MAPPING PLANT DIVERSITY AND COMPOSITION ACROSS NORTH CAROLINA PIEDMONT FOREST LANDSCAPES USING LIDAR-HYPERSPECTRAL REMOTE SENSING

Christopher R. Hakkenberg

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Ecology in the Curriculum for the Environment and Ecology.

Chapel Hill
2017

Approved by:

Conghe Song

Robert K. Peet

Dean L. Urban

Peter S. White

Aaron Moody

Xiaodong Chen

ABSTRACT

Christopher R. Hakkenberg: Mapping plant diversity and composition across North Carolina Piedmont forest landscapes using LiDAR-hyperspectral remote sensing
(Under the direction of Conghe Song)

Forest modification, from local stress to global change, has given rise to efforts to model, map, and monitor critical properties of forest communities like structure, composition, and diversity. Predictive models based on data from spatially-nested field plots and LiDAR-hyperspectral remote sensing systems are one particularly effective means towards the otherwise prohibitively resource-intensive task of consistently characterizing forest community dynamics at landscape scales. However, to date, most predictive models fail to account for actual (rather than idealized) species and community distributions, are unsuccessful in predicting understory components in structurally and taxonomically heterogeneous forests, and may suffer from diminished predictive accuracy due to incongruity in scale and precision between field plot samples, remotely-sensed data, and target biota of varying size and density. This three-part study addresses these and other concerns in the modeling and mapping of emergent properties of forest communities by shifting the scope of prediction from the individual or taxon to the whole stand or community. It is, after all, at the stand scale where emergent properties like functional processes, biodiversity, and habitat aggregate and manifest. In the first study, I explore the relationship between forest structure (a proxy for successional demographics and resource competition) and tree species diversity in the North Carolina Piedmont, highlighting the empirical basis and potential for utilizing forest structure from LiDAR in predictive models of tree species diversity. I

then extend these conclusions to map landscape pattern in multi-scale vascular plant diversity as well as turnover in community-continua at varying compositional resolutions in a North Carolina Piedmont landscape using remotely-sensed LiDAR-hyperspectral estimates of topography, canopy structure, and foliar biochemistry. Recognizing that the distinction between correlation and causation mirrors that between knowledge and understanding, all three studies distinguish between prediction of pattern and inference of process. Thus, in addition to advancing mapping methodologies relevant to a range of forest ecosystem management and monitoring applications, all three studies are noteworthy for assessing the ecological relationship between environmental predictors and emergent landscape patterns in plant composition and diversity in North Carolina Piedmont forests.

For Vida

ACKOWLEDGEMENTS

I am humbled to have been able to surround myself with so many fine people these past few years. No doubt, this dissertation project would not have been possible without the support of my advisor, Conghe Song, and a dedicated dissertation committee straddling the realms of Biology and Geography. I'm constantly reminded not just of the quantity of their combined knowledge, but of the quality of their guidance and mentorship. They gave me great independence to explore, to fail, to question my assumptions, and grapple with the implications of my conclusions. I am truly indebted to them – both for their intellectual contributions to the field which established the theoretical foundation from which this work could commence and, selfishly, for their commitment to me, and the goals of this project. Indeed, there is a degree of solemnity in this academic tradition, not just in passing the torch of knowledge, but in imparting the skills with which to wield it.

My classmates and colleagues, with whom I broke bread and corralled countless dogs, were another vibrant source of inspiration. These include Matt Dannenberg, Clare Fieseler, Bianca Lopez, Chris Payne, Dennis Tarasi, Sam Tessel, Sierra Woodruff, and Qi Zhang. Others, informal mentors and collaborators like Jim Clark, Josh Gray, Reuben Valbuena, Alan Weakley, Jack Weiss, and Kai Zhu selflessly offered their time and expertise. And while responsibility for the errors in this document should be placed squarely at my feet, where there is success, it is shared.

This work was partly supported under a NASA Earth and Space Science Fellowship, the Center for the Study of the American South, a Kevin Satisky and Judith Thorn Summer Research Fellowship, and several other sources of funding. In addition, I would like to thank the USDA Forest Service for supplying Forest Inventory and Analysis (FIA) data, Bruce D. Cook and others

at NASA Goddard for provision of G-LiHT data, the Duke Forest field office for facilitating access to the Blackwood Division, as well as field assistants for the field sampling campaign: Jeffery Berger, Joshua Bradley, Per Frisk, Alec Keiper, Nicole Murnighan, Matthew Seibert, and Stephanie Trillo-ordonez. Lastly, but most importantly, my profound thanks go out to my adventurous and hilarious family. I feel enveloped by their love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The pace of species loss in the Anthropocene has far surpassed background extinction rates, and is only forecasted to accelerate in the near future (Davis & Shaw 2001; Hooper et al. 2012). Perhaps nowhere is this pernicious trend more evident than in the world's forests. Forests cover approximately 30% of the global land surface, yet account for almost half of terrestrial carbon, and three quarters of terrestrial biodiversity (Kindermann et al. 2008; FAO 2010). But in recent decades, global forests have been profoundly modified, degraded, and destroyed due to the synergistic anthropogenic forces of habitat change, climate change, overexploitation, invasive alien species, and pollution (MEA 2005; Rudel et al. 2005). When these interacting stressors overwhelm ecosystem resilience to environmental and climatic variability, forest ecosystems are at increased risk of harm, including ecological regime shifts and hyperdynamism in ecosystem process rates (Laurance 2002; Walther et al. 2002). Studies have demonstrated the central role of plant diversity for maintaining stability in ecosystem functioning in forest landscapes effected by anthropogenic degradation (Hooper et al. 2002; Folke et al. 2004; Civitello et al. 2015). In fact, a fundamental extrinsic value of biodiversity lies in its potential to provide functionally redundant system components (e.g. species) to regulate ecosystem processes and buffer against anthropogenically-induced environmental change (Elmqvist et al. 2003).

In forest ecosystems plants are, perhaps not surprisingly, the dominant lifeform and driver of ecosystem processes (Diaz & Cabido 2001; McGill et al. 2006). Not only do plants comprise the vast majority of biomass, they are likewise a key indicator of overall diversity by providing

habitat and serving as nutritional resources for specialist and generalist consumers throughout the trophic chain (Gaston 1992; Myers et al. 2000; Raxworthy et al. 2003; Zak et al. 2003). The central role of plants in constraining forest ecosystem functioning as well as taxonomic, phylogenetic, functional, and biochemical diversity has driven global efforts to model aggregate, emergent properties of forest communities like diversity, species composition, canopy structure, and fucntional traits for applications ranging from landscape-scale ecosystem management, conservation, and habitat modeling, to regional and global modeling of biodiversity and ecosystem function (Gillespie et al. 2008; Abelleira Martinez et al. 2016; Jetz et al. 2016).

Field-based methods for mapping forest community properties can be highly accurate for detecting species or characterizing communities at small spatial scales, however they tend to be extremely limited in extent (Condit 1995). For this, correlative species distribution models (SDMs) and community distribution models (CDMs), have gained prominence in recent years as a means towards modeling forest community dynamics at landscape to global scales (Ferrier & Guisan 2006; Pearman et al. 2008; Elith & Leathwick 2009). These models exploit statistical relationships between species/community occurrence and environmental conditions constraining the niche space of focal taxa to make predictions of expected distributions based on known abiotic conditions (Evans et al. 2011; Cord et al. 2014). Predictive models based purely on abiotic GIS layers like topography or climate attempt to translate species' fundamental niche space - an abstract conception of the abiotic conditions required to maintain positive population growth rate - as a probabilistic and spatially-explicit geographic entity (Pearman et al. 2008). In other words, niche models predict where species should be, rather than where they actually are - their realized niche - due to biotic interactions and stochastic events like disturbance. However, if the assumption of

niche conservatism underlying these models is flawed and biotic components completely ignored, niche modeling based purely on coarse abiotic layers may prove inadequate or even misleading.

One effective method to return biological realism to niche modeling is through the incorporation of remotely-sensed data to simultaneously identify underlying environmental gradients constraining the fundamental niche, and directly observe the presence of specific entities occupying their realized niche (Guisan & Zimmermann 2000; Cord et al. 2013). Though remotely-sensed data may be limited in accuracy and precision due to technological and logistical limitations in data acquisition, processing, and analysis, it nonetheless excels for providing a consistent, repeatable, and synoptic - or "wall-to-wall" - census of biophysical factors within the scene extent. In fact, in many cases remotely-sensed data may surpass field-based studies for describing large-scale forest patterns susceptible to spatially-structured environmental filters otherwise concealed amid the complexity of local-scale forest dynamics (Asner et al. 2015).

Remote sensing, and specifically multi-band optical remote sensing, provides spectrally-resolved data across the electromagnetic spectrum on the reflectance characteristics of the target medium (Song et al. 2015). In the case of continuous cover forest vegetation, aerial or satellite sensors detect spectral reflectance signatures based on narrowband light absorption and scattering features of the forest canopy, including foliage and nonphotosynthetic vegetation (NPV) like woody stems and branches (Asner 1998). These spectral signatures reflect the net outcome of plant morphology, biochemistry, and structure that can be diagnostic of species traits and community properties in the context of local environmental conditions (Curran 1989). Provided structural variation (Knyazikhin et al. 2013) or intra-specific trait disparities (Price 1994) don't overwhelm the signal, reflectance spectra can be an effective means for detecting a host of upper canopy features including the composition of emergent individuals and the impact of environmental stress

(Ustin & Gamon 2010; Ollinger 2011). However, in structurally-complex, continuous-cover forests, optical imagery fails to provide robust data beyond (or below) the upper canopy, and thus offers little to no information on the composition and structural properties of individuals obscured from the sensor's view. To correct for this lacuna and model the entirety of a stand's structural elements, from the canopy surface to the shadowed understory, LiDAR has found increasing prominence as a compliment to optical imagery in the remote sensing of vegetation.

Unlike passive optical imagery, active remote sensing systems like LiDAR employ laser pulses that bounce off elements throughout the canopy. By calculating the pulse's return time to the sensor, the height of forest structural elements can be estimated (Lesky et al. 2002). LiDAR's ability to provide detailed and accurate data on the three-dimensional structure of the forest environment has had a profound effect on the remote sensing of vegetation (Kampe et al. 2010; Asner et al. 2012; Cook et al. 2013). This advancement is not merely a case of technology driving the science. Instead, it reflects a long-held recognition that structure, along with function and composition, is one of the primary components of forest ecosystems (Noss 1990; Franklin et al. 2002; McElhinny et al. 2005), and a particularly expedient method for quantifying forest temporal dynamics like stand age, successional stage, and other demographic trends (Peet & Christensen 1980; Peet & Christensen 1987; Oliver & Larson 1996; Smith et al. 1997). As a proxy for characterizing a stand's local history and the spatial distribution of resources, forest structure - and structural heterogeneity in particular – also serves as an effective predictor of local levels of biodiversity. In fact, several studies have found significant correlations between attributes of forest structure and local biodiversity among trees (Neumann & Starlinger 2001; Chiarucci & Bonini 2005; Bacaro et al. 2008), understory herbs (Berger & Puettmann 2000; Cook 2015), birds

(MacArthur & MacArthur 1961; Moen & Gutiérrez 1997), insects (Recher et al. 1996), and mammals (Sullivan et al. 2001).

Another potential shortcoming with niche modeling, and with SDMs in particular, worth reviewing is the inherent incongruity in the characteristic scales of field plot data (based on sampling size, design, and intensity), remotely-sensed data (such as, spatial resolution and geolocational precision), and that of target biota (e.g. size, density, and coverage extent). This mismatch in scale need not necessarily be problematic. For example, in monospecific forest communities, or in open woodlands where individual tree crowns are spatially distinct and discrete, remotely-sensed pixels may represent a single endmember; that is, pixels may consist entirely of a single plant of sufficient size or cover. However, when scene elements are smaller than the pixel resolution, or when target biota are diverse, intertwined, or otherwise obscured from view, the task of reconciling the disparate scales of data and model may become problematic. Spectral mixing and spatial mismatch between data sources is especially pervasive in structurally and taxonomically heterogeneous forests.

Thus, to better align grain and extent among field plots, remotely-sensed data, and target biota in multi-strata multi-species communities, studies have shifted the scale of interest from the individual or taxon to that of the whole stand or community (Bunting et al. 2010; Wolter & Townsend 2011). In complex forests, this means foregoing a species-centric approach and instead embracing community-scale models (Ferrier & Guisan 2006). After all, it is at these aggregate scales that the emergent properties of communities such as functional processes, biodiversity, and habitat manifest (McGill et al. 2006; Anderson & Ferree 2010). This alignment of disparate scales likewise mitigates against ecological fallacies that may occur when scale mismatch confounds inference (Ess & Sudweeks 2001). For example, when used to reconstruct entire communities,

stacked SDMs (multiple, single-species models), tend to underestimate biotic processes like competition that contribute to observed compositional variance in plot data, and thereby grossly inflate species richness values (Guisan & Rahbek 2011; Clark et al. 2014). Provided predictive models adequately represent the full spectrum of compositional and structural variance observed in field plots, a stand-scale approach enables researchers to model communities holistically, thereby circumventing the problem of failing to predict obscured, and thus undetected, individuals like understory herbs (Ohmann et al. 2011).

Broadly speaking, this study was motivated by the desire to bridge scales of analysis of ecological phenomena from the plot to the landscape, with the ultimate goal to contribute to the better integration of the fields of community ecology, landscape ecology, and remote sensing. Analyses are intended to uncover phenomenological patterns and underlying drivers of forest community properties to map turnover in vegetation composition and diversity across forest landscapes of the North Carolina Piedmont. Spatially nested field sampling, together with burgeoning technologies of LiDAR-hyperspectral remote sensing and emerging statistical techniques like nonparametric predictive community modeling, allow for forest community mapping at scales far finer than that of the nominal resolution of the remotely-sensed data, and at extents far larger than what field sampling would expediently support. Efforts to predict and infer the scale-dependent relationships among remotely-sensed parameters and forest community properties will only grow in the years to come as ecologists anticipate annual NEON data products and the deployment of a fleet of new LiDAR and hyperspectral satellites for use in applications ranging from local-scale land management to global-scale ecosystem modeling.

The study is split into three parts. In the first section (Ch. 2), I seek to determine whether forest structure can serve as a significant predictor of tree species diversity in the forests of the

North Carolina Piedmont. If so, which structural attributes are most strongly correlated with diversity, and how effective are they when used in concert within a generalized predictive model? To accomplish this goal, I analyze Spearman correlations between 15 measures of forest structure and five indices of tree species diversity based on a set of 972 geographically-distributed Forest Inventory and Analysis (FIA) plots located within North Carolina Piedmont forests (Gray et al. 2012). Next, I combine all predictors in a nonparametric support vector regression (SVR) model to predict tree species diversity in the North Carolina Piedmont based on structure alone, without accounting for other known predictors of diversity, such as environment, soil conditions, and site history. Beyond the theoretical implications of unraveling the underlying relationship between structure (as a surrogate for successional stage) and tree species diversity, this study is designed and implemented to determine the empirical basis for successfully utilizing forest structure from LiDAR remote sensing in predictive models of tree species diversity over large geographic regions.

In the second section (Ch. 3), I employ lessons from Ch. 2 to map multi-scale vascular plant species richness in a compositionally- and structurally-complex North Carolina Piedmont forest landscape. Specifically, I use spatially-nested field plots in conjunction with active and passive remotely-sensed data from the Goddard LiDAR, hyperspectral, thermal (G-LiHT) airborne sensor in feature-selected nonparametric models to predict wall-to-wall vascular plant species richness in the Duke Forest Blackwood Division study site at $0.01m^2$, $0.1m^2$, $1m^2$, $10m^2$, $100m^2$, $400m^2$, and $900m^2$ scales. Due to the inherent scale-dependence in diversity patterns (e.g. species-area relationships) and those of remotely-sensed data (e.g. resolution and spatial precision) in relation to plant size and density, post-hoc analyses focus explicitly on the role of spatial scale in constraining the relationship between remotely-sensed variables and species richness values. In

addition to spatially-explicit uncertainty maps and accuracy assessment of nonparametric models, hierarchical parametric models are used to infer the relationship between remotely-sensed data and drivers of diversity. This study provides insights into multi-scale landscape turnover of plant diversity, species-area relationships, and remotely-sensible correlates of plant richness.

In the final section (Ch. 4), I use multi-scale nested vegetation sampling data in conjunction with remotely-sensed data from the G-LiHT airborne sensor to map plant composition in the Duke Blackwood forest landscape. I compare two distinct approaches to the mapping of forest composition: community-unit classification, where communities are mapped as discrete geographical entities at varying compositional resolutions, and compositional gradient regression, where intergrading assemblages are depicted as continuous compositional gradients. In so doing, I provide a mapped example of the community-continua concept. For both the discrete and continuous compositional maps, landscape turnover in composition is consistently represented as the RGB color translation of its coordinates in ordination space. Finally, I employ post-hoc analysis of remotely-sensed predictor variables to determine the complementarity of LiDAR-hyperspectral sensors to predict the primary dimensions of vascular plant composition of the North Carolina Piedmont forest site.

CHAPTER 2

FOREST STRUCTURE PREDICTS TREE SPECIES DIVERSITY IN THE NORTH CAROLINA PIEDMONT*

2.1. Introduction

Concerns over global environmental change and biodiversity loss have driven efforts to model the spatial distribution of taxonomic diversity (Ackerly et al. 2010; Hooper et al. 2012). Despite significant progress in recent decades in the modeling of large-scale patterns in biodiversity, direct mapping is still greatly limited by technological and cost constraints (Gaston 2000; Rodrigues & Brooks 2007; Asner & Martin 2009). To improve the estimation of the spatial distribution of diversity over large continuous areas, scalable proxy variables that link ground measurements with remotely-sensed data must be identified and assessed (Kreft & Jetz 2007; Kane et al. 2010; He et al. 2015). Candidate variables should be remotely sensible, ecologically relevant, common in current vegetation inventory databases, as well as temporally dynamic and scalable to larger landscapes (Turner et al. 2003; Anderson & Ferree 2010). This study investigates the utility of employing one such suite of candidate variables, namely those based on forest structure, to predict tree species diversity over large areas of temperate forest.

Forest structure reflects abiotic conditions and site history, including competition and stochastic disturbance events, at multiple spatial scales that affect the three-dimensional distribution of biomass in a forest stand. Several potential mechanisms underlie the relationship

* This chapter previously appeared as an article in The Journal of Vegetation Science. The original citation is as follows: Hakkenberg, C.R., Song, C., Peet, R.K., & White, P.S. 2016. Forest structure as a predictor of tree species diversity in the North Carolina Piedmont. *Journal of Vegetation Science* 27: 1151–1163.

between forest structure and tree species diversity, though no single explanation is diagnostic in assigning causality. On one hand, higher species counts increase the expectation of greater functional diversity in species' traits. On the other hand, structural complexity begets fine-scale environmental heterogeneity conducive to the partitioning of niche space and the creation of new habitat for a greater diversity of species (Franklin 1988; Palmer & Maurer 1997). These explanations are by no means exhaustive, and without experimental manipulation, teasing apart the causal mechanisms driving these relationships remains problematic.

Observational studies on the effect of stand structure on plant diversity have primarily focused on understory herbaceous species (see Cook 2015 for a review), and to a lesser extent, on tree species diversity. Chiarucci and Bonini (2005), for example, found plot-scale species richness to be significantly related to tree stem density (R^2=0.35; $p<0.01$) and total basal area (R^2=0.05; $p<0.01$), though non-structural, regional predictors like elevation had the highest predictive power ($R^2 = 0.64$). Bacaro et al. (2008) were able to use a host of local and regional predictors to explain 83% of the total variation in woody plant species richness in Tuscan forests, though forest structure accounted for only 1% of non-shared variation in woody richness. Other studies, mainly based on small sample sizes, have tended to find an insignificant or weak correlation between forest structure and woody diversity (Aiba & Kitayama 1999; Neumann & Starlinger 2001). In light of these inconsistent findings and the emergence of remote-sensing technologies like LiDAR as tools for forest biodiversity mapping based on structural surrogates (Rodrigues & Brooks 2007; Wolf et al. 2012; Camathias et al. 2013), further work is needed to evaluate the empirical basis of forest structure as a generalizable predictor of tree species diversity.

In this study, we use an extensive set of Forest Inventory and Analysis (FIA) plots, geographically distributed across the North Carolina Piedmont region (47500 km^2) to determine

whether forest structure is significantly correlated with tree species diversity in the temperate forests of this region. We then examine which structural attributes are most strongly correlated with tree species diversity, and how effective they are when used in concert in a generalized predictive model. To address these questions, we first focus on characterizing the relationship between individual structural attributes and indices of species diversity. Next, we assess the ability of a generalized machine learning algorithm trained solely with structural variables to predict tree species diversity. Finally, we evaluate cross-validated model performance using the entire NC Piedmont plot database as well as subsets differentiated by stand origin and forest type.

2.2. Methods

2.2.1. Study Area

The study area is located in the central portion of the EPA level 3 "Piedmont" ecoregion and bounded by the state borders of North Carolina (Fig. 2.1) (Griffith et al. 2002). The heavily-forested Piedmont ecoregion, stretching from Northern Virginia to central Alabama, separates the mountainous Appalachians to the northwest from the flat Coastal Plain to the southeast. Once largely dominated by agriculture and grazing lands

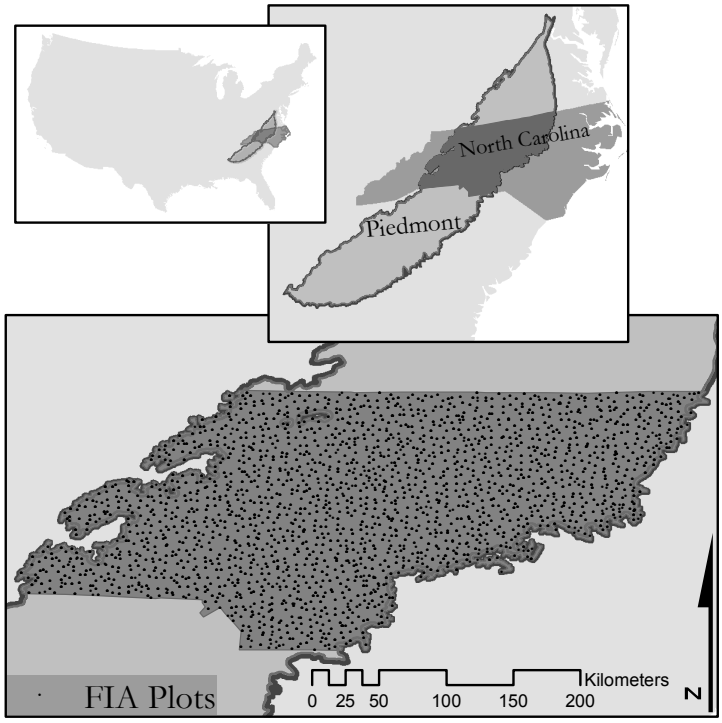


Figure 2.1. Piedmont NC study area and FIA plots.

(including upland hardwood forests), the NC Piedmont has largely reverted to old-field successional pine and hardwood forest (Peet & Christensen 1988).

Plant species richness levels in the NC Piedmont tend to be strongly correlated with soil nutrient content, soil moisture and parent material (Peet & Christensen 1980; Peet & Christensen 1988; Peet et al. 2014). At local scales (100-1000 m^2), the highest levels of species richness occur in riparian communities, in moist but well-drained sites (Matthews et al. 2011). Upland forest species richness, on the other hand, is driven primarily by soil chemistry and especially base cations, phosphorus availability, and soil moisture (Peet & Christensen 1980). Despite the large spatial extent of the study area, the narrow range of variability in elevational and latitudinal range limits the influence of climatic factors, thereby allowing us to focus exclusively on the relationship between structure and tree diversity (National Climatic Data Center 2011).

2.2.2. Forest Inventory Data

Vegetation plot data were drawn from the USDA Forest Service's Forest Inventory and Analysis (FIA) database, the largest and most comprehensive forest inventory and monitoring program in the US, with remeasurement of all plots occurring on a 5-10 year cycle. (Gray et al. 2012). Under the nationally-consistent inventory design, plot locations are geographically-distributed throughout a base grid, with individual plots selected to assign one plot for each 2428 ha

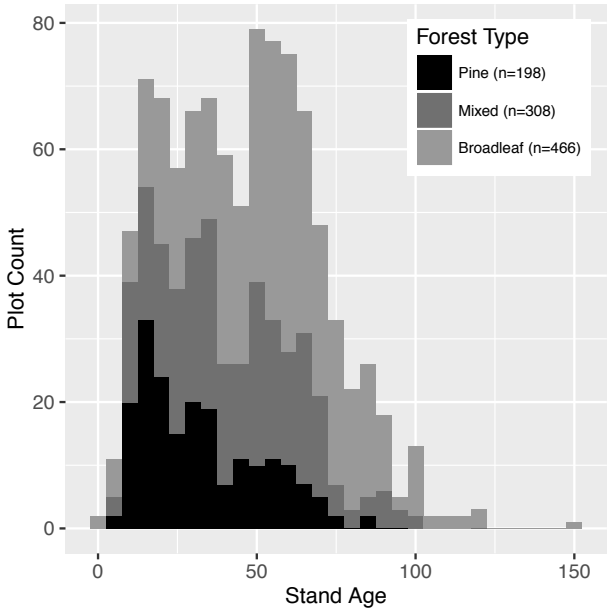


Figure 2.2. Histogram of FIA plots by stand age and forest type

hexagon across the US (O'Connell et al. 2015). Within each plot, trees (>12.7 cm DBH) were measured on each of four 7.3m radius subplots (O'Connell et al. 2015). Stem counts were converted to a per-unit-area metric using the FIA's subplot expansion factors, at a spatial scale capable of capturing the range of variability in stand structure (Clebsch & Busing 1989; Busing & White 1993; Bechtold & Patterson 2005). Due to concerns over landowner privacy, as well as plot integrity and vandalism of plot locations on public lands, the precise locations of FIA plots are "fuzzed", such that plot locational precision is limited to one mile, and "swapped", where up to 20% of private plot location coordinates are swapped with similar plots within the same county (O'Connell et al. 2015). To ensure that plot locations retain their anonymity while still conforming to the specific spatial extent of our study area, the Forest Service assisted in the provision of the final plot list used in this analysis (USDA 2014, database accessed July, 2015).

After removing plots experiencing timber removal, 972 plots from the most recent remeasurement period (2005-2013) were retained for analysis – containing 89 tree species and across a range of stand ages (Fig. 2.2, Appendix 1). Thereafter, the full dataset was subset according to stand origin and forest type. Stand origin refers to FIA designated "natural stands" (837 plots) and those with "clear evidence of artificial regeneration … by planting or artificial seeding" (135 plots) (O'Connell et al. 2015). Forest types represent dominant leaf habit and are derived from a parsimonious three-way split among FIA's forest type groups, a designation of community type based on a hierarchical clustering algorithm (O'Connell et al. 2015). FIA designated forest types used in this study include: (1) 198 pine plots, 97% of which are in the *Pinus taeda*/*Pinus echinata* group; (2) 308 broadleaf plots, with 88% in the *Quercus*/*Carya* group; and (3) 466 mixed plots, representing a broad range of taxa (Fig. 2.2, Appendix 1). Nomenclature follows (USDA 2016).

Table 2.1. Indices of taxonomic diversity

Species Richness (SR)	$SR = S$	where *S* is the number of species	(2.1)
Shannon Index (H)	$H = -\sum_{i=1}^{S} p_i \ln p_i$	where S is the number of species, and p_i is the proportion of individuals (p) belonging to the ith species (Shannon & Weaver 1949)	(2.2)
Simpson index (D)	$D = 1 - \sum_{i=1}^{s} p_i^2$	a variation of Simpson's index (Simpson 1949) where p_i is the proportion of individuals in a community of s species	(2.3)
Pielou's J (J)	$J = \frac{H}{H_{max}}$	where H is the Shannon Index value and H_{max} equals the maximum possible value of H for a given sample (Pielou 1966)	(2.4)
Rarified SR	SR per 24 stems (i.e. median plot density)	rarefaction and extrapolation calculated using the iNEXT package in R (Hsieh et al. 2016)	(2.5)

2.2.3. Indices of Taxonomic Diversity

We used five commonly-used and ecologically-interpretable indices of taxonomic diversity, that each emphasize a different aspect of species diversity (Peet 1974; Magurran 1988): species richness (SR), the Shannon's *H* and Simpson's *D* indices of entropy, Pielou's *J* measure of evenness, and rarified species richness (rarified SR) (Table 2.1). Despite the parsimony in tallying species for a given area, species richness fails to account for the distribution of relative abundances, making inference and comparison across communities of varying densities problematic. Therefore the Shannon index (*H*) - which is more sensitive to rarer species - and the Simpson index (*D= 1-Simpson's 1949 index;* Simpson 1949) - which responds more to abundant species - were included to represent two points in a spectrum of relative sensitivity to species number versus relative evenness (Hill 1973; Peet 1974; Heip et al. 1998). Pielou's J (*J*; Pielou 1966) was used to approximate evenness of species presence (Jost 2010).

Rarified SR is included to provide an estimate of species richness independent of a density-dependent sampling effect, a statistical artifact whereby the more individuals sampled, the greater

the number of expected species (Bunge & Fitzpatrick 1993). Rarefaction enables a comparison of species richness between stands of varying density by re-sampling individuals to simulate species accumulation curves at equivalent stem density (Gotelli & Colwell 2001). Rarified SR values were calculated from a coverage-based sampling curve based on a Monte Carlo re-sampling procedure on all plots (Chao & Jost 2012). Interpolated (rarified) and extrapolated (predicted) richness values were then estimated from this sampling curve based on an expectation of equivalent density corresponding to the median density of all plots (i.e. 24 individual stems or 0.14 stems per m^2) (Chao & Jost 2012).

2.2.4. Indices of Forest Structure

Structural indices quantify attributes such as abundance and size variation in the standing biomass in the horizontal plane (e.g. stem density and basal area), as well as the vertical dimension (e.g. canopy heights, foliar profile, and stratification) (Davis & Roberts 2000; Gadow et al. 2012). A host of numerical methods exist for quantifying and indexing stand structure, though no single authoritative set of criteria exists, making direct comparison problematic (Neumann & Starlinger 2001; Staudhammer & LeMay 2001). For this study, we selected three classes of easily-measured structural metrics with precedence in the forest ecology literature (Lexerød & Eid 2006): (1) summary statistics, (2) size heterogeneity, and (3) size distribution statistics (Table 2.2).

Table 2.2. Indices of forest structure. ("BA" - basal area; "HT" - height; "qmd" - quadratic mean diameter; "ST" - stems; "-" – unitless)

Index Class	Attribute	Abbreviation	Unit	Mean	Sd	Min	Max
Summary statistics	Maximum	maxBA	cm^2	1748.68	1227.37	136.85	9816.88
		maxHT	m	26.5	6.67	7.32	47.24
	Mean	qmd	cm^2	25.55	6.77	12.95	72.26
		meanHT	m	18.27	3.86	6.9	33.84

	Density	densityST	stems/m^2	0.04	0.02	0	0.15
Size Heterogeneity	Coefficient of variation (cv)	cvBA	-	0.76	0.3	0.02	2.36
		cvHT	-	0.24	0.09	0.01	0.6
	Gini Coefficient (GC)	GCBA	-	0.36	0.11	0.01	0.68
		GCHT	-	0.13	0.05	0	0.28
Size Distributions Statistics	Skewness	skewBA	-	1.25	0.86	-0.86	6.35
		skewHT	-	0.1	0.63	-2.83	3.53
	Kurtosis	kurtosisBA	-	1.45	4.08	-2.75	42.17
		kurtosisHT	-	-0.56	1.7	-2.75	15.4
	Weibull (WB)	WBshape	-	1.98	3.79	0.65	107.84
		WBscale	-	589.03	350.98	134.17	5257.4

Basic summary statistics of forest structure include those based on individual stems such as maximum height (maxHT) and maximum basal area (maxBA), and those based on plot-wide statistics such as quadratic mean diameter (qmd), mean height (meanHT) and stem density (densityST). Two measures of structural heterogeneity were selected to indicate dispersion in basal area and height – attributes that have been found to correlate with stand micro-habitats (Acker et al. 1998) and distinguish between successional stages (Spies & Franklin 1991). The coefficient of variation (cv) was chosen owing to its ability to measure relative rather than absolute variation, thereby allowing direct cross-comparison among stands (Weiner & Thomas 1986). The Gini coefficient (GC) is a measure of total inequality in a stand, or the "relative mean difference", calculated as:

$$GC = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\mu} \quad (2.6)$$

where x_i and x_j represent individual stems i and j, respectively; n denotes total individuals and μ represents mean stem size (Damgaard & Weiner 2000). GC ranges from zero, when all

individuals are equal in size, to a theoretical maximum of one when a stand is dominated by a single individual. Being a statistic of dispersion normalized by the stand mean, GC has the desired property of being independent of density and total BA, thus allowing for comparison between stands (Knox et al. 1989; Valbuena et al. 2012).

The final class of stand structural measures describes higher order moments of the stands' size distribution, a common indirect method for estimating life table information and stand age structure (Lorimer & Krug 1983). Kurtosis, which estimates the degree of peakedness in the distribution, indicates the extent to which a stand is dominated by modal size classes that loosely track the distribution of age cohorts. Skewness, which measures the degree of asymmetry in the distribution about its mean, has been associated with differences in degree of competition (Knox et al. 1989). We also adopt the Weibull function to describe stand size distribution for its flexible statistical properties and its long history in forestry applications (Weibull 1951; Leak 1964; Bailey & Dell 1973). While there is not a distinct biological basis for this function, it derives its utility from its superior performance in fitting a wide variety stand-size distributions, despite having just two parameters (Rennolls et al. 1985; Jaworski & Podlaski 2011). For a distribution starting at zero, the Weibull probability distribution function is expressed as:

$$f(x;\lambda;k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{k}\right)^{k}} & x \geq 0, \\ 0 & x < 0, \end{cases} \tag{2.7}$$

where x is the size or age class, λ is the scale parameter, and k is the shape parameter such that $k>0$ and $\lambda>0$ (Weibull 1951). The scale parameter controls the extent of the horizontal displacement of the mean value, while the shape parameter determines the form the curve assumes. For $k<1$, f(x) resembles a negative monotonic reverse-J curve. As k increases, the distribution transforms to positively-skewed unimodal ($1 < k < 3.6$), approximately normal ($k = 3.6$), and finally

to a negatively-skewed unimodal distribution ($k > 3.6$). Parameters of the Weibull distribution and their corresponding standard errors were estimated by non-parametric bootstrapping. These three measures of the stand size distribution are attractive because, as continuous variables, they circumvent the issue of information loss and dependency on subjective choice via thresholding and size-class binning (Staudhammer & LeMay 2001; Lexerød & Eid 2006; Valbuena et al. 2012).

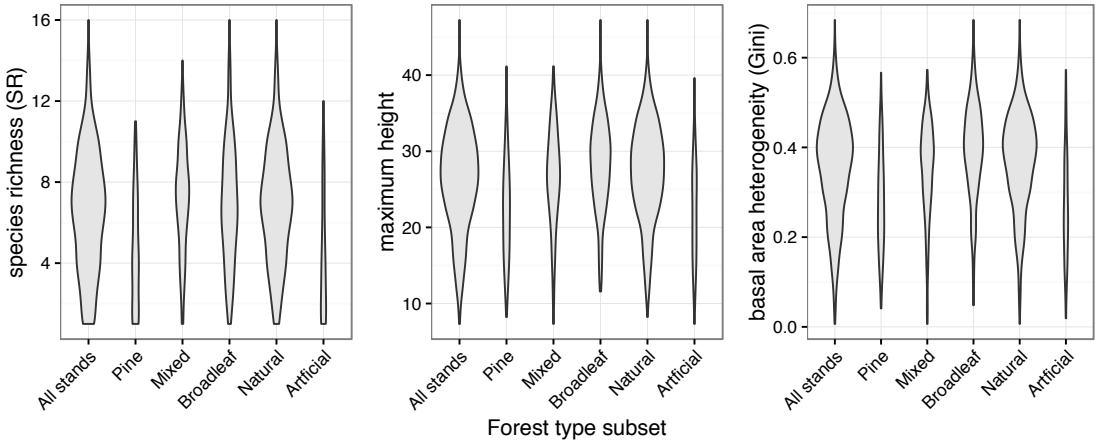


Figure 2.3. Distribution of selected response and predictor variables among all plots and subset by forest type and regeneration method. Violin plots depict distribution of field plots by species richness (SR), maximum height, and basal area heterogeneity (Gini).

