< 0.31632 192 100 Hammock (0.021 0 0.48 0 0.031 0.32 0.083 0 0.021 0.047)
         24) std1st< 5.617505 95  19 Hammock (0.011 0 0.8 0 0.042 0.12 0.011 0 0 0.021)
           48) Frac22>=10.34305 77   5 Hammock (0 0 0.94 0 0 0.052 0.013 0 0 0) *
           49) Frac22< 10.34305 18  11 Mixed (0.056 0 0.22 0 0.22 0.39 0 0 0 0.11)
             98) std1st< 3.927075 8   4 Hammock (0 0 0.5 0 0.5 0 0 0 0 0)
              196) caP10>=8.92 4   0 Hammock (0 0 1 0 0 0 0 0 0 0) *
```

```
      197) caP10< 8.92 4   0 Marsh (0 0 0 0 1 0 0 0 0 0) *
     99) std1st>=3.927075 10   3 Mixed (0.1 0 0 0 0 0.7 0 0 0 0.2) *
  25) std1st>=5.617505 97  47 Mixed (0.031 0 0.16 0 0.021 0.52 0.15 0 0.041 0.072)
    50) mu1st< 15.29255 59  39 Mixed (0.034 0 0.27 0 0.034 0.34 0.17 0 0.068 0.085)
     100) stdGround>=0.265778 25  11 Mixed (0.08 0 0.2 0 0 0.56 0 0 0 0.16) *
     101) stdGround< 0.265778 34  23 Hammock (0 0 0.32 0 0.059 0.18 0.29 0 0.12 0.029)
       202) FracGround< 30.4306 14   4 Hammock (0 0 0.71 0 0 0.21 0.071 0 0 0) *
       203) FracGround>=30.4306 20  11 Plantation (0 0 0.05 0 0.1 0.15 0.45 0 0.2 0.05) *
    51) mu1st>=15.29255 38   8 Mixed (0.026 0 0 0 0 0.79 0.13 0 0 0.053) *
 13) stdGround>=0.31632 203 116 Swamp (0.34 0.0049 0.03 0 0.0099 0.19 0 0 0 0.43)
  26) muCHM< 14.0278 57  16 Baygall (0.72 0.018 0.088 0 0.035 0.12 0 0 0 0.018)
    52) Frac22< 12.58255 47   7 Baygall (0.85 0 0 0 0.043 0.085 0 0 0 0.021) *
    53) Frac22>=12.58255 10   5 Hammock (0.1 0.1 0.5 0 0 0.3 0 0 0 0)
     106) grP75mP25Frac< 30.19925 7   2 Hammock (0.14 0.14 0.71 0 0 0 0 0 0 0) *
     107) grP75mP25Frac>=30.19925 3   0 Mixed (0 0 0 0 0 1 0 0 0 0) *
  27) muCHM>=14.0278 146  60 Swamp (0.19 0 0.0068 0 0 0.21 0 0 0 0.59)
    54) p951st>=30.105 22  10 Baygall (0.55 0 0 0 0 0.36 0 0 0 0.091)
     108) stdCanopy>=10.39745 9   1 Baygall (0.89 0 0 0 0 0 0 0 0 0.11) *
     109) stdCanopy< 10.39745 13   5 Mixed (0.31 0 0 0 0 0.62 0 0 0 0.077) *
    55) p951st< 30.105 124  40 Swamp (0.13 0 0.0081 0 0 0.19 0 0 0 0.68)
     110) grP75< 0.525 23  10 Mixed (0.087 0 0.043 0 0 0.57 0 0 0 0.3)
       220) FracShrub< 10.2555 15   4 Mixed (0.067 0 0.067 0 0 0.73 0 0 0 0.13) *
       221) FracShrub>=10.2555 8   3 Swamp (0.12 0 0 0 0 0.25 0 0 0 0.62) *
     111) grP75>=0.525 101  24 Swamp (0.14 0 0 0 0 0.099 0 0 0 0.76)
       222) stdCHM>=6.40041 12   6 Baygall (0.5 0 0 0 0 0.33 0 0 0 0.17)
         444) caP25>=6.19 6   0 Baygall (1 0 0 0 0 0 0 0 0 0) *
         445) caP25< 6.19 6   2 Mixed (0 0 0 0 0 0.67 0 0 0 0.33) *
       223) stdCHM< 6.40041 89  14 Swamp (0.09 0 0 0 0 0.067 0 0 0 0.84) *
7) FracGround>=52.19205 202 115 Sandhill (0.069 0.02 0.0099 0.005 0.005 0.005 0.42 0.02 0.43 0.02)
