Ecology, 99(2), 2018, pp. 474–487
© 2017 by the Ecological Society of America

Mapping multi-scale vascular plant richness in a forest landscape with integrated LiDAR and hyperspectral remote-sensing

C. R. Hakkenberg,[1,5] K. Zhu,[2] R. K. Peet,[1,3] and C. Song[1,4]

[1]*Curriculum for the Environment and Ecology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599 USA*
[2]*Department of Environmental Studies, University of California at Santa Cruz, Santa Cruz, California 95064 USA*
[3]*Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599 USA*
[4]*Department of Geography, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599 USA*

Abstract. The central role of floristic diversity in maintaining habitat integrity and ecosystem function has propelled efforts to map and monitor its distribution across forest landscapes. While biodiversity studies have traditionally relied largely on ground-based observations, the immensity of the task of generating accurate, repeatable, and spatially-continuous data on biodiversity patterns at large scales has stimulated the development of remote-sensing methods for scaling up from field plot measurements. One such approach is through integrated LiDAR and hyperspectral remote-sensing. However, despite their efficiencies in cost and effort, LiDAR-hyperspectral sensors are still highly constrained in structurally- and taxonomically-heterogeneous forests - especially when species' cover is smaller than the image resolution, intertwined with neighboring taxa, or otherwise obscured by overlapping canopy strata. In light of these challenges, this study goes beyond the remote characterization of upper canopy diversity to instead model total vascular plant species richness in a continuous-cover North Carolina Piedmont forest landscape. We focus on two related, but parallel, tasks. First, we demonstrate an application of predictive biodiversity mapping, using nonparametric models trained with spatially-nested field plots and aerial LiDAR-hyperspectral data, to predict spatially-explicit landscape patterns in floristic diversity across seven spatial scales between 0.01–900 m^2. Second, we employ bivariate parametric models to test the significance of individual, remotely-sensed predictors of plant richness to determine how parameter estimates vary with scale. Cross-validated results indicate that predictive models were able to account for 15–70% of variance in plant richness, with LiDAR-derived estimates of topography and forest structural complexity, as well as spectral variance in hyperspectral imagery explaining the largest portion of variance in diversity levels. Importantly, bivariate tests provide evidence of scale-dependence among predictors, such that remotely-sensed variables significantly predict plant richness only at spatial scales that sufficiently subsume geolocational imprecision between remotely-sensed and field data, and best align with stand components including plant size and density, as well as canopy gaps and understory growth patterns. Beyond their insights into the scale-dependent patterns and drivers of plant diversity in Piedmont forests, these results highlight the potential of remotely-sensible essential biodiversity variables for mapping and monitoring landscape floristic diversity from air- and space-borne platforms.

Key words: essential biodiversity variables (EBVs); hyperspectral remote-sensing; LiDAR; North Carolina Piedmont; predictive vegetation mapping; remote sensing; scale-dependence; spectral variation; structural complexity; vascular plant species richness.

Introduction

Human activities have been implicated in the severe degradation of the Earth's ecosystems, especially the elimination of species whose absence can alter the functioning of ecosystems and the extrinsic benefits they provide to society (Hooper et al. 2012). Species loss is an especially problematic trend, as species diversity has been found to bolster community resilience to disturbance (Naeem and Li 1997), strengthen resistance to infectious disease and pathogens (Civitello et al. 2015), and increase net ecosystem productivity (Tilman et al. 1996). In forest ecosystems, plants are critical to ecosystem function, and in their capacity to provide habitat and resources throughout the trophic chain, serve as effective cross-taxon surrogates for overall levels of biodiversity (Zak et al. 2003, Rocchini et al. 2015). The recognition of the significance of plant diversity for conservation priority setting and the modeling of habitat, biodiversity, and ecosystem function has propelled efforts to identify remotely-sensed essential biodiversity variables (EBVs) to map and monitor its spatial distribution (Pettorelli et al. 2016).