2.2.5. Data Analysis

Structural predictors and diversity response variables capture a large spectrum of variation within and between forest types throughout the study area (Fig. 2.3). Data analyses were designed to (1) test the statistical relationship between attributes of forest structure and tree-species diversity based on Spearman's ρ correlation coefficients, a rank-based measure of association that facilitates application to non-normal data distributions, and (2) evaluate model fit of a series of support vector regression (SVR) predictive models of tree species diversity via 10-fold cross-validation. SVR was chosen over other parametric and machine learning models because of its ability to model nonlinearities in complex datasets, while balancing between high accuracy in response variable prediction and generalizability to unseen data (Vapnik & Vapnik 1998; Schölkopf & Smola 2002).

SVR is an extension of support vector machines (SVM), first developed in statistical learning theory as a machine learning method for classification that has since been extended to prediction and regression (Cortes & Vapnik 1995; Cristianini & Shawe-Taylor 2000). SVMs aim to translate a nonlinear problem to a linear one by using kernel functions to fit a model that maps the original low-dimensional input space into a higher dimensional feature space (Vapnik & Vapnik 1998). The global optimum solution is unique and minimizes overfitting by optimally balancing between the accuracy of the model based on cross-validated training and validation data (Schölkopf & Smola 2002).

Support vector regression (SVR) was performed using a Gaussian radial kernel parameter (γ) and two empirically-determined hyperparameters (C and ε) derived from training data (Appendix 2). Variable importance for the SVR model was assessed via the R^2 statistic of a loess smoother that compares singular predictors to an intercept-only null model (Kuhn 2008). Model fit between observed and predicted values was assessed using 10-fold cross-validation, an out-of-sample model evaluation procedure whereby random subsamples are iteratively withheld from model training and then used as an independent validation of model results. While out-of-sample validations may result in a lower goodness of fit than in-sample tests, they are unbiased and provide a better measure of a model's predictive accuracy, especially when generalized to independent data (Hastie et al. 2009). All statistical analyses were performed using the software R, v. 3.2.1 (R Core Team 2016), with SVR model training and variable importance calculated using the caret package (Kuhn 2015) and coverage-based rarified species richness performed using the iNEXT package (Hsieh et al. 2016).

2.3 Results

2.3.1. Forest Structural Attributes as Predictors of Tree Species Diversity

Spearman correlations reveal several significant relationships between forest structure and tree species diversity (Fig. 2.4). Among the 15 structural attributes examined in this study, maximum size (maxBA and maxHT), size heterogeneity (cv and the Gini coefficient), and the shape of basal area distributions (e.g. skewBA and Weibull shape) emerged as the structural attributes most correlated with tree species diversity (Fig. 2.5). Stem density exhibited the highest degree of correlation with those indices most susceptible to sampling effect (e.g. SR), and was non-significant with indices explicitly designed to be independent of it (e.g. rarefied SR). Measures representing plot-wide means (e.g. meanHT, qmd and Weibull scale) and the shape of the tree height distributions (e.g. skewHT and kurtosisHT) generally registered weak or non-significant correlations across diversity indices. Despite some redundancy among covariant structural attributes, all variables were included in final models in order to allow comparison between related measures (Appendix 3). The highest covariance among structural features occurred between those describing alternate yet associated aspects of stem size (e.g. maxHT ~ maxBA; ρ=0.81) or where the indices were themselves related (e.g. cvBA ~ GCBA; ρ=0.86).

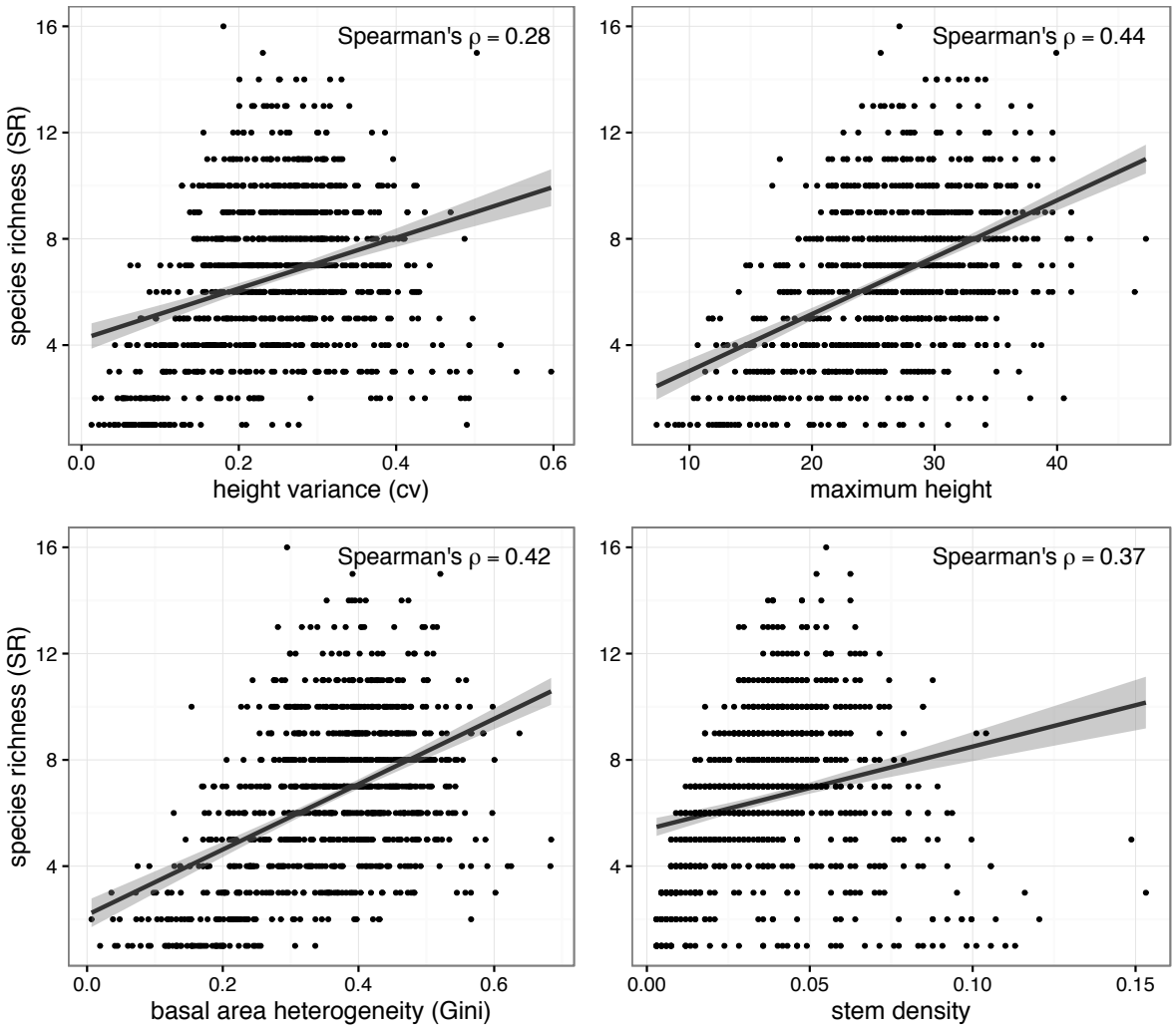


Figure 2.4. Scatterplots of species richness versus selected predictor variables. Linear regression line (black) and 95% confidence intervals (grey).

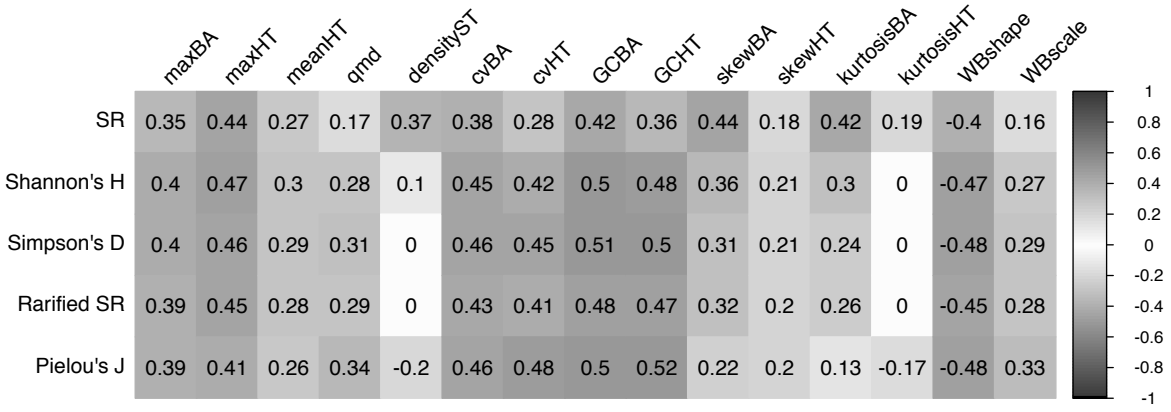


Figure 2.5. Correlation matrix of forest structure versus tree species diversity. Spearman correlation coefficients. Non-zero values significant at $p<0.05$.

2.3.2. Models of Tree Plant Diversity based on Forest Structure

Support vector regression (SVR) models run on the full dataset (n=972 plots) explained 55% of the variance in predicted SR values, and 60%, 61%, 44% and 59% of variance in predicted values for Shannon's *H*, Simpson's *D*, rarified SR, and Pielou's *J*, respectively (Fig. 2.6). When subset by stand origin (natural and artificially planted/seeded) and forest type (pine, mixed, and broadleaf), predictive accuracy was highest for artificially-regenerated and pine-dominated stands. This trend was especially pronounced for diversity indices most sensitive to relative abundance, like Pielou's *J* (R^2=0.71 and R^2=0.66 for artificially-regenerated and pine stands, respectively). In general, tree species diversity in the broadleaf ($0.21 < R^2 < 0.52$) and mixed categories ($0.39 < R^2 < 0.56$) exhibited the lowest degree of predicted accuracy. Relative variable importance of structural attributes based on the full SVR model revealed a relationship between individual structural predictors and diversity indices

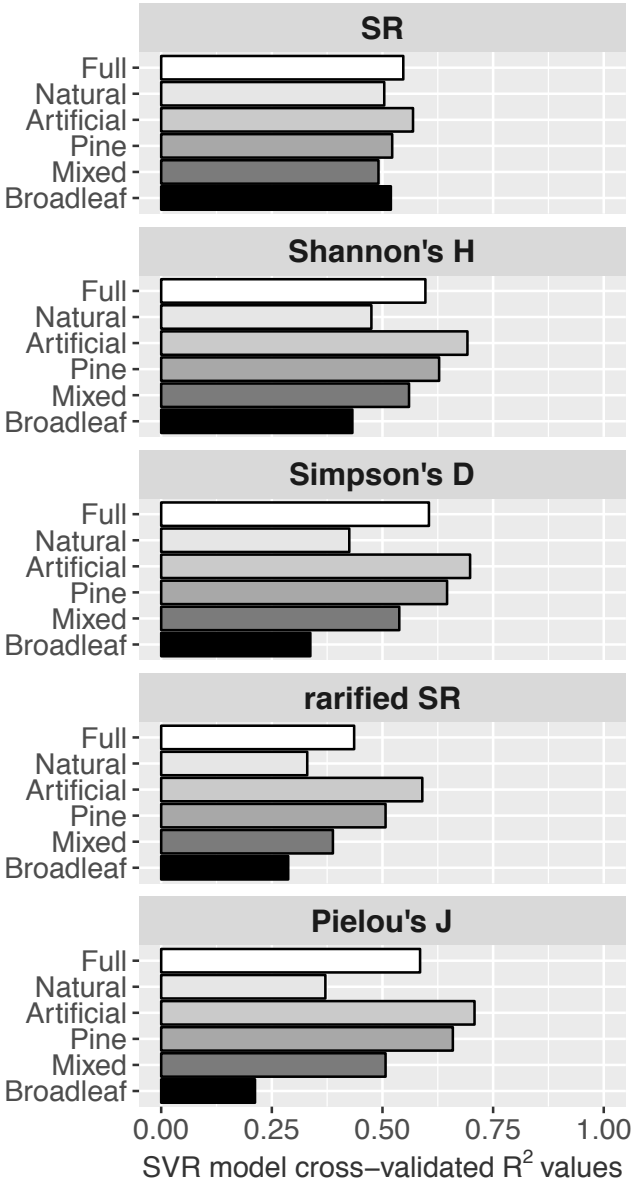


Figure 2.6. Cross-validated adjusted R2 of SVR models. All values significant at p<0.05. Full dataset and subset by stand origin - Natural (n=837) vs. Artificial (n=135) - and forest type - Pine (n=198), Mixed (n=466), and Broadleaf (n=308).

resembling those observed with Spearman correlations, with size dispersion generally exhibiting the largest influence on predictive models, followed by maximum size and Weibull shape (Fig. 2.7).

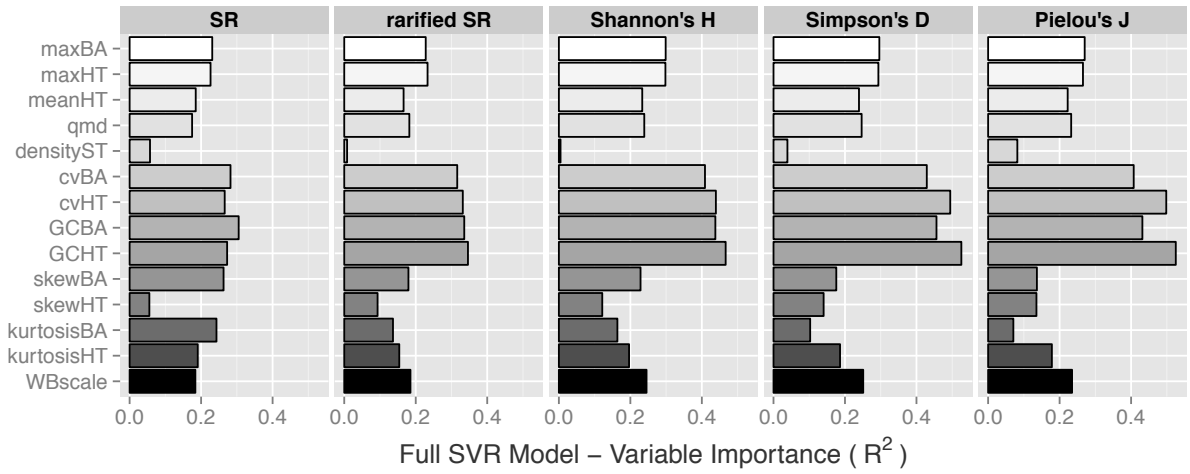


Figure. 2.7. Importance values based on the R^2 statistic of a loess smoother comparing singular predictors to an intercept-only null model. Predictors assessed for the full dataset.

2.4. Discussion

2.4.1. Forest Structural Attributes and the Structure-Diversity Relationship

2.4.1.1. Summary statistics: maximum, mean and density

Spearman correlation coefficients confirm the first hypothesis that measureable components of forest structure are significantly correlated with a range of taxonomic diversity indices, though the strength of this relationship varies greatly among structural attributes. Corroborating results from other studies, we found a significant correlation between diversity and stand maximum values like maxHT and maxBA across all diversity indices (Aiba & Kitayama 1999; Wolf et al. 2012). While stand maximum values partly reflect the growth traits of constituent species whose presence may not necessarily be driving species, they can also be characteristic of underlying site factors that drive productivity and diversity levels, such as site index and time since

stand-replacing disturbance (Franklin 1988; Peet & Christensen 1988). Stand-wise means such as qmd, meanHT, and WBscale, in contrast, were only weakly significant predictors of tree species diversity, partially reflecting information loss due to averaging across individuals.

Of particular note in assessing these results is the shifting role of stem density across diversity indices. While density may vary with species richness simply by virtue of sampling effect, it may likewise reflect biologically meaningful patterns in successional stage and resource availability that can influence community assembly processes (Weiner & Thomas 1986; Goodburn & Lorimer 1999; Gotelli & Colwell 2001). Results confirmed our expectation that the magnitude of correlation between density and diversity tracks the relative influence of species number versus evenness across the spectrum of diversity indices. Thus, while density had the highest correlation with species richness, it was not significantly correlated with rarified SR, species evenness, Shannon's H, or Simpson's D - all measures designed to be independent from this density-dependent sampling effect.

2.4.1.2. Size heterogeneity: cv and Gini

Indices of structural heterogeneity like the coefficient of variation (cv) and the Gini Coefficient (GC) were found to be among the most important structural predictors of tree species diversity (Damgaard & Weiner 2000). Because they are continuous measures, they avoid information loss and bias when subjectively apportioning stems into size-class bins (Lexerød & Eid 2006; Valbuena et al. 2012). The positive correlation between size heterogeneity and tree species diversity is partly attributable to the fact that higher taxonomic diversity could be expected to promote a greater diversity in species traits, like growth rate and size, which then manifest as a greater dispersion in stem sizes. An alternative, complementary hypothesis for this close relationship is premised instead upon the observation of increasing structural heterogeneity

through succession and its impact on resource levels and competition (Franklin 1988; Palmer & Maurer 1997). More specifically, the development of increasing structural complexity through succession affects patterns of diversity and composition by driving absolute changes in resource levels as well as the spatial variability of the individuals exploiting those resources (Peet 1992; Halpern & Spies 1995).

In keeping with the Piedmont old-field model of succession, disturbance and competition drive temporal dynamics in stand structure, characterized by a progression from structurally simple to more complex, multi-cohort stands (Peet & Christensen 1987). This pattern of increasing structural complexity through time reflects findings from other locations that confirm structural heterogeneity as an effective proxy for stand age and population structure (Lorimer & Krug 1983; Uuttera et al. 1997; Lexerød & Eid 2006). Spatially heterogeneous resource distributions, in turn, drive changes in light penetration and root competition that may increase the probability for recruitment of a variety of species not necessarily present in the stand overstory, including suppressed understory individuals as well as recruits newly dispersed or emergent from the seed bank (Canham et al. 1994; Montgomery & Chazdon 2001). Increased microsite heterogeneity drives niche-space partitioning and other community assembly processes that allow for greater species packing within a given area (Shmida & Wilson 1985).

2.4.1.3. Size distributions: skewness, kurtosis, and Weibull shape

Skewness and kurtosis, the third and fourth statistical moments of a stand's size distribution, serve to index patterns in population structure that drive emergent stand-level properties like species diversity (Knox et al. 1989; Lorimer & Frelich 1997). These measures of size distributions have been found to effectively distinguish single-cohort, early-successional forests from multi-cohort, mid- or late-successional ones (Mohler et al. 1978; Lorimer & Krug

1983; Knox et al. 1989). Whereas even-aged stands tend toward unimodal Gaussian size-class distributions, uneven-aged stands exhibit a negative monotonic diameter distribution, culminating in the balanced diameter distribution of the highly skewed reverse-J form (Goodburn & Lorimer 1999; Rubin et al. 2006). The reverse-J reflects later successional canopy dynamics, namely when a highly abundant gap-regenerated younger cohort coincides with remaining large-canopy dominants that, despite their scarcity, retain a high proportion of stand basal area (Leak 1964; Bailey & Dell 1973).

Skewness and WBshape (a measure of skewness of the Weibull distribution) were highly correlated with species diversity, with left-skewed distributions characterized by positive skewness and negative Weibull shape values. In either case, the left-skewed distributions of uneven-aged forests tend toward higher tree species diversity levels. These results match findings from other studies in the North Carolina Piedmont reporting species richness to peak in late succession following gap development and understory re-initiation when strands simultaneously possess shade-tolerant, climax species and early-successional colonizers (Peet & Christensen 1988). Kurtosis, on the other hand, appeared to have only limited utility in representing cohort structure, primarily because high kurtosis values can be suggestive both of structurally-simple, single-cohort stands, as well as multi-cohort, later successional stands characterized by the reverse-J basal-area distribution or fat-tailed, unimodal distributions. These confounding patterns are partly responsible for the weak to non-significant correlation between kurtosis values and tree species diversity.

2.4.2. Predictive Models of Diversity based on Forest Structure

The second aim of this study was to assess the utility of using forest structural attributes in a predictive model of tree species diversity. Using 10-fold cross-validation to test model fit and

generalizability, we found SVR models trained solely with structural attributes explain more than half of the variance observed in tree species diversity across the full Piedmont dataset. Individual structural attributes were selected for their potential to reflect components of stand structure readily measured on the ground and at larger spatial scales using remote sensing products including LiDAR, Interferometric Synthetic Aperture Radar (InSAR), and optical time series data (Hyde et al. 2006; Bergen et al. 2009; Song et al. 2015). At present, airborne LiDAR scanners are capable of detecting canopy structural elements with <10cm vertical and horizontal precision, and explain 50-95% of the variance in many of the ground-measured structural attributes used in this study, including mean height, maximum height, stem density, and basal area (Næsset 2002; Cook et al. 2013). In this way, the utility of using forest structure as a parameter for landscape-scale predictive mapping of tree species diversity is derived from its ability to indirectly account for variation in diversity levels otherwise caused by latent but difficult to measure factors like soil conditions and site history (Beier & de Albuquerque 2015). These results confirm the empirical relationship between structure and diversity, and thus, the basis for its use as an effective surrogate in modeling species diversity – one that would likely be improved greatly with the inclusion of topographic, environmental, and land cover data.

While most studies that have investigated the relationship between forest structure and plant diversity have focused on herbaceous understory plants, those focusing explicitly on tree diversity tended to find only weak or insignificant relationships (Bagnaresi 2002; Ishii et al. 2004; Cook 2015). Inconsistencies between these findings and our own may reflect methodological factors driving results and their interpretation. For example, studies based on a small sample size of large, closely-monitored plots may fail to capture the spectrum of variability in community types expected at larger scales (Aiba & Kitayama 1999; Neumann & Starlinger 2001; Cook 2015).

Weak correlations found in studies using large national inventories to investigate regional patterns of diversity may reflect a design focusing on maximizing plant diversity prediction through the inclusion of other types of environmental predictors (Chiarucci & Bonini 2005; Bacaro et al. 2008). Depending on the variance partitioning method, covariance between structure and other more significant environmental predictors could obscure the predictive power of structural variables when used alone, a factor noted by Bacaro et al. (2008). Finally, weak correlations found in previous studies could result from inherent limitations of the statistical models used, such that assumptions about linearity, normality, and collinearity in parametric models may constrain their predictive ability and generalizability to independent data (Vapnik & Vapnik 1998; Dormann et al. 2013). Although distribution-free kernel methods like SVR are not immune to multi-collinearity, redundancy in collinear predictors is minimized when the data are translated into a higher-dimensional feature space (Schölkopf & Smola 2002; Toloşi & Lengauer 2011). What the black box SVR model loses in lack of transparency of parameter estimation and standard errors, it more than regains in flexibility and statistical power.

2.4.3. Structure and Diversity by Stand Origin and Forest Type

SVR models were able to predict tree species diversity with far higher accuracy in forests that possess clear evidence of artificial regeneration via planting or seeding, compared with natural forests where no such evidence exists. While this result partly reflects ecologically meaningful patterns that result from management practices, it may also reflect the imbalanced sample sizes between the two classes (881 natural vs. 142 artificial plots), such that the natural regeneration class possessed far larger variation in community type and site history than the artificial stand category. Although stands with evidence of timber removal were censored from the dataset, the artificially-regenerated forest category still retains a large portion of managed stands, such as

single-cohort, monospecific pine plantations (e.g. *Pinus taeda*), and to a lesser extent mixed-species, multi-cohort stands. Structurally-simple planted stands tend towards monospecificity, whereas multiple-cohort stands with distinct vertical layering between cohorts are more likely to consist of multiple species (Oliver & Larson 1996; Smith et al. 1997). The close relationship between structure and diversity among artificially-regenerated stands reflects the fact that small changes in structure likely coincide with compositional change. Successional stands have been found to show a strong pattern of low diversity during the early self-thinning phase of development, followed by a steady increase with the opening of the canopy (Peet 1992). In the case of an even-aged, monospecific *Pinus*-dominated overstory, the establishment of a single new understory hardwood in a canopy gap would double species richness – a dramatic change in diversity levels, readily observed in the structural signal. While a broad range of silvicultural practices exist that maintain multi-strata, multi-species stands, natural forests in our dataset tend, by comparison, to a broader range of compositional types and structural forms that tend to be less correlated with structural attributes (Smith et al. 1997).

When subset by forest type and leaf habit, tree species diversity in the pine category was predicted with the highest accuracy. As with artificially-regenerated stands, pine-dominated forests in the Piedmont tend to be early-successional, single-cohort, structurally simple, and near monospecific. Unlike dense-canopy, early-successional pine stands, those undergoing gap formation open themselves to recruitment, thereby increasing their likelihood for higher species richness (Peet 1992). As with natural forests, the broadleaf and mixed forest categories include a wider range of forest types and highlight some of the inherent limitations to modeling species diversity based on structure in more species-rich broadleaf forests.

2.5. Conclusion

Based on our analysis of a large, geographically-distributed sample of FIA plots from the North Carolina Piedmont, we found that measures of forest structure were able to predict a substantial portion of the variance in tree species diversity without accounting for other known predictors of diversity in Piedmont forests such as topo-edaphic conditions and site history. Among all structural predictors, maximum height, basal area size inequality (Gini coefficient), and skewness of the basal area distribution (Weibull shape) exhibited the strongest correlations with indices of diversity. SVR models were able to explain and predict more than 50% of the variance in species richness across all subsets, with models of species evenness and heterogeneity in planted/seeded stands and pine forests greatly surpassing this mark. However, a large portion of unexplained variance remains after accounting for structure, and the extent to which this can be explained by deterministic ecological processes linked to environment and stand history remains a future research priority. Beyond the theoretical implications of unraveling primary patterns underlying tree species diversity, these findings highlight the empirical basis and potential for using remotely-sensed structural data as an effective predictor for modeling tree diversity over large geographic regions.

CHAPTER 3

MODELING MULTI-SCALE PLANT SPECIES RICHNESS IN A PIEDMONT NORTH CAROLINA LANDSCAPE USING LIDAR-HYPERSPECTRAL REMOTE-SENSING

3.1. Introduction

Human activities have been implicated in the severe degradation of the Earth's ecosystems, especially the elimination of species whose absence possess the potential to alter the functioning of ecosystems and the extrinsic benefits they provide to society (Loreau et al. 2001; Hooper et al. 2012; Newbold et al. 2015). Species loss is a particularly problematic trend as species diversity has been found to bolster community resilience to disturbance (Naeem & Li 1997; Hooper et al. 2002), strengthen resistance to infectious disease and pathogens (Civitello et al. 2015), and increase net ecosystem productivity (Tilman et al. 1996; Aarssen 2004). Plants are a particularly important indicator group for monitoring overall diversity levels due to their capacity to provide habitat and resources throughout the trophic chain (Myers et al. 2000; Raxworthy et al. 2003). In forest ecosystems, plants comprise the majority of total biomass, and increases in the levels of plant diversity have been found to be positively correlated with productivity (Liang et al. 2016), as well as that of specialist and generalist consumers (Gaston 1992; Zak et al. 2003). The recognized importance of plant diversity in the context of overall biodiversity has driven efforts to model and map its spatial distribution over landscapes for applications ranging from habitat and conservation management to regional and global modeling of biodiversity and ecosystem function (Rocchini et al. 2007; Gillespie et al. 2008).

One promising technology in this pursuit is hyperspectral imaging, or image spectroscopy, which provides highly spectrally-resolved data on the reflectance properties of forest canopies (Ghiyamat & Shafri 2008; Im & Jensen 2008). Hyperspectral imaging exploits the fact that spectral reflectance can be diagnostic of phenotypic traits like foliar morphology and biochemistry (Curran 1989). Even when the mechanistic relationships between spectral reflectance and the contingencies of trait - environment interactions are not fully resolved, correlational relationships between narrowband spectral features (only partially observable using broadband instruments) and foliar chemistry allow for the detection of a host of canopy properties including stand composition and underlying environmental conditions (Ustin & Gamon 2010; Ollinger 2011). While intra-species trait variation and canopy structure can confound consistent spectral retrievals (Castro-Esau et al. 2006; Knyazikhin et al. 2013), to the extent that inter-species variation in foliar biochemistry manifests as heterogeneity in the spectral reflectance signal in a local neighborhood, spectral variance has been observed to correlate with taxonomic diversity (Rocchini et al. 2010; Cavender-Bares et al. 2016).

While hyperspectral imagery excels in detecting subtle spectral variation in the upper canopy surface, it is neither well-suited for the detection of mid- and understory plants, nor the derivation of canopy structural properties that have been found to be associated with species richness in forest stands (Buddenbaum et al. 2013; Hakkenberg et al. 2016). For the goal of characterizing forest structure, LiDAR is particularly adept (Lesky et al. 2002). Unlike optical imagery, LiDAR is largely immune from saturation at high biomass values (Parker et al. 1999; Næsset & Økland 2002; White et al. 2013). When combined in a single platform, active LiDAR and passive hyperspectral sensors provide complementary data on canopy properties that simultaneously characterize multiple remotely-sensible environmental data domains, including

canopy height from LiDAR first returns, topo-physiognomy from LiDAR ground returns, biophysical structure from LiDAR all returns, and canopy biochemistry from image spectroscopy (Anderson et al. 2008; Dalponte et al. 2008; Torabzadeh et al. 2014). In fact, combined LiDAR-hyperspectral datasets have been found to more accurately characterize canopy composition and distribution than either used separately (Asner et al. 2008; Feilhauer & Schmidtlein 2009).

Hyperspectral-LiDAR systems are adept in the direct remote detection of species and that of species diversity through the direct detection of structural and biochemical heterogeneity (Rocchini et al. 2010). Moreover, they excel as a data source for the mapping of geophysical gradients from which multi-species niche models, and therefore landscape patterns in biodiversity, can be constrained (Elith & Leathwick 2009; Ohmann et al. 2011). When paired with field plots, remotely-sensed data on canopy reflectance and structure allows for robust prediction of landscape turnover in plant diversity and provides a basis for inference concerning the underlying drivers and merging patterns of diversity at spatial extents far larger than traditional field sampling would allow. Most studies in biodiversity modeling focus on empirical relationships between remotely-sensed data and levels of species diversity (Cayuela et al. 2006; Rocchini et al. 2007; Simonson et al. 2012; Camathias et al. 2013; Higgins et al. 2014). Other studies have employed these empirical relationships in a predictive mapping format (Ohmann & Gregory 2002; Leutner et al. 2012; Wolf et al. 2012; Fricker et al. 2015). With some notable exceptions (e.g. Gould 2000; Schmidtlein & Sassin 2004; Simonson et al. 2012), most remotely-sensed diversity mapping studies have focused on woody canopy species, at a single spatial scale.

Mechanisms driving plant diversity are scale-dependent, with no one scale exhaustive in explaining multi-scale processes (Levin 1992). Indeed, multiple factors simultaneously effect the accuracy of diversity models, ensuring that spatial coherence in the characteristic scales of field

plot data (sampling size, design, and intensity), remotely-sensed data (spatial resolution and geolocational precision), species diversity patterns (species-area relationships), and that of target biota (size, density, and coverage extent) are critical to model output. In light of these concerns, this study seeks to explicitly investigate remotely-sensible drivers of plant diversity across multiple spatial scales to contribute to the operationalization of wall-to-wall landscape management and conservation planning for biodiversity in anticipation of annual NEON data products and the deployment of a fleet of new LiDAR and hyperspectral satellites (Kampe et al. 2010; Cook et al. 2013; Dubayah et al. 2014). Specifically, we employ spatially-nested field plots in conjunction with data from the Goddard LiDAR, hyperspectral, thermal (G-LiHT) airborne sensor to map vascular plant species richness at seven spatial scales in a compositionally- and structurally-complex Piedmont forest landscape in North Carolina (NC), USA. To accomplish this goal, this study focuses on three primary tasks. First, we assess how the predictive power of nonparametric models of plant species richness using feature-selected, remotely-sensed data, changes across seven spatial scales. Second, we employ these models, trained with remotely-sensed data and spatially-nested field plots, to predict plant species richness across the study area. Finally, we determine the remotely-sensible metrics that correlate with plant species richness in the Piedmont study site and how they change with spatial scale.

3.2. Methods

3.2.1 Study site

The 2.8 km^2 study site is located in Duke Forest Blackwood Division, NC, USA consists of Piedmont secondary old-field successional pine and mature hardwood forests, following selective cutting, agriculture, and grazing in the 19^{th} and early 20^{th} century (Peet & Christensen

1988). The mapped study area focuses solely on natural and semi-natural forests, and thus excludes areas of plantation forest, clear-cut, and built infrastructure (Fig. 3.1). Field plot locations are based on a stratified random design, such that individual plot locations were randomly pre-determined within the constraints of stratified bands along an east-west and a north-south topographic gradient (Fig. 3.1). The ensuing aspect-elevation combinations ensure field plots span the primary physiognomic types in the study area, including upland, riparian, and bottomland forests, which in aggregate comprise 0.1% of the entire study area. In all field plots, species presence was recorded following Carolina Vegetation Survey protocols for all vascular plant species in $0.01m^2$, $0.1m^2$, $1m^2$, $10m^2$, $100m^2$ spatially-nested subplots, and assessed at the $400m^2$ and $900m^2$ scales (Peet et al. 1998). In total, 36 $900m^2$ plots were sampled, each with eight subplots at $<100m^2$ scales (288 plots total) and four subplots per plot at $100m^2$ (144 plots total). Sub-meter geo-locational precision of all plot and subplot vertices was achieved based on triangulation of Ground Control Points (GCPs) measured in the field with tape measures and in a GIS environment using highly-precise, fine

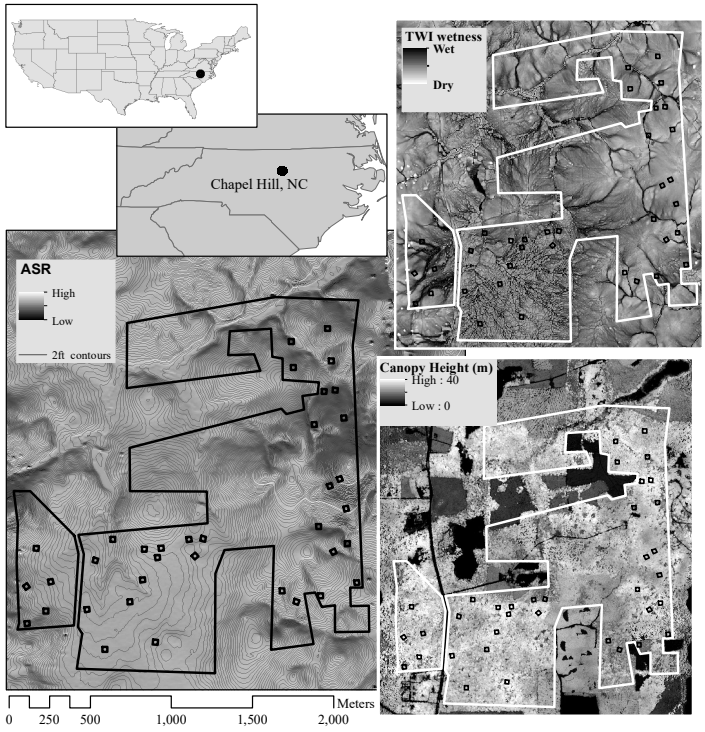


Figure 3.1. Duke Forest Blackwood study area extent. The study area excludes areas of plantation forest, clear-cuts, and human habitation. Topographic wetness index (TWI) is a proxy for soil moisture (upper right). Canopy height (lower right) is derived from G-LiHT LiDAR. Elevation (lower left) is featured with ASR (annualized solar radiation) for shading and 2ft (0.6m) contours, with squares representing plot location and extent.

resolution (0.15m) Digital Orthophoto Quadrangle (DOQ) imagery (NC OneMap Geospatial Portal 2016). Botanical nomenclature follows Weakley (2015). The resultant plot data are available on Vegbank (http://vegbank.org; Peet et al. 2012).

3.2.2. Derived remotely-sensed predictors

Aerial remotely-sensed data come from NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) airborne imager at a 2m ($4m^2$) spatial resolution and 12-bit radiometric resolution (Cook et al. 2013). The G-LiHT airborne imager utilizes commercial, off-the-shelf sensors a Hyperspec imaging spectrometer (Headwall Photonics, Fitchburg, MA, USA) with a 407–1,007 nm spectral range and a ≤5 nm full width half maximum (FWHM) spectral resolution, as well as a VQ-480 (Riegl USA, Orlando, FL, USA) airborne laser scanning (ALS) system, with a mean return density of up to 50 laser pulses/m^2 and 10 cm diameter footprint at the nominal operating altitude of 335m (Cook et al. 2013). Remotely-sensed data used in this study was collected on Oct. 25, 2013, during late-season, leaf-on conditions when inter-species phenological differences could aid in taxonomic discrimination. Thermal data was not available for our study area.