 14) FracGroundCHM>=13.60545 86  36 Sandhill (0.14 0.047 0.012 0.012 0.012 0 0.13 0.047 0.58 0.023)
  28) stdGround>=0.2362425 15   3 Baygall (0.8 0 0 0 0.067 0 0 0 0 0.13) *
  29) stdGround< 0.2362425 71  21 Sandhill (0 0.056 0.014 0.014 0 0 0.15 0.056 0.7 0)
    58) AAD>=2.79167 15  10 Plantation (0 0.2 0 0 0 0 0.33 0.27 0.2 0)
     116) caP90< 20.03 10   6 Ruderal (0 0.3 0 0 0 0 0 0.4 0.3 0)
       232) caP100mP75Frac< 23.0631 6   3 Fringe (0 0.5 0 0 0 0 0 0 0.5 0)
         464) Frac1st>=93.26195 3   0 Fringe (0 1 0 0 0 0 0 0 0 0) *
         465) Frac1st< 93.26195 3   0 Sandhill (0 0 0 0 0 0 0 0 1 0) *
       233) caP100mP75Frac>=23.0631 4   0 Ruderal (0 0 0 0 0 0 0 1 0 0) *
     117) caP90>=20.03 5   0 Plantation (0 0 0 0 0 0 1 0 0 0) *
    59) AAD< 2.79167 56   9 Sandhill (0 0.018 0.018 0.018 0 0 0.11 0 0.84 0) *
 15) FracGroundCHM< 13.60545 116  43 Plantation (0.017 0 0.0086 0 0 0.0086 0.63 0 0.32 0.017)
  30) caP100mP75< 5.665 105  32 Plantation (0.0095 0 0.0095 0 0 0.0095 0.7 0 0.27 0.0095)
    60) caP75< 10.92 23   0 Plantation (0 0 0 0 0 0 1 0 0 0) *
    61) caP75>=10.92 82  32 Plantation (0.012 0 0.012 0 0 0.012 0.61 0 0.34 0.012)
     122) Frac1stGround< 49.15255 21   1 Plantation (0 0 0 0 0 0 0.95 0 0.048 0) *
     123) Frac1stGround>=49.15255 61  31 Plantation (0.016 0 0.016 0 0 0.016 0.49 0 0.44 0.016)
       246) caP50mP25Frac>=14.07245 23   7 Plantation (0.043 0 0.043 0 0 0 0.7 0 0.17 0.043) *
       247) caP50mP25Frac< 14.07245 38  15 Sandhill (0 0 0 0 0 0.026 0.37 0 0.61 0)
         494) caP100mP75< 2.095 8   0 Plantation (0 0 0 0 0 0 1 0 0 0) *
         495) caP100mP75>=2.095 30   7 Sandhill (0 0 0 0 0 0.033 0.2 0 0.77 0)
           990) caP50mP25< 1.11 4   1 Plantation (0 0 0 0 0 0.25 0.75 0 0 0) *
           991) caP50mP25>=1.11 26   3 Sandhill (0 0 0 0 0 0 0.12 0 0.88 0) *
  31) caP100mP75>=5.665 11   2 Sandhill (0.091 0 0 0 0 0 0 0 0.82 0.091) *
```

Appendix B.3 Integrated Decision Tree

```
n= 1000

node), split, n, loss, yval, (yprob)
      * denotes terminal node

  1) root 1000 900 Baygall (0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1)
    2) muDeriv1064< 0.932652 105  10 Lake (0 0 0 0.9 0.086 0 0 0 0.0095 0)
      4) grP100mP0< 1.535 96   3 Lake (0 0 0 0.97 0.021 0 0 0 0.01 0) *
      5) grP100mP0>=1.535 9   2 Marsh (0 0 0 0.22 0.78 0 0 0 0 0) *
    3) muDeriv1064>=0.932652 895 795 Baygall (0.11 0.11 0.11 0.0056 0.1 0.11 0.11 0.11 0.11 0.11)
      6) NDI578i643< -12.37815 401 301 Baygall (0.25 0.005 0.22 0.0025 0.025 0.24 0.017 0 0.0025 0.24)
      12) vegNDI1515i2216< -16.7446 290 196 Mixed (0.31 0.0034 0.3 0 0.021 0.32 0.024 0 0 0.021)
        24) stdGround>=0.3372265 119  36 Baygall (0.7 0 0.017 0 0.017 0.22 0 0 0 0.05)