Manuscript received 31 August 2017; revised 25 October 2017; accepted 20 November 2017. Corresponding Editor: Samantha K. Chapman.
[5] E-mail: chrishakkenberg@gmail.com

Though limited in scope and laborious to collect, field plot data have served almost exclusively as the basis for ecological studies into the patterns and drivers of forest biodiversity over the past century (Condit 1995, Peet et al. 2012*a*, Anderson-Teixeira et al. 2015). Remote-sensing data, in comparison, is coarsely resolved and limited by technological constraints and environmental conditions like cloud cover, sun angle, and topographic shadow (Song et al. 2015). However, while lacking the fine-scale precision of most ground-based observations, airborne and satellite remote sensing offers researchers access to spatially-continuous data collection over large portions of the Earth surface at a fraction of the effort and cost of ground-based sampling (Palmer et al. 2002, Asner and Martin 2009). Global remote sensing data archives are vast and growing swiftly, allowing for the collection of disparate types of data on the Earth surface at multiple spatial and temporal resolutions, capable of sampling in remote or difficult to access areas, and reconstructing historical trends in vegetation dynamics in otherwise under-sampled regions (Townshend et al. 2012, Yu et al. 2015). Indeed, when coupled with field plots, remote-sensing is uniquely capable of providing a synoptic and consistent observation platform from which to detect and model the spatio-temporal dynamics of forest biodiversity at landscape- to global-scales (Gillespie et al. 2008).

One promising remote-sensing technology enabling the spatially continuous characterization of emergent properties of forest communities like plant diversity is airborne image spectroscopy, or hyperspectral imaging (Ustin et al. 2009). Hyperspectral imagers are adept at detecting finely partitioned, narrowband portions of the electromagnetic spectrum capable of distinguishing subtle variations in phenotypic traits like foliar morphology and biochemistry (Curran 1989). Even where the mechanistic relationship between spectral reflectance and the contingencies of trait-environment interactions is not fully resolved, correlational relationships between spectral features and foliar chemistry allow for the detection of a host of canopy properties including stand composition and underlying environmental conditions (Ollinger 2011).

In sparse canopies with non-overlapping crowns, hyperspectral remote-sensing has been used to identify disparate taxa based on field and lab-measured spectral libraries, and thereby tally tree species richness from the bottom-up (Pu and Landry 2012, Roth et al. 2015). However, this approach depends on relatively consistent environmental and illumination conditions and the expectation that inter-species spectral variation exceeds that of intra-species differences - an assumption not always met in practice (Price 1994). Even where hyperspectral imagery is able to distinguish among species, detection is restricted to the upper canopy, and therefore ineffective for the identification of understory plants.

For the goal of distinguishing properties beyond (and below) closed canopies, light detection and ranging (LiDAR) is a particularly effective supplement to hyperspectral data that is less susceptible to issues of signal saturation known to afflict optical imagery in areas of dense canopy and high biomass (Kane et al. 2010). Aerial LiDAR measures the distance between an airborne sensor and target objects, thereby providing a three-dimensional point cloud of ground contours (i.e. topography) and above-ground elements like canopy structure (Dubayah and Drake 2000). When combined in a single platform, integrated LiDAR and hyperspectral sensors (hereafter "LiDAR-hyperspectral") provide complementary data capable of simultaneously characterizing multiple geophysical gradients constraining species' niche space (Cord et al. 2013).

Most studies of remote-sensing of biodiversity have focused on empirical relationships between remotely-sensed data and the levels of species diversity observed in field plots (Chiarucci and Bonini 2005, Cayuela et al. 2006, Wolf et al. 2012, Rocchini et al. 2015). Envisioning biodiversity as the cumulative result of aggregated multi-species niches, these statistical relationships can be used to guide predictive models of the spatial distribution of plant diversity, most often focusing on canopy species at a single spatial scale (Feilhauer and Schmidtlein 2009, Féret and Asner 2014). But plant diversity manifests throughout the vertical profile of a forest stand and its spatial distribution varies as a function of scale, as do the parameters used to predict it (Levin 1992, Crawley and Harral 2001). Indeed, full-stand vascular plant diversity mapping is critical in these mixed species, multi-strata stands - precisely where efforts to model it are most technologically and logistically constrained when species' cover is smaller than the remotely-sensed pixel resolution, intertwined with neighboring taxa, or obscured from overhead optical sensors due to dense canopies (Franklin 1988).