All derived remotely-sensed predictor variables used in this study emanate from four G-LiHT Level 3 products: (1) a digital terrain model (DTM), (2) all LiDAR returns, (3) a canopy height model (CHM), and (4) a 114-band atmospherically-corrected surface reflectance image stack (Table 3.1). Owing to the data redundancy in the hyperspectral image stack, full-spectrum and narrow-band indices were used in final analyses (Ustin et al. 2009; Asner & Martin 2009). For the former, principal components analysis (PCA) was used to reduce the full 114-band hyperspectral dataset to three orthogonal layers (PCA 1-3), which cumulatively explained 99.7%

of variance among all bands. While in the latter case, narrow-band indices with established precedence in the literature were employed (Table 3.1).

Table 3.1. Remotely-sensed predictor variables. All remotely-sensed predictors represent aggregates of 2x2m pixels resampled to a 30x30m (900m2) output resolution.

Category	Predictor	Abbv.	Equation/Source
LiDAR topography (last returns)	Average Solar Radiance (annual)	ASR	(ESRI 2016)
	Deviation from mean elevation	DEV	$DEV = \frac{z_0 - \bar{z}}{SD}$; where z_0 is the elevation of the focal pixel, $\bar{z}$ and SD are mean and standard deviation of elevation in a 222m window (De Reu et al. 2013)
	Height above EGM96 (Earth Gravitational Model 1996) geoid	elev	DTM (Cook et al. 2013)
	Slope in degrees	slope	DTM (Cook et al. 2013)
	Topographic Position Index	TPI	$TPI = z_0 - \bar{z}$; where z_0 is the elevation of the focal pixel and $\bar{z}$ is mean elevation in a 222m window (De Reu et al. 2013)
	Topographic Wetness Index	TWI	$TWI = ln\left(\frac{a}{tan\beta}\right)$; where a is the local upslope area and β is slope (Beven & Kirkby 1979)
LiDAR canopy height (first returns)	Canopy height model	CHM	CHM (Cook et al. 2013)
LiDAR canopy structure (all returns)	All return heights	all_returns	LiDAR returns (Cook et al. 2013)
	Tree return heights †	tree_returns	LiDAR returns (Cook et al. 2013)
	Understory return heights †	understory_ returns	LiDAR returns (Cook et al. 2013)
	Anthocyanin Reflectance Index 1	ARI1	$ARI1 = \frac{1}{\rho_{550}} - \frac{1}{\rho_{700}}$ (Gitelson et al. 2002)

Hyperspectral foliar reflectance	Anthocyanin Reflectance Index 2	ARI2	$ARI2 = \rho_{800}\left[\frac{1}{\rho_{550}} - \frac{1}{\rho_{700}}\right]$ (Gitelson et al. 2002)
	Carotenoid Reflectance Index 1	CRI1	$CRI1 = \frac{1}{\rho_{510}} - \frac{1}{\rho_{550}}$ (Gitelson et al. 2002)
	Carotenoid Reflectance Index 2	CRI2	$CRI2 = \frac{1}{\rho_{510}} - \frac{1}{\rho_{700}}$ (Gitelson et al. 2002)
	Normalized Difference Vegetation Index (narrowband)	NDVI	$NDVI = \frac{\rho_{800}-\rho_{670}}{\rho_{800}+\rho_{670}}$ (Haboudane et al. 2004)
	Principal Component Axis 1	PCA1	PCA axis 1 (ρ_{407}:ρ_{1007})
	Principal Component Axis 2	PCA2	PCA axis 2 (ρ_{407}:ρ_{1007})
	Principal Component Axis 3	PCA3	PCA axis 3 (ρ_{407}:ρ_{1007})
	Photochemical Reflectance Index	PRI	$PRI = \frac{\rho_{531}-\rho_{570}}{\rho_{531}+\rho_{570}}$ (Gamon et al. 1995)
	Red Edge Position Index	REPI	REPI=max[(ρ_n+1-ρ_n)/10] where 690 ≤ n ≥ 750 (Souza et al. 2010)
	Simple Ratio Index (narrowband)	SRI	$SRI = \frac{\rho_{800}}{\rho_{670}}$ (Haboudane et al. 2004)

† Tree returns defined as all returns above 1.37m

To facilitate pixel aggregation (e.g. sub-pixel heterogeneity measures) and test for scale-dependence among remotely-sensed predictors of species richness, all remotely-sensed derived geophysical variables were aggregated based on mean, minimum, maximum, and standard deviation of the 4m^2 raw pixels at four spatial resolutions – 16m^2, 100m^2, 400m^2, and 900m^2. Annualized solar radiation (ASR), deviation from mean elevation (DEV), elevation (elev), slope, and topographic position index (TPI) were aggregated as mean values only, while for CHM, the mean, minimum, maximum, standard deviation, skewness and kurtosis were computed. All other variables were aggregated by mean, minimum, maximum, and standard deviation. For all

subsequent analyses, the spatial resolution of the remotely sensed predictor data best corresponding to that of the spatial scale of the species richness response variable was employed. For example, for the four smallest spatial scales of plant richness ($0.01m^2$, $0.1m^2$, $1m^2$, and $10m^2$) we used derived remotely sensed layers at a $16m^2$ resolution, the finest derived pixel resolution capable of incorporating heterogeneity values of $4m^2$ pixels. Owing to sub-meter locational accuracy of pixels and subplots, aggregate predictor pixels thus need not be at an identical spatial resolution with sampled subplots, provided they are fully (and centrally) subsumed. For larger spatial scales, remotely-sensed predictors were aggregated and geo-locationally aligned at the scale corresponding with the plant richness response variable: $100m^2$ (25 2m pixels), $400m^2$ (100 2m pixels), and $900m^2$ (225 2m pixels). All derived products and subsequent analyses were calculated using the software R, v. 3.3.1 (R Core Team 2016), with the exception of ASR, which was calculated using the ArcMap version 10.4 AASR plug-in (ESRI 2016).

3.2.3. Data Analysis

Field plot training data and derived remotely-sensed predictors were used to predict vascular plant species richness at seven spatial scales using random forest (RF) regression models. RF is a nonparametric, machine-learning technique that is not subject to distributional assumptions and robust to multi-collinearity and over-fitting (Breiman 2001; Toloşi & Lengauer 2011). Each RF model in our parameterization is determined from an ensemble of 2000 trees, each grown from a bootstrapped sample of the predictors, with the remaining observations left out-of-bag (OOB) for model assessment. To maximize the predictive accuracy of each of the seven RF models, feature selection of predictor variables was performed on the full model (all predictors), whereby the worst performing variables were withheld from each successive model based on importance value, the residual sum of squares after splitting on each respective variable. The best performing

of the *p*-1 models, where *p* is the total number of features (*p*), was selected based on a comparison of the cross-validated accuracy of all *p*-1 models. Accuracy estimates for feature-selected RF models are based on 10-fold cross-validation, a model evaluation procedure that assesses model fit and generalizability based on iteratively withheld random samples of the full dataset into quasi-independent training and validation subsets (Appendix 7). The best-performing, feature-selected models at each of the seven scales was then inversed to predict species richness for all pixels throughout the study area. Predicted values are based on the mean vote of all OOB observations, and rescaled to fit the range of species richness observed in the field (Zhang & Lu 2012). Uncertainty maps, on the hand, represent the coefficient of variation of all 2000 OOB votes (Singh et al. 2015). All random forest models were run with the randomForest package (Liaw & Wiener 2002) in R, v. 3.3.1 (R Core Team 2016).

Feature-selected predictor variables in final RF models reflect those input parameters that, by exploiting potentially complex, highly non-linear interactions, result in generalizable models optimized for predictive accuracy. Despite their ability to produce high accuracy predictions, black-box nonparametric models are relatively uninformative for inference into the role of specific parameters (Evans et al. 2011). To determine the sign and magnitude of the relationship between key parameters and species richness, we ran two post-hoc tests. First, we assessed the strength of correlation using Spearman correlation coefficients, a nonparametric rank-based measure of association that facilitates application to non-normal data distributions. Key parameters are defined as those variables selected in one of the seven feature-selected random forest regression models that are significantly correlated with species richness at more than one spatial scale. Some partially-redundant variables were excluded (e.g. TWI at 4m versus 8m). While Spearman coefficients are helpful in assessing broader, non-directional correlational patterns between values,

they are less helpful in assessing significance and sign in the directional relationship between predictor and response from which inference can be made.

For this latter task, after having determined there to be no consistently significant spatial auto-correlation across scales based on Moran's I and Mantel's *t* values, we assessed predictors' strength across spatial scales using simple univariate Bayesian GLMs (Legendre 1993; Lennon 2000). For subplot i, we modeled species richness y_i, a non-negative integer, as a Poisson distribution,

$$y_i \sim Poi(\theta_i) \tag{3.1}$$

where the mean species richness θ_i is linked to linear predictors,

$$\log \theta_i = \beta_0 + \beta_1 x_i \tag{3.2}$$

Here β_0 is the intercept, β_1 is the slope corresponding to x_i, the remote-sensing covariate of interest in the subplot i. Under a Bayesian framework, we used vague uniform priors,

$$\beta_0, \beta_1 \sim U(-10^{-3}, 10^3) \tag{3.3}$$

While vague priors in simple (i.e. non-hierarchical) Bayes models yield near identical means and confidence intervals compared with frequentist GLMs, the stochastic generation of an empirical posterior in parameter models better facilitates the pragmatic concern to assess patterns of variance in the sign and magnitude (slope or β_1) of regression parameters for comparison across scales (Clark 2003; Clark et al. 2005). GLMs were corrected for over-dispersion and zero-inflation when necessary. To ensure convergence, Markov chain Monte Carlo (MCMC) were run for 20,000 iterations, after which the first 10,000 were withheld for burn in based on trace plots.

3.3. Results

3.3.1. Predicting plant species richness across spatial scales

In total, 208 distinct vascular plant taxa were identified in field plots (Appendix 5). Field plots, stratified along two predominant environmental gradients, were designed to span the spectrum of compositional variability of the relatively taxonomically diverse and structurally heterogeneous study area. Species richness ranged from 0 at the smallest subplot scales ($0.01m^2$) to 100, recorded at the $900m^2$ full plot scale (Table 3.2). Plots exhibited a wide range of successional-structural conditions with diameter at breast height (DBH) ranging from the minimum recorded 0.1cm to 101.5cm (mean DBH=7.6cm; DBH standard deviation=11.7cm).

Table 3.2. Species richness summary statistics for vegetation sampling in the summer of 2015 in the Blackwood Division of Duke Forest, Durham, North Carolina.

scale (m^2)	mean	sd	min	max
0.01	0.3	0.6	0	4
0.1	1.6	1.5	0	9
1	4.8	3.4	0	18
10	13.5	7.5	0	35
100	34.4	10.2	15	63
400	53.2	13.8	28	84
900	61.4	15.6	35	100

Based on field plot data and feature-selected remotely-sensed predictors, we predicted species richness at seven spatial scales over the extent of the entire study site using bias-corrected random forest regression models (Fig. 3.2). The value of predicted pixels reflects the mean of all 2000 random forest votes, while per-pixel uncertainty is expressed as the standard deviation of those votes (Fig. 3.3). Uncertainty captures spatially-structured variance unexplained in statistical models, as well as data, model and geo-locational errors. Alternatively, the mean and standard deviation of random forest out-of-bag votes can be combined into a single red-green-blue (RGB) color model such that species richness predictions track a gradient from red to green (low to high)

simultaneously with uncertainty values loaded on the blue color gun, such that the greater the uncertainty, the more the red coloration approaches magenta while the green approaches cyan (Fig. 3.4). 10-fold cross-validation of predictive diversity models indicates a distinct trend in accuracy across spatial scales, such that predictive accuracy generally increases with spatial scale, from the smallest 0.01m^2 subplot (R^2=0.14) up to 100m^2 (R^2=0.68), after which it declines slightly at 400m^2 (R^2=0.64) and 900m^2 (R^2=0.59) (Table 3.3). Predicted richness totals likewise increase with spatial scale, reflecting expectations from species-area relationships.

Table 3.3. Final species richness predictive models – parameters and results

scale (m^2)	CV R^2	Feature selected predictor variables*
0.01	0.141	elev (mean), TWI (2m min), all_sd (mean), CRI1 (mean), CHM (max), TWI (4m mean), PCA2 (min), REPI (mean), PCA2 (mean)
0.1	0.174	elev (mean), TWI (8m min), all returns (sd), CHM (max), understory returns (sd)
1	0.253	elev (mean), PCA1 (min), ASR (mean), NDVI (min), all returns (skew), CHM (skew), all returns (sd), NDVI (max), slope (mean), TWI (2m min)
10	0.353	elev (mean), PCA1 (min), TPI (222m mean), NDVI (max), all returns (skew)
100	0.676	SRI (mean), elev (mean), NDVI (mean), all returns (skew), PCA1 (max), PCA2 (max), DEV (222m mean), SRI (min), NDVI (min), TPI (222m mean), PCA1 (mean), CRI1 (min), ASR (mean), DEV (162m mean), understory returns (sd), CRI1 (mean), PCA2 (mean), TWI (8m max), TWI (4m mean), TWI (2m max)
400	0.641	NDVI (mean), SRI (mean), PCA2 (mean), SRI (mean), NDVI (min), PCA1 (max), elev (mean), DEV (222m mean)
900	0.592	PCA2 (min), PCA2 (mean), elev (mean), PCA1 (max)

* In order of variable importance. See Table 3.1 for abbreviations

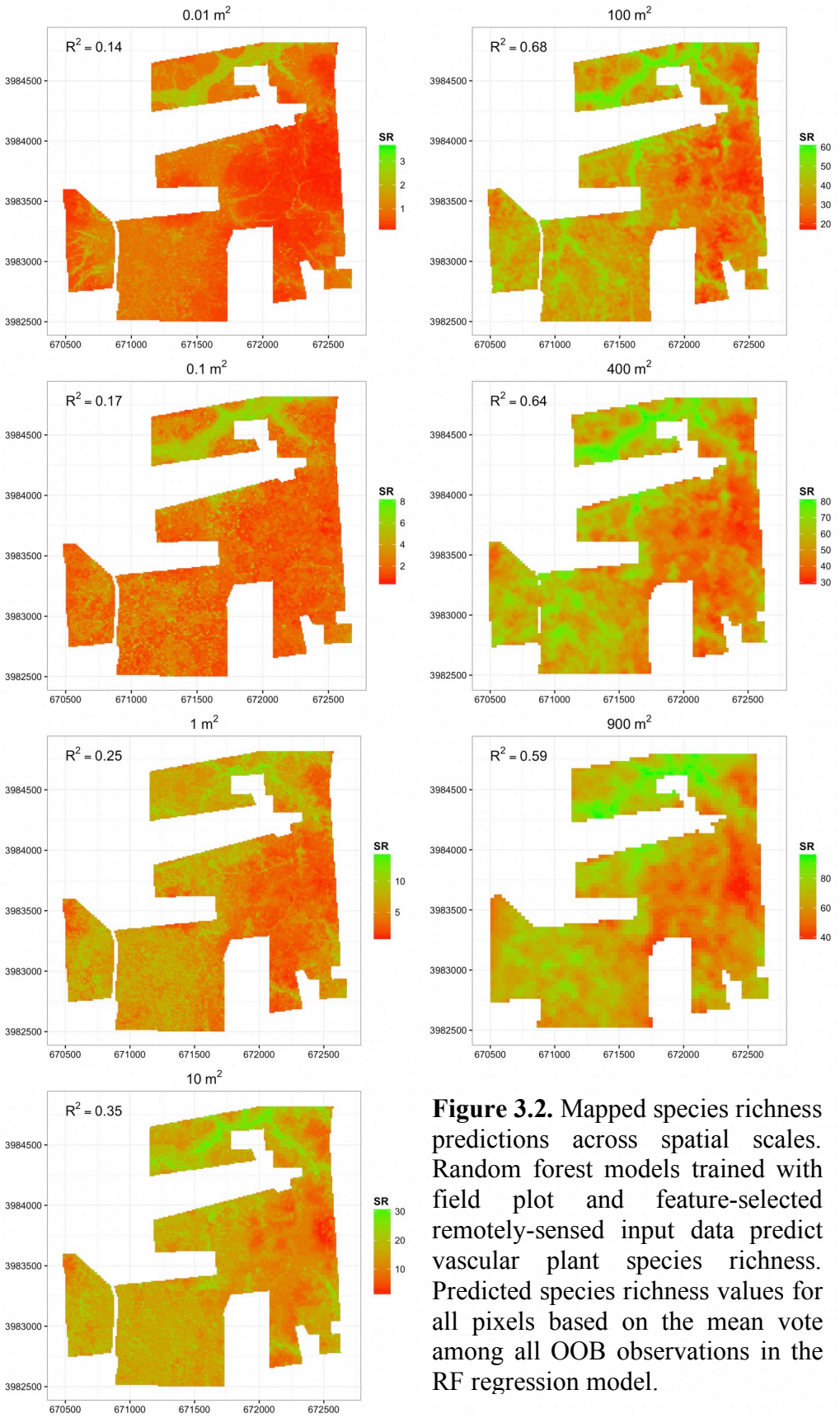


Figure 3.2. Mapped species richness predictions across spatial scales. Random forest models trained with field plot and feature-selected remotely-sensed input data predict vascular plant species richness. Predicted species richness values for all pixels based on the mean vote among all OOB observations in the RF regression model.

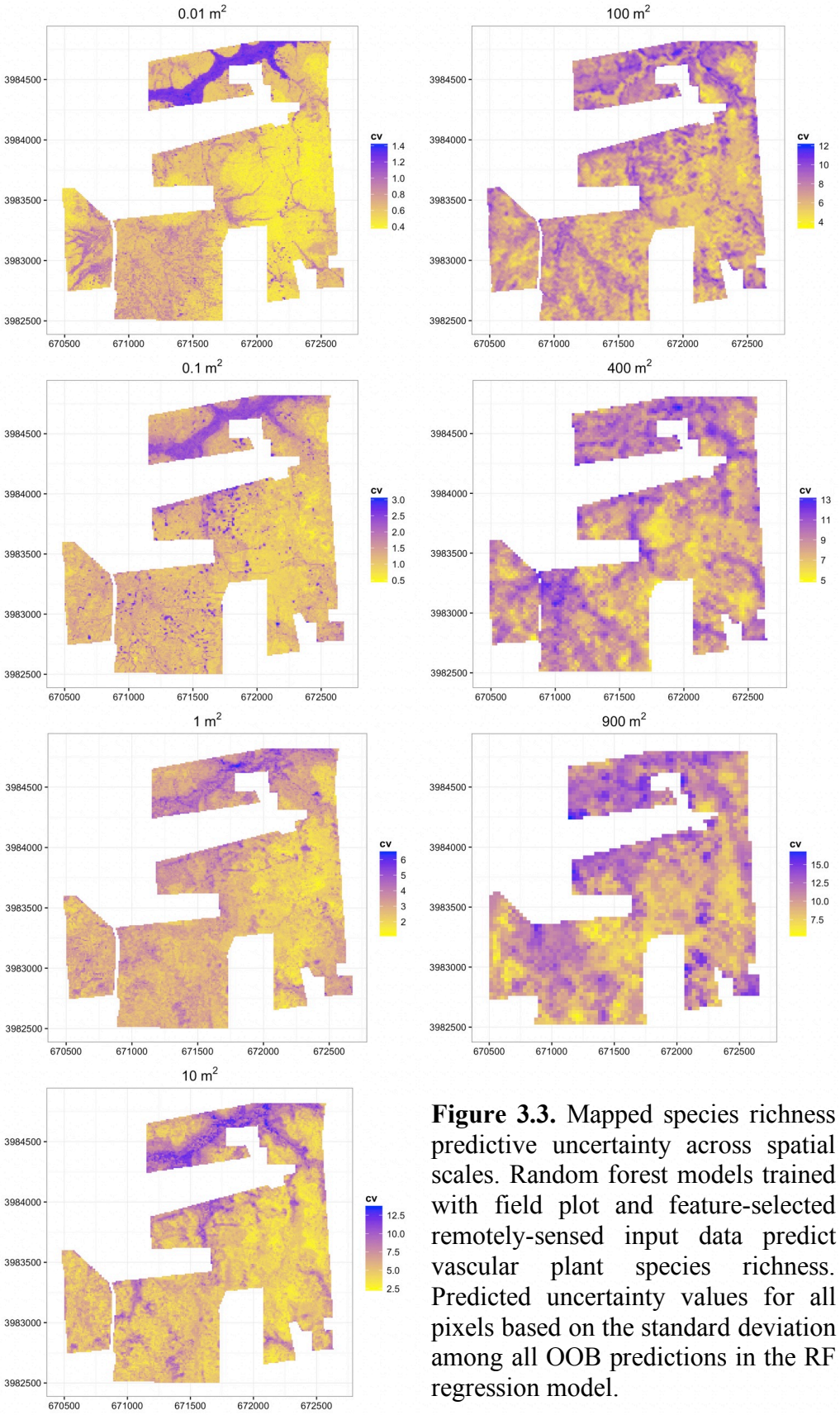


Figure 3.3. Mapped species richness predictive uncertainty across spatial scales. Random forest models trained with field plot and feature-selected remotely-sensed input data predict vascular plant species richness. Predicted uncertainty values for all pixels based on the standard deviation among all OOB predictions in the RF regression model.

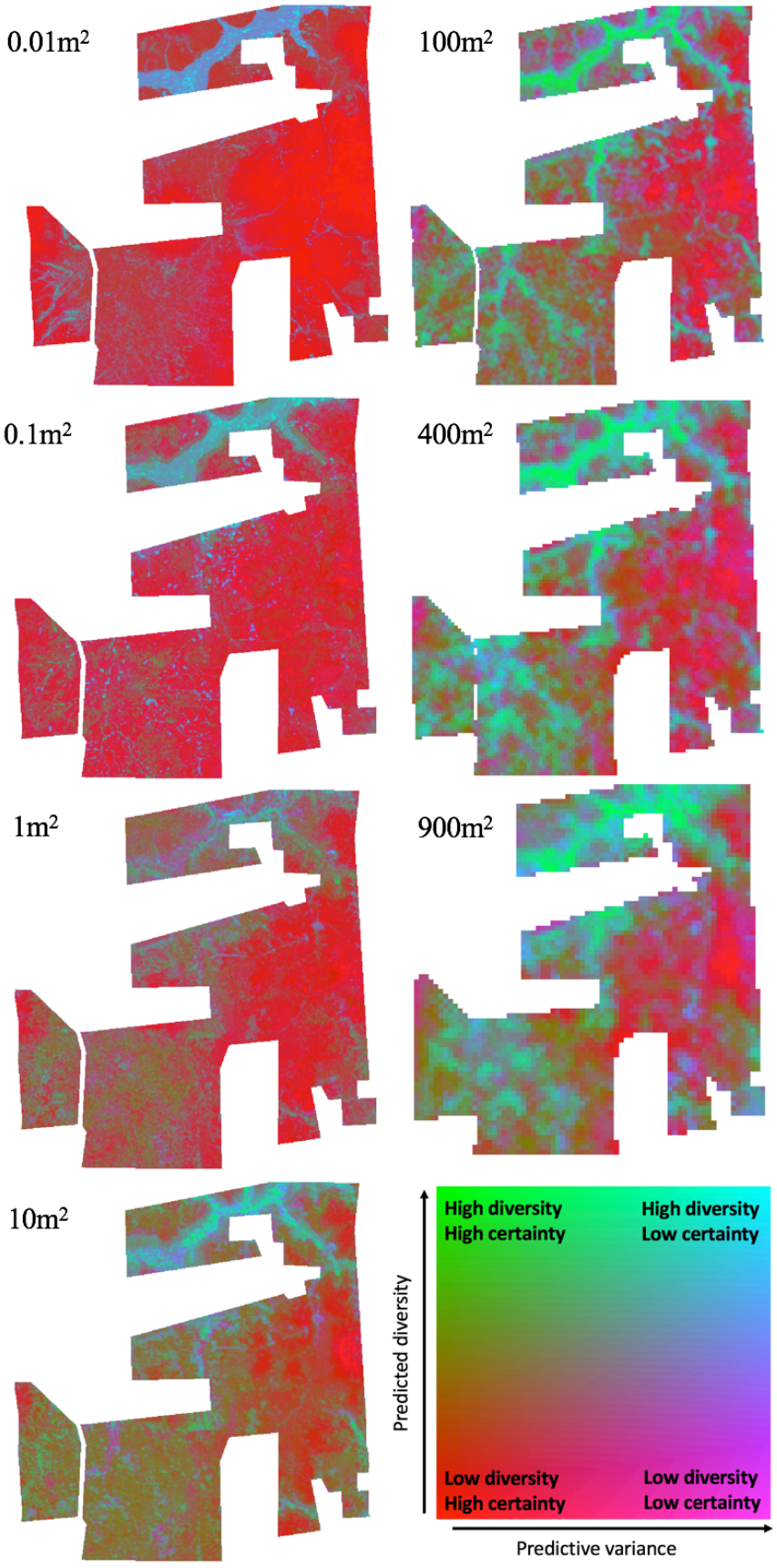


Figure 3.4. Combined diversity-uncertainty maps across spatial scales. Map output reflects results from Figs. 2-3. The spectrum of species richness predictions depicted from red (low) to green (high), while uncertainty mapped from white (low) to blue (high). When combined in an RBG pallet, low uncertainty values, have no effect on coloring of relative species richness, while they approach magenta (low SR) or cyan (high SR when uncertainty is high.

3.3.2. Remotely-sensed predictors of species diversity

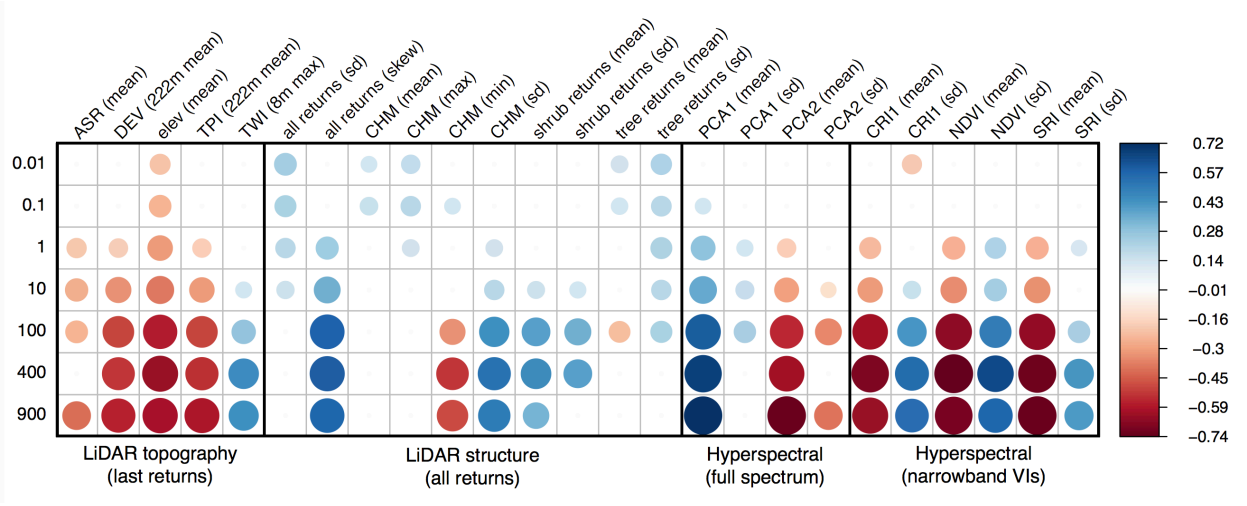


Figure 3.5. Spearman correlation matrix for selected remotely-sensed variables versus species richness values across spatial scales. When statistically significant ($\rho \neq 0$; $p < 0.01$), color and circle size vary with the extent and magnitude of the Spearman correlation coefficient, otherwise blank. See Appendix 8 for precise values.

The magnitude of Spearman correlation coefficients between individual remotely-sensed predictors and species richness values exhibit a general pattern of increasing strength of relationship (or, predictive power) with increasing spatial scale, especially in the data domains of LiDAR topography and hyperspectral data (Fig. 3.5; Appendix 8). The primary exception to this pattern comes from remotely-sensed indicators of forest height and structure derived from LiDAR (first and all returns), where the effect of scale (magnitude and direction) varies differentially among individual predictors (Fig. 3.5; Appendix 8).

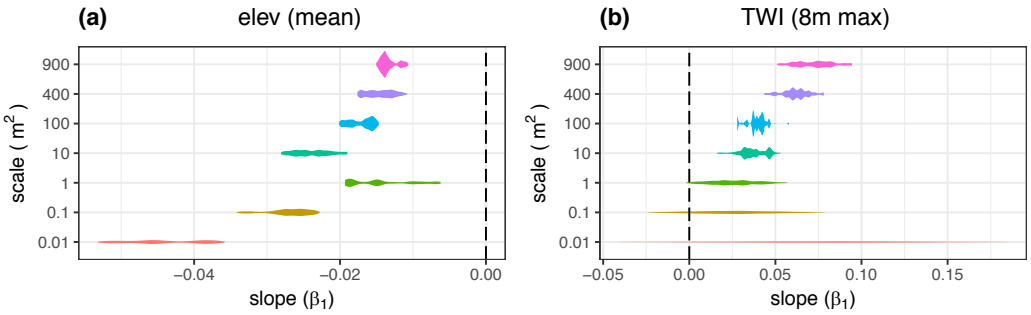


Figure 3.6. Plant richness ~ LiDAR topography slope parameter posteriors across scales. Posterior distributions reflect slope values between predictors and plant richness values at seven spatial scales.

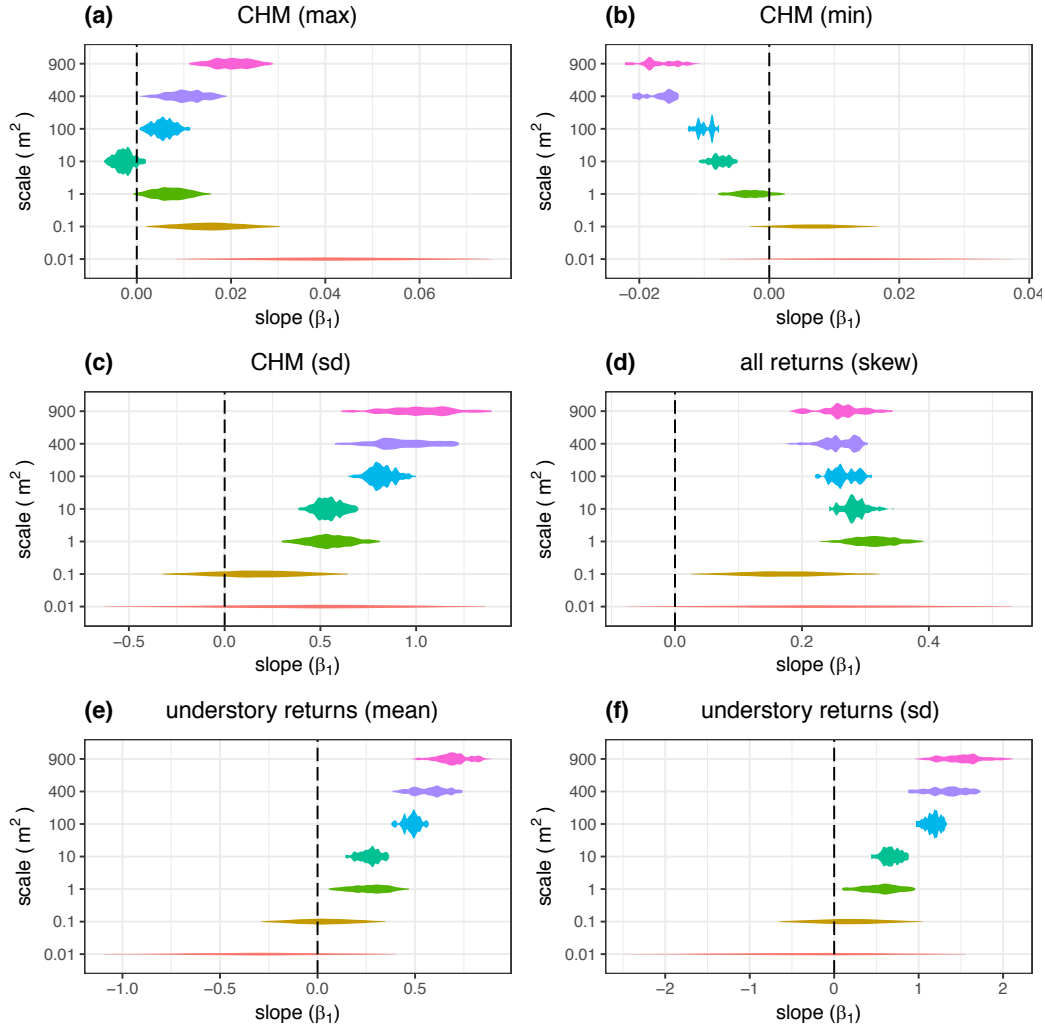


Figure 3.7. Plant richness ~ LiDAR structure slope parameter posteriors across scales. Posterior distributions reflect slope values between predictors and plant richness values at seven spatial scales.

Univariate Bayesian GLMs reflect the role of topography in driving species richness patterns, such that elevation is consistently significantly negatively correlated across scales, while topographic wetness index (TWI) is significantly positive, though relationships below $10m^2$ were not significant (Fig. 3.6a-b). Canopy height (CHM_max) is significant and positively correlated with richness at the smallest and largest spatial scales, but insignificant at intermediate scales (Fig. 3.7a). GLMs reveal a mirrored, diverging pattern between the minimum CHM and variance in CHM (sd) across spatial scales (Fig. 3.7b-c). While both models are non-significant at the two smallest spatial scales ($0.01m^2$ and $0.1m^2$), minimum CHM grows increasingly positive with increasing spatial scale (i.e. at larger spatial scales, lower canopy minimums are associated with higher plant richness totals), whereas increasing canopy height variance is linked to increasing species richness. The skewness of all returns (whereby increasingly positive values reflect higher proportions of biomass in the lower-canopy) fails as a significant predictor of richness at the smallest spatial scales, yet exhibits an increasingly positive relationship at larger spatial scales (Fig. 3.7d). Understory returns reflect this same general pattern, such that the greater the proportion of LiDAR returns between the ground and 1.37m, the greater the chance for higher species richness totals (Fig. 3.7e-f).

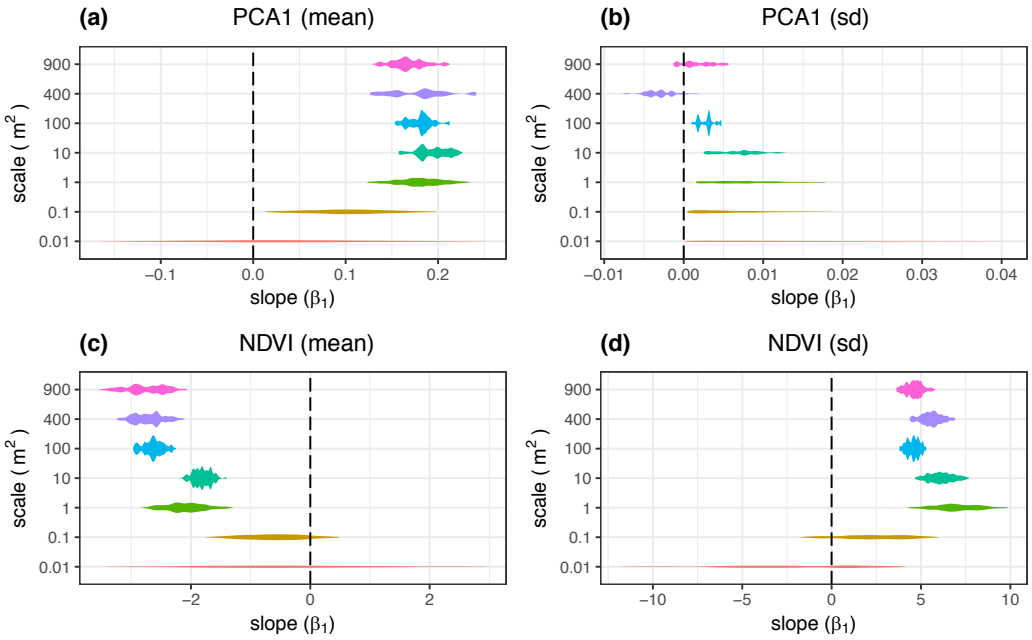


Figure 3.8. Plant richness ~ hyperspectral slope parameter posteriors across scales. Posterior distributions reflect slope values between predictors and plant richness values at seven spatial scales.