          48) NDI1650i1685>=-0.6650525 83   9 Baygall (0.89 0 0 0 0.024 0.048 0 0 0 0.036) *
          49) NDI1650i1685< -0.6650525 36  14 Mixed (0.25 0 0.056 0 0 0.61 0 0 0 0.083)
            98) FracMarshGreen>=37.98185 11   5 Baygall (0.55 0 0.18 0 0 0.091 0 0 0 0.18) *
            99) FracMarshGreen< 37.98185 25   4 Mixed (0.12 0 0 0 0 0.84 0 0 0 0.04)
            198) Frac22>=17.91175 3   0 Baygall (1 0 0 0 0 0 0 0 0 0) *
            199) Frac22< 17.91175 22   1 Mixed (0 0 0 0 0 0.95 0 0 0 0.045) *
        25) stdGround< 0.3372265 171  87 Hammock (0.041 0.0058 0.49 0 0.023 0.4 0.041 0 0 0)
          50) caP100mP0< 20.495 92  19 Hammock (0.033 0 0.79 0 0.043 0.12 0.011 0 0 0)
          100) muVegDeriv2036< 2.246195 81   9 Hammock (0.025 0 0.89 0 0 0.074 0.012 0 0 0)
            200) caP90mP50< 4.155 70   3 Hammock (0.029 0 0.96 0 0 0 0.014 0 0 0) *
            201) caP90mP50>=4.155 11   5 Mixed (0 0 0.45 0 0 0.55 0 0 0 0)
              402) muDeriv1685>=-2.9831 5   0 Hammock (0 0 1 0 0 0 0 0 0 0) *
              403) muDeriv1685< -2.9831 6   0 Mixed (0 0 0 0 0 1 0 0 0 0) *
          101) muVegDeriv2036>=2.246195 11   6 Mixed (0.091 0 0.091 0 0.36 0.45 0 0 0 0)
            202) caP50mP10< 3.145 6   2 Marsh (0.17 0 0.17 0 0.67 0 0 0 0 0) *
            203) caP50mP10>=3.145 5   0 Mixed (0 0 0 0 0 1 0 0 0 0) *
          51) caP100mP0>=20.495 79  22 Mixed (0.051 0.013 0.14 0 0 0.72 0.076 0 0 0)
          102) muDeriv668< 1.87541 8   0 Hammock (0 0 1 0 0 0 0 0 0 0) *
          103) muDeriv668>=1.87541 71  14 Mixed (0.056 0.014 0.042 0 0 0.8 0.085 0 0 0)
            206) NDI1545i2106>=-38.32745 7   4 Hammock (0 0.14 0.43 0 0 0 0.43 0 0 0)
              412) FracGround>=15.3989 4   1 Hammock (0 0.25 0.75 0 0 0 0 0 0 0) *
              413) FracGround< 15.3989 3   0 Plantation (0 0 0 0 0 0 1 0 0 0) *
            207) NDI1545i2106< -38.32745 64   7 Mixed (0.062 0 0 0 0 0.89 0.047 0 0 0)
              414) FracMarshGreen>=61.56465 3   0 Baygall (1 0 0 0 0 0 0 0 0 0) *
              415) FracMarshGreen< 61.56465 61   4 Mixed (0.016 0 0 0 0 0.93 0.049 0 0 0) *
      13) vegNDI1515i2216>=-16.7446 111  20 Swamp (0.09 0.009 0.009 0.009 0.036 0.018 0 0 0.009 0.82)
        26) mu1st< 6.68162 16   7 Baygall (0.56 0.062 0 0.062 0.25 0 0 0 0 0.062)
          52) stdR528>=1.47904 9   0 Baygall (1 0 0 0 0 0 0 0 0 0) *
          53) stdR528< 1.47904 7   3 Marsh (0 0.14 0 0.14 0.57 0 0 0 0 0.14) *
        27) mu1st>=6.68162 95   5 Swamp (0.011 0 0.011 0 0 0.021 0 0 0.011 0.95) *
      7) NDI578i643>=-12.37815 494 394 Ruderal (0 0.2 0.026 0.0081 0.16 0.0081 0.19 0.2 0.2 0.0061)
      14) Frac1st>=93.88985 288 190 Ruderal (0 0.33 0.0035 0.01 0.28 0 0.0035 0.34 0.031 0)