In light of these challenges and in anticipation of a rapid increase in annual LiDAR-hyperspectral data products (Kampe et al. 2010, Cook et al. 2013), this study focuses on two related, but parallel tasks: First, we train feature-selected, nonparametric models with nested field plot data and derived LiDAR and hyperspectral variables from NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) airborne imager to model and map multi-scale vascular plant species richness in a compositionally- and structurally-complex forest landscape in the North Carolina Piedmont. Multi-scale predictive models, and associated uncertainty estimates, are spatially explicit and continuous throughout the study site. Next, we assess individual remotely-sensed predictors of vascular plant diversity to infer how remotely-sensed predictors of underlying structural, ecological, and environmental gradients affect plant richness predictions across spatial scales. Beyond the insight they provide into spatially-explicit, multi-scale patterns of landscape plant richness, predictive models of diversity provide a basis from which to infer specific scale-dependent drivers of vascular plant species richness. While the tasks

are interrelated, the approaches are not. For prediction of pattern, accuracy and generalizability is of paramount concern; whereas for inference into process, transparency in parameter estimates is most important (Lo et al. 2015, Houlahan et al. 2016).

Methods

Vegetation sampling

The 2.8 km^2 study site in the Duke Forest Blackwood Division, NC, USA consists primarily of Piedmont secondary old-field successional pine and mature hardwood forests following selective cutting, agriculture, and grazing in the 19th and early 20th century (Peet and Christensen 1988). Owing to an explicit focus on natural and semi-natural forests, the study site excludes areas of plantation forest, clear-cut, and built infrastructure (Fig. 1). Specific field plot locations were based on a stratified random sampling design, randomly pre-determined within the constraints of stratified bands along topographic gradients in the study area (Fig. 1; upper left). Thus conceived, field plots span the spectrum of compositional variation and physiognomic types of the relatively taxonomically diverse and structurally heterogeneous study area to include upland, riparian, and bottomland forest community types (Hakkenberg et al. 2018). In all thirty-six 900 m^2 field plots, species presence was recorded following Carolina Vegetation Survey protocols for all vascular plant species in eight 0.01, 0.1, 1, and 10 m^2 spatially nested subplots, four 100 m^2 subplots, one 400 m^2 subplot, as well as at the 900 m^2 full plot scale (Appendix S1; Peet et al. 1998, 2012*a*). Thus, among all full plots sampled (together comprising 0.1% of the entire study area), species cover values were recorded in 288 subplots at 0.01–10 m^2 scales, 144 subplots at 100 m^2, 36 subplots at 400 m^2, and 36 full plots at the 900 m^2 scale. Botanical nomenclature follows Weakley (2015), and the resultant plot data are available on Vegbank (http://vegbank.org accession code: VB.pj.11067.DUKEBLACKWOODNA) (Peet et al. 2012*b*).

Remotely-sensed predictors

Aerial remotely-sensed data come from NASA Goddard's LiDAR, Hyperspectral and Thermal (G-LiHT) airborne imager at 2 × 2 m spatial resolution (Cook et al. 2013). The G-LiHT airborne imager utilizes commercial, off-the-shelf sensors including 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 spectral resolution, as well as a VQ-480 (Riegl USA, Orlando, FL, USA) airborne LiDAR scanning 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 335 m (Cook et al. 2013). Thermal data were not available for our study area.

Remotely-sensed data was collected on Oct. 25, 2013, during late-season, leaf-on conditions when inter-species phenological differences could aid in taxonomic discrimination. Due to the stringent requirement for a high degree of geo-locational precision in fitting nested field plots with fine-resolution (4 m^2) remotely-sensed imagery, all efforts were made to achieve sub-meter spatial locational accuracies for plot and subplot footprints. Accordingly, the precise location of plot and subplot vertex coordinates were calculated by triangulating the distance between field-measured plot corners to Ground Control Points visible in fine resolution (0.15 m) Digital Orthophoto Quadrangle imagery (NC OneMap Geospatial Portal 2016).