Hyperspectral predictors reveal two contrasting trends between aggregated pixel means and variance measures. The mean of PCA 1 (whose sign is arbitrary) exhibits a consistently strong relationship with species richness across all but the smallest spatial scales (Fig 3.8a). Spectral variance of PCA 1 is non-significant or otherwise only weakly correlated with species richness at all other scales (Fig. 3.8b). On the other hand, mean values for narrowband vegetation indices (VIs) tend towards an increasing negative relationship with richness at larger spatial scales, while the standard deviation of the aggregated pixel values (i.e. the spectral variance term) grows increasingly positive with increasing spatial scale (e.g. NDVI in Fig. 3.8c-d). Both relationships are non-significant at the smallest spatial scales. Mean values of the primary components (PC axes) of full-spectrum imagery expectedly reveal orthogonal trends (see Fig. 3.5; Appendix 8) between the PC axes 1 and 2.

3.4. Discussion

3.4.1. Mapping vascular plant species richness at different spatial scales

In this study, we find evidence of scale-dependence in the predictive power of species richness models based on remotely-sensed data (Table 3.3). Cross-validated R^2s increase steadily from the smallest scale (R^2=0.14) to the maximum (R^2=0.68 at 100m^2), before declining slightly to the largest scales of 400m^2 and 900m^2 (R^2=0.64, R^2=0.59, respectively). These results corroborate findings from other studies that find increasing predictive power when modeling species richness at larger spatial scales and/or at coarser resolutions (Rocchini et al. 2007; Fricker et al. 2015). That increasing model power as a function of increasing spatial scale assumes a modal, rather than monotonic, form indicates a local maxima and characteristic scale for predicting species richness at 100m^2 in our study area - a result of the idiosyncratic relationship between plant size and density as well as the characteristics of the G-LiHT remotely-sensed imagery, including spatial precision and resolution.

Models are evaluated based on 10-fold cross-validation, a relatively conservative estimate of model fit balancing the desire to maximize the impact of scarce training data against the need to minimize the potential for over-fit predictions (Hastie et al. 2009). Our results compare favorably with similar studies, though caution is advised with direct comparisons, especially when studies differ widely in the statistics used to quantify model fit or predictive power. For example, Cayuela et al. (2006) explained 44% (GLM deviance reduction, D^2 =0.44) of variability in Fisher's alpha diversity of tropical tree species in Mexico with remotely-sensed data and interpolated environmental data. Simonson et al (2012) explained 50% (linear regression, R^2 = 0.5) of woody species diversity (Shannon index) with LiDAR data in Mediterranean woodlands. Hernández-Stefanoni et al. (2014) used LiDAR data to explain 46% of variance in plant species richness (CV

$R^2 = 0.46$) in Yucatan Mexico. And Fricker et al. (2015) accounted for 35% (adjusted $R^2 = 0.35$) in tree species richness, and up to 52% (adjusted $R^2 = 0.52$) for understory trees and shrubs alone using high resolution optical imagery fused with LiDAR on Barro Colorado Island, Panama. However, even with comparable statistics, inter-site (and inter-biome) comparisons are problematic when response parameters differ (e.g. all plant richness versus woody species only), and when the eccentricities of the remotely-sensed data coupled with the those of the ecological properties of the study site, as well as those of error in data and model, preclude direct comparison.

When inversed, and used in a predictive mode, species richness maps reveal several distinct landscape patterns, and provide a distinct visual representation of how species-area relationships (SARs) manifest at landscape scales (Fig. 3.2). By far the most dominant visual pattern is that of soil moisture, as predicted plant richness clearly tracks known topographic gradients, including the riparian zone to the north of the study site, as well as upland-bottomland transition from the center right to the lower middle zones. With the increase in spatial scale (and spatial resolution of input imagery and output maps), the distinct boundaries of the hydrological flows gradually dissipate. Uncertainty maps show that high relative uncertainty values likewise align with known topographic patterns (e.g. riparian areas) at small spatial scales, that are visibly diminished at large spatial scales (Fig. 3.3). To better quantify relationships between remotely-sensed environmental gradients and predicted richness, post-hoc parametric tests are better suited.

3.4.2. Spatial scale constrains remotely-sensed predictors of plant species richness

Nonparametric predictive models excel in their ability to employ highly non-linear, non-intuitive relationships among covariates upon which high accuracy predictions can be made (De'ath 2007). In this study, we employ a data mining approach to maximize explained variance when generalized to quasi-independent data via cross-validation. Model-selection for predictive

maps is thus driven by agnostic feature-selection rather than being guided by ecological intuition (Evans et al. 2011). However, lacking explicit parameters to test, these black box models fail to provide a robust basis for inference into the relationships between predictor and response (Prasad et al. 2006). For this, post-hoc parametric tests can be performed to uncover underlying relationships between the response, plant species richness across all seven spatial scales, and key derived, remotely-sensed predictor variables (Olden et al. 2008; Evans et al. 2011).

Post-hoc analysis of Spearman correlation coefficients among those key predictors (those remotely-sensed variables that consistently rank highest in featured-selected nonparametric models) reveals a distinct pattern of increasing strength of correlation with increasing spatial scale, particularly in the LiDAR topography and hyperspectral data domains. In most cases, the largest correlation coefficients occur at $400m^2$ or $900m^2$. With the exception of CRI1 (sd), which actually registers a significant negative correlation at $0.01m^2$, before changing sign (to positive) at larger spatial scales, all predictors in these two remotely-sensed data domains retain a consistent sign (directionality) and generally larger degree of correlation with increasing spatial scale. LiDAR structure, on the other hand, appears to be an exception as Spearman coefficients do not appear to exhibit any one consistent pattern across scales.

3.4.2.1. LiDAR last returns – topography

Terrain variables derived from LiDAR last (ground) returns confirm the predominant role of topography in driving landscape patterns of plant diversity. Predictive maps of plant richness across scales confirm a clearly visible elevation gradient and visible stream channels and bottomland hydrology in topographic wetness maps (Fig. 3.1, upper right). Elevation is notable for being highly significant with plant richness across all spatial scales (Spearman ranges from $\rho = -0.21$ at $0.01m^2$ to -0.63 at $400m^2$; $p<0.01$) (Fig. 3.6a). While TWI is inversely correlated with

elevation and not significant at the smallest spatial scales, it nonetheless grows increasingly positive with spatial scale (Fig 3.6b). These results support findings from other studies that found remotely-derived abiotic factors related to topographic and edaphic properties as significant predictors of plant species richness (Fricker et al. 2015; Zellweger et al. 2016). Similar patterns have been found in empirical studies in Piedmont forests, where bottomland and riparian communities tend towards higher species richness values compared with their upland counterparts (Matthews et al. 2011).

3.4.2.2. LiDAR first returns - canopy height

Our results find several LiDAR-derived canopy height estimates to be significant predictors in plant diversity models, though effects vary differentially with scale. Specifically, results support the hypothesis that canopy height (CHM mean) is positively correlated with plant richness – a pattern supported by broad-scale empirical findings (Hakkenberg et al. 2016; Marks et al. 2016). Though LiDAR first returns from the upper canopy tend to underestimate canopy height because they are more likely to emanate from trees' upper shoulders rather than the uppermost lead, this bias is largely systematic, and thus doesn't greatly affect the determination of relative heights or sub-canopy structure as derived from LiDAR all returns (Sexton et al. 2009; Kane et al. 2010). Interestingly, the relationship between maximum canopy height and plant richness is not significant at the intermediate scales of $1m^2$ and $10m^2$ (Fig. 3.7a). This trend may be explained by the characteristic scale of the large crowns of the upper canopy in the study site, such that the projected ground area for those regions of maximum height at the finest pixel resolution ($16m^2$) may be dominated by that tree's trunk, thus precluding establishment by herbs or saplings. At the largest spatial scales ($400m^2$ and $900m^2$), richness levels increase as the vertical projection of laterally-extended canopy fall well beyond the stem (especially for wide crown

broadleaf trees), and thus lead to positive, if modest, increases in plant diversity levels. At the smallest spatial scales, on the other hand, where mean richness at $0.01m^2$ <1, the presence of a trunk actually has an inflationary effect on richness prediction.

Unlike canopy maximum height, canopy minimum height (CHM min) shows a monotonic trend with increasing spatial scale, from non-significance towards significant, and increasingly negative relationships (Fig. 3.7b). Depending on the scale, canopy minimum may denote shorter, immature trees or full gaps with no cover, where canopy height values approach zero. That this relationship grows more negative with increasing spatial scale is indicative of the characteristic scale of canopy gaps and understory regrowth. Specifically, understory regrowth (and, by extension, higher diversity levels) result from the increased light environment in and near pronounced canopy minimums (e.g. gaps detectable with LiDAR) (Clebsch & Busing 1989; Garbarino et al. 2012). Because speciose regrowth may or may not occur within the bounds of vertical projection of fine-resolution LiDAR-determined gaps due to micro-environmental limitations and/or errors in spatial precision, fine scale correlations are non-significant (Spies et al. 1990; Chen et al. 1995). However, when aggregated to $900m^2$, CHM minimums become a sufficient proxy for heterogeneity in the light environment underneath gap-filled canopy that may increase the potential for species rich understory growth (Canham et al. 1994; Ferris et al. 2000).

3.4.2.3. LiDAR all returns - canopy heterogeneity

LiDAR returns are effective tools for wall-to-wall modeling of the three-dimensional structure of canopy and sub-canopy elements (Lesky et al. 2002). Forest structure, the three dimensional distribution of biomass, reflects the net outcome of environmental conditions, biotic competition, disturbance, and stochastic events on plants in a forest stand (Peet & Christensen 1987; Neumann & Starlinger 2001). Forest structure has been used to characterize successional

stage which, through various mechanisms, has been found to correlate with temporal trends in plant species richness in Piedmont forests (Peet & Christensen 1988). Specifically, several studies have found significant correlations between measures of forest structural heterogeneity such as variance in basal area and vertical layering with overall tree species diversity (Chiarucci & Bonini 2005; Marks et al. 2016). Our results generally corroborate these findings, and can be explained with the 'diversity begets diversity" hypothesis (Diamond 1988; Palmer et al. 2007). According to this line of reasoning, increasing structural heterogeneity drives the spatial partitioning of light, water and nutrient resources, thus constraining the breadth of species' environmental niche space and facilitating complementarity in resource use between species with different resource demands and acquisition traits (Franklin 1988; Halpern & Spies 1995). However, this relationship may be correlational, and not necessarily causative. For example, greater plant diversity could be expected to promote a larger range of species traits, like growth rate and size, which then manifest as a greater degrees of structural complexity (Hakkenberg et al. 2016).

Standard deviation in the canopy height model is one effective means for characterizing heterogeneity in the upper canopy surface. While CHM_sd was not significantly correlated with plant species richness at the smallest scales, with increasing spatial scale, CHM_sd is significant, and increasingly positively correlated with richness (Fig. 3.7c). At the smallest scales, variance in the CHM may or may not precisely correspond with the ground location of canopy gaps and differential light penetration thought to drive understory richness. Instead, areas of increased CHM variation may be spatially offset from where the increased light resources of gaps actually reach the forest floor. At larger spatial scales, on the other hand, canopy height variation drives heterogeneity in the resources environment at a scale large enough to incorporate both closed

canopy and open gaps, which in aggregate are hypothesized to drive increases total plant richness levels (Pickett & White 1985; Canham et al. 1994).

Skewness of all LiDAR returns registered the highest degree of correlation with plant richness among all LiDAR-derived structural indicators (Spearman $\rho = 0.61$ at $400m^2$; Appendix 8). Skewness in the vertical distribution of LiDAR returns is an indicator of the degree to which biomass is concentrated in the lower canopy (positive skewness) versus the upper canopy (negative skewness). While this measure is insignificant at smaller spatial scales, at larger scales it grows increasingly positive indicating that top-heavy canopies (e.g. single layer canopies) in the study area tend towards lower diversity, while stands with a more bottom-heavy distribution of biomass (e.g. those having robust lower and understory vegetation) tend towards higher diversity levels (Fig. 3.7d). These results are in keeping with those from other studies finding biomass skewness as an effective proxy for species diversity, especially as a means for estimating the presence of smaller size classes which tends towards higher abundance levels, and thus by virtue of sampling effect, higher richness (Knox et al. 1989; Lorimer & Frelich 1997).

As understory individuals tend to drive overall plant richness values, it follows that understory returns (LiDAR returns below 1.37m) should be expected to be critical in the detection and prediction of plant richness. Unlike canopy measures, mean understory height and the standard deviation of these understory heights, shows an increasingly positive relationship with species richness with increasing spatial scale (Fig. 3.7e-f). At the largest spatial scales, higher understory returns simply connote the presence of a robust understory. While this relationship is not significant at the smallest spatial scales, with an increase in scale, the relationship becomes significant, and increasingly positive. Owing to the rarity of any LiDAR return between the forest

floor and 1.37m especially in the more depauperate upland areas, the presence of any understory returns increases the likelihood of higher richness totals at larger spatial scales.

3.4.2.4. Hyperspectral imagery

Hyperspectral imagery is effective for detecting inter-species trait variation in canopy structure, water content, and foliar biochemicals such as photosynthetic pigments like chlorophyll and carotenoids as well as structural compounds like lignin and cellulose (Curran 1989; Asner 1998). Owing to the daunting size and inherent redundancy of the 114 band hyperspectral image layers, hyperspectral imagery was reduced to two derived products: full spectrum indices (PCA 1-3) and narrowband vegetation indices (VIs). While the former benefits from incorporation of ~99% of variance among all hyperspectral image layers, PCA indices are unit-less, relational, and limited to the extent of the training data space. Narrowband VIs, on the other hand, each capture just a small portion of the reflectance spectrum but excel as interpretable indicators of known traits (e.g. CRI – as an index of carotenoid pigments) based on empirical optimal band selection (Ustin et al. 2009).

While reflectance spectra are helpful for remote species discrimination (Naidoo et al. 2012; Roth et al. 2015), intra-specific variation in spectral signatures may outweigh inter-specific variation, especially due to varying illumination conditions as well as idiosyncrasies of differential environmental stress on biochemical plant pigments (Price 1994; Nagendra 2001). Owing to the problematic nature of estimating richness based on tallying disparate identified taxa with resolved spectral libraries, spectral variability has been proposed as a surrogate for species richness that is agnostic to the identity of individual taxa (Palmer et al. 2002; Rocchini et al. 2007; Rocchini et al. 2010). The spectral variation hypothesis (SVH), posits a positive relationship between spectral variation due to foliar biochemical diversity and phylogenetically-conserved trait variation among

taxa which can aid remotely-sensed biodiversity modeling efforts (Cavender-Bares et al. 2016). Our results confirm the utility of the SVH, especially for narrowband VIs, though full spectrum PCAs provide mixed results.

Variance in the principal components of the full spectrum imagery (e.g. PCA1_sd) was only significantly correlated at intermediate scales, and when significant, the relationship was not strong (ρ = 0.13 at 1m^2, $p<0.01$; ρ = 0.23 at 100m^2, $p<0.01$). This result contracts with the aggregated PCA1 mean, which was among the most highly correlated variables with species richness (ρ = 0.72 at 900m^2, $p<0.01$). On the other hand, among narrowband VIs, we see a near universal monotonic trend of increasing correlation strength with spatial scale. For narrowband VIs, variance measures (e.g. sd) are consistently positively correlated with plant richness (Fig. 3.5, Fig. 3.8c-d). The scale-dependence of the SVH, and specifically its increasing utility at coarser spatial scales, has been documented in other studies (Palmer et al. 2002; Rocchini et al. 2007). While the positive correlation of sub-pixel variance in VIs with plant richness supports the SVH, mean narrowband VI values are near-universally negatively correlated with plant richness.

This result contrasts with those from several studies that found a positive relationship between NDVI and plant richness, especially in arid environments where mere presence of chlorophyll could significantly predict richness (Levin et al. 2007). However, in the continuous, multi-strata canopy of the study area, the opposite seems to be the case. Interestingly, VI means in this study area may not be informative for detecting patterns of foliar greenness (which VIs like NDVI were originally conceived to do) when they easily saturate in dense and continuous canopy (Haboudane et al. 2004). Conversely, the negative relationship may instead reflect patterns of an absence of greenness, and specifically the confounding effect of canopy structure in the optical signal when mutual shading of leaves within tree crowns subsume a substantial portion of pixel's

endmember fractions (White et al. 2010). Supporting this conclusion, measures of structural complexity like CHM_sd - which are positively correlated with richness (ρ = 0.53 at 400m^2; $p<0.01$) - are negatively correlated with mean VI values (e.g. CHM_sd ~ NDVI (mean) ρ = -0.4; $p<0.01$).

3.5. Conclusion

While there is no inherently correct scale at which to observe the mechanism and patterns of diversity, incongruity in scale between field data (e.g. sampling size, design, and intensity), remotely sensed data (e.g. spatial resolution and geolocational precision), and that of target biota (e.g. size, density, and coverage extent) can confound conclusions on the scale-dependent relationships driving these patterns. In this study, we model the wall-to-wall distribution of vascular plant species richness across seven scales to assess how change in spatial grain affects the significance and direction of the relationship between remotely-derived predictors and plant richness. To the best of our knowledge, this study is novel in mapping total plant species richness, including those understory taxa obscured from passive sensors, across multiple spatial scales in a compositionally and structurally complex forest. Predictive maps of multi-scale plant richness provide a realistic representation of species-area relationships in relation to known environmental gradients, and has immediate application for conservation and habitat modeling at landscape scales. Results confirm the scale-dependence of remotely-sensed predictors of species diversity in relation to plant size and density in relation to environmental factors like canopy gaps and understory growth. The degree to which these results can be readily extrapolated to other study systems is not yet fully resolved, and this lacuna invites comparison among disparate sites to ascertain the generalizability of using remotely-derived data products to predict multi-scale

biodiversity patterns in other ecosystems. With the increasing availability of highly accurate and information-rich remotely data, ecologists will increasingly be able to test hypotheses regarding the phenomenological patterns and underlying drivers of biodiversity at previously unobtainable scales.

CHAPTER 4

MODELING PLANT COMMUNITY-CONTINUA IN A PIEDMONT FOREST LANDSCAPE WITH LIDAR-HYPERSPECTRAL REMOTE SENSING

4.1. Introduction

Efforts to model ecosystem functioning, map habitat quality, and monitor biodiversity in temperate forests hinge upon accurate knowledge of the taxonomic composition of those communities (Running & Coughlan 1988; Tilman et al. 1997; Anderson & Ferree 2010). While field-based methods for mapping forest composition can be highly accurate at small spatial scales, they are extremely resource-intensive at larger scales (Condit 1995). However, when paired with field plot data, aerial remote-sensing offers an efficient, repeatable, and synoptic platform with which to explicitly characterize species distributions and forest composition at multiple scales (Schmidtlein et al. 2007; Anderson et al. 2008; Roth et al. 2015).

Remotely-sensed imaging spectroscopy, or hyperspectral imaging, is one particularly powerful tool for mapping spatial variation in forest composition (Asner 1998; Clark et al. 2005). Hyperspectral sensors are capable of distinguishing biochemical absorption features of foliage at different narrowband wavelengths that can be used to infer species-specific spectral signatures (Curran 1989). However, despite its strengths, the direct estimation of leaf-level properties based on imaging spectroscopy alone remains problematic. Several studies have noted how intra-specific variation in leaf optical properties can overwhelm attempts at species discrimination based on in-situ reflectance measurements and spectral libraries (Kodani et al. 2002; Ghiyamat & Shafri 2008). The ability to translate canopy reflectance to foliar chemistry, diagnostic of species-specific

spectral signatures is further complicated by a number of factors including the phenological stage of the vegetation, environmental conditions (including climate, nutrient availability, and biotic interactions), epiphyll cover, and, importantly, canopy structure (Okin et al. 2001; Asner & Martin 2009). Canopy structural attributes - including height, stratification, plant architecture, and leaf orientation – shape the fundamental means by which photons interact with multiple surfaces in a forest canopy and are thus implicated as one of the most pernicious sources of noise in the retrieval of a consistent spectral signature in dense canopy (Asner 1998; Torabzadeh et al. 2014).

To mitigate the confounding effect of canopy structure, active remote sensing using LiDAR has been found an effective complement to passive (optical) instruments for modeling canopy structural properties (Ustin & Gamon 2010). When used in combination, hyperspectral-LiDAR sensors can simultaneously characterize both foliar biochemistry and canopy biophysics (Dalponte et al. 2008). Several studies have found combined LiDAR-hyperspectral datasets outperform either used on its own (Hill & Thomson 2005; Leutner et al. 2012), though time lags between LiDAR and hyperspectral flights as well as imperceptive calibration and data fusion methods may diminish these improvements (Torabzadeh et al. 2014). In recent years, combined LiDAR-hyperspectral sensors have seen increasing prominence, allowing for pixel-level fusion and calibration of both sensors in a single platform (e.g. Kampe et al. 2010; Asner et al. 2012; Cook et al. 2013).

Integrated LiDAR-hyperspectral systems have proven competent at species detection, particularly in sparse canopies (Naidoo et al. 2012). However, the task of unpacking the taxonomic identity of the entire vascular plant community is greatly complicated in forests with relatively high species richness, especially those consisting of multiple layers of overlapping strata (Thenkabail et al. 2004; Clark et al. 2005). In closed canopy forests, attempts at inferring composition are limited largely to emergent and upper-canopy exposed stems, as understory

elements may be completely obscured from optical sensors. Owing to the recognition of uncertainty in identifying sub-canopy taxa, as well as individuals smaller than the resolution of the remotely-sensed imagery, one promising approach to the mapping of landscape turnover in plant composition is to shift the scale of analysis to that of whole communities, rather than individuals or species (Ferrier & Guisan 2006; Schmidtlein et al. 2007). This stand-level approach embraces sub-pixel heterogeneity and exploits the correlative relationships between the remotely-sensed signal and the optical-structural properties of aggregate stand traits to predict total stand composition, inclusive of those understory individuals out of reach of optical sensors (Gamon 2008; Ustin & Gamon 2010; Leutner et al. 2012).

Community-level predictive maps have tended towards the depiction of forest composition either as a patch mosaic of discrete community polygons (Foster & Townsend 2002; Bunting et al. 2010) or as a continuous compositional gradient overlaid on a pixel-based raster grid (e.g. Schmidtlein & Sassin 2004; Feilhauer & Schmidtlein 2009; Middleton et al. 2012). This distinction in the translation of abstract compositional space into a spatially-explicit geographic space has its historical roots in the community-continua concept in community ecology (Daubenmire 1966; Whittaker 1967). And while no one approach is optimal for all applications, the choice of depicting communities as discrete units versus continuous turnover in composition has important implications for the utilization and interpretation of compositional maps (Hakkenberg et al. 2017).

In light of the need to operationalize the mapping of forest composition for diverse applications from land management to ecosystem modeling, this study employs spatially-nested field plots in conjunction with active and passive remotely-sensed data from the Goddard LiDAR, hyperspectral, thermal (G-LiHT) airborne sensor to map vascular plant composition in a compositionally- and structurally-complex Piedmont forest landscape in NC, USA. In so doing,

we are guided by three key questions (1) How do community-units compare with ordination gradients as the basis for the predictive mapping of landscape turnover in plant composition? (2) How does compositional resolution affect the accuracy and precision of community characterization in geographical and ordination space? and finally (3) How effective is the fusion of LiDAR and hyperspectral sensors in predicting the primary components of vascular plant composition in the Piedmont study area?

4.2. Methods

4.2.1. Study site

The 2.8 km^2 study area encompasses natural and semi-natural forests in the Blackwood Division of Duke Forest, Orange County, NC, excluding all areas of plantation forest, clear-cuts, and human habitation (Fig. 4.1). Duke Forest is a research forest located in the central portion of the EPA level 3 "Piedmont" ecoregion (Griffith et al. 2002). The heavily forested Piedmont region occupies the southern remnants of the Taconic orogeny, where mountain-building events and subsequent weathering has left the Piedmont as a

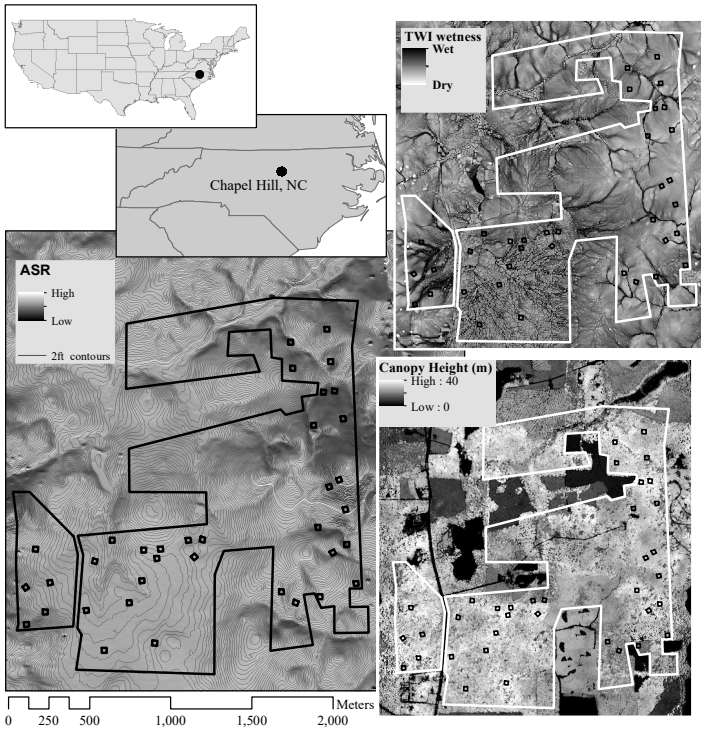


Figure 4.1. Duke Forest Blackwood study area extent. The study area excludes areas of plantation forest, clear-cuts, and human habitation. Topographic wetness index (TWI) is a proxy for soil moisture (upper right). Canopy height (lower right) is derived from G-LiHT LiDAR. Elevation (lower left) is featured with ASR (annualized solar radiation) for shading and 2ft (0.6m) contours, with squares representing plot location and extent.

spatially complex geologic mosaic. Soils exhibit a high degree of edaphic and hydrological variability, closely tracking substrate composition and topographic position, with sandy sediment-derived soils in floodplains and predominantly clay soils in the uplands (Stone et al., 1985). Following selective cutting, agriculture, and grazing in the 19^{th}-20^{th} century, the Duke Forest Blackwood Division is today composed largely of secondary old-field successional pine and mature hardwood forests (Peet & Christensen 1988). The study area possesses modest topographic relief, with elevations ranging from 127m to 228m. Temperatures in this section of the Piedmont range from a mean monthly minimum of 3.8 °C in January to 31.1°C in July, with mean annual precipitation of 1072 mm (NCDC, 2011).

4.2.2. Field data

Vegetation plot locations throughout the study area were randomized within the constraints of topographically stratified study design, ensuring roughly equal representation among plots by aspect and elevation across an east-west and a north-south topographic aspect gradient (Fig. 4.1). Accordingly, plots span the range of physiognomic forest types in the study area, inclusive of upland, riparian, and bottomland forest ecosystems. Vegetation sampling of spatially-nested subplots was performed at each site at $0.01m^2$, $0.1m^2$, $1m^2$, $10m^2$, $100m^2$, $400m^2$, and $900m^2$ scales (Appendix 4). Sampling of nested subplots allow for efficient collection of cover data for all vascular plants that is robust to problems of scale and spatial autocorrelation (Peet et al. 1998; Peet et al. 2012). Within each nested subplot from 0.01m2 to 400m2, cover was estimated for all vascular plant species using Carolina Vegetation Survey categorical cover-class codes (see Peet et al. 1998, 2012), after which class percentage values were estimated based on the midpoint value for each cover class. Using this nested sampling design, each plot's four $100m^2$ subplot cover values were averaged to provide mean cover for the entire $900m^2$ plot, based on Carolina

Vegetation Survey protocols adapted for square plots (Peet et al. 1998). In total, 36 30x30m plots (and 144 nested 100m^2 subplots at 5 spatial scales) were recorded – a number deemed adequate considering their large size in comparison to the total mapped area of Duke Blackwood (0.1% of the entire area). Botanical nomenclature follows Weakley (2015). The resultant plot data are available in Vegbank (http://vegbank.org; Peet et al. 2012).

Due to the importance of geo-locational precision in fitting field plots with high resolution (2x2m) remotely-sensed imagery, all efforts were made to achieve sub-meter spatial locational accuracies for plot extents. Thus, due to GPS signal interference in dense forest canopy, plot and subplots locations were determined based on triangulation of Ground Control Points (GCPs) visible in fine resolution (0.15m) Digital Orthophoto Quadrangle (DOQ) imagery (NC OneMap Geospatial Portal 2016). The distance of all plot vertices to GCPs in the fine resolution ortho-rectified imagery was determined in the field with tape measures and in a GIS environment to triangulate actual plot vertex coordinates based on the DOQ.

4.2.3. Remotely-sensed data

Hyperspectral and LiDAR data covering the study area were collected with NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) airborne imager (Cook et al. 2013). The G-LiHT airborne imager was designed to enable the integration of co-registered data of similar grain size at the instrument level, rather than via post-hoc fusion of instrument data flown on different platforms and acquired on different dates. G-LiHT consists of a scanning LiDAR, profiling LiDAR, imaging spectrometer, Global Positioning System and Inertial Navigation System (GPS-INS) and time server, data acquisition computer, and downwelling irradiance spectrometer. The system utilizes commercial, off-the-shelf instrumentation for use with local general aviation aircraft in an attempt to reduce costs and simplify worldwide deployment (Cook

et al. 2013). The VQ-480 (Riegl USA, Orlando, FL, USA) airborne laser scanning (ALS) system possesses a 60° field of view, a mean return density of up to 50 laser pulses/m^2, and 10 cm diameter footprint at the nominal operating altitude of 335m. The Hyperspec imaging spectrometer (Headwall Photonics, Fitchburg, MA, USA) operates in the 407–1,007 nm spectral region, providing imagery at a ≤5 nm full width half maximum (FWHM) spectral resolution, 2m spatial resolution, and 12-bit radiometric resolution. Data for our study site were collected during leaf-on conditions during Oct. 2013.

All remotely-sensed predictor variables are derived from four G-LiHT Level 3 (L3) products mapped as a 2m resolution raster grid: (1) a digital terrain model (DTM), (2) all LiDAR returns, (3) a canopy height model (CHM), and (4) a 114-band atmospherically-corrected hyperspectral surface reflectance image stack (Cook et al. 2013). The four data products were used to generate three categories of derived products, namely: (1) LiDAR topography, (2) LiDAR canopy structure, and (3) hyperspectral foliar reflectance (Table 4.1). All derived LiDAR topography layers are calculated from the DTM at a 2m resolution, with the exception of the topographic wetness index (TWI), which was also calculated at 4m and 8m resolutions to capture scale-dependence in water flow models (Beven & Kirkby 1979). LiDAR canopy structure was derived from CHM product as well as all LiDAR returns (Cook et al. 2013). To capture foliar chemistry, full-spectrum and narrow-band indices were used (Ustin et al. 2009; Asner & Martin 2009). Principal components analysis (PCA) was used to reduce the full 114-band hyperspectral dataset to three orthogonal layers (PCA 1-3), which cumulatively explained 99.7% of variance among all bands. In addition, narrow-band indices with established precedence in the literature were employed, including the Anthocyanin Reflectance Index 1 (ARI1), Anthocyanin Reflectance Index 2 (ARI2), Carotenoid Reflectance Index 1 (CRI1), Carotenoid Reflectance Index 2 (CRI2),

Photochemical Reflectance Index (PRI), Red Edge Position Index (REPI), Simple Ratio Index (SRI), and Normalized Difference Vegetation Index (NDVI) (Table 4.1).

All derived geophysical variables at 2m (and 4m and 8m resolutions for TWI) were aggregated to a 30m grain size that coincides with the extent of field plots and was deemed the smallest acceptable minimum mapping unit for assessing forest composition in the Piedmont. Specifically, annualized solar radiation (ASR), deviation from mean elevation (DEV), elevation (elev), slope, and topographic position index (TPI) are calculated only as means while for CHM, mean, minimum, maximum, standard deviation, skewness, and kurtosis was computed. All other variables were aggregated by mean, minimum, maximum, and standard deviation. Pixel aggregation reflects the dominant signal (e.g. mean), and sub-plot / sub-pixel variation in that signal including extremes (e.g. min, max, and standard deviation). All derived products and subsequent analyses were calculated using the software R, v. 3.3.1 (R Core Team 2016), with the exception of annual average solar radiation (AASR) calculated using the ArcMap version 10.4 AASR plug-in (ESRI 2016).

4.2.4. Data Analysis

4.2.4.1. Community-unit classification

Unsupervised classification was performed to determine the most parsimonious crisp partitionings of field plots using the Optimal Partitioning of Similarity Relations (optpart) clustering function (Roberts 2016b). Optpart is a non-hierarchical, iterative re-allocation algorithm that seeks to maximize the partana ratio, a measure of within-cluster similarity versus among-cluster similarity for a given number of clusters. Optpart has been found to consistently rank among the highest performing clustering algorithms, based on goodness-of-clustering evaluators across datasets and dissimilarity matrices (Aho et al. 2008; Roberts 2015). Because no one *a priori*

number of classes exists that optimizes the trade-offs between sensitivity and specificity for a given dataset, we first evaluated relative performance in class differentiability from two to eight clusters – with two clusters being the minimum number of possible clusters, and more than eight groupings deemed excessive for so small an area (Kaufman & Rousseeuw 2005). The number of clusters used for community-unit classification across the two eight-cluster combinations was determined based on local maxima of partana ratio and average silhouette width, the mean similarity of each plot to other plots in its cluster versus its similarity to the most similar cluster (Rousseeuw 1987; Roberts 2016b). Based on these criteria, the two- and seven-class partitions were selected to guide final community-unit classification due to their large relative partana ratios (1.44 and 1.47, respectively) and local maxima in average silhouette widths (0.13 and 0.16, respectively) (Fig. 4.2).

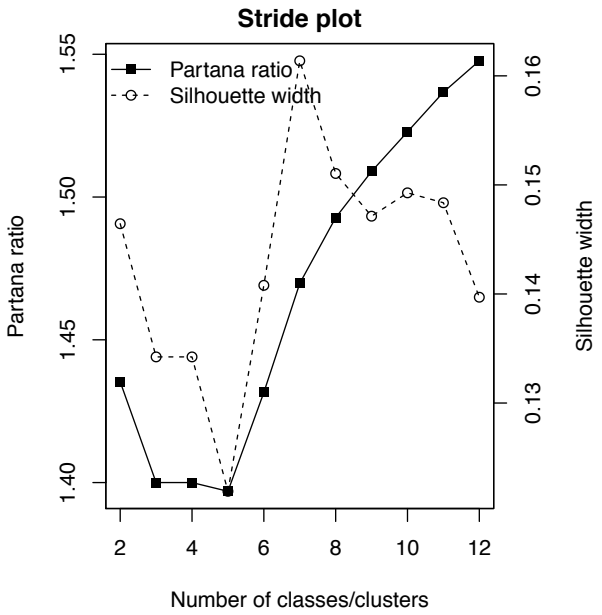


Figure 4.2. Mapped species richness predictions across spatial scales. Random forest regression models trained with field plot and feature-selected remotely-sensed input parameters predict vascular plant species richness across seven spatial scales. Predicted species richness values for all pixels based on the mean vote among all OOB observations in the RF regression model.

Following community-unit clustering, community labels and diagnostic species were apportioned to all clusters to reference the central taxonomic concept and physiognomy of each community type. Diagnostic species were likewise assessed based on indicator species scores, canopy dominance, and expert opinion for each community-unit. Indicator species (the product of the relative frequency and relative average abundance in clusters) were determined using the indval

function in labdsv package in R (Roberts 2016a), while canopy dominance was based on cover when present (Chytrý et al. 2002).

Table 4.1. Remotely-sensed predictor variables. All remotely-sensed predictors represent aggregates of 2x2m pixels resampled to a 30x30m (900m2) output resolution.