        28) stdDeriv2326< 1.63329 73   8 Marsh (0 0.055 0 0.041 0.89 0 0 0.014 0 0)
          56) muVegDeriv668< 1.86629 6   2 Fringe (0 0.67 0 0.17 0 0 0 0.17 0 0) *
          57) muVegDeriv668>=1.86629 67   2 Marsh (0 0 0 0.03 0.97 0 0 0 0 0) *
        29) stdDeriv2326>=1.63329 215 118 Ruderal (0 0.43 0.0047 0 0.07 0 0.0047 0.45 0.042 0)
```

```
   58) muDeriv528< 5.8747 94  13 Fringe (0 0.86 0.011 0 0.011 0 0.011 0.074 0.032 0)
    116) muVegDeriv578>=-3.248385 88   8 Fringe (0 0.91 0.011 0 0.011 0 0 0.034 0.034 0) *
    117) muVegDeriv578< -3.248385 6   2 Ruderal (0 0.17 0 0 0 0 0.17 0.67 0 0) *
   59) muDeriv528>=5.8747 121  31 Ruderal (0 0.091 0 0 0.12 0 0 0.74 0.05 0)
    118) muDeriv2216< 0.446008 21   7 Marsh (0 0.29 0 0 0.67 0 0 0.048 0 0)
     236) DEMslopeK21>=9.0028 6   1 Fringe (0 0.83 0 0 0 0 0 0.17 0 0) *
     237) DEMslopeK21< 9.0028 15   1 Marsh (0 0.067 0 0 0.93 0 0 0 0 0) *
    119) muDeriv2216>=0.446008 100  11 Ruderal (0 0.05 0 0 0 0 0 0.89 0.06 0)
     238) muR1314< 38.28815 13   7 Sandhill (0 0.23 0 0 0 0 0 0.31 0.46 0)
      476) stdDeriv643< 1.728095 7   3 Ruderal (0 0.43 0 0 0 0 0 0.57 0 0)
       952) caP100mP75Frac< 23.3621 3   0 Fringe (0 1 0 0 0 0 0 0 0 0) *
       953) caP100mP75Frac>=23.3621 4   0 Ruderal (0 0 0 0 0 0 0 1 0 0) *
      477) stdDeriv643>=1.728095 6   0 Sandhill (0 0 0 0 0 0 0 0 1 0) *
     239) muR1314>=38.28815 87   2 Ruderal (0 0.023 0 0 0 0 0 0.98 0 0) *
15) Frac1st< 93.88985 206 114 Plantation (0 0.0097 0.058 0.0049 0.0049 0.019 0.45 0.0097 0.43 0.015)
  30) NDI1585i1730< -0.3178745 112  30 Plantation (0 0 0.062 0 0.0089 0.027 0.73 0 0.17 0)
   60) stdVegDeriv643< 1.466715 82   8 Plantation (0 0 0.012 0 0.012 0.024 0.9 0 0.049 0) *
   61) stdVegDeriv643>=1.466715 30  15 Sandhill (0 0 0.2 0 0 0.033 0.27 0 0.5 0)
    122) CanopyDensity>=63.9873 6   0 Hammock (0 0 1 0 0 0 0 0 0 0) *
    123) CanopyDensity< 63.9873 24   9 Sandhill (0 0 0 0 0 0.042 0.33 0 0.62 0)
     246) caP100mP50< 4.175 8   1 Plantation (0 0 0 0 0 0 0.88 0 0.12 0) *
     247) caP100mP50>=4.175 16   2 Sandhill (0 0 0 0 0 0.062 0.062 0 0.88 0) *
  31) NDI1585i1730>=-0.3178745 94  24 Sandhill (0 0.021 0.053 0.011 0 0.011 0.11 0.021 0.74 0.032)
   62) FracMixedDark>=6.00907 12   8 Hammock (0 0.17 0.33 0 0 0.083 0.17 0 0 0.25)
    124) Frac22>=21.269 4   0 Hammock (0 0 1 0 0 0 0 0 0 0) *
    125) Frac22< 21.269 8   5 Swamp (0 0.25 0 0 0 0.12 0.25 0 0 0.38) *
   63) FracMixedDark< 6.00907 82  12 Sandhill (0 0 0.012 0.012 0 0 0.098 0.024 0.85 0)
    126) vegNDI1104i1249< -5.840945 5   2 Plantation (0 0 0 0.2 0 0 0.6 0.2 0 0) *
    127) vegNDI1104i1249>=-5.840945 77   7 Sandhill (0 0 0.013 0 0 0 0.065 0.013 0.91 0) *
```