All derived remotely-sensed predictor variables emanate from four G-LiHT Level 3 products: a digital terrain model, a canopy height model (CHM), the full LiDAR point cloud, and a 114-band atmospherically-corrected surface reflectance hyperspectral image stack (Table 1). The digital terrain model uses LiDAR last returns - those returns remaining after all others have been reflected by above-ground elements - to estimate a detailed model of ground surface contours. This, in turn, was used to generate secondary measures of topography including elevation, slope, annualized solar radiation (ASR), topographic position index (TPI, a measure of relative concavity in hillslope position), deviation from mean elevation (DEV, equivalent to TPI normalized to local surface roughness), and topographic wetness index (TWI, an index of soil moisture along hillslope catenas) (Beven and Kirkby 1979, De Reu et al. 2013). Owing to the inherent redundancy of the hyperspectral layers, the imagery was reduced to two classes of derived products: standardized principal component axes (PCA 1-3) and narrowband vegetation indices (VIs) (Asner and Martin 2009, Ustin et al. 2009). While the former benefits from incorporation of ~98% of variance among all hyperspectral image layers, standardized PCA indices are unit-less, relational, and limited to the extent of the training data space. Narrowband VIs, on the other hand, each detect only a small portion of the reflectance spectrum but excel as interpretable indicators of known traits (e.g. the carotenoid reflectance index, an index of carotenoid pigments) based on empirical spectral band selection (Gitelson et al. 2002, Ustin et al. 2009).

To facilitate the incorporation of spectral variability measures and test for scale-dependence among remotely-sensed predictors of plant richness, all remotely-sensed derived variables were aggregated at four spatial resolutions – 16, 100, 400, and 900 m^2. Elevation, slope, ASR, DEV and TPI were aggregated as mean values, while for the CHM, mean, minimum, maximum, standard deviation, skewness and kurtosis was computed (see Table 1). All other 4 m^2 raw pixel variables were aggregated by mean, minimum, maximum, and standard deviation. To compare how LiDAR-detected microtopography (including imprecise or spurious ground returns) impacted hydrological flow routing and