Category	Predictor	Abbv.	Equation/Source
LiDAR topography (last returns)	Average Solar Radiance (annual)	ASR	(ESRI 2016)
	Deviation from mean elevation	DEV	$DEV = \frac{z_0 - \bar{z}}{SD}$; where z_0 is the elevation of the focal pixel, $\bar{z}$ and SD are mean and standard deviation of elevation in a 222m window (De Reu et al. 2013)
	Height above EGM96 (Earth Gravitational Model 1996) geoid	elev	DTM (Cook et al. 2013)
	Slope in degrees	slope	DTM (Cook et al. 2013)
	Topographic Position Index	TPI	$TPI = z_0 - \bar{z}$; where z_0 is the elevation of the focal pixel and $\bar{z}$ is mean elevation in a 222m window (De Reu et al. 2013)
	Topographic Wetness Index	TWI	$TWI = ln\left(\frac{a}{tan\beta}\right)$; where a is the local upslope area and β is slope (Beven & Kirkby 1979)
LiDAR canopy height (first returns)	Canopy height model	CHM	CHM (Cook et al. 2013)
LiDAR canopy structure (all returns)	All return heights	all_returns	LiDAR returns (Cook et al. 2013)
	Tree return heights †	tree_returns	LiDAR returns (Cook et al. 2013)
	Understory return heights †	understory_returns	LiDAR returns (Cook et al. 2013)
Hyperspectral foliar reflectance	Anthocyanin Reflectance Index 1	ARI1	$ARI1 = \frac{1}{\rho_{550}} - \frac{1}{\rho_{700}}$ (Gitelson et al. 2002)
	Anthocyanin Reflectance Index 2	ARI2	$ARI2 = \rho_{800}\left[\frac{1}{\rho_{550}} - \frac{1}{\rho_{700}}\right]$ (Gitelson et al. 2002)
	Carotenoid Reflectance Index 1	CRI1	$CRI1 = \frac{1}{\rho_{510}} - \frac{1}{\rho_{550}}$ (Gitelson et al. 2002)

Carotenoid Reflectance Index 2	CRI2	$\text{CRI2} = \frac{1}{\rho_{510}} - \frac{1}{\rho_{700}}$ (Gitelson et al. 2002)
Normalized Difference Vegetation Index (narrowband)	NDVI	$\text{NDVI} = \frac{\rho_{800} - \rho_{670}}{\rho_{800} + \rho_{670}}$ (Haboudane et al. 2004)
Principal Component Axis 1	PCA1	PCA axis 1 (ρ_{407}: ρ_{1007})
Principal Component Axis 2	PCA2	PCA axis 2 (ρ_{407}: ρ_{1007})
Principal Component Axis 3	PCA3	PCA axis 3 (ρ_{407}: ρ_{1007})
Photochemical Reflectance Index	PRI	$\text{PRI} = \frac{\rho_{531} - \rho_{570}}{\rho_{531} + \rho_{570}}$ (Gamon et al. 1995)
Red Edge Position Index	REPI	REPI=max[(ρ_n+1-ρ_n)/10] where 690 ≤ n ≥ 750 (Souza et al. 2010)
Simple Ratio Index (narrowband)	SRI	$\text{SRI} = \frac{\rho_{800}}{\rho_{670}}$ (Haboudane et al. 2004)

† Tree returns defined as all returns above 1.37m

Field plot groupings, corresponding to each of the designated community-units, were used to train a random forests (RF) classification model parameterized with all remotely-sensed predictors (Table 4.1) over 2000 separate trees using the randomForest package in R (Liaw & Wiener 2002). RF is a nonparametric modelling approach well-suited to high dimensional, "small n large p" ecological data, whose variables exhibit nonlinear and complex interactions (Pal 2005). The RF algorithm utilizes an ensemble of classification trees to produce highly accurate and unbiased predictions based on votes across bootstrap replicates that are largely immune to over-fitting (Prasad et al. 2006). While RF classification is easily adapted to fuzzy prediction based on the distribution of out-of-bag (OOB) predictions for each observation, our goal was to predict a single, 'crisp' class membership for all pixels in the study area. Therefore, each pixel was classified according to the majority OOB 'votes' across all trees (Cutler et al. 2007). Class weights, based

on the respective proportions of field plots in each community-unit, were applied to training data to reduce prediction bias in underrepresented classes (Chen et al. 2004). Classification accuracy of the predicted map was assessed using a weighted Cohen's kappa coefficient (κ), a measure of inter-rater agreement that takes into account the possibility of the agreement occurring by chance and is weighted to adjust for expectation in the frequency of observations (Cohen 1968), as well as confusion matrices from which overall agreement (OA), user's agreement (UA), and producer's agreement (PA) was calculated.

4.2.4.2. Compositional ordination

Ordination methods are an expedient tool for transforming large species-by-plot matrices into a reduced-dimensional space. This data reduction technique enables a plot's composition, relative to other training plots, to be referenced by its coordinates (mean and variance) in ordination space. In this study, we used Nonmetric Multidimensional Scaling (NMDS), a technique with precedence in the remote sensing and community ecology literature (Thessler et al. 2005). NMDS preserves the ordering relationship among samples in ordination space based on their ranked dissimilarity in compositional space (Legendre & Legendre 2012). Final scores are arbitrary in an absolute sense, but meaningful as relative indictors of *n*-dimensional compositional dissimilarity (where n+1 represents the number of total species in a species-by samples data matrix of relative abundances) in a more tractable k-dimensional ordination space (where $k < n-1$). NMDS was chosen over alternate ordination methods because it makes no assumptions about dimensionality, linearity or the shape of species-response curves to gradients (Kruskal 1964). Because the NMDS procedure seeks to align sample data in ordination space according to ecological dissimilarity, we derived a compositional distance matrix among plots using Bray–Curtis dissimilarity, a distance matrix for species data that controls for bias due to joint absences and double weights joint

presence as a strong indicator of resemblance (Legendre & Legendre 2012). Before running the NMDS on vegetation samples, rare species occurring on <5% of plots were removed from the dataset (Appendix 5).

To determine the optimum number of ordination axes, a step-down procedure was performed to determine goodness-of-fit based on 180 preliminary NMDS runs (30 runs each on 1-6 dimensions). With extension to higher dimensions only producing marginal improvements in goodness-of-fit, the three-axis solution was deemed an acceptable solution balancing model parsimony (low dimensionality) and explanatory power, with NMDS 1-3 explaining 76% of all variation in vascular plant composition (R^2=0.76) (Appendix 9). In addition, a three-dimensional (3D) ordination space allows for an interpretable 3D visualization of compositional space, readily translated to an RGB color scale. Because the numerical algorithm of NMDS cannot guarantee a global solution, the model was run for 2000 iterations and the solution with the lowest dissimilarity between ordination and Bray-Curtis distances (stress=0.2) was selected as the final model. As there is no intrinsic ordering to the final NMS ordination axes, the final model was rotated using PCA to align it along orthogonal axes of maximum floristic variation (Clarke 1993). All procedures in the NMDS analysis, including the derivation of dissimilarity matrices, were performed using the ecodist package (Goslee & Urban 2007) in R, v. 3.3.1 (R Core Team 2016).

4.2.4.3. Compositional gradient modeling

Continuous gradient modeling of forest composition was performed by running three parallel RF regression models, each parameterized with remotely-sensed predictors and trained to field data, to predict continuous values for NMDS axes 1-3 at a scale of 30m across the study site. While the RF algorithm is robust to multi-collinearity and over-fitting, prediction accuracy can be affected when the number of features (p) is significantly higher than the number of samples (n)

(Toloşi & Lengauer 2011). Thus, for each of the three models (predicting NMDS 1-3 scores), feature selection was performed to reduce the total number of predictors. To accomplish this, full models utilizing all predictors were run for each of the three continuous response variables (i.e. NMDS 1-3) and predictive accuracy was recorded based on 10-fold cross-validatio, an OOB model evaluation procedure that iteratively withholds random subsamples of the training data for use as an independent validation of model fit. Thereafter, features were iteratively withheld, until convergence on a final set of predictors that maximizes cross-validated prediction accuracy (Appendix 10). Feature-selected models, trained with all field plot data and remotely-sensed inputs, were then used in a random forest regression model to predict NMDS axis scores for all pixels throughout the study area. Predicted values are based on the mean vote (rather than the majority vote in the case of classification) from all OOB observations, and rescaled to fit the range of NMDS values in the training dataset (Zhang & Lu 2012). As a visual aid, predicted NMDS 1-3 scores were scaled to an 8-bit dynamic range and displayed as RGB, respectively. Accuracy estimates for final models are based on 10-fold cross-validated accuracy of the training set, and uncertainty maps were derived from the standard deviation of all per-pixel OOB votes (Singh et al. 2015). Because significance in the statistical relationship between remotely-sensed predictors and NMDS axes 1-3 is obscured in the black box RF model, post-hoc Spearman correlations, a rank-based measure of association that facilitates application to non-normal data distributions, were run for all predictors. The final Spearman correlation matrix includes all remaining predictors (after feature selection and removal of duplicates) used in RF regression models, significant ($\rho \neq 0$; $p<0.01$) with at least one of the NMDS axes.

4.3. Results

4.3.1. Community characterization

Sampled field plots, stratified along the two predominant environmental gradients, span the range of compositional variability of the relatively taxonomically diverse and structurally heterogeneous study area. In total, 208 vascular plant taxa were identified in field plots, among which 144 remained after removing rare species occurring on less than 5% of all plots (Appendix 5). Among all 8860 woody stems recorded, diameter at breast height (DBH) ranged from the minimum recorded 0.1cm to 101.5cm (mean DBH=7.6cm; DBH standard deviation=11.7cm). Mean vascular plant species richness (SR) was 61 species per plot (range: 35-100 species). Among all plots, *Acer rubrum, Liquidambar styraciflua, Liriodendron tulipifera, Quercus alba*, and *Pinus taeda* possessed the highest cover values (in descending order) and *Acer rubrum, Carya glabra, Cornus florida, Muscadinia rotundifolia, Nyssa sylvatica, Prunus serotina,* and *Toxicodendron radicans* occurred on all plots.

For efficient, intuitively-interpretable data visualization, the three-dimensional coordinates of sampled plots and predicted pixels in NMDS ordination space are reduced to a single, semantically-meaningful color referencing its compositional identity. Thereafter, all colored points and convex hulls (Fig. 4.3) as well as mapped pixels (Fig. 4.4) consistently translate their location in three-dimensional ordination space as an RBG color combination, such that NMDS 1 maps onto the red color gun, NMDS 2 onto green, and NMDS 3 on blue. Unlike points and pixels, whose RGB coloration is derived directly from NMDS 1-3 values, community-unit convex hulls assume the color representing the central concept (and centroid coordinate) of all constituent plot points (Fig. 4.3). Community labels reference the physiognomy of each community type (e.g.

Upland sub-xeric vs. Bottomland hydric) and its central taxonomic concept by means of diagnostic species (Table 4.2).

Table 4.2. Community-unit characterization

Classification	Label	Diagnostic species
2 class	Upland sub-xeric (USX)	*Quercus montana, Juniperus virginiana, Amelanchier arborea, Vaccinium pallidum*
	Mixed Mesic (MM)	*Liriodendron tulipifera, Ulmus alata, Polystichum acrostichoides, Euonymus americanus, Aesculus sylvatica*
7 class	Upland sub-xeric (USX)	*Quercus montana, Juniperus virginiana, Amelanchier arborea, Vaccinium pallidum*
	Midslope sub-xeric (MSX)	*Carya pallida, Oxydendrum arboreum, Smilax bona-nox, Chimaphila maculata*
	Midslope mesic (MSM)	*Liriodendron tulipifera, Carya tomentosa, Hylodesmum nudiflorum, Uvularia perfoliata*
	Pine mixed (PM)	*Pinus taeda, Liquidambar styraciflua, Muscadinia rotundifolia, Vaccinium pallidum*
	Riparian (R)	*Carpinus caroliniana, Aesculus sylvatica, Polystichum acrostichoides, Galium tinctorium*
	Toe-slope mesic (TM)	*Quercus falcata, Carya ovata, Rubus pensilvanicus, Arisaema triphyllum*
	Bottomland hydric (BH)	*Quercus michauxii, Carpinus caroliniana, Smilax rotundifolia, Carex flaccosperma*

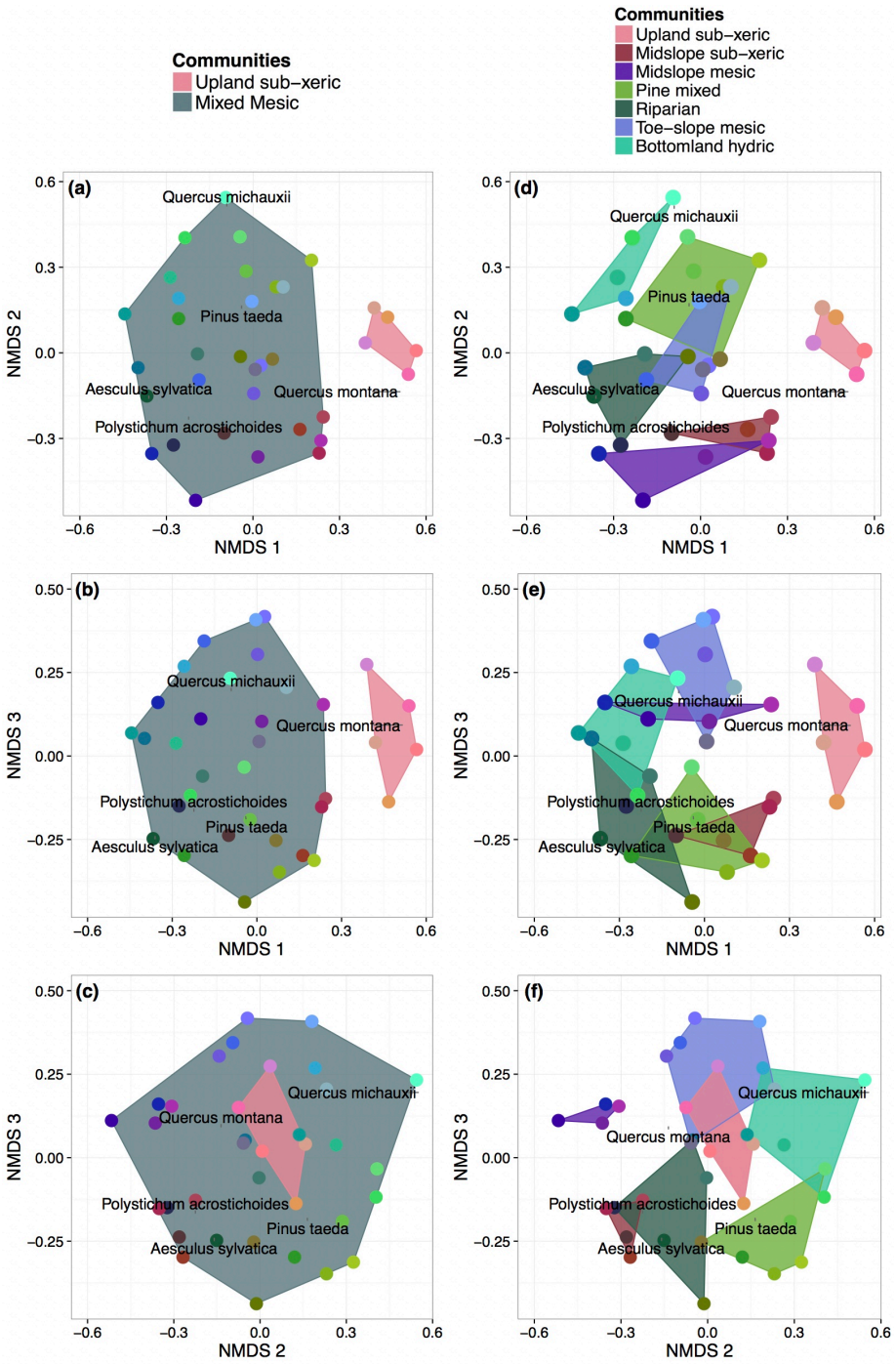


Figure 4.3. Field plots classified into two and seven community-unit clusters in NMDS space. (a-c) Two-cluster community-unit grouping, and (d-f) seven-cluster community-unit grouping in three 2-dimensional slices of 3-dimensional NMDS space. Chart colors reflect the compositional identity of plots/community types in ordination space. The color of field plots (points) and community-units (convex hulls) result from RGB color combinations of point/hull coordinates in 3-dimensional NMDS space, translated into red (R), green (G), and blue (B), respectively. Unlike field plot points, community-unit convex hulls assume the color of that class's centroid, and thus reflect it's central concept in NMDS space as an RGB color combination.

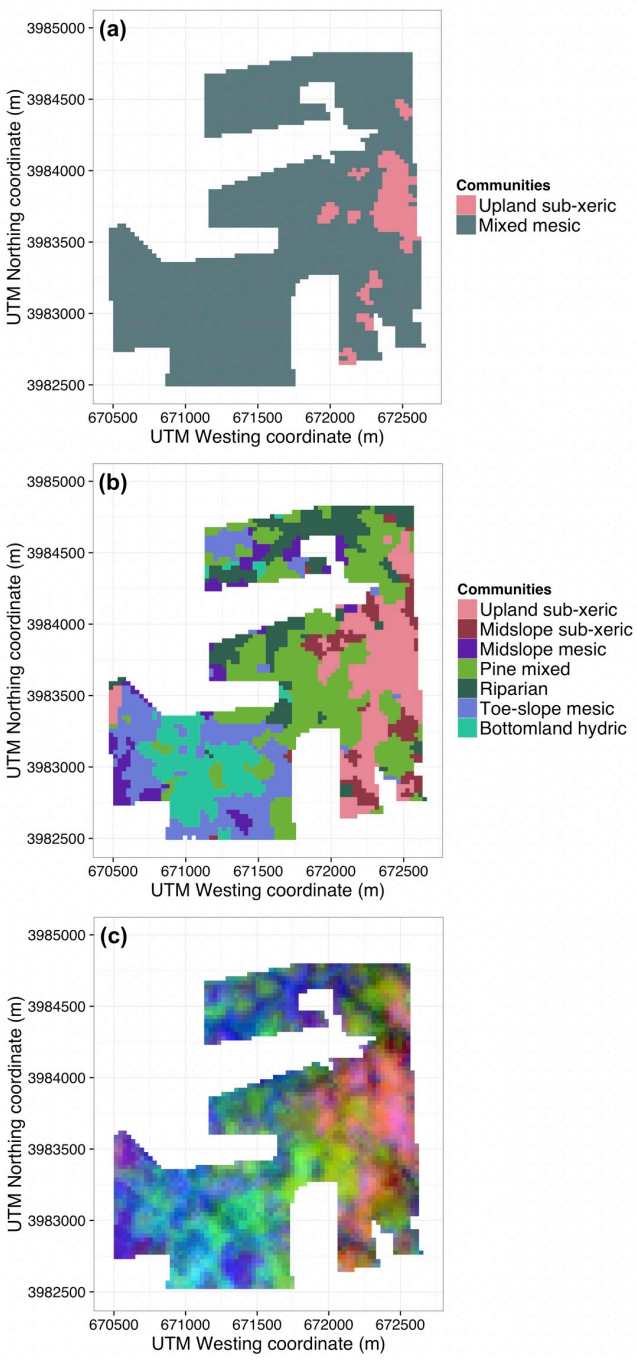


Figure 4.4. Predictive map of vascular plant composition in Duke Forest Blackwood. (a) Two community-unit patch mosaic; (b) seven community-unit patch mosaic; (c) continuous gradient map of ordinated NMDS axes 1-3, with each axis mapped onto red, green, and blue color guns, respectively. Each map is post-processed for visual clarity (See Appendix 4.5. for pre-processed predictive maps). RGB coloration is a reflection of the compositional identity of each pixel in 3-dimensional NMDS space such that NMDS 1 maps onto the red (R), NMDS 2 onto green (G), and NMDS 3 on blue (B). The RGB combination directly references the predicted location in ordination space occupied by that pixel. Colors are consistent across maps and with Fig. 4.3. See Appendix 11 for pre-processed maps.

4.3.2. Mapping vascular plant composition

For predictive mapping of landscape turnover in community-units, two- and seven-class partitionings of field plots were deemed suitable, though not optimal, solutions for balancing the simultaneous goals of model parsimony and compositional resolution (see Methods 2.4.1.). Trained with field plot and remotely-sensed data, random forest classification models predict compositional turnover for all pixels in the study site at a spatial resolution of 30m and a compositional resolution of two and seven community-units (Fig. 4.4a-b). Overall accuracy of two community-unit classification was 97% (weighted $\kappa = 0.87$), while the seven community-unit classification achieved an overall accuracy of 57% (weighted $\kappa = 0.60$) (Tables 4.3 - 4.4).

Table 4.3. Two-class confusion matrix. Total accuracy as true positive rate and Cohen's weighted Kappa coefficient (lower right). User's agreement (UA), producer's agreement (PA). (See Table 4.2 for class label acronyms).

Class	MM	USX	**UA**
MM	30	0	**100%**
USX	1	4	**80%**
PA	**97%**	**100%**	**97% (κ=0.87)**

Table 4.4. Seven-class confusion matrix. Total accuracy as true positive rate and Cohen's weighted Kappa coefficient (lower right). User's agreement (UA), producer's agreement (PA). (See Table 4.2 for class label acronyms).

Class	USX	MSX	MSM	PM	R	TM	BH	**UA**
USX	4	1	0	0	0	0	0	**80%**
MSX	0	3	0	1	0	0	0	**75%**
MSM	0	0	1	1	0	2	0	**25%**
PM	0	0	1	4	0	0	1	**67%**
R	0	0	0	1	4	0	1	**67%**
TM	0	0	1	0	0	3	1	**60%**
BH	1	0	0	1	1	1	1	**20%**
PA	**80%**	**75%**	**33%**	**50%**	**80%**	**50%**	**25%**	**57% (κ=0.60)**

Continuous compositional gradient maps were produced using the same set of remotely-sensed predictors, but using a random forest regression model to predict continuous ordination scores for all new samples (Fig. 4.4c and Fig. 4.5a-c). Based on 10-fold cross-validation accuracy assessment, remotely-sensed predictors account for 72%, 51%, and 57% of the variance in NMDS axes 1-3, respectively (Table 4.5). The standard deviation of the distribution of all RF votes provides a per-pixel measure of predictive uncertainty, which captures spatially-structured variance unexplained in statistical models, as well as data, model and geo-locational errors. (Fig. 4.5d-f).

Table 4.5. RF regression model and parameters (post feature-selection)

NMDS axis	CV R2	Feature selected predictor variables*
1	0.723	elev (mean), CRI2 (mean), PCA1 (mean), SRI (mean), PCA1 (min), NDVI (min), SRI (min), TWI (8m max)
2	0.512	all returns (skew), PCA2 (min), CHM (sd), TWI (8m min), tree returns (sd), PCA1 (max), PRI, CRI1 (sd), slope (mean), PCA1 (mean), PRI (sd), PCA3 (mean), PCA1 (sd), CHM (max), CRI1 (max), ASR (mean), TWI (4m mean)
3	0.570	REPI (mean), PCA3 (mean), CRI2 (mean), PCA3 (min), slope (mean), CRI1 (max), PCA3 (max)

* see Table 4.1 for abbreviations

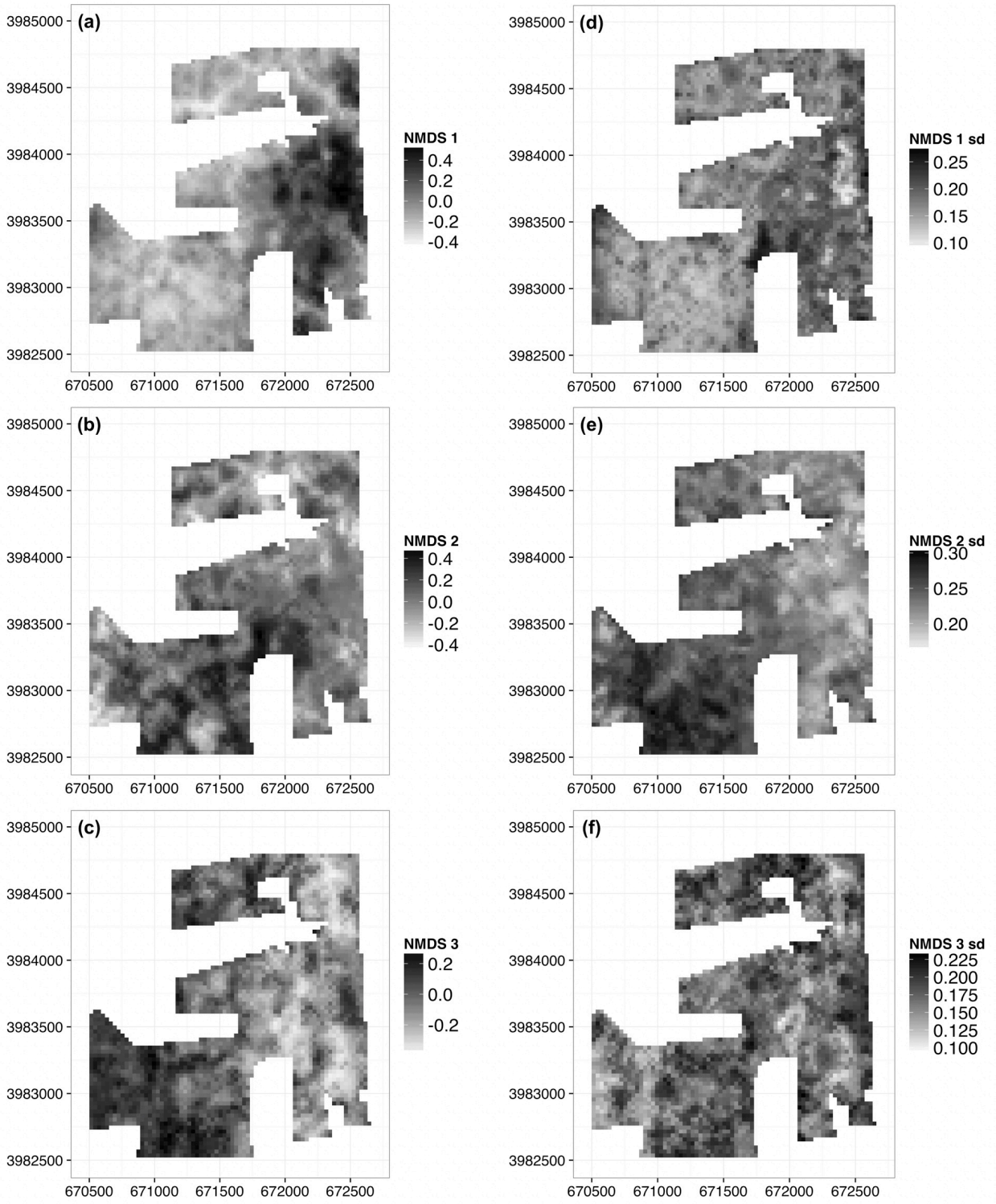


Figure 4.5. Mapped predictive mean and standard deviation of NMDS 1-3. Pixel-level prediction of (a) NMDS 1, (b) NMDS 2, and (c) NMDS 3, reflect the mean vote among all OOB observations in the RF regression model. Uncertainty estimated as the standard deviation of all 2000 RF OOB votes per pixel for (d) NMDS 1, (e) NMDS 2, and (f) NMDS 3 predictions.

4.3.3. Remotely-sensed predictors of forest composition

Environmental biplots depict the magnitude and direction of correlations between NMDS 1-3 and the three primary RS data domains: (1) topography via LiDAR ground returns, (2) canopy structure derived from all LiDAR returns, and (3) foliar biochemistry as inferred from full spectrum and narrow-band image spectroscopy (Fig. 4.6). With the exception of the positive NMDS 2 / negative NMDS 3 quadrant, LiDAR-hyperspectral derived predictors explain a substantial amount of variance across all sectors of the ordination space. Spearman's ρ correlation coefficients were used to quantify the relationship between LiDAR-hyperspectral predictors and ordination scores of all field plots (Fig. 4.7). Spearman coefficients indicate that NMDS axis 1 is largely driven by topographic gradients (e.g. elev and TWI), while NMDS axis 2 is significantly correlated with measures of canopy structure (e.g. standard deviation of tree returns and maximum canopy height). Hyperspectral imagery explains compositional variance across all three NMDS axes, with PCA axes 1-3 and narrow-band vegetation indices each exhibiting significant correlations with NMDS axes 1-3.

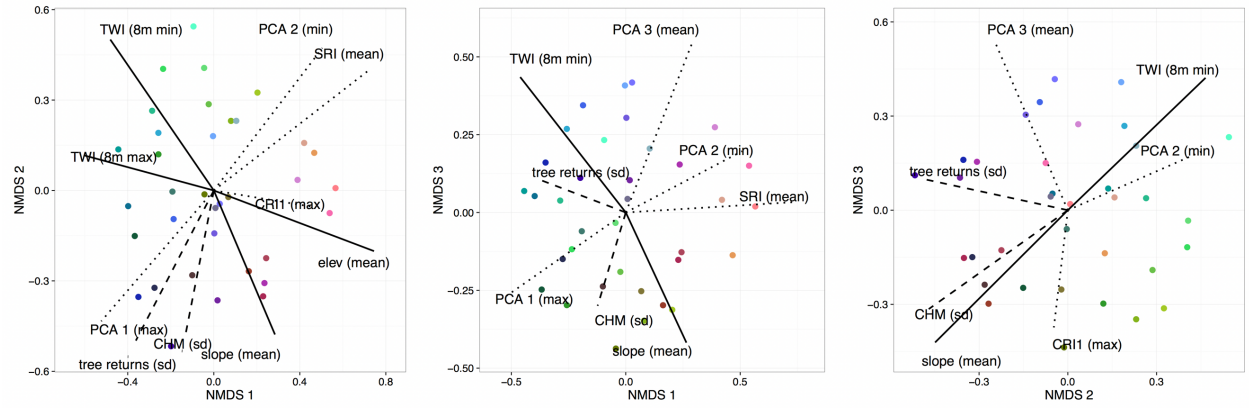


Figure 4.6. NMDS 1-3 Environmental Biplots. Field plot (point) color represents its 3-dimensional NMDS coordinates as an RGB color combination. Vectors represent statistically-significant ($\rho \neq 0$; $p<0.01$) remotely-sensed predictors of NMDS axes 1-3 from the three RS data domains: LiDAR topography (solid line), LiDAR canopy structure (dashed line), and hyperspectral reflectance (dotted line). The direction and magnitude of each vector is proportional to the Spearman coefficient of correlation with each respective NMDS axis.

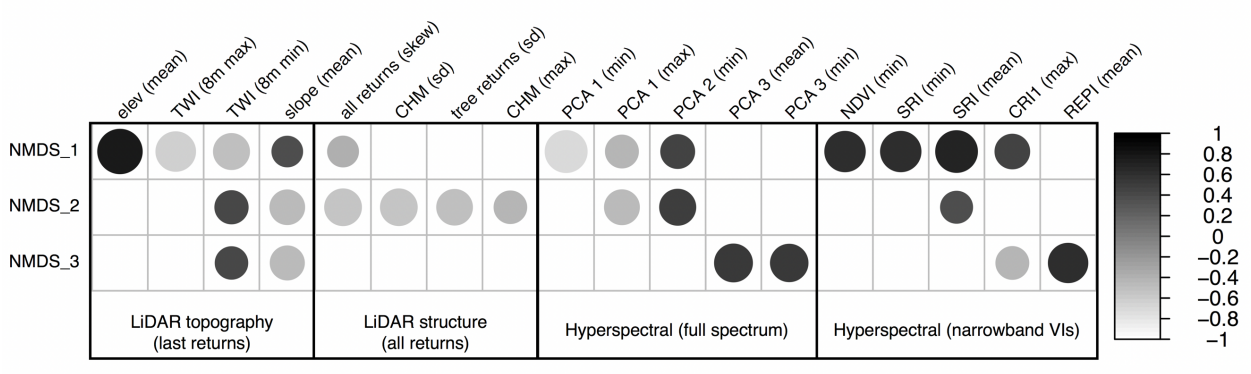


Figure 4.7. Spearman correlation matrix NMDS 1-3 ~ RS variables. Includes the six predictors most correlated with each of NMDS axes 1-3 ($\rho \neq 0$; $p < 0.01$). Shade and circle size vary with the extent and magnitude of the Spearman correlation coefficient.

4.4. Discussion

4.4.1. From mapping canopy individuals to modeling total stand composition

In monospecific communities, or in open woodlands where individual tree crowns are spatially distinct and discrete, a representation of forest composition as patch-based polygons may be relatively precise, at least for canopy species (Naidoo et al. 2012; Roth et al. 2015). However, this task is greatly complicated in multi-strata, multi-species communities where variability in leaf pigmentation and plant architecture couple to confound attempts at inferring the taxonomic identity of individuals via consistent spectral signatures (Price 1994; Ghiyamat & Shafri 2008). Owing to the uncertainties inherent in modeling total vascular plant composition in structurally and taxonomically heterogeneous forests, as well as the recognition that emergent properties of ecosystems such as functional processes, biodiversity, and habitat aggregate at the stand scale, studies have shifted focus from individuals to entire stands (McGill et al. 2006; Ferrier & Guisan 2006; Anderson & Ferree 2010). Despite the coarser spatial resolution, aggregation to the stand scale allows for a host of new derived predictors, such as those which rely on sub-pixel heterogeneity among the 225 2m pixels used to predict a single 30m response pixel. In addition,

stand scale prediction allows for the modeling of total stand composition, including individuals out of reach of optical sensors, as well as the spatially-explicit characterization compositional variance and predictive uncertainty (Ustin et al. 2009; Singh et al. 2015).

4.4.2. Community–continua in compositional and geographical space

Two primary approaches to the predictive mapping of forest communities populate the literature: the depiction of community-units as a patch mosaic of discrete vector-based polygons (e.g. Anderson et al. 2008; van Ewijk et al. 2011; Wolter & Townsend 2011) and gradient mapping of continuous compositional variation in space as a pixel-based raster grid (e.g. Schmidtlein & Sassin 2004; Feilhauer & Schmidtlein 2009; Middleton et al. 2012). These alternate approaches reflect the dialectic encompassed by the community-continua concept in community ecology. The dualistic conception of community-continua refers to the community-unit hypothesis, which describes ecological communities as distinct and repeatable entities, and the continuum theory, which envisions communities as idiosyncratic assemblages of overlapping yet independent species distributions that track environmental gradients and reflect stochastic events of dispersal and disturbance (Daubenmire 1966; Whittaker 1967). While each approach is best suited to its own unique set of applications, the two are not necessarily mutually exclusive, and can even complement each other.

Part of this divergence in the characterization of landscape turnover in composition may be attributed to ambiguity concerning the role of community-units as abstract components of a vegetation classification hierarchy versus their representation as spatially-explicit geographic entities (Austin 2013; Hakkenberg et al. 2017). As a conceptual category, community-units define the central concept of resolved community types or categories (Chytrý et al. 2002; Peet & Roberts 2013). Based on numerical analysis of representative field plots, community classification employs

community-units as descriptive categories and the primary components of a hierarchically-organized vegetation classification system (Jennings et al. 2009). However, as geographic entities, community-units are the spatially-explicit representations of those abstract categories which are posited to exist in replicate over the landscape (Austin 2013). While both the abstract and geographic community models recognize that species assemblages are fundamentally continuous in space and dynamic through time, they expediently assert the existence of community-units as practical abstractions of *n*-dimensional species assemblages.

Continuous gradient modeling, on the other hand, explicitly embraces the Gleasonian conception of spatial continuity in vegetation composition (as a continuous response variable) as well as the idiosyncratic nature of individual assemblages (modeling uncertainty as variance in the expected value or posterior). Although gradient ordination maps are still an abstraction of *n*-dimensional compositional space to *k* reduced ordination axes, because they circumvent the artefactual boundaries of discrete compositional categories, they embrace heterogeneity in multivariate analyses and lend themselves to modeling applications where the depiction of realistic landscape vegetation patterning is paramount, such as predictive habitat modeling and functional trait mapping (McGarigal & Cushman 2005; Cushman et al. 2010). While this study focused on first three NMDS axes primarily for its potential depiction in RGB color space (where NMDS 1-3 collectively explain 76% of compositional variance), advanced ecosystem modeling applications could clearly benefit from the inclusion of additional axes to better account for unexplained variance.

4.4.3. Compositional resolution of community-units

In this study, we utilize multi-scale vegetation sampling to extrapolate vascular plant species cover at 900m^2. Field plot locations were determined using a random stratified sampling

design to reflect the spectrum of variation in land-use history and environmental conditions of forests in the study site. Field plots define the extent of the compositional space to be modeled and serve as the sampling basis from which it can be most parsimoniously partitioned into manageable community-units (Kaufman & Rousseeuw 2005). Based on the comparison of *relative* dissimilarity among plots' species cover values, two and seven community-unit clusterings were deemed optimal solutions to the challenge of balancing model parsimony and compositional resolution (Roberts 2016b). However, as no single unambiguous optimization exists for the unsupervised classification of field plots, community-units, thus defined, are expedient as indicators of relative dissimilarity in the context of the focal landscape, yet somewhat arbitrary in an objective sense (Aho et al. 2008).

Viewed in ordination space, the two-class partitioning of field plots reveals a highly distinct and spatially discrete set of five points whose individual color (representing each plot's coordinates in ordination space) closely matches the mean color of the Upland sub-xeric community-unit (Fig. 4.3a-c). Classified maps reveal that the distinct aggregation of the two-class solution in ordination space is mirrored in geographical space, as the location of the outlying Upland sub-xeric class is constrained to the higher elevation areas on the east side of the study area (Fig. 4.4a). By contrast, the second unit, generically labeled Mixed mesic, takes on a grayish hue starkly deviating from its constituent point colors as a result of averaging the compositionally variegated remaining plot points (Fig. 4.3d-f). This color mismatch between the Mixed mesic class and its constituent plots is a visual confirmation of what its relatively low mean silhouette width (0.13 versus 0.40 for Upland sub-xeric) otherwise indicates: the Mixed mesic class is perhaps too generic (and the two-class solution too coarse) for effective ecological characterization.

Expanding the field plot partitioning from two to seven classes has the immediate impact of splitting the Mixed mesic class into six component parts, whose community-unit hull colors more closely resemble that of their constituent plot points. When applied for predictive mapping, the seven-unit classification reveals a more reticulated pattern of interlocked patches that track known cover types and physiognomic contours of the landscape (Fig. 4.4b). Visual observation confirms that the botanically-defined Bottomland hydric category corresponds with the fine-scale dendritic venation characteristic of topographically-defined bottomlands visible in the same location (Fig. 4.8). While compositionally similar to the flat bottomlands, steep stream channels constrain the distribution of botanically-defined riparian communities in the north of the study area. Classified pine communities form an "H" shape in the center of the study area, which are clearly visible in leaf-off reference imagery (Fig. 4.8). That the Upland sub-xeric class remains unchanged at the finer seven community-unit resolution, reaffirms both its distinction from other classes, as well as the hierarchically nested topology of all field plots in the unsupervised classification.

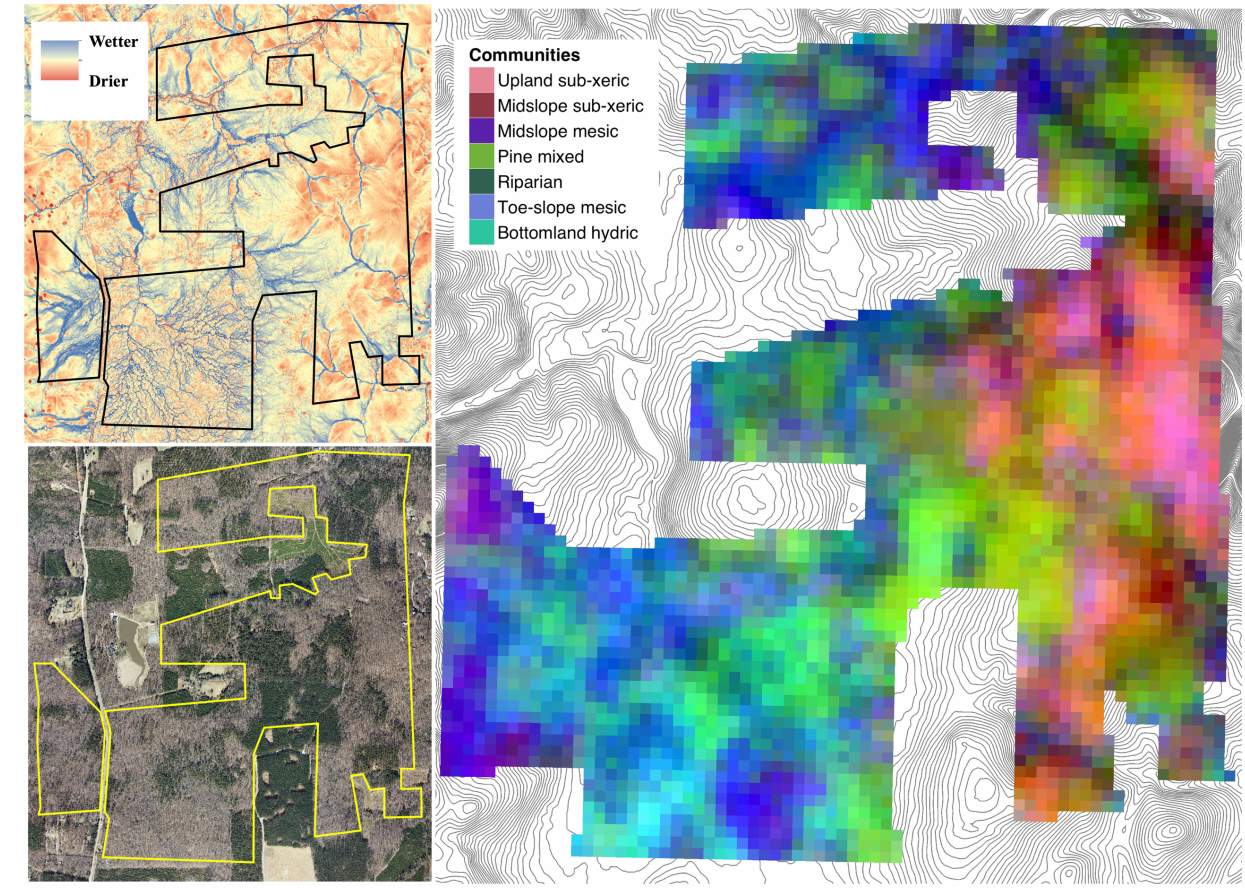


Figure 4.8. RBG compositional gradient map with community-unit legend.. The topographic wetness index (TWI) (upper left) predicts relative soil moisture. Leaf-off broadband optical imagery is particularly effective at highlighting evergreen vs deciduous cover, which in the study area translates to pine vs broadleaf cover (NC OneMap Geospatial Portal 2016). See Appendix 11 for pre-processed maps.

While finer compositional resolution allows for a more precise articulation of forest composition (as indicated by relatively higher mean silhouette widths), when applied in the context of predictive mapping, this improvement in precision comes at the expense of accuracy such that cross-validated OOB overall accuracies fall from 97% for two classes to 57% for seven (Tables 4.3-4.4). Despite the advantages of resolved compositional categories, predictive mapping based on finer partitionings of compositional space increases the potential for overlap and confusion such as when pixels are understandably, but erroneously, classified to the nearest community category. Fuzzy classification, whereby pixel class membership is framed in terms of relative membership

probability across all classes, partly circumvents this potential for confusion but still fails to address the underlying issue inherent in segmenting an *n*-dimensional continuous entity to a few somewhat arbitrary defined, discrete categories (Gopal & Woodcock 1994).

4.4.4. Gradient mapping of compositional continua

The recognition that community assemblages are fundamentally continuous in space and dynamic through time, coupled with the need for spatially-explicit characterization of landscape gradients, has given rise to predictive mapping not of community-units, but of continuous ordination scores (Schmidtlein & Sassin 2004; Thessler et al. 2005; Feilhauer & Schmidtlein 2009; Middleton et al. 2012; Gu et al. 2015). Given a representative group of samples whose location in ordination space exhibits a statistically significant relationship with continuously measured variables such as environment or reflectance, regression can be effectively employed to interpolate the mean and variance of predicted samples (Gu et al. 2015; Singh et al. 2015). Instead of classification accuracy, the power and generalizability of the regression model can be assessed based on cross-validated accuracy assessment. In this study, 10-fold cross-validation indicates that remotely-sensed predictors were able to account for 72%, 51%, and 57% of the variance in compositional ordination space for NMDS axes 1-3, respectively (Table 4.5).

This exercise in enhancing the compositional realism of modelled landscapes highlights some of its shortcomings. Although NMDS is robust to bias, it is ultimately a data visualization technique and results are only interpretable within the well-defined constraints of the field data used to construct the original space (Legendre & Legendre 2012). Individual loadings on NMDS axes are acutely sensitive to training data and highly unstable with the addition of a new samples. Thus, NMDS scores, and the subsequent coloration of the gradient map, are only meaningful as indicators of relative difference among training data. Continuous compositional gradient maps

lend themselves to the parameterization of spatially-explicit ecosystem models. However, lacking supplemental information, they fail to convey readily-interpretable information on the identity of the compositional gradients being mapped. In the context of land management, consistent coloration and the inclusion of a community-unit legend can be helpful to guide navigation of the otherwise semantically-devoid color space (Fig 4.8.). In this capacity, community-unit categories serve as complementary and compatible components of a continuous gradient map, and as landmarks for the interpretation of *n*-dimensional compositional gradients.

4.4.5. Remotely-sensed environmental correlates of vascular plant composition

Nonparametric data mining approaches have been praised for their ability to identify and explore highly non-linear, non-intuitive relationships, upon which parametric statistical tests can reveal the sign and magnitude of specific covariates (Evans et al. 2011). In this study, post-hoc analysis of remotely-sensed predictor variables confirms the complementarity of the three primary remotely-sensed environmental data domains - (1) topo-physiognomy from LiDAR ground returns, (2) biophysical structure derived from LiDAR all returns, and (3) foliar biochemistry inferred from hyperspectral canopy reflectance - in explaining compositional variation across the compositional space. LiDAR-derived terrain variables (e.g. elevation, and TWI), confirm the predominant role of topography, a proxy for soil moisture and nutrients, as an indirect driver of forest composition along NMDS axis 1. A clearly visible elevation gradient and visible stream channels in the NMDS 1 predictive map confirms this interpretation (Fig. 4.8). These results support findings from other studies that have found vegetation patterns in the Piedmont's rolling topography to track catenal formations, characterized by sandy, nutrient-poor soils in the uplands and moist, clayey soils, high in organic matter content and cation exchange capacity of the bottomland shrink-swell clays and the more silty, periodically submerged riparian areas (Peet et

al. 2014). Despite these edaphic and hydrologic differences, riparian and bottomland communities both tend towards higher stem density and species richness compared with their upland counterparts (Matthews et al. 2011).

Indicators of forest structure derived from all LiDAR returns (e.g. the standard deviation of LiDAR tree returns and maximum canopy height) almost exclusively predict variance along NMDS 2. While these metrics of forest structure are collinear with aspects of topography and canopy reflectance (Appendix 12), that NMDS 2 is so distinctly characterized by forest structure demonstrates the added utility of LiDAR canopy structure measurements for integrating the signal of successional processes for predicating composition. Of particular note in this regard is the visible distinction in the NMDS 2 prediction between the homogenous canopy structure of the secondary pine compared to the more heterogeneous mature hardwood mixed class.

Unlike topography and structure, which each load heavily onto a single NMDS axis, hyperspectral imagery explains compositional variance across all three NMDS axes. Part of this relationship stems from the correlation between hyperspectral and LiDAR-derived terrain and structural predictors which capture overlapping information on the same underlying environmental gradients. For example, narrow-band vegetation indices expressing leaf greenness like NDVI and SRI are likewise indirect indicators of underlying soil moisture gradients (e.g. NDVI mean ~ TWI max: ρ =-0.49; $p<0.01$) that directly regulate leaf physiology, and specifically, greenness reflectance. Unsurprisingly, TWI and NDVI both load heavily onto the topo-edaphically constrained NMDS axis 1 (ρ =-0.61 and ρ =0.68, respectively; Appendix 12).

These results confirm the proficiency of LiDAR sensors in detecting the primary environmental gradients constraining the distribution of individual species in a stand, whether by modeling topo-edaphic gradients (ground returns) or resource heterogeneity related to canopy

structure (all returns). Image spectroscopy, on the other hand, excels in the direct detection of physiological leaf traits via their reflectance spectra. Rather than predicting presence based on the assumption of species-environment relationships (the fundamental niche), hyperspectral sensors capture the net result of complex abiotic and biotic interactions (i.e. the realized niche). Despite some degree of collinearity in the hyperspectral and LiDAR remotely-sensed data domains, the two sets of predictors interact in complex ways as evidenced by the retention of both sets of variables in final models after feature selection.

4.5. Conclusion

With evidence accumulating for the role of local and global environmental change in altering ecological processes in forest ecosystems, predictive mapping of forest composition will be of increasing utility across a range of ecological applications. While integrated LiDAR-hyperspectral systems have proven competent at species detection, particularly for monospecific stands or in sparse woodland canopies, this task is highly problematic in closed-canopy, multi-strata, high-diversity forests. In this study, we shift emphasis from detecting the optical-structural properties diagnostic of individual crowns to those of entire communities. In so doing, we reflect a growing recognition of the scale-dependence of aggregate plant traits in driving ecosystem processes, while embracing uncertainty in the detection and prediction of constituent individuals. To the best of our knowledge, this study is novel in modeling landscape turnover in the vascular plant composition of complex forests at nested compositional resolutions. In demonstrating this translation of an abstract compositional space into a concrete geographical space, we provide a mapped example of the community-continua concept. While community-unit classification and compositional gradient regression each have applications to which they are best suited, provided

a rigorous numerical approach, the two methods are consistent, compatible, and when used jointly provide finer resolution and deeper understanding.

CHAPTER 5

CONCLUSIONS

Well-designed vegetation sampling protocols, coupled with highly resolved remotely-sensed data and advanced statistical modeling techniques are increasingly enabling the modeling and mapping of forest pattern and process at landscape scales (Turner 2005). Complex, multi-dimensional forest community properties like those of function, diversity, and composition can be imagined as an *n*-dimensional abstraction, analogous to principal components, that when combined in nonparametric models with equally complex predictor data, such as derived LiDAR-hyperspectral layers, serve as sufficient (if not always transparent) tools for robust prediction. Envisioned as such, nonparametric models are capable of predicting forest community properties like compositional identity and vascular plant species richness simultaneously at scales far finer than that of the nominal resolution of the remotely-sensed data, and at extents far larger than what field sampling would expediently support. While the maps produced in this three-part study reflect forest properties from a single date, the methods employed are applicable (recommended, in fact) for multi-temporal monitoring projects in anticipation of the deployment of a fleet of new LiDAR and hyperspectral airborne sensors and spaceborne satellites (Kampe et al. 2010; Dubayah et al. 2014).

Broadly speaking, this study was designed and implemented to develop methodologies in field sampling and Earth observation that exploit the strengths of active and passive remote sensing to uncover phenomenological patterns and underlying drivers of forest community properties at landscape scales. In the first study, results show a consistent and significant relationship between

indices of tree species diversity and several attributes of forest structure, especially maximum height, basal area size inequality, and skewness of the basal area distribution. Combining structural predictors in a nonparametric support vector regression (SVR) approach, predictive models - trained solely with structural attributes - explained 44-61% of the variance in tree species diversity in the full Piedmont dataset, and 22-71% of the variance in subsets defined by stand origin and forest type. These results confirm the utility of forest structure to predict tree species diversity in the North Carolina Piedmont without accounting for other known predictors of diversity, such as environment, soil conditions, and site history. Beyond the theoretical implications of unraveling the underlying relationship between structure (as a surrogate for successional stage) and tree species diversity, these findings highlight the empirical basis and potential for utilizing forest structure from LiDAR in predictive models of tree species diversity over large geographic regions.

In the second study, feature-selected nonparametric models based on spatially-nested plot data and over 100 aggregate, derived remotely-sensed input variables from the G-LiHT airborne sensor were used to predict wall-to-wall vascular plant species richness at $0.01m^2$, $0.1m^2$, $1m^2$, $10m^2$, $100m^2$, $400m^2$, and $900m^2$ scales in a compositionally- and structurally-complex Piedmont forest landscape. Results showed a general pattern of increasing predictive power with spatial scale, with cross-validated predictive accuracy ranging from 14% ($0.01m^2$) to 68% ($100m^2$). Due to the inherent scale-dependence in diversity patterns, as well as those between remote sensing data resolution and spatial precision in relation to plant size and density, post-hoc analyses focus explicitly on the role of spatial scale in constraining the relationship between remotely-sensed variables and species richness values. Results confirm topography derived from LiDAR ground returns (a surrogate for soil moisture and nutrients), and forest structure estimated from LiDAR all returns (a proxy for successional stage and resource heterogeneity) especially informative for

predicting multiple species' fundamental niche space driving vascular plant richness. Hyperspectral imagery, on the other hand, and spectral variability in particular, performed best at direct detection of foliar biochemical diversity - another effective indicator of canopy species diversity. In addition to insights into multi-scale landscape turnover of plant diversity, species-area relationships, and remotely-sensible correlates of plant richness, this study is, to the best of our knowledge, novel for mapping total plant richness across multiple scales in a compositionally- and structurally-complex forest landscape.

In the third study, field plot and G-LiHT data were used to map plant composition as community-continua. Results confirm the expectation of a trade-off between precision and accuracy such that an increase in compositional resolution results in a concomitant decline in overall classification accuracy. Continuous gradient maps, on the other hand, model composition not as a categorical but as a continuous response - explaining 51-72% of the variation in the first three dimensions of a nonmetric multidimensional scaling (NMDS) ordination space, respectively. Post-hoc analysis of remotely-sensed predictor variables confirms the complementarity of LiDAR-hyperspectral sensors to simultaneously predict forest composition based on the primary environmental gradients constraining species' fundamental niche, while directly detecting the foliar spectral signatures that indicate their realized niche. These results demonstrate the effectiveness of LiDAR-hyperspectral sensors to detect diagnostic optical-structural properties across the primary dimensions of vascular plant composition of the Carolina Piedmont forest site and holds promise for applications spanning local land management to biogeographic modeling.

Plant traits - evolutionary adaptations related to the uptake, use, and allocation of resources - mediate complex biotic and abiotic interactions in ecosystems (Reich et al. 2003; McGill et al.

2006; Shipley et al. 2006). While they vary widely, global plant traits cluster along ecologically interpretable primary axes of form and function, such as the plant size, leaf economics, or r-K colonization-exploitation strategies (Díaz et al. 2015). Because remotely-sensed vegetation indices specifically designed to ascertain functional traits correlate strongly with ecosystem-wide nutrient cycling rates, trait mapping has seen increasing prominence as a means towards bottom-up ecosystem process modeling (Asner et al. 2015; Abelleira Martinez et al. 2016; Jetz et al. 2016). Most recently, debate has focused on the utility of employing explicitly trait-based remotely-sensed predictors for the mapping of community properties (Gu et al. 2015; Singh et al. 2015; Abelleira Martinez et al. 2016). In this context, it should be noted that in all three studies featured here, response variables are agnostic to plant trait and ecosystem function.

Indeed, traits matter. But measuring them consistently across taxa and between communities can be highly challenging as intra-specific trait variation can rival that of inter-specific variation, and generally varies widely by species, trait, and community type (Albert et al. 2010; Hulshof & Swenson 2010). If the goal is simply prediction of community properties like composition and diversity, inference into trait-taxa relationships is non-essential. Nor is it necessary, in this limited case, to use compositional mapping as an intermediate step to functional trait mapping when models directly linking traits with remotely-sensed data would clearly be more parsimonious (Abelleira Martinez et al. 2016). Even if the response changes, say from community diversity to abundance-weighted traits, the approach needn't. Provided models are soundly constructed, if the goal is predication and not inference, understanding is not required. Alternatively, success is defined solely by predictive power.

Predictive models make a statement about an unknown quantity based on a simplified description of how that system works (Houlahan et al. 2016). They are parsimonious

representations of complex phenomena constructed from a combination of informed hypotheses, relevant data, and a robust statistical framework (Wenger & Olden 2012). The predictive models employed in this research are correlative, not mechanistic and thus if they are explanatory, it is because they are sufficiently accurate and generalizable (Urban et al. 2016). We assessed model accuracy and generalizability via cross-validation - an efficient approach to producing a pseudo-independent validation data set when training data is otherwise scarce (Arlot & Celisse 2010). Based on this model selection and assessment criterion, predictive models perform sufficiently well when they explain a significant portion of the variance in the validation set. Results are thus most immediately applicable to describe ecological patterns within the study sites, and interpolation is made within the constraints of the ecological and geographical space of the training data. But this statement on interpolation is immediately suggestive of its corollary: what is the utility of this knowledge for extrapolation outside the study sites? Without further investigation, it is impossible to say which components are transferable, and to what extent.

The complexity inherent in the task of modeling aggregate forest community properties should not be understated, and can be enough to overwhelm simple linear, parametric models. Nonparametric models, on the other hand, exploit non-intuitive, non-linear interactions between multi-dimensional remotely-sensed data and field plot data, that have proven to excel in predictive accuracy and generalizability (Schölkopf & Smola 2002; Prasad et al. 2006; Cutler et al. 2007). But lacking explicit parameters to test, "black box" nonparametric models fail to provide a robust basis for inference into the relationship between remotely-sensed environmental predictors and forest community properties (Prasad et al. 2006). For this parallel, but ultimately separate task of inferring ecological process from underlying pattern, we employ post-hoc, parametric tests of individual predictor variables. Resulting conclusions from these tests should be framed within the

epistemological goals of inference and updated understanding. That is, results are interpreted in the context of institutional knowledge (informed through a history of experiment and summarized in the literature) and conditionalized on new data.

A persistent theme throughout this three-part study is the delicate balance between prediction and inference, or analogously, between knowledge and understanding. Belief in conclusions based on knowledge separates the two. In all three studies, I acknowledge this delicate balance, applying specific statistical models based on explicit goals, be they prediction or inference. Indeed, correlation does not imply causation. A variable may be a good predictor not because it fulfills a mechanistic role in a causal relationship, but because it expediently captures enough information on "true" drivers that it proves itself sufficient to predict. Conclusions based on models require interpretation, and are thus coded by layers of belief informed by our prior understanding of the ecological system. When one adds to this uncertainty and error in data and model, between observed and latent variables, it would seem robust inference, and belief in our conclusions, a shifting target. Ultimately, progress in science is best contextualized within the scope of this dialectic between knowledge generation and updated understanding.

APPENDIX 1: FIA NORTH CAROLINA PIEDMONT TREE SPECIES LIST

Genus	Species	Common Name
Acer	*floridanum*	Florida maple
Acer	*negundo*	Boxelder
Acer	*pensylvanicum*	Striped maple
Acer	*rubrum*	Red maple
Acer	*saccharinum*	Silver maple
Acer	*saccharum*	Sugar maple
Ailanthus	*altissima*	Tree of Heaven
Albizia	*julibrissin*	Silktree
Amelanchier	*spp.*	Serviceberry spp.
Asimina	*triloba*	Pawpaw
Betula	*lenta*	Sweet birch
Betula	*nigra*	River birch
Carpinus	*caroliniana*	American hornbeam
Carya	*cordiformis*	Bitternut hickory
Carya	*glabra*	Pignut hickory
Carya	*illinoinensis*	Pecan
Carya	*laciniosa*	Shellbark hickory
Carya	*ovalis*	Red hickory
Carya	*ovata*	Shagbark hickory
Carya	*tomentosa*	Mockernut hickory
Castanea	*mollissima*	Chinese chestnut
Catalpa	*bignonioides*	Southern catalpa
Celtis	*laevigata*	Sugarberry
Celtis	*occidentalis*	Hackberry
Cercis	*canadensis*	Eastern redbud
Cornus	*florida*	Flowering dogwood
Diospyros	*virginiana*	Common persimmon
Fagus	*grandifolia*	American beech
Fraxinus	*americana*	White ash
Fraxinus	*caroliniana*	Carolina ash
Fraxinus	*pennsylvanica*	Green ash
Fraxinus	*profunda*	Pumpkin ash
Gleditsia	*triacanthos*	Honeylocust
Ilex	*opaca*	American holly
Juglans	*cinerea*	Butternut
Juglans	*nigra*	Black walnut
Juniperus	*virginiana*	Eastern redcedar
Liquidambar	*styraciflua*	Sweetgum
Liriodendron	*tulipifera*	Tulip poplar
Magnolia	*acuminata*	Cucumbertree

Magnolia	*fraseri*	Fraser magnolia
Magnolia	*macrophylla*	Bigleaf magnolia
Magnolia	*tripetala*	Umbrella magnolia
Magnolia	*virginiana*	Sweetbay
Morus	*alba*	White mulberry
Morus	*rubra*	Red mulberry
Nyssa	*biflora*	Swamp tupelo
Nyssa	*sylvatica*	Blackgum
Ostrya	*virginiana*	Eastern hophornbeam
Oxydendrum	*arboreum*	Sourwood
Paulownia	*tomentosa*	Princess tree
Pinus	*echinata*	Shortleaf pine
Pinus	*palustris*	Longleaf pine
Pinus	*pungens*	Table Mountain pine
Pinus	*rigida*	Pitch pine
Pinus	*strobus*	Eastern white pine
Pinus	*taeda*	Loblolly pine
Pinus	*virginiana*	Virginia pine
Platanus	*occidentalis*	American sycamore
Populus	*deltoides*	Eastern cottonwood
Prunus	*americana*	American plum
Prunus	*avium*	Sweet cherry
Prunus	*serotina*	Black cherry
Quercus	*alba*	White oak
Quercus	*coccinea*	Scarlet oak
Quercus	*falcata*	Southern red oak
Quercus	*laurifolia*	Laurel oak
Quercus	*lyrata*	Overcup oak
Quercus	*marilandica*	Blackjack oak
Quercus	*michauxii*	Swamp chestnut oak
Quercus	*montana*	Chestnut oak
Quercus	*nigra*	Water oak
Quercus	*pagoda*	Cherrybark oak
Quercus	*phellos*	Willow oak
Quercus	*rubra*	Northern red oak
Quercus	*shumardii*	Shumard oak
Quercus	*stellata*	Post oak
Quercus	*velutina*	Black oak
Robinia	*pseudoacacia*	Black locust
Salix	*nigra*	Black willow
Sassafras	*albidum*	Sassafras
Taxodium	*distichum*	Baldcypress
Tilia	*americana*	American basswood

Tsuga	*canadensis*	Eastern hemlock
Tsuga	*caroliniana*	Carolina hemlock
Ulmus	*alata*	Winged elm
Ulmus	*americana*	American elm
Ulmus	*pumila*	Siberian elm
Ulmus	*rubra*	Slippery elm

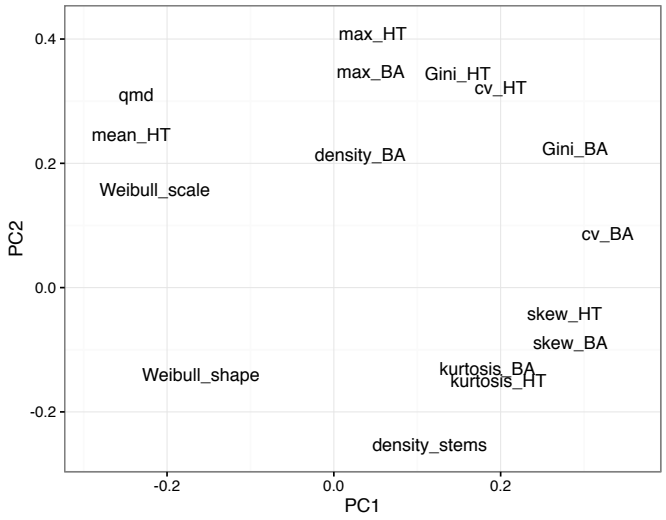
APPENDIX 2: SVR EMPIRICAL TUNING HYPERPARAMETERS*

Response	Block	ε	Cost
Species richness (SR)	Artificial	0.091	1
	Broadleaf	0.113	1
	Full	0.103	1
	Mixed	0.076	1
	Natural	0.099	0.5
	Pine	0.117	1
Shannon's *H*	Artificial	0.091	1
	Broadleaf	0.110	0.5
	Full	0.103	0.5
	Mixed	0.076	1
	Natural	0.109	1
	Pine	0.117	1
Simpson's *D*	Artificial	0.091	1
	Broadleaf	0.110	0.5
	Full	0.103	1
	Mixed	0.076	1
	Natural	0.109	1
	Pine	0.117	1
Rarified SR	Artificial	0.091	1
	Broadleaf	0.110	0.25
	Full	0.103	1
	Mixed	0.076	0.5
	Natural	0.109	1
	Pine	0.117	1
Pielou's *J*	Artificial	0.091	1
	Broadleaf	0.110	0.5
	Full	0.103	1
	Mixed	0.076	0.5
	Natural	0.102	0.5
	Pine	0.111	1

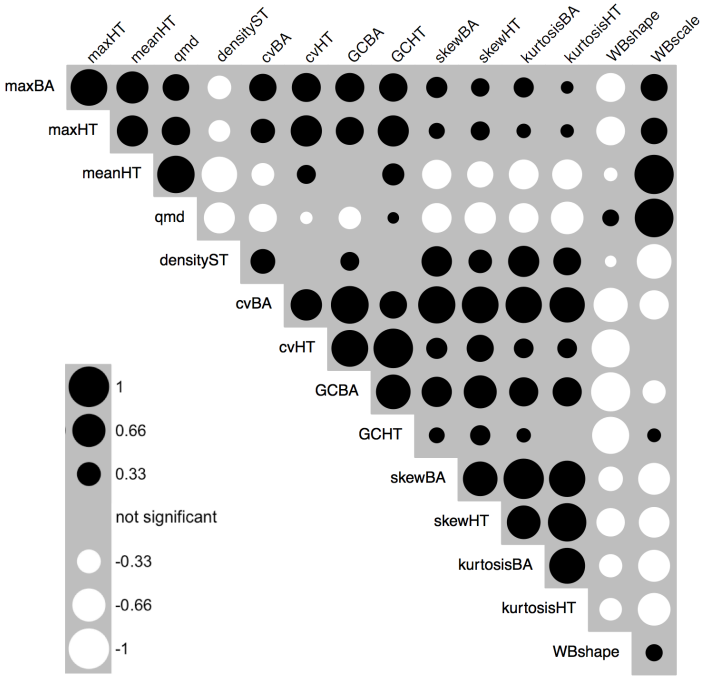
* All models run using a Gaussian radial kernel parameter

APPENDIX 3. RELATIONSHIPS AMONG STRUCTURAL VARIABLES

Relationship among structural variables represented as scatterplot of ordination scores between PCA axis 1 and 2, with proximity in the PCA 1-2 space indicating degree of similarity among predictors.

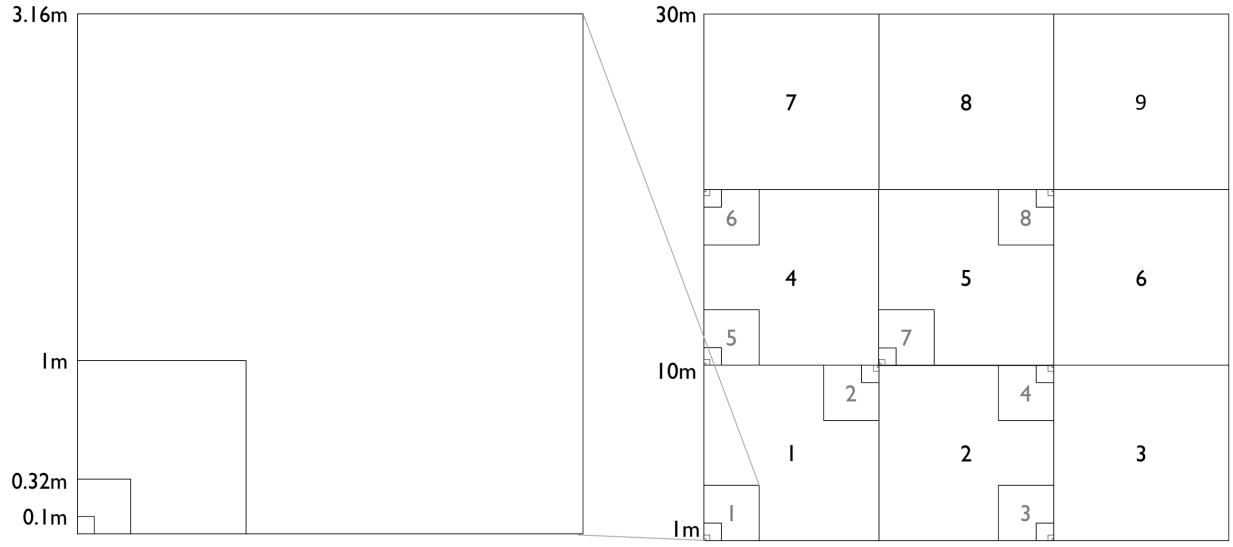


Raw Spearman correlation matrix among structural predictors. Circle size scaled to magnitude of correlation coefficient. Correlations significant at $p<0.05$, otherwise blank.



APPENDIX 4. SPATIALLY NESTED PLOT DESIGN

Nested sampling of 30x30m plots at 0.01m2, 0.1m2, 1m2, 10m2, 100m2, 400m2, and 900m2 scales.



APPENDIX 5. DUKE BLACKWOOD FIELD PLOT TAXA LIST

208 distinct vascular plant taxa were identified in field plots. Where species identity was undetermined, but its uniqueness relative to other species in the genus was confirmed, a generic species name is given (e.g. *Carex* sp. 3).