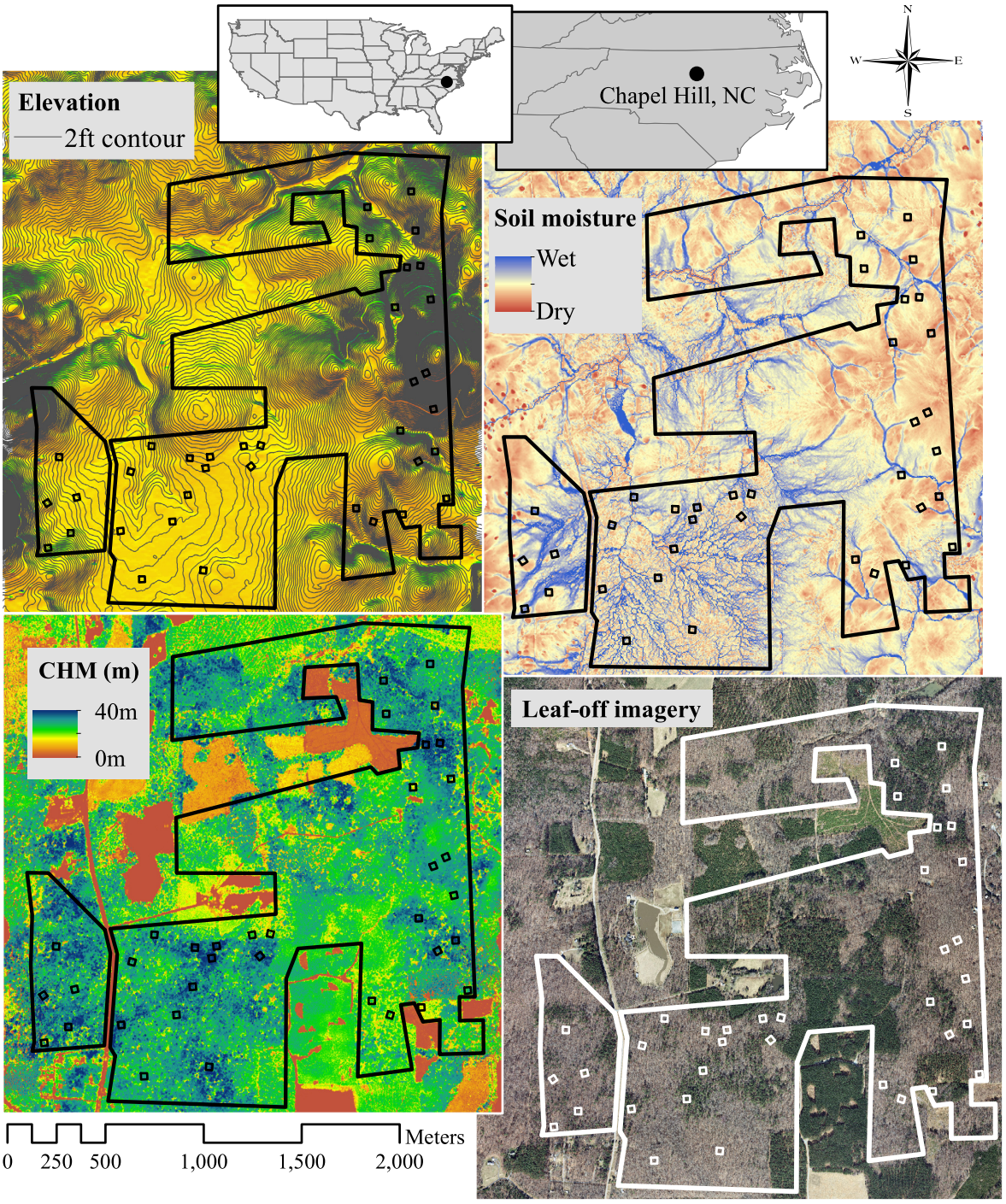

FIG. 1. Duke Forest Blackwood Study Site. The footprint of 900 m^2 nested field plots (hollow squares) fall along a N-S and an E-W transect within the bounds of study area (white boundary line) designed to exclude plantation forests, clear-cuts, and human habitation. Leaf-off imagery (lower right), canopy height model (lower left), elevation contours (upper left), and topographic wetness index (TWI), a proxy for soil moisture (upper right).

landform analysis, TWI was assessed at three scales (2, 4, and 8 m), while landform indices like TPI and DEV were assessed within a window of 162 and 222 m before aggregation at the aforementioned four resolutions. For all subsequent analyses, the spatial resolution of the remotely-sensed predictors best corresponding to the spatial scale of each plant richness response variable was employed. For example, for the four smallest spatial

TABLE 1. LiDAR and hyperspectral remote-sensing predictor variables*.

Category and 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 222 m 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 222 m 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 and Kirkby 1979)
LiDAR canopy height (first returns)		
Canopy height model	CHM	CHM (Cook et al. 2013)
LiDAR canopy structure (all returns)		
Returns per pulse	rpp	LiDAR returns (Cook et al. 2013)
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‡		
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 Axes 1–3	PCA1-3	Full spectrum PCA axis 1–3 (ρ_{407}:ρ_{1007})
Red Edge Position Index	REPI	$\max[(\rho_n + 1 - \rho_n)/10]$ where $690 \leq n \geq 750$ (Haboudane et al. 2004)
Simple Ratio Index (narrowband)	SRI	$SRI = \frac{\rho_{800}}{\rho_{670}}$ (Haboudane et al. 2004)

*All 2 × 2 m remotely-sensed predictors aggregated at a 30 × 30 m (900 m^2) output resolution.
†Tree vs. understory returns defined as all returns above and below 1.37 m, respectively.
‡ρ_n; where n is wavelength in nm.

scales of plant richness (0.01, 0.1, 1, and 10 m^2) we aggregated remotely-sensed data at a 16 m^2 resolution - the finest derived pixel resolution capable of incorporating heterogeneity values for multiple 4 m^2 pixels. Owing to the sub-meter locational accuracy of pixels and subplots, aggregate pixels need not be at the identical spatial resolution as sampled subplots, provided they are fully (and centrally) contained. For assessing diversity at larger spatial scales, remotely-sensed predictors were aggregated at the scale corresponding with the plant richness response variable of interest: at 100 m^2 (25 pixels), 400 m^2 (100 pixels), and 900 m^2 (225 pixels) scales, respectively. 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 AASR plug-in in ArcMap version 10.4 (ESRI 2016).

Data analysis

Field plot training data and derived remotely-sensed variables were used to predict vascular plant species richness using random forest (RF) regression models at each of the seven spatial scales. RF is a nonparametric, machine-learning technique that is not subject to distributional assumptions and is relatively robust to multi-collinearity and over-fitting (Breiman 2001). All seven RF models were determined from an ensemble of 2,000 trees, each grown from a bootstrapped sample of predictors. For RF models at each of the seven spatial scales, feature selection of predictor variables was performed to select among all predictor variables the combination able to maximize predictive accuracy and generalizability outside of the constraints of training data. Specifically, model fit was assessed via 10-fold cross-validation, whereby random samples of the full dataset are iteratively withheld and split into ten quasi-independent training and validation subsets to test the fit (adjusted R^2, or "adjR^2") of predicted values vs. actual values. The feature selection procedure was initiated with the full model (all predictors), with the worst performing single variable based on importance value (or the residual sum of squares after splitting on each respective variable) withheld from each successive model. For any one spatial scale, the single model possessing the highest cross-validated accuracy (CV adjR^2) among all $p-1$ models, where p is the total number of features, was selected (Appendix S2). Feature-selected models were then used to predict vascular plant richness based on derived remotely-sensed variables for all pixels

throughout the study area. Predicted values were ultimately determined from the mean vote of all 2,000 out-of-bag (OOB) observations, rescaled to fit the range of plant richness observed in the field, while uncertainty was expressed as the coefficient of variation of all 2,000 OOB votes in the posterior distribution (Zhang and Lu 2012). All random forest models were run with the randomForest package (Liaw and Wiener 2002) in R, v. 3.3.1 (R Core Team 2016).

Feature-selected predictors in final RF models represent those variables that, through potentially complex and highly non-linear interactions, are critical to the prediction of plant richness in the study site. Despite their ability to produce high accuracy predictions, black-box nonparametric models are relatively uninformative for inference into the role of specific predictors (Prasad et al. 2006, De'ath 2007). Thus for this parallel, but ultimately separate, task of inferring ecological process from underlying pattern, we employed post-hoc tests of 15 key predictor variables deemed critical to nonparametric model performance by virtue of their common recurrence in final RF models after feature-selection (Evans et al. 2011). Specifically, we performed a series of univariate generalized linear models (GLMs) to compare slope parameters (β_1) across spatial scales to assess how spatial scale affects the direction and magnitude of key predictors of plant richness. Sample sizes for models at each spatial scale follow from the number of plots/subplots assessed at each scale respectively (e.g. $n = 288$ at 0.01–10 m^2 scales, $n = 144$ at 100 m^2, and $n = 36$ at 400 and 900 m^2).

Owing to the fact that species counts were overdispersed, with variance decoupled from the mean, we modeled univariate relationships using negative binomial GLMs in the MASS package in R (Venables and Ripley 2002). At some spatial scales, models exhibited inflated zero counts (three scales) and spatial autocorrelation (four scales). For these scales, zero-inflated negative binomial and spatial negative binomial GLMs were likewise assessed and superimposed on negative binomial model results (see Appendices S3 and S4). However, to maintain model consistency across scales, the most parsimonious model, the negative binomial GLM, was selected as our final model. To ensure model convergence, covariates were standardized (subtracting the mean and dividing by the standard deviation) prior to running GLMs. With the exception of hyperspectral indices, whose precise values are best interpreted as relational, model coefficients were then converted back from standard deviation units to ensure interpretation of parameters at the original, biologically-meaningful scale (Zhu et al. 2015).