Acer floridanum	*Lonicera standishii*
Acer rubrum	*Lorinseria areolata*
Aesculus sylvatica	*Lycopus sp.*
Ailanthus altissima	*Lysimachia tonsa*
Amelanchier arborea	*Maianthemum racemosum*
Amphicarpaea bracteata	*Micranthes virginiensis*
Apocynum cannabinum	*Microstegium vimineum*
Arisaema triphyllum	*Mitchella repens*
Asclepias sp.	*Morus rubra*
Asplenium platyneuron	*Muscadinia rotundifolia*
Athyrium asplenioides	*Nabalus altissimus*
Boehmeria cylindrica	*Nabalus sp. 2*
Botrypus virginianus	*Nyssa sylvatica*
Campsis radicans	*Onoclea sensibilis*
Cardamine diphylla	*Ophioglossum pycnostichum*
Carex cephalophora	*Osmunda spectabilis*
Carex flaccosperma	*Osmundastrum cinnamomeum*
Carex sp. 3	*Ostrya virginiana*
Carex sp. 4	*Oxalis stricta*
Carex sp. 5	*Oxalis violacea*
Carex sp. 6	*Oxydendrum arboreum*
Carpinus caroliniana	*Parathelypteris noveboracensis*
Carya glabra	*Parthenocissus quinquefolia*
Carya ovata	*Passiflora lutea*
Carya pallida	*Paulownia tomentosa*
Carya tomentosa	*Phryma leptostachya*
Celtis occidentalis	*Physalis sp.*

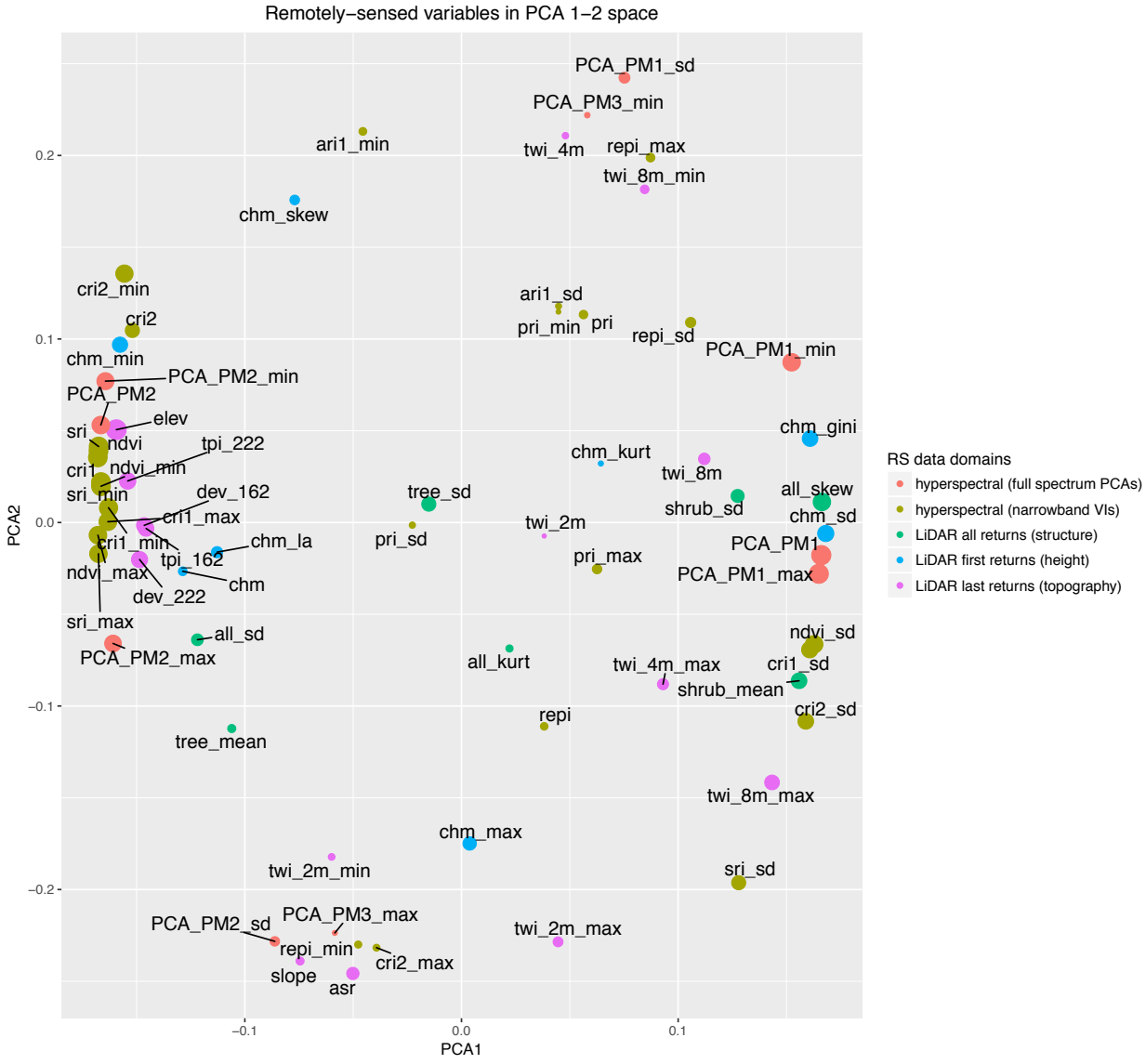
Cercis canadensis	*Pinus echinata*
Chelone glabra	*Pinus taeda*
Chimaphila maculata	*Pinus virginiana*
Chionanthus virginicus	*Poaceae sp. 1*
Chrysogonum virginianum	*Poaceae sp. 2*
Cirsium pumilum	*Poaceae sp. 3*
Clematis viorna	*Poaceae sp. 4*
Cornus florida	*Podophyllum peltatum*
Crataegus pruinosa	*Polygonatum biflorum*
Crataegus sp. 2	*Polystichum acrostichoides*
Cunila origanoides	*Populus deltoides*
Cynoglossum virginianum	*Potentilla canadensis*
Cypripedium acaule	*Prunus serotina*
Daucus carota	*Quercus alba*
Desmodium paniculatum	*Quercus coccinea*
Desmodium sp. 2	*Quercus falcata*
Dichanthelium boscii	*Quercus michauxii*
Dichanthelium commutatum	*Quercus montana*
Dichanthelium dichotomum	*Quercus phellos*
Dichanthelium laxiflorum	*Quercus rubra*
Dichanthelium polyanthes	*Quercus stellata*
Dichanthelium sp. 6	*Quercus velutina*
Dioscorea villosa	*Ranunculus recurvatus*
Diospyros virginiana	*Rhamnus caroliniana*
Diphasiastrum digitatum	*Rhododendron periclymenoides*
Elaeagnus umbellata	*Robinia pseudoacacia*
Elephantopus carolinianus	*Rosa multiflora*
Elephantopus sp. 3	*Rubus pensilvanicus*
Elephantopus sp. 4	*Rudbeckia laciniata*
Elephantopus tomentosus	*Ruellia caroliniensis*
Endodeca serpentaria	*Salvia lyrata*
Euonymus americanus	*Sanicula canadensis*

Eupatorium serotinum	*Sassafras albidum*
Euphorbia corollata	*Saururus cernuus*
Eutrochium sp.	*Sceptridium sp.*
Fagus grandifolia	*Scutellaria elliptica*
Fraxinus americana	*Scutellaria integrifolia*
Galium circaezans	*Scutellaria sp.*
Galium obtusum	*Silene virginica*
Galium pilosum	*Smilax bona-nox*
Galium tinctorium	*Smilax glauca*
Galium triflorum	*Smilax herbacea*
Geranium maculatum	*Smilax pulverulenta*
Geum canadense	*Smilax rotundifolia*
Gleditsia triacanthos	*Smilax sp. 6*
Goodyera pubescens	*Solidago rugosa*
Hedeoma pulegioides	*Solidago sp. 2*
Heuchera americana	*Stellaria pubera*
Hexastylis arifolia	*Styrax americanus*
Hibiscus moscheutos	*Styrax grandifolius*
Hieracium gronovii	*Symphyotrichum sp.*
Hieracium venosum	*Teucrium canadense*
Houstonia purpurea	*Thalictrum thalictroides*
Hylodesmum nudiflorum	*Thaspium barbinode*
Hypericum sp.	*Thelypteris sp.*
Hypopitys monotropa	*Thyrsanthella difformis*
Ilex ambigua	*Toxicodendron radicans*
Ilex decidua	*Trillium catesbaei*
Ilex opaca	*Trillium cuneatum*
Ilex sp. 6	*Ulmus alata*
Ilex verticillata	*Ulmus americana*
Ilex vomitoria	*Ulmus rubra*
Impatiens capensis	*Uvularia perfoliata*
Ipomoea sp.	*Uvularia sessilifolia*

Itea virginica	*Vaccinium pallidum*
Juglans nigra	*Vaccinium sp. 4*
Juniperus virginiana	*Vaccinium stamineum*
Lespedeza cuneata	*Vaccinium tenellum*
Lespedeza procumbens	*Viburnum prunifolium*
Ligustrum sinense	*Viburnum rafinesqueanum*
Lindera benzoin	*Viburnum rufidulum*
Liparis liliifolia	*Viburnum sp. 4*
Liquidambar styraciflua	*Viola affinis*
Liriodendron tulipifera	*Viola palmata*
Lobelia sp.	*Viola sp. 3*
Lonicera japonica	*Vitis sp.*
Lonicera sempervirens	*Zizia trifoliata*

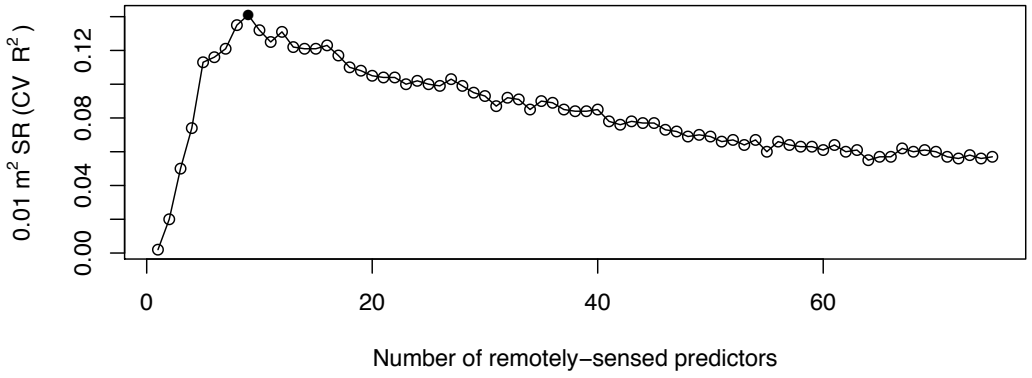
APPENDIX 6. REMOTELY-SENSED PREDICTORS IN PCA SPACE

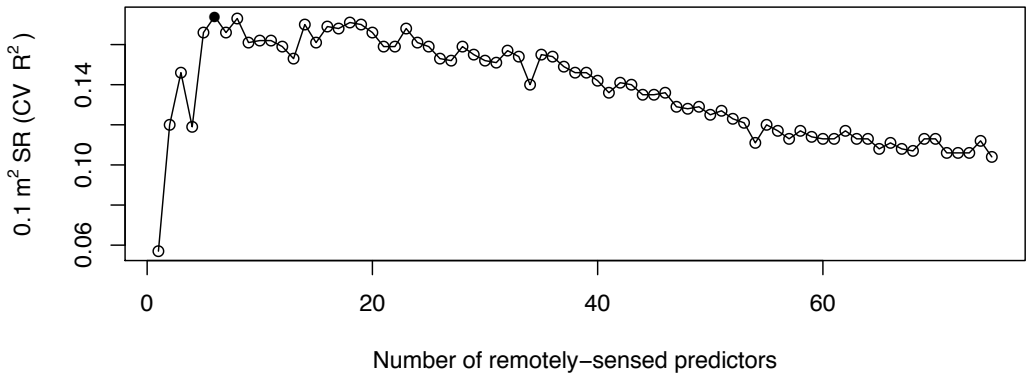
Relationship among remotely-sensed variables represented as scatterplot of ordination scores between PCA axis 1 and 2, with proximity in the PCA 1-2 space indicating degree of similarity among predictors. Size of circle refers to its relative degree of correlation with plant species richness across all seven scales. Color refers to RS data domains: LiDAR topography, LiDAR structure, hyperspectral full spectrum, and hyperspectral narrowband VIs. See Table 1 for description of abbreviations.

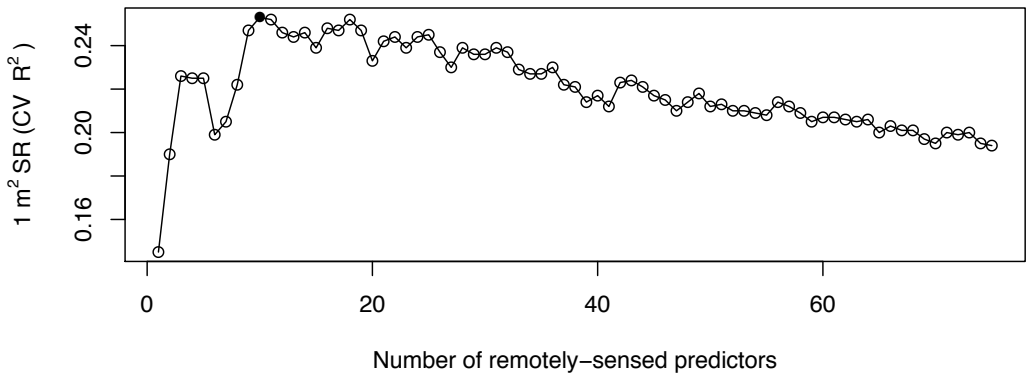


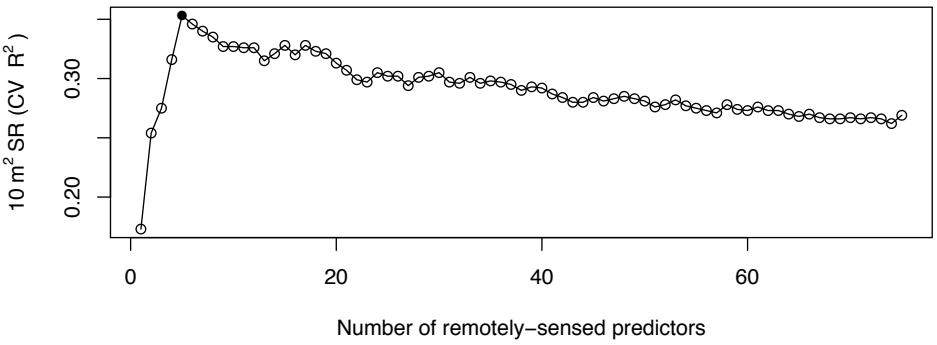
APPENDIX 7. SPECIES RICHNESS - RANDOM FOREST FEATURE SELECTION

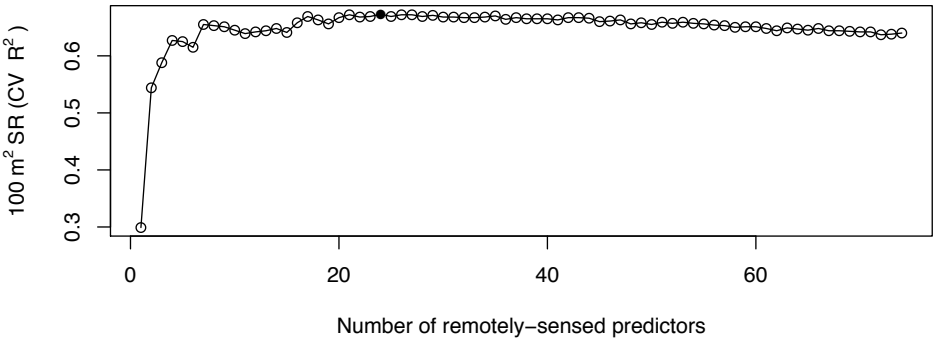
Charts reflect the effect of feature selection on model fit and generalizability by means of 10-fold cross-validation. The features (combination of predictors) that maximize cross-validated accuracy were selected for use in final RF models (see Methods 3.4.2.)

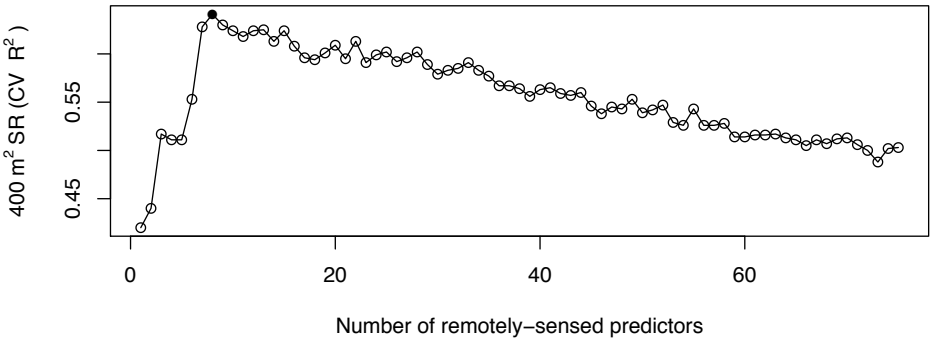


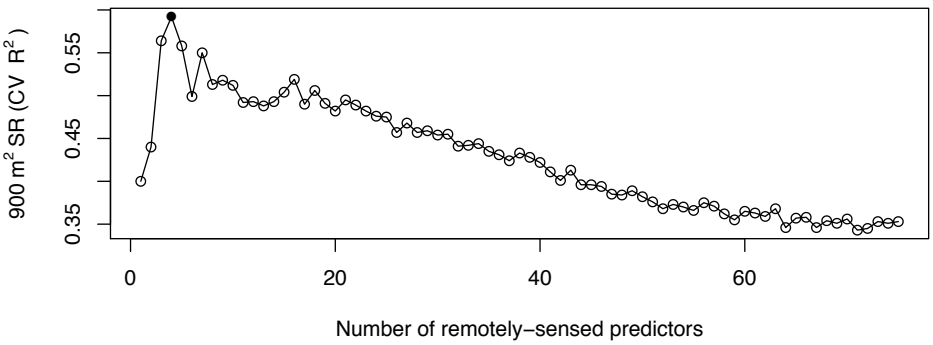





10 m^2 SR (CV R^2)
0.20
0.30
0
20
40
60
Number of remotely−sensed predictors


100 m^2 SR (CV R^2)
0.3
0.4
0.5
0.6
0
20
40
60
Number of remotely−sensed predictors


400 m^2 SR (CV R^2)
0.45
0.55
0
20
40
60
Number of remotely−sensed predictors


900 m^2 SR (CV R^2)
0.35
0.45
0.55
0
20
40
60
Number of remotely−sensed predictors

APPENDIX 8. SPEARMAN CORRELATION MATRIX FOR REMOTELY-SENSED VARIABLES VERSUS SPECIES RICHNESS ACROSS SPATIAL SCALES.

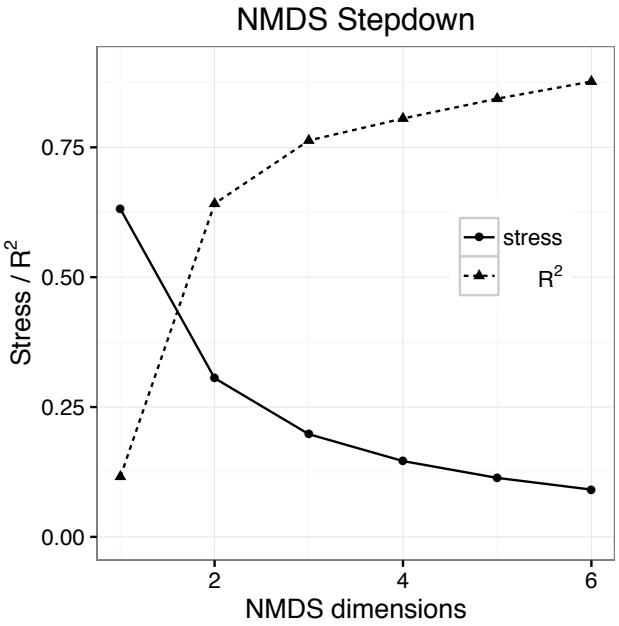
scale (m^2)	ASR (mean)	DEV (222m mean)	elev (mean)	TPI (222m mean)	TWI (8m max)	all returns (sd)	all returns (skew)	CHM (mean)	CHM (max)
0.01	0	0	-0.208	0	0	0.247	0	0.128	0.176
0.1	0	0	-0.244	0	0	0.23	0	0.161	0.197
1	-0.194	-0.178	-0.314	-0.175	0	0.196	0.255	0	0.146
10	-0.26	-0.329	-0.382	-0.312	0.131	0.155	0.349	0	0
100	-0.247	-0.492	-0.574	-0.493	0.286	0	0.586	0	0
400	0	-0.527	-0.632	-0.537	0.456	0	0.602	0	0
900	-0.401	-0.567	-0.608	-0.596	0.444	0	0.58	0	0

scale (m^2)	CHM (min)	CHM (sd)	shrub returns (mean)	shrub returns (sd)	tree returns (mean)	tree returns (sd)	PCA1 (mean)	PCA1 (sd)
0.01	0	0	0	0	0.146	0.216	0	0
0.1	0.128	0	0	0	0.14	0.202	0.127	0
1	0	0.144	0	0	0	0.228	0.297	0.135
10	0	0.195	0.149	0.134	0	0.196	0.372	0.169
100	-0.328	0.449	0.403	0.354	-0.225	0.231	0.612	0.239
400	-0.524	0.533	0.46	0.403	0	0	0.682	0
900	-0.487	0.515	0.34	0	0	0	0.719	0

scale (m^2)	PCA2 (mean)	PCA2 (sd)	CRI1 (mean)	CRI1 (sd)	NDVI (mean)	NDVI (sd)	SRI (mean)	SRI (sd)
0.01	0	0	0	-0.191	0	0	0	0
0.1	0	0	0	0	0	0	0	0
1	-0.175	0	-0.23	0	-0.257	0.22	-0.252	0.123
10	-0.298	-0.12	-0.311	0.156	-0.34	0.249	-0.338	0
100	-0.548	-0.35	-0.616	0.425	-0.652	0.506	-0.643	0.24
400	-0.618	0	-0.685	0.562	-0.739	0.653	-0.711	0.435
900	-0.729	-0.399	-0.635	0.551	-0.697	0.578	-0.731	0.416

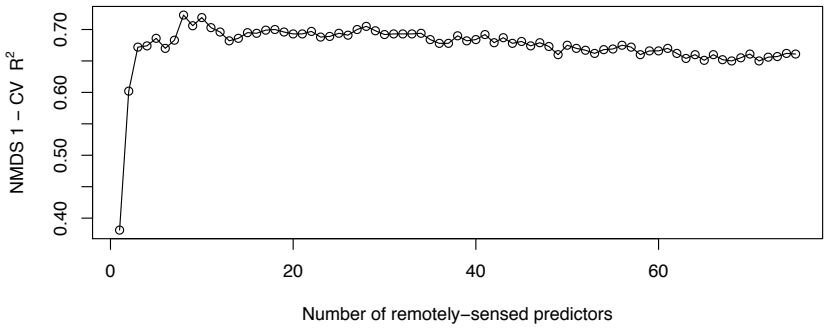
APPENDIX 9. NMDS STEP-DOWN PLOTS

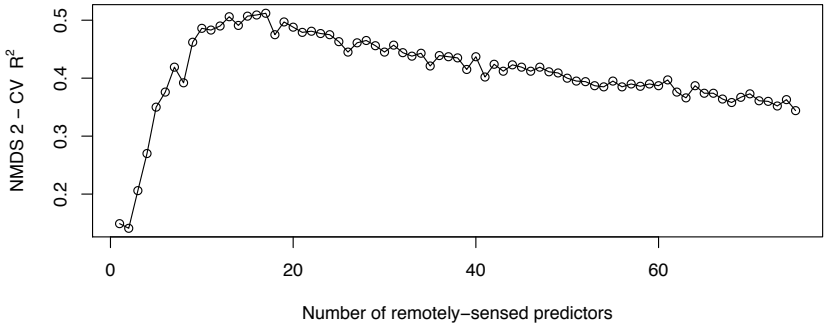
A-priori dimensionality of the NMDS ordination was evaluated using a stress metric (and corresponding R^2), which characterizes the agreement between the rank-order of distances in the original data compared to their respective scores in ordination space.

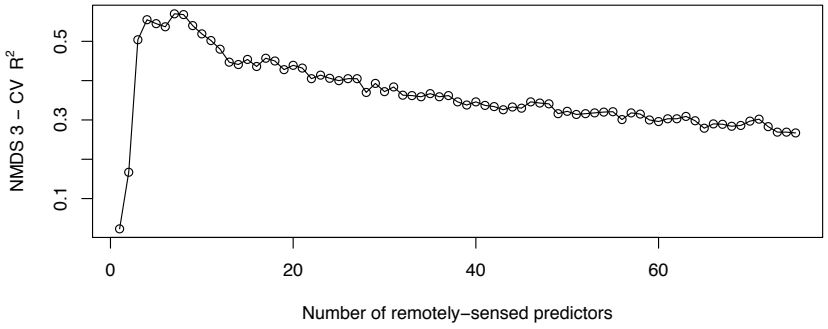


APPENDIX 10. NMDS RANDOM FOREST FEATURE SELECTION.

Charts reflect the effect of feature selection on model fit and generalizability by means of 10-fold cross-validation. The features (combination of predictors) that maximize cross-validated accuracy were selected for use in final RF models (see Methods 2.4.3.)

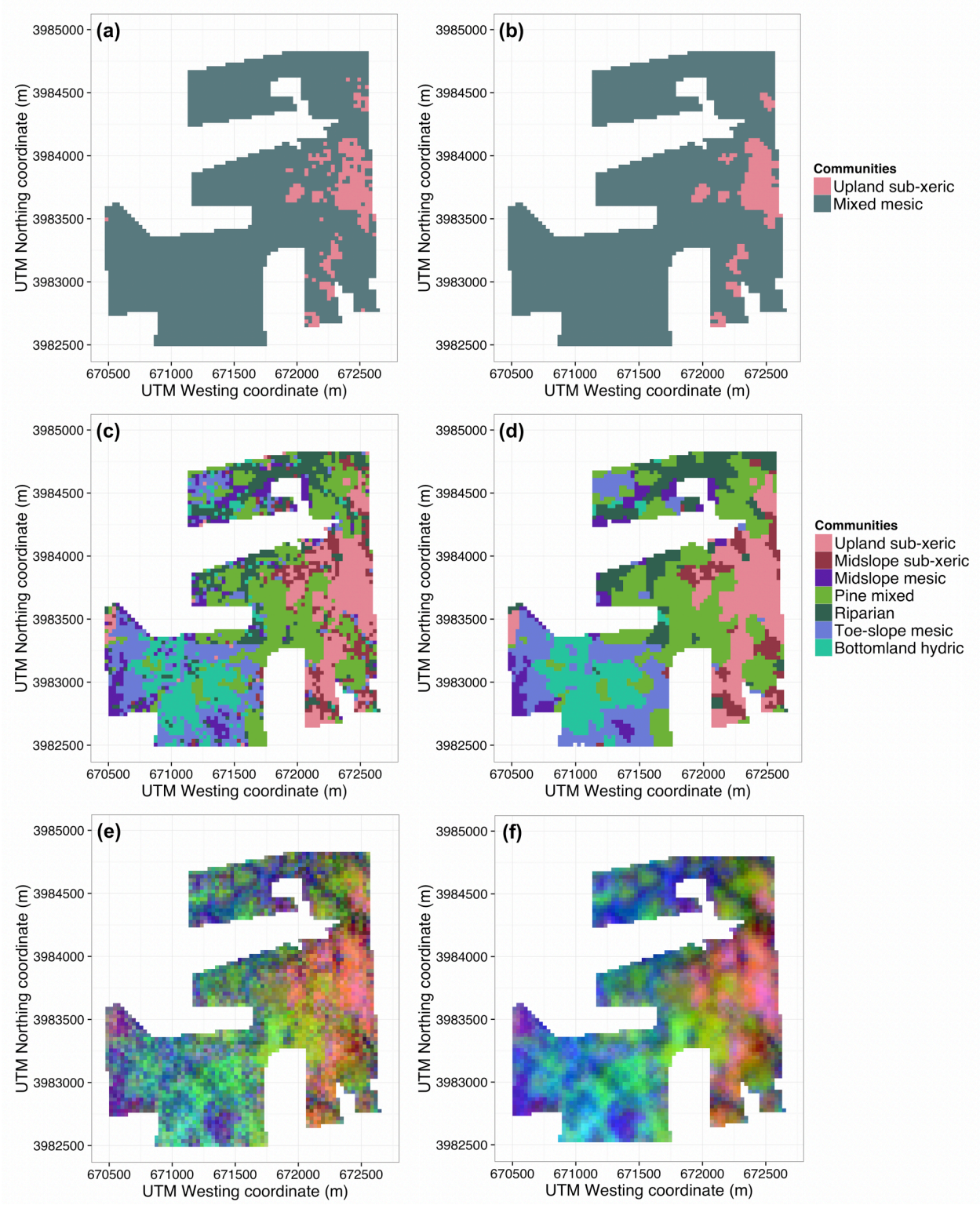






APPENDIX 11. NMDS MAP POST-PROCESSING.

For ease of interpretation and to minimize salt-and-pepper effects, final predictive maps of vascular plant composition are processed using a 3x3 pixel modal filter. Maps depict (a) pre-processed two community-unit classification, (b) post-processed two community-unit classification, (c) pre-processed seven community-unit classification, (d) post-processed seven community-unit classification, (e) pre-processed compositional gradient predictions, and (f) post-processed compositional gradient predictions. Unlike majority filters which are highly sensitive to the values of all adjacent cells, modal filters only smooth the values of singletons pixels, defined as focal pixels unique in a 3x3 pixel window. Filtered maps aid in visual interpretation but are not appropriate for modeling applications.


(a)
(b)
(c)
(d)
(e)
(f)
UTM Northing coordinate (m)
UTM Westing coordinate (m)
3985000
3984500
3984000
3983500
3983000
3982500
670500
671000
671500
672000
672500
Communities
Upland sub-xeric
Mixed mesic
Communities
Upland sub-xeric
Midslope sub-xeric
Midslope mesic
Pine mixed
Riparian
Toe-slope mesic
Bottomland hydric

APPENDIX 12. SPEARMAN CORRELATION MATRIX AMONG ALL REMOTELY-SENSED PREDICTORS

Matrix trace and non-significant correlations (p>0.01) left blank.

		LiDAR topography					LiDAR structure						Hyperspectral (full spectrum)									Hyperspectral (narrow-bands Vis)					
		ASR (mean)	elev (mean)	slope (mean)	TWI (4m mean)	TWI (8m max)	TWI (8m min)	all returns (skew)	CHM (max)	CHM (sd)	tree returns (sd)	PCA 1 (mean)	PCA 1 (max)	PCA 1 (min)	PCA 1 (sd)	PCA 2 (min)	PCA 3 (mean)	PCA 3 (max)	PCA 3 (min)	CRI 1 (max)	CRI 1 (sd)	CRI 2 (mean)	NDVI (min)	PRI (mean)	PRI (sd)	REPI (mean)	SRI (mean)	SRI (min)
LiDAR topography	ASR (mean)							-0.43				-0.75	-0.59	-0.50		0.36	0.53	0.42										
	elev (mean)			0.49	-0.38	-0.50	-0.66							-0.46					-0.34	0.45	-0.37	0.41	0.55	-0.37			0.47	0.56
	slope (mean)		0.49		-0.59	-0.35	-0.91							-0.34						0.46		0.46	0.42			-0.36		0.42
	TWI (4m mean)		-0.38	-0.59		0.61	0.63							0.34						-0.53		-0.51						
	TWI (8m max)		-0.50	-0.35	0.61		0.46					0.48	0.38	0.43						-0.39		-0.47				-0.42		
	TWI (8m min)		-0.66	-0.91	0.63	0.46								0.44						-0.59		-0.57	-0.49					-0.49
LiDAR structure	all returns (skew)	-0.43								0.88	0.34	0.65	0.68	0.35		-0.75					0.71		-0.58				-0.80	-0.59
	CHM (max)										0.89											-0.41						
	CHM (sd)							0.88				0.57	0.67		0.43	-0.71					0.64		-0.42				-0.62	-0.42
	tree returns (sd)							0.34	0.89										-0.36			-0.44					-0.36	
Hyperspectral (full spectrum)	PCA 1 (mean)	-0.75			0.48		0.65		0.57				0.82		0.52	0.69	0.52	-0.54	-0.38	-0.44		-0.40	0.45	-0.45	-0.54			-0.70	-0.54
	PCA 1 (max)	-0.59			0.38		0.68		0.67			0.82		0.36	0.52	-0.68					0.47		-0.47				-0.60	-0.47
	PCA 1 (min)	-0.50	-0.46	-0.34	0.43	0.44	0.35					0.69	0.36							-0.53	-0.65	-0.56		-0.62	-0.58	0.54	-0.37	-0.56	-0.58
	PCA 1 (sd)									0.43		0.52	0.52															
	PCA 2 (min)	0.36						-0.75		-0.71		-0.54	-0.68									-0.66	0.54				0.66	0.54
	PCA 3 (mean)	0.53										-0.38		-0.53					0.68	0.62				-0.60	0.39	0.37		
	PCA 3 (max)	0.42									-0.36	-0.44		-0.65				0.68		0.38		0.50	-0.37					
	PCA 3 (min)		-0.34															0.62		-0.36				-0.38	0.45			
Hyperspectral (narrow-bands Vis)	CRI 1 (max)	0.45	0.46	-0.53	-0.39	-0.59						-0.40		-0.56					0.38	-0.36			0.82	0.49			0.60	0.50
	CRI 1 (sd)	-0.37				0.71				0.64		0.45	0.47					-0.66					-0.45	-0.73			-0.75	-0.73
	CRI 2 (mean)	0.41	0.46	-0.51	-0.47	-0.57		-0.41			-0.44	-0.45		-0.62					0.50		0.82	-0.45		0.60			0.63	0.61
	NDVI (min)	0.55	0.42		-0.49	-0.58		-0.42				-0.54	-0.47	-0.58		0.54					0.49	-0.73	0.60				0.70	1.00
	PRI (mean)	-0.37												0.54					-0.60	-0.37	-0.38				-0.77			
	PRI (sd)													-0.37					0.39		0.45			-0.77				
	REPI (mean)		-0.36																0.37									
	SRI (mean)	0.43	0.47		-0.42		-0.80			-0.62	-0.36	-0.70	-0.60	-0.56		0.66						0.60	-0.75	0.63	0.70			0.71
	SRI (min)		0.56	0.42		-0.49	-0.59			-0.42		-0.54	-0.47	-0.58		0.54						0.50	-0.73	0.61	1.00		0.71	

REFERENCES

Aarssen, L.W. 2004. Interpreting co-variation in species richness and productivity in terrestrial vegetation: Making sense of causations and correlations at multiple scales. *Folia Geobotanica* 39: 385–403.

Abelleira Martinez, O.J., Fremier, A.K., Gunter, S., Ramos Bendana, Z., Vierling, L., Galbraith, S.M., Bosque-Perez, N.A., & Ordonez, J.C. 2016. Scaling up functional traits for ecosystem services with remote sensing: concepts and methods. *Ecology and Evolution* 6: 4359–4371.

Acker, S., Sabin, T., Ganio, L., & McKee, W. 1998. Development of old-growth structure and timber volume growth trends in maturing Douglas-fir stands. *Forest Ecology and Management* 104: 265–280.

Ackerly, D.D., Loarie, S.R., Cornwell, W.K., Weiss, S.B., Hamilton, H., Branciforte, R., & Kraft, N.J.B. 2010. The geography of climate change: implications for conservation biogeography. *Diversity and Distributions* 16: 476–487.

Aho, K., Roberts, D., & Weaver, T. 2008. Using geometric and non-geometric internal evaluators to compare eight vegetation classification methods. *Journal of Vegetation Science* 19: 549–562.

Aiba, S., & Kitayama, K. 1999. Structure, composition and species diversity in an altitude-substrate matrix of rain forest tree communities on Mount Kinabalu, Borneo. *Plant Ecology* 140: 139–157.

Albert, C.H., Thuiller, W., Yoccoz, N.G., Soudant, A., Boucher, F., Saccone, P., & Lavorel, S. 2010. Intraspecific functional variability: Extent, structure and sources of variation. *Journal of Ecology* 98: 604–613.

Anderson, M.G., & Ferree, C.E. 2010. Conserving the stage: climate change and the geophysical underpinnings of species diversity. *PLOS ONE* 5: e11554.

Anderson, J.E., Plourde, L.C., Martin, M.E., Braswell, B.H., Smith, M.L., Dubayah, R.O., Hofton, M.A., & Blair, J.B. 2008. Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment* 112: 1856–1870.

Arlot, S., & Celisse, A. 2010. A survey of cross-validation procedures for model selection *. *Statistics Surveys* 4: 40–79.

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

Asner, G.P., Jones, M.O., Martin, R.E., Knapp, D.E., & Hughes, R.F. 2008. Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment* 112: 1912–1926.

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.

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

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.

Austin, M.P. 2013. Vegetation and environment: discontinuities and continuities. In Franklin, E. van der M. and J. (ed.), *Vegetation Ecology*, pp. 71–106. John Wiley and Sons, Ltd., Chichester, West Sussex, UK.

Bacaro, G., Rocchini, D., Bonini, I., Marignani, M., Maccherini, S., & Chiarucci, A. 2008. The role of regional and local scale predictors for plant species richness in Mediterranean forests. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 142: 630–642.

Bagnaresi, U. 2002. Stand structure and biodiversity in mixed, uneven-aged coniferous forests in the eastern Alps. *Forestry* 75: 357–364.

Bailey, R., & Dell, T. 1973. Quantifying diameter distributions with the Weibull function. *Forest Science* 19: 97–104.

Bechtold, W.A., & Patterson, P.L. 2005. *The Enhanced Forest Inventory and Analysis Program — National Sampling Design and Estimation Procedures. USDA Forest Service, General Technical Report SRS-80.*

Beier, P., & de Albuquerque, F.S. 2015. Environmental diversity as a surrogate for species representation. *Conservation Biology* 29: 1401–1410.

Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., & Radeloff, V.C. 2009. Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences* 114: 1–13.

Berger, A., & Puettmann, K. 2000. Overstory Composition and Stand Structure Influence Herbaceous Plant Diversity in the Mixed Aspen Forest of Northern Minnesota. *The American Midland Naturalist* 143: 111–125.

Beven, K.J., & Kirkby, M.J. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal* 24: 43–69.

Breiman, L. 2001. Random forests. *Machine Learning* 45: 5–32.

Buddenbaum, H., Seeling, S., & Hill, J. 2013. Fusion of full-waveform lidar and imaging

spectroscopy remote sensing data for the characterization of forest stands. *International Journal of Remote Sensing* 34: 4511–4524.

Bunge, J., & Fitzpatrick, M. 1993. Estimating the Number of Species: A Review. *Journal of the American Statistical Association* 88: 364–373.

Bunting, P., Lucas, R.M., Jones, K., & Bean, A.R. 2010. Characterisation and mapping of forest communities by clustering individual tree crowns. *Remote Sensing of Environment* 114: 2536–2547.

Busing, R.T., & White, P.S. 1993. Effects of area on old-growth forest attributes: implications for the equilibrium landscape concept. *Landscape Ecology* 8: 119–126.

Camathias, L., Bergamini, A., Kuchler, M., Stofer, S., & Baltensweiler, A. 2013. High-resolution remote sensing data improves models of species richness. *Applied Vegetation Science* 16: 539–551.

Canham, C.D., Finzi, A.C., Pacala, S.W., & Burbank, D.H. 1994. Causes and consequences of resource heterogeneity in forests: interspecific variation in light transmission by canopy trees. *Canadian Journal of Forest Research* 24: 337–349.

Castro-Esau, K.L., Sanchez-Azofeifa, G.A., & Rivard, B. 2006. Comparison of spectral indices obtained using multiple spectroradiometers. *Remote Sensing of Environment* 103: 276–288.

Cavender-Bares, J., Meireles, J.E., Couture, J.J., Kaproth, M.A., Kingdon, C.C., Singh, A., Serbin, S.P., Center, A., Zuniga, E., Pilz, G., & Townsend, P.A. 2016. Associations of leaf spectra with genetic and phylogenetic variation in oaks: Prospects for remote detection of biodiversity. *Remote Sensing* 8: 1–17.

Cayuela, L., Benayas, J.M.R., Justel, A., & Salas-Rey, J. 2006. Modelling tree diversity in a highly fragmented tropical montane landscape. *Global Ecology and Biogeography* 15: 602–613.

Chao, A., & Jost, L. 2012. Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology* 93: 2533–2547.

Chen, J., Franklin, J.F., & Spies, T.A. 1995. Growing-Season Microclimatic Gradients from Clearcut Edges into Old-Growth Douglas-Fir Forests. *Ecological Applications* 5: 74.