Results

In total, 208 distinct vascular plant taxa were identified in field plots (Appendix S5), with plant richness ranging from 0 at the smallest subplot scales (0.01–10 m^2) to 100 species at 900 m^2 (Table 2). The study site exhibited a wide range of successional-structural conditions, with diameter at breast height ranging from the minimum recorded 0.1 cm to 101.5 cm (mean = 7.6 cm; standard deviation = 11.7 cm). Trained with field plot and derived remotely-sensed data, RF regression models predicted vascular plant species richness with greater accuracy at increasing spatial scales, from the ten-predictor 0.01 m^2 model (CV adjR^2 = 0.19) to the five-predictor 900 m^2 model (CV adjR^2 = 0.70) (Table 2; Fig. 2). Final, feature-selected models reflect landscape patterns

Table 2. Plant richness: plot summaries and predictive model results.

Scale (m^2)	SR (mean)	SR (sd)	SR (range)	CV adjR^2	RMSE	Feature-selected predictor variables*
0.01	0.3	0.6	0–4	0.19	0.26	elev (mean), all_sd (mean), PCA3 (sd), TWI (2 m min), CHM (max), ARI (sd), PCA3 (max), PCA3 (mean), CRI1 (sd), REPI (mean)
0.1	1.6	1.5	0–9	0.15	0.62	elev (mean), PCA3 (mean), understory_returns (sd), all returns (sd), CHM (max), PCA1 (max), TWI (8 m min), TWI (4 m mean), ARI (sd),
1	4.8	3.4	0–18	0.29	1.40	elev (mean), PCA1 (min), ASR, SRI (max)
10	13.5	7.5	0–35	0.40	2.81	elev (mean), NDVI (max), TPI (222 m), PCA1 (min), SRI (max)
100	34.4	10.2	15–63	0.69	2.27	PCA2 (mean), SRI (mean), elev (mean), NDVI (mean), all_returns (skew), PCA2 (max), DEV (222 m), TPI (222 m), PCA2 (min), NDVI (min), ASR, SRI (min), CRI1 (min), DEV (162 m), CRI1 (mean), understory_returns (sd), TWI (8 m max), TWI (4 m mean), TWI (2 m min),
400	53.2	13.8	28–84	0.58	3.58	NDVI (mean), SRI (mean), NDVI (min), SRI (min), DEV (222 m), PCA2 (mean), elev, CRI1 (max), CHM (sd), PCA2 (min), CRI1 (mean), DEV (162 m), CRI2 (min), all_returns (sd), understory_returns (sd), TPI (22 m), NDVI (max), TWI (8 m max)
900	61.4	15.6	35–100	0.70	4.23	PCA2 (mean), PCA2 (min), DEV (222 m), elev (mean), CHM (sd)

*See Table 1 for abbreviations.