Chen, C., Liaw, A., & Breiman, L. 2004. Using random forest to learn imbalanced data. *Journal of Machine Learning Research* 1–12.

Chiarucci, A., & Bonini, I. 2005. Quantitative floristics as a tool for the assessment of plant diversity in Tuscan forests. *Forest Ecology and Management* 212: 160–170.

Chytrý, M., Tichý, L., Holt, J., & Botta-Dukát, Z. 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* 13: 79–90.

Civitello, D.J., Cohen, J., Fatima, H., Halstead, N.T., Liriano, J., Mcmahon, T.A., Ortega, C.N.,

Sauer, E.L., Sehgal, T., Young, S., & Rohr, J.R. 2015. Biodiversity inhibits parasites: Broad evidence for the dilution effect. *Pnas* 112: 8667–8671.

Clark, J.S. 2003. Uncertainty and variability in demography and population growth: A hierarchical approach. *Ecology* 84: 1370–1381.

Clark, J.S., Gelfand, A.E., Woodall, C.W., & Zhu, K. 2014. More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications* 24: 990–999.

Clark, M.L., Roberts, D.A., & Clark, D.B. 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment* 96: 375–398.

Clarke, K.R. 1993. Non-parametric Multivariate Analyses of Changes in Community Structure. *Austral Ecology* 18: 117–143.

Clebsch, E., & Busing, R. 1989. Secondary Succession , Gap Dynamics , and Community Structure in a Southern Appalachian Cove Forest. *Ecology* 70: 728–735.

Cohen, J. 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological bulletin* 70: 213–220.

Condit, R. 1995. Research in large, long-term tropical forest plots. *Trends in Ecology & Evolution* 10: 18–22.

Cook, J.E. 2015. Structural effects on understory attributes in second-growth forests of northern Wisconsin, USA. *Forest Ecology and Management* 347: 188–199.

Cook, B.D., Corp, L.A., Nelson, R.F., Middleton, E.M., Morton, D.C., McCorkel, J.T., Masek, J.G., Ranson, K.J., Ly, V., & Montesano, P.M. 2013. NASA goddard's LiDAR, hyperspectral and thermal (G-LiHT) airborne imager. *Remote Sensing* 5: 4045–4066.

Cord, A.F., Klein, D., Mora, F., & Dech, S. 2014. Comparing the suitability of classified land cover data and remote sensing variables for modeling distribution patterns of plants. *Ecological Modelling* 272: 129–140.

Cord, A.F., Meentemeyer, R.K., Leitão, P.J., & Václavík, T. 2013. Modelling species distributions with remote sensing data: Bridging disciplinary perspectives. *Journal of Biogeography* 40: 2226–2227.

Cortes, C., & Vapnik, V. 1995. Support-vector networks. *Machine Learning* 20: 273–297.

Cristianini, N., & Shawe-Taylor, J. 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, Cambridge, UK.

Curran, P.J. 1989. Remote sensing of foliar chemistry. *Remote Sensing of Environment* 30: 271–278.

Cushman, S., Evans, J., McGarigal, K., & Kiesecker, J. 2010. *Toward Gleasonian landscape ecology: from communities to species, from patches to pixels*. United States Department of Agriculture / Forest Service Rocky Mountain Research Station, Fort Collins, CO.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. 2007. Random forests for classification in ecology. *Ecology* 88: 2783–92.

Dalponte, M., Bruzzone, L., & Gianelle, D. 2008. Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing* 46: 1416–1427.

Damgaard, C., & Weiner, J. 2000. Describing size inequality in plant size or fecundity. *Ecology* 81: 1139–1142.

Daubenmire, R. 1966. Vegetation: identification of typal communities. *Science* 151: 291–298.

Davis, F., & Roberts, D. 2000. Stand structure in terrestrial ecosystems. In *Methods in ecosystem science*, pp. 7–30. Springer Science & Business Media.

Davis, M., & Shaw, R. 2001. Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science* 292: 673–679.

De'ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88: 243–251.

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P., Van Meirvenne, M., Verniers, J., & Crombé, P. 2013. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* 186: 39–49.

Diamond, J. 1988. Factors controlling species diversity: overview and synthesis. *Annals of the Missouri Botanical Garden* 75: 117–129.

Diaz, S., & Cabido, M. 2001. Vive la différence: plant functional diversity matters to ecosystem processes. *Trends in Ecology & Evolution* 5347: 646–655.

Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I.C., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P.B., Moles, A.T., Dickie, J., Gillison, A.N., Zanne, A.E., Chave, J., Wright, S.J., Sheremet'ev, S.N., Jactel, H., Christopher, B., Cerabolini, B., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J.S., Günther, A., Falczuk, V., Rüger, N., Mahecha, M.D., & Gorné, L.D. 2015. The global spectrum of plant form and function. *Nature* 529: 1–17.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., & Lautenbach, S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27–46.

Dubayah, R., Goetz, S.J., Blair, J.B., Hansen, T.E.F.M., Healey, S.P., Hofton, M.A., Hurtt, G.C., Kellner, J., Luthcke, S.B., & Swatantran, A. 2014. The Global Ecosystem Dynamics Investigation. In *American Geophysical Union: Fall Meeting*, p. U14A–07.

Elith, J., & Leathwick, J. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, …* 40: 677–697.

Elmqvist, T., Folke, C., Nyström, M., Peterson, G., Bengtsson, J., Walker, B., & Norberg, J. 2003. Response diversity, ecosystem change, and resilience. *Frontiers in Ecology and the Environment* 1: 488–494.

ESRI. 2016. ArcGIS Desktop: Release 10.4.

Ess, C., & Sudweeks, F. 2001. *Culture, technology, communication : towards an intercultural global village*. Suny Press.

Evans, J.S., Murphy, M.A., Holden, Z.A., & Cushman, S.A. 2011. Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology*, pp. 139–159.

FAO. 2010. Global Forest Resources Assessment 2010. Main Report. *FAO Forestry paper* 378.

Feilhauer, H., & Schmidtlein, S. 2009. Mapping continuous fields of forest alpha and beta diversity. *Applied Vegetation Science* 12: 429–439.

Ferrier, S., & Guisan, A. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* 43: 393–404.

Ferris, R., Peace, A., Humphrey, J., & Broome, A. 2000. Relationships between vegetation, site type and stand structure in coniferous plantations in Britain. *Forest Ecology and Management* 136: 35–51.

Folke, C., Carpenter, S., & Walker, B. 2004. Regime Shifts, Resilience, and Biodiversity in Ecosystem Management. *Annual review of ecology, evolution, and systematics* 35: 557–581.

Foster, J., & Townsend, P. 2002. Mapping Forest Composition in the Central Appalachians Using AVIRIS: Effects of Topography and Phenology. In *Proc. 11th JPL Airborne Earth Science Workshop*,

Franklin, J. 1988. Structural and functional diversity in temperate forests. In *Biodiversity*, pp. 166–175. National Academy Press, Washington DC.

Franklin, J.F., Spies, T. a, Pelt, R.V., Carey, A.B., Thornburgh, D. a, Berg, D.R., Lindenmayer, D.B., Harmon, M.E., Keeton, W.S., Shaw, D.C., Bible, K., & Chen, J. 2002. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forests as an example. *Forest Ecology and Management* 155: 399–423.

Fricker, G.A., Wolf, J. a., Saatchi, S.S., & Gillespie, T.W. 2015. Predicting spatial variations of

tree species richness in tropical forests from high resolution remote sensing. *Ecological Applications* 25: 1776–1789.

Gadow, K., Zhang, C.Y., Wehenkel, C., Pommerening, A., Corral-Rivas, J., Korol, M., Myklush, S., Hui, G.Y., Kiviste, A., & Zhao, X.H. 2012. Forest Structure and Diversity. In Pukkala, T. & von Gadow, K. (eds.), *Continuous Cover Forestry*, pp. 29–84. Springer Netherlands, Dordrecht.

Gamon, J. 2008. Tropical remote sensing – opportunities and challenges. In Kalacska, M. & Sanchez-Azofeifa, G. (eds.), *Hyperspectral remote sensing of tropical and subtropical forests*, pp. 297–304. CRC Press Taylor & Francis Group, Boca Raton, FL.

Gamon, J.A., Field, C.B., Goulden, M.L., Griffin, K.L., Hartley, A.E., Joel, G., Penuelas, J., & Valentini, R. 1995. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications* 5: 28–41.

Garbarino, M., Borgogno Mondino, E., Lingua, E., Nagel, T. a., Dukić, V., Govedar, Z., & Motta, R. 2012. Gap disturbances and regeneration patterns in a Bosnian old-growth forest: a multispectral remote sensing and ground-based approach. *Annals of Forest Science* 69: 617–625.

Gaston, K.J. 2000. Global patterns in biodiversity. *Nature* 405: 220–7.

Gaston, K.J. 1992. Regional Numbers of Insect and Plant-Species. *Functional Ecology* 6: 243–247.

Ghiyamat, A., & Shafri, H.Z.M. 2008. A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing* 31: 1837–1856.

Gillespie, T., Foody, G., Rocchini, D., Giorgi, A., & Saatchi, S. 2008. Measuring and modelling biodiversity from space. *Progress in Physical Geography* 32: 203–221.

Gitelson, A. a, Zur, Y., Chivkunova, O.B., & Merzlyak, M.N. 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and photobiology* 75: 272–281.

Goodburn, J.M., & Lorimer, C.G. 1999. Population structure in old-growth and managed northern hardwoods: an examination of the balanced diameter distribution concept. *Forest Ecology and Management* 118: 11–29.

Gopal, S., & Woodcock, C. 1994. Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy-Sets. *Photogrammetric Engineering and Remote Sensing* 60: 181–188.

Goslee, S.C., & Urban, D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22: 1–19.

Gotelli, N., & Colwell, R. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters* 4: 379–391.

Gould, W. 2000. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological Applications* 10: 1861–1870.

Gray, A.N., Brandeis, T.J., Shaw, J.D., McWilliams, W.H., & Miles, P.D. 2012. Forest Inventory and Analysis Database of the United States of America (FIA). *Biodiversity and Ecology* 4: 225–231.

Griffith, G.E., Omernik, J., Comstock, J.A., Schafale, M.P., McNab, W.H., Lenat, D.R., MacPherson, T.F., Glover, J.B., & Shelburne, V.B. 2002. *Ecoregions of North Carolina and South Carolina,(color poster with map, descriptive text, summary tables, and photographs): Reston, Virginia. US Geological Survey (map scale 1: 1,500,000)*.

Gu, H., Singh, A., & Townsend, P.A. 2015. Detection of gradients of forest composition in an urban area using imaging spectroscopy. *Remote Sensing of Environment* 167: 168–180.

Guisan, A., & Rahbek, C. 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography* 38: 1433–1444.

Guisan, A., & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., & Strachan, I.B. 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* 90: 337–352.

Hakkenberg, C.R., Song, C., Peet, R.K., & White, P.S. 2016. Forest structure as a predictor of tree species diversity in the North Carolina Piedmont. *Journal of Vegetation Science* 27: 1151–1163.

Hakkenberg, C.R., Tarasi, D.D., & Peet, R.K. 2017. Community / continuum in biogeography. *International Encyclopedia of Geography: People, the Earth, Environment, and Technology*.

Halpern, C., & Spies, T. 1995. Plant Species Diversity in Natural and Managaed Forests of the Pacific Northwest. *Ecological Applications* 5: 913–934.

Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Berlin/Heidelberg, Germany.

He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., Turner, W., Wegmann, M., & Pettorelli, N. 2015. Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation* 1: 4–18.

Heip, C.H.., Herman, P.M.J., & Soetaert, K. 1998. Indices of Diversity and Evenness. *Oceanis* 24: 61–87.

Higgins, M.A., Asner, G.P., Martin, R.E., Knapp, D.E., Anderson, C., Kennedy-Bowdoin, T.,

Saenz, R., Aguilar, A., & Joseph Wright, S. 2014. Linking imaging spectroscopy and LiDAR with floristic composition and forest structure in Panama. *Remote Sensing of Environment* 154: 358–367.

Hill, M. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427–432.

Hill, R., & Thomson, A. 2005. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing* 26: 3763–3779.

Hooper, D.U., Adair, E.C., Cardinale, B.J., Byrnes, J.E.K., Hungate, B. a, Matulich, K.L., Gonzalez, A., Duffy, J.E., Gamfeldt, L., & O'Connor, M.I. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486: 105–8.

Hooper, D., Solan, M., & Symstad, A. 2002. Species diversity, functional diversity and ecosystem functioning. In Loreau, M. (ed.), *Biodiversity and Ecosystem Functioning*, pp. 195–281.

Houlahan, J.E., Mckinney, S.T., Anderson, T.M., & Mcgill, B.J. 2016. The priority of prediction in ecological understanding. *Oikos* 1–7.

Hsieh, T.C., Ma, K.H., & Chao, A. 2016. iNEXT: iNterpolation and EXTrapolation for species diversity.

Hulshof, C.M., & Swenson, N.G. 2010. Variation in leaf functional trait values within and across individuals and species: an example from a Costa Rican dry forest. *Functional Ecology* 24: 217–223.

Hyde, P., Dubayah, R., Walker, W., Blair, J.B., Hofton, M., & Hunsaker, C. 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment* 102: 63–73.

Im, J., & Jensen, J. 2008. Hyperspectral remote sensing of vegetation. *Geography Compass* 2: 1943–1961.

Ishii, H.T., Tanabe, S., & Hiura, T. 2004. Exploring the relationships among canopy structure, stand productivity, and biodiversity of temperature forest ecosystems. *Forest Science* 50: 342–355.

Jaworski, A., & Podlaski, R. 2011. Modelling irregular and multimodal tree diameter distributions by finite mixture models: an approach to stand structure characterisation. *Journal of Forest Research* 17: 79–88.

Jennings, M.D., Faber-Langendoen, D., Loucks, O.L., Peet, R.K., & Roberts, D. 2009. Standards for associations and alliances of the US. National Standards Classification Vegetation. *Ecological Monographs* 79: 173–199.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P., Schaepman, M.E., Schildhauer, M.P., Schneider,

F.D., Schrodt, F., Stahl, U., & Ustin, S.L. 2016. Monitoring plant functional diversity from space. *Nature Plants* 2: 16024.

Jost, L. 2010. The Relation between Evenness and Diversity. *Diversity* 2: 207–232.

Kampe, T., 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: 1–24.

Kane, V.R., McGaughey, R.J., Bakker, J.D., Gersonde, R.F., Lutz, J. a., & Franklin, J.F. 2010. Comparisons between field- and LiDAR-based measures of stand structural complexity. *Canadian Journal of Forest Research* 40: 761–773.

Kaufman, L., & Rousseeuw, P.J. 2005. *Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics)*. John Wiley & Sons.

Kindermann, G., Obersteiner, M., Sohngen, B., Sathaye, J., Andrasko, K., Rametsteiner, E., Schlamadinger, B., Wunder, S., & Beach, R. 2008. Global cost estimates of reducing carbon emissions through avoided deforestation. *Proceedings of the National Academy of Sciences* 105: 10302–10307.

Knox, R.G., Peet, R.K., & Christensen, N.L. 1989. Population dynamics in loblolly pine stands: changes in skewness and size inequality. *Ecology* 70: 1153–1166.

Knyazikhin, Y., Schull, M.A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., Marshak, A., Latorre Carmona, P., Kaufmann, R.K., Lewis, P., Disney, M.I., Vanderbilt, V., Davis, A.B., Baret, F., Jacquemoud, S., Lyapustin, A., & Myneni, R.B. 2013. Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Sciences of the United States of America* 110: 1–8.

Kodani, E., Awaya, Y., Tanaka, K., & Matsumura, N. 2002. Seasonal patterns of canopy structure, biochemistry and spectral reflectance in a broad-leaved deciduous Fagus crenata canopy. *Forest Ecology and Management* 167: 233–249.

Kreft, H., & Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the United States of America* 104: 5925–30.

Kruskal, J.B. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29: 115–129.

Kuhn, M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28: 1–26.

Kuhn, M. 2015. caret: Classification and Regression Training. R package version 6.0-41.

Laurance, W. 2002. Hyperdynamism in fragmented habitats. *Journal of Vegetation Science* 13: 595–602.

Leak, W. 1964. An Expression of Diameter Distribution For Unbalanced, uneven-aged stands and forests. *Forest Science* 10: 39–50.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659–1673.

Legendre, P., & Legendre, L. 2012. *Numerical ecology: Third English edition*. Elsevier, Amsterdam, Netherlands.

Lennon, J.J. 2000. Red-shifts and red herrings in geographical ecology. *Ecography* 23: 101–113.

Lesky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. 2002. Lidar remote sensing for ecosystem studies. *BioScience* 52: 19–30.

Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C., Dech, S., & Wegmann, M. 2012. Modelling forest α-diversity and floristic composition - on the added value of LiDAR plus hyperspectral remote sensing. *Remote Sensing* 4: 2818–2845.

Levin, S.A. 1992. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture Author(s): Simon A. Levin Source: *Ecology* 73: 1943–1967.

Levin, N., Shmida, A., Levanoni, O., Tamari, H., & Kark, S. 2007. Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. *Diversity and Distributions* 13: 692–703.

Lexerød, N.L., & Eid, T. 2006. An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management* 222: 17–28.

Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E.-D., McGuire, D., Bozzato, F., Pretzsch, H., De-Miguel, S., Paquette, A., Hérault, B., Scherer-Lorenzen, M., Barrett, C.B., Glick, H.B., Hengeveld, G.M., Nabuurs, G.-J., Pfautsch, S., Viana, H., Vibrans, A.C., Ammer, C., Schall, P., Verbyla, D., Tchebakova, N., Fischer, M., Watson, J. V., Chen, H.Y.H., Lei, X., Schelhaas, M.-J., Lu, H., Gianelle, D., Parfenova, E.I., Salas, C., Lee, E., Lee, B., Kim, H.S., Bruelheide, H., Coomes, D.A., Piotto, D., Sunderland, T., Schmid, B., Gourlet-Fleury, S., Sonké, B., Tavani, R., Zhu, J., Brandl, S., Vayreda, J., Kitahara, F., Searle, E.B., Neldner, V.J., Ngugi, M.R., Baraloto, C., Frizzera, L., Bałazy, R., Oleksyn, J., Zawiła-Niedźwiecki, T., Bouriaud, O., Bussotti, F., Finér, L., Jaroszewicz, B., Jucker, T., Valladares, F., Jagodzinski, A.M., Peri, P.L., Gonmadje, C., Marthy, W., O'Brien, T., Martin, E.H., Marshall, A., Rovero, F., Bitariho, R., Niklaus, P.A., Alvarez-Loayza, P., Chamuya, N., Valencia, R., Mortier, F., Wortel, V., Engone-Obiang, N.L., Ferreira, L. V., Odeke, D.E., Vasquez, R.M., & Reich, P.B. 2016. Positive biodiversity–productivity relationship predominant in global forests. *Science* 354: 196.

Liaw, A., & Wiener, M. 2002. Classification and Regression by randomForest. *R News* 2: 18–22.

Loreau, M., Naeem, S., & Inchausti, P. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* 294: 804–808.

Lorimer, C.G., & Frelich, L.E. 1997. A Structural Alternative to Chronosequence Analysis for

Uneven-Aged Northern Hardwood Forests. *Journal of Sustainable Forestry* 6: 347–366.

Lorimer, C., & Krug, A. 1983. Diameter distributions in even-aged stands of shade-tolerant and midtolerant tree species. *American Midland Naturalist* 109: 331–345.

MacArthur, R., & MacArthur, J. 1961. On Bird Species Diversity. *Ecology* 42: 594–598.

Magurran, A. 1988. *Ecological diversity and its measurement*. Springer Netherlands, Dordrecht.

Marks, C.O., Muller-Landau, H.C., & Tilman, D. 2016. Tree diversity, tree height and environmental harshness in eastern and western North America. *Ecology Letters* 743–751.

Matthews, E.R., Peet, R.K., & Weakley, A.S. 2011. Classification and description of alluvial plant communities of the Piedmont region, North Carolina, USA. *Applied Vegetation Science* 14: 485–505.

McElhinny, C., Gibbons, P., Brack, C., & Bauhus, J. 2005. Forest and woodland stand structural complexity: Its definition and measurement. *Forest Ecology and Management* 218: 1–24.

McGarigal, K., & Cushman, S. a. 2005. The gradient concept of landscape structure. In *Issues and perspectives in landscape ecology.*, pp. 112–119. Cambridge University Press, Cambridge, UK.

McGill, B.J., Enquist, B.J., Weiher, E., & Westoby, M. 2006. Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution* 21: 178–185.

MEA. 2005. Ecosystems and human well-being. Health synthesis. *Ecosystems* 5: 1–100.

Middleton, M., Narhi, P., Arkimaa, H., Hyvonen, E., Kuosmanen, V., Treitz, P., & Sutinen, R. 2012. Ordination and hyperspectral remote sensing approach to classify peatland biotopes along soil moisture and fertility gradients. *Remote Sensing of Environment* 124: 596–609.

Moen, C., & Gutiérrez, R. 1997. California Spotted Owl Habitat Selection in the Central Sierra Nevada. *The Journal of wildlife management* 61: 1281–1287.

Mohler, C.L., Marks, P.L., & Sprugel, D.G. 1978. Stand structure and allometry of trees during self-thinning of pure stands. *Journal of Ecology* 66: 599–614.

Montgomery, R., & Chazdon, R. 2001. Forest structure, canopy architecture, and light transmittance in tropical wet forests. *Ecology* 82: 2707–2718.

Myers, N., Mittermeier, R. a, Mittermeier, C.G., da Fonseca, G. a, & Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–8.

Naeem, S., & Li, S. 1997. Biodiversity enhances ecosystem reliability. *Nature* 390: 507–509.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80: 88–99.

Næsset, E., & Økland, T. 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment* 79: 105–115.

Nagendra, H. 2001. Using remote sensing to assess biodiversity, review article. *International Journal of Remote Sensing* 22: 2377–2400.

Naidoo, L., Cho, M. a., Mathieu, R., & Asner, G. 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing* 69: 167–179.

National Climatic Data Center. 2011. *SO 1981-2010 Climate Normals. NOAA Climatic Data Center*.

NC OneMap Geospatial Portal. 2016. NC OneMap Geospatial Portal.

Neumann, M., & Starlinger, F. 2001. The significance of different indices for stand structure and diversity in forests. *Forest ecology and Management* 145: 91–106.

Newbold et al., T. 2015. Global effects of land use on local terrestrial biodiversity. *Nature* 520: 45–50.

Noss, R.F. 1990. Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conservation Biology* 4: 355–364.

O'Connell, B.M.., LaPoint, E.B.., Turner, J.A.., Ridley, T., Pugh, S.A.., Wilson, A.M.., Waddell, K.L.., & Conkling, B.L. 2015. *The Forest Inventory and Analysis Database: Database description and user guide version 6.0.2 for Phase 2.*

Ohmann, J.L., & Gregory, M.J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest- neighbor imputation in coastal Oregon, U.S.A. *Canadian Journal of Forest Research* 32: 725–741.

Ohmann, J.L., Gregory, M.J., Henderson, E.B., & Roberts, H.M. 2011. Mapping gradients of community composition with nearest-neighbour imputation: Extending plot data for landscape analysis. *Journal of Vegetation Science* 22: 660–676.

Okin, G.S., Roberts, D.A., Murray, B., & Okin, W.J. 2001. Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment* 77: 212–225.

Olden, J.D., Lawler, J.J., & Poff, N.L. 2008. Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* 83: 171–193.

Oliver, C.D., & Larson, B.C. 1996. *Forest stand dynamics.* Wiley and Sons, New York, NY.

Ollinger, S. V. 2011. Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist* 189: 375–394.

Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26: 217–222.

Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S., & Wohlgemuth, T. 2002. Quantitative tools for perfecting species lists. In *Environmetrics*, pp. 121–137.

Palmer, M.W., & Maurer, T.A. 1997. Does diversity beget diversity? A case study of crops and weeds. *Journal of Vegetation Science* 8: 235–240.

Palmer, M.W., Peet, R.K., Reed, R.A., Xi, W., & White, P.S. 2007. a Multiscale Study of Vascular Plants in a North Carolina Piedmont Forest. *Ecology* 88: 2674.

Parker, G., Cohen, W.., Harding, D., Lefsky, M. a, & Shugart, H.H. 1999. Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA. *Remote Sensing of Environment* 67: 83–98.

Pearman, P.B., Guisan, A., Broennimann, O., & Randin, C.F. 2008. Niche dynamics in space and time. *Trends in Ecology and Evolution* 23: 149–158.

Peet, R.K. 1992. Community structure and ecosystem properties. In Glenn-Lewin, D.C., Peet, R.K., & Veblen, T. (eds.), *Plant succession: theory and prediction*, pp. 102–115. Chapman and Hall, London.

Peet, R.K. 1974. The measurement of species diversity. *Annual Review of Ecology and Systematics* 5: 285–309.

Peet, R.K., & Christensen, N.L. 1988. Changes in species diversity during secondary forest succession on the North Carolina piedmont. In During, J.J., Werger, M.J.A., & Willems, J.H. (eds.), *Diversity and Pattern in Plant Communities*, pp. 233–245.

Peet, R.K., & Christensen, N. 1987. Competition and Tree death. *BioScience* 37: 586–595.

Peet, R.K., & Christensen, N. 1980. Hardwood forest vegetation of the North Carolina piedmont. *Verőffentlichungen des Geobotani Institutes der ETH, Stiftung Rübel, Zürich* 69: 14–39.

Peet, R., Lee, M., Jennings, M., & Faber-Langendoen, D. 2012. VegBank – a permanent, open-access archive for vegetation-plot data. *Biodiversity & Ecology* 4: 233–241.

Peet, R.K., Palmquist, K.A., & Tessel, S.M. 2014. Herbaceous layer species richness of southeastern forests and woodlands: Patterns and causes. In F.S. Gilliam & M.R. Roberts, M.. (ed.), *The herbaceous layer in forests of eastern North America, Chapter 10. Second edition*, pp. 255–276. Oxford University Press, Oxford, UK.

Peet, R.K., & Roberts, D.W. 2013. Classification of Natural and Semi-natural Vegetation. In *Vegetation Ecology: Second Edition*, pp. 28–70.

Peet, R., Wentworth, T., & White, P. 1998. A flexible, multipurpose method for recording vegetation composition and structure. *Castanea* 63: 262–274.

Petchey, O.L., & Gaston, K.J. 2002. Functional diversity (FD), species richness and community composition. *Ecology Letters* 5: 402–411.

Pickett, S.T.A., & White, P.S. 1985. Patch dynamics: a synthesis. In *In The ecology of natural disturbance and patch dynamics*, pp. 371–384.

Pielou, E.C. 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* 13: 131–144.

Prasad, A.M., Iverson, L.R., & Liaw, A. 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 9: 181–199.

Price, J.C. 1994. How unique are spectral signatures? *Remote Sensing of Environment* 49: 181–186.

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A., & Townsend Peterson, A. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426: 837–841.

Recher, H.F., Majer, J.D., & Ganesh, S. 1996. Eucalypts, arthropods and birds: on the relation between foliar nutrients and species richness. *Forest Ecology and Management* 85: 177–195.

Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M., Walters, M.B., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M., & Walters, M.B. 2003. The evolution of plant functional variation: traits, spectra, and strategies. *International Journal of Plant Sciences* 164: S143–S164.

Rennolls, K., Geary, D.N., & Rollinson, T.J.D. 1985. Characterizing Diameter Distributions by the use of the Weibull Distribution. *Forestry* 58: 57–66.

Roberts, D.W. 2016a. labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.8-0.

Roberts, D.W. 2016b. optpart: Optimal Partitioning of Similarity Relations. R package version 2.2-0.

Roberts, D.W. 2015. Vegetation classification by two new iterative reallocation optimization algorithms. *Plant Ecology* 216: 741–758.

Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M., Nagendra, H., Oldeland, J., Ricotta, C., Southworth, J., & Neteler, M. 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecological Informatics* 5: 318–329.

Rocchini, D., Ricotta, C., & Chiarucci, A. 2007. Using satellite imagery to assess plant species

richness : The role of multispectral systems. *Applied Vegetation Science* 10: 325–331.

Rodrigues, A.S.L., & Brooks, T.M. 2007. Shortcuts for Biodiversity Conservation Planning: The Effectiveness of Surrogates. *Annual Review of Ecology, Evolution, and Systematics* 38: 713–737.

Roth, K.L., Roberts, D.A., Dennison, P.E., Alonzo, M., Peterson, S.H., & Beland, M. 2015. Differentiating plant species within and across diverse ecosystems with imaging spectroscopy. *Remote Sensing of Environment* 167: 135–151.

Rousseeuw, P.J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20: 53–65.

Rubin, B.D., Manion, P.D., & Faber-Langendoen, D. 2006. Diameter distributions and structural sustainability in forests. *Forest Ecology and Management* 222: 427–438.

Rudel, T.K., Coomes, O.T., Moran, E., Achard, F., Angelsen, A., Xu, J., & Lambin, E. 2005. Forest transitions: Towards a global understanding of land use change. *Global Environmental Change* 15: 23–31.

Running, S.W., & Coughlan, J.C. 1988. A general model of forest ecosystem processes for regional applications. *Ecological Modelling* 42: 125–154.

Schmidtlein, S., & Sassin, J. 2004. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing of Environment* 92: 126–138.

Schmidtlein, S., Zimmermann, P., Schupferling, R., & Weiss, C. 2007. Mapping the floristic continuum: Ordination space position estimated from imaging spectroscopy. *Journal of Vegetation Science* 18: 131–140.

Schölkopf, B., & Smola, A. 2002. *Learning with Kernels—Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge, MA.

Sexton, J.O., Bax, T., Siqueira, P., Swenson, J.J., & Hensley, S. 2009. A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern North America. *Forest Ecology and Management* 257: 1136–1147.

Shannon, C.E., & Weaver, W. 1949. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL.

Shipley, B., Lechowicz, M.J., Wright, I., & Reich, P.B. 2006. Fundamental trade-offs generating the worldwide leaf economics spectrum. *Ecology* 87: 535–541.

Shmida, A., & Wilson, M. V. 1985. Biological Determinants of Species Diversity. *Journal of Biogeography* 12: 1–20.

Simonson, W.D., Allen, H.D., & Coomes, D.A. 2012. Use of an Airborne Lidar System to Model Plant Species Composition and Diversity of Mediterranean Oak Forests. *Conservation*

Biology 26: 840–850.

Simpson, E.H. 1949. Measurement of diversity. *Nature* 163: 688.

Singh, A., Serbin, S.P., McNeil, B.E., Kingdon, C.C., & Townsend, P.A. 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecological Applications* 25: 2180–2197.

Smith, D.M., Larson, B.C., Kelty, M.J., & Ashton, P.M.S. 1997. The practice of silviculture: applied forest ecology.

Song, C., Chen, J.M., Hwang, T., Gonsamo, A., Croft, H., Zhang, Q., Dannenberg, M., Zhang, Y., Hakkenberg, C., & Li, J. 2015. Ecological characterization of vegetation using multi-sensor remote sensing in the solar reflective spectrum. In Thenkabail, P.S. (ed.), *Remote Sensing Handbook, Vol 2. Land Resources: Monitoring, Modeling, and Mapping*, pp. 533–575. Taylor and Francis, London, UK.

Souza, A.A., Galvão, L.S., & Santos, J.R. 2010. Relationships between Hyperion-derived vegetation indices, biophysical parameters, and elevation data in a Brazilian savannah environment. *Remote Sensing Letters* 1: 55–64.

Spies, T.A., & Franklin, J.F. 1991. The Structure of Natural Young, Mature, and Old-Growth Douglas-Fir Forests in Oregon and Washington. In *USDA Forest Service General Technical Report PNW-GTR-285*, pp. 91–111. Portland, USA.

Spies, T.A., Franklin, J.F., & Klopsch, M. 1990. Canopy gaps in Douglas-fir forests of the Cascade Mountains. *Canadian Journal of Forest Research* 20: 649–658.

Staudhammer, C.L., & LeMay, V.M. 2001. Introduction and evaluation of possible indices of stand structural diversity. *Canadian Journal of Forest Research* 31: 1105–1115.

Sullivan, T., Sullivan, D., & Lindgren, P. 2001. Stand structure and small mammals in young lodgepole pine forest: 10-year results after thinning. *Ecological Applications* 11: 1151–1173.

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., & Van Der Meer, B. 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment* 91: 354–376.

Thessler, S., Ruokolainen, K., Tuomisto, H., & Tomppo, E. 2005. Mapping gradual landscape-scale floristic changes in Amazonian primary rain forests by combining ordination and remote sensing. *Global Ecology and Biogeography* 14: 315–325.

Tilman, D., Knops, J.M.H., Wedin, D., Reich, P.B., & Siemann, E. 1997. The influence of functional diversity and composition on ecosystem processes. *Science* 277: 1300–1302.

Tilman, D., Wedin, D., & Knops, J. 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* 379: 718–720.

Toloşi, L., & Lengauer, T. 2011. Classification with correlated features: Unreliability of feature ranking and solutions. *Bioinformatics* 27: 1986–1994.

Torabzadeh, H., Morsdorf, F., & Schaepman, M.E. 2014. Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems - A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 97: 25–35.

Turner, M.G. 2005. What Is the State of the Science? *Annual Review of Ecology, Evolution, and Systematics* 36: 319–344.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution* 18: 306–314.

Urban, M. C., et al. 2016. Improving the forecast for biodiversity under climate change. *Science (New York, N.Y.)* 353: 1293–1310.

USDA. 2014. *Forest Service Forest Inventory and Analysis Program. FIA DataMart v. 5.1.*

USDA. 2016. *The PLANTS Database (http://plants.usda.gov, 8 January 2016)*. Greensboro, NC, USA.

Ustin, S.L., & Gamon, J.A. 2010. Remote sensing of plant functional types. *The New Phytologist* 186: 795–816.

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.

Uuttera, J., Maltamo, M., & Hotanen, J. 1997. The structure of forest stands in virgin and managed peatlands: a comparison between Finnish and Russian Karelia. *Forest Ecology and Management* 96: 125–138.

Valbuena, R., Packalén, P., Martı´n-Fernández, S., & Maltamo, M. 2012. Diversity and equitability ordering profiles applied to study forest structure. *Forest Ecology and Management* 276: 185–195.

van Ewijk, K.Y., Treitz, P.M., & Scott, N.A. 2011. Characterizing Forest Succession in Central Ontario using Lidar-derived Indices. *Photogrammetric Engineering and Remote Sensing* 77: 261–269.

Vapnik, V.N., & Vapnik, V. 1998. *Statistical learning theory. Vol. 1*. Wiley and Sons, New York, NY.

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.M., Hoegh-Guldberg, O., & Bairlein, F. 2002. Ecological responses to recent climate change. *Nature* 416: 389–395.

Weakley, A.S. 2015. *Flora of the Southern and Mid-Atlantic States. Working draft of 21 May 2015.* Univ. of North Carolina Herbarium (NCU), Chapel Hill.

Weibull, W. 1951. A statistical distribution function of wide applicability. *ASME Journal of Applied Mechanics* 18: 293–297.

Weiner, J., & Thomas, S. 1986. Size variability and competition in plant monocultures. *Oikos* 2: 211–222.

Wenger, S.J., & Olden, J.D. 2012. Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* 3: 260–267.

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.

White, J.C., Wulder, M. a, Varhola, A., Vastaranta, M., Coops, N.C., Cook, B.D., Pitt, D., & Woods, M. 2013. *A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach*.

Whittaker, R.H. 1967. Gradient Analysis of Vegetation. *Biological Reviews* 42: 207–264.

Wolf, J., Fricker, G., Meyer, V., Hubbell, S., Gillespie, T., & Saatchi, S. 2012. Plant Species Richness is Associated with Canopy Height and Topography in a Neotropical Forest. *Remote Sensing* 4: 4010–4021.

Wolter, P.T., & Townsend, P. a. 2011. Multi-sensor data fusion for estimating forest species composition and abundance in northern Minnesota. *Remote Sensing of Environment* 115: 671–691.

Zak, D.R., Holmes, W.E., White, D.C., Peacock, A.D., & Tilman, D. 2003. Plant diversity, soil microbial communities, and ecosystem function: Are there any links? *Ecology* 84: 2042–2050.

Zellweger, F., Baltensweiler, A., Ginzler, C., Roth, T., Braunisch, V., Bugmann, H., & Bollmann, K. 2016. Environmental predictors of species richness in forest landscapes: Abiotic factors versus vegetation structure. *Journal of Biogeography* 43: 1080–1090.

Zhang, G., & Lu, Y. 2012. Bias-corrected random forests in regression. *Journal of Applied Statistics* 39: 151–160.