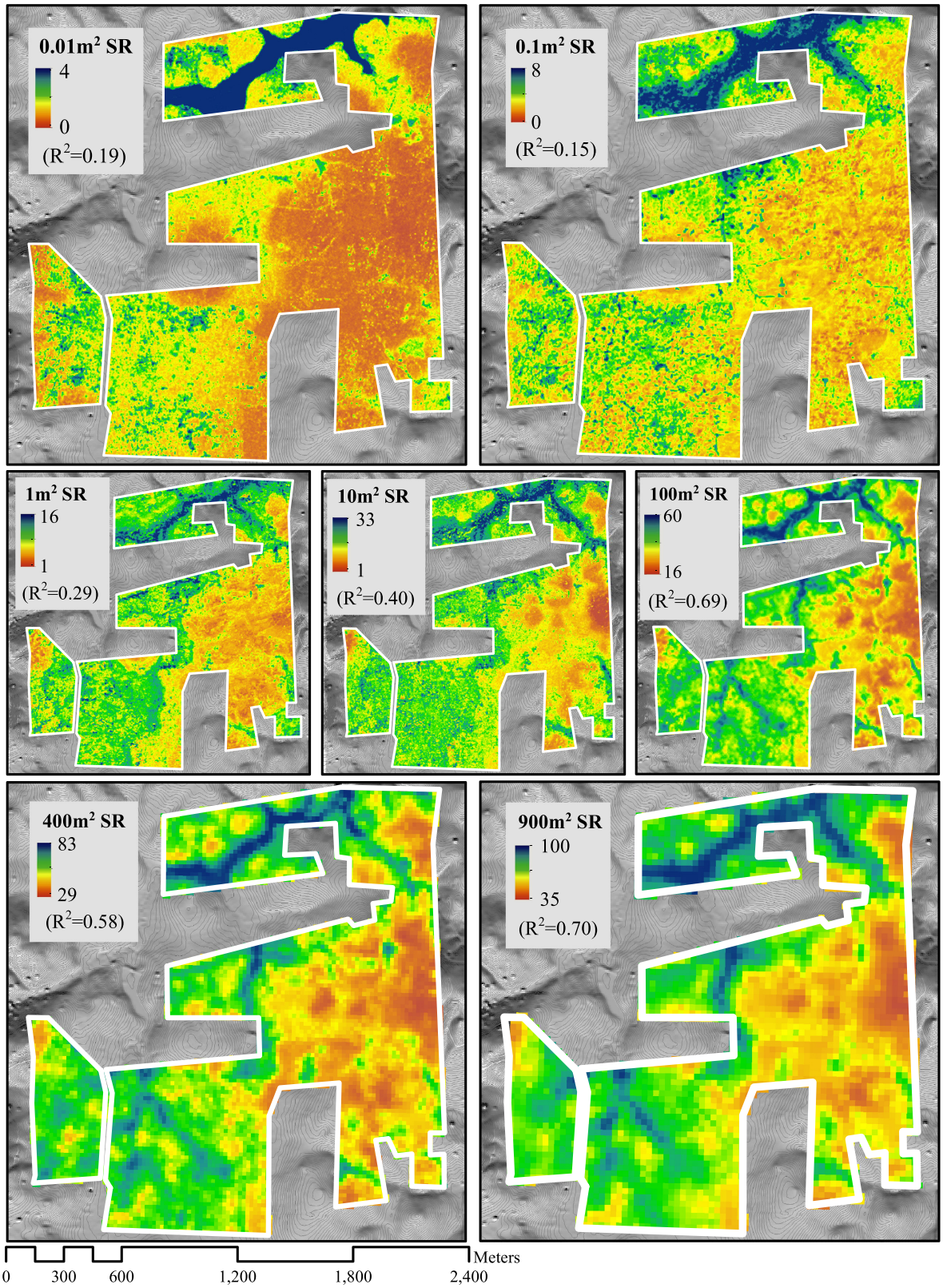

FIG. 2. Vascular plant species richness maps at seven spatial scales. Maps reflect predicted vascular plant species richness (SR) from random forest regression models trained with derived remotely-sensed layers and spatially-nested field plots. Model fit is based on 10-fold cross-validated adjusted R^2.

in plant diversity in relation to remotely-sensed environmental gradients including topo-physiognomy from LiDAR ground returns, canopy height from LiDAR first returns, biophysical structure from all LiDAR returns, as well as canopy biochemistry and foliar morphology from image spectroscopy (Table 1). Pixel-level uncertainty maps, on the other hand, highlight errors in data, model, and geo-locational alignment, as well as spatially-structured variance unexplained in statistical models that highlight those regions in the landscape where added sampling effort or additional covariates could most improve model performance (Appendix S6).

LiDAR-derived estimates of topo-physiognomy were consistently selected as key predictors in all predictive models (Table 2). Among them, elevation and TPI, a measure of hillslope concavity, exhibited a consistently significant relationship with richness across all spatial scales, while TWI, a proxy for soil moisture, was significant at scales between 100 m^2 - 900 m^2 (Fig. 3a–c). LiDAR estimates of canopy height and structure were likewise prominent among remotely-sensed predictors (Table 2). Maximum canopy height was significantly (positively) correlated with plant richness at the 900 m^2 scale, while the minimum and standard deviation of the canopy height model (CHM) were significant predictors of plant richness at all scales above 10 m^2 (Fig. 3d–f). Among selected LiDAR proxies for canopy structure, the skewness of all returns and the mean understory height both suggest a significant (positive) relationship between lower canopy growth and plant richness that manifested most significantly at spatial scales larger than 10 m^2 (Fig. 3h, i). The standard deviation of all returns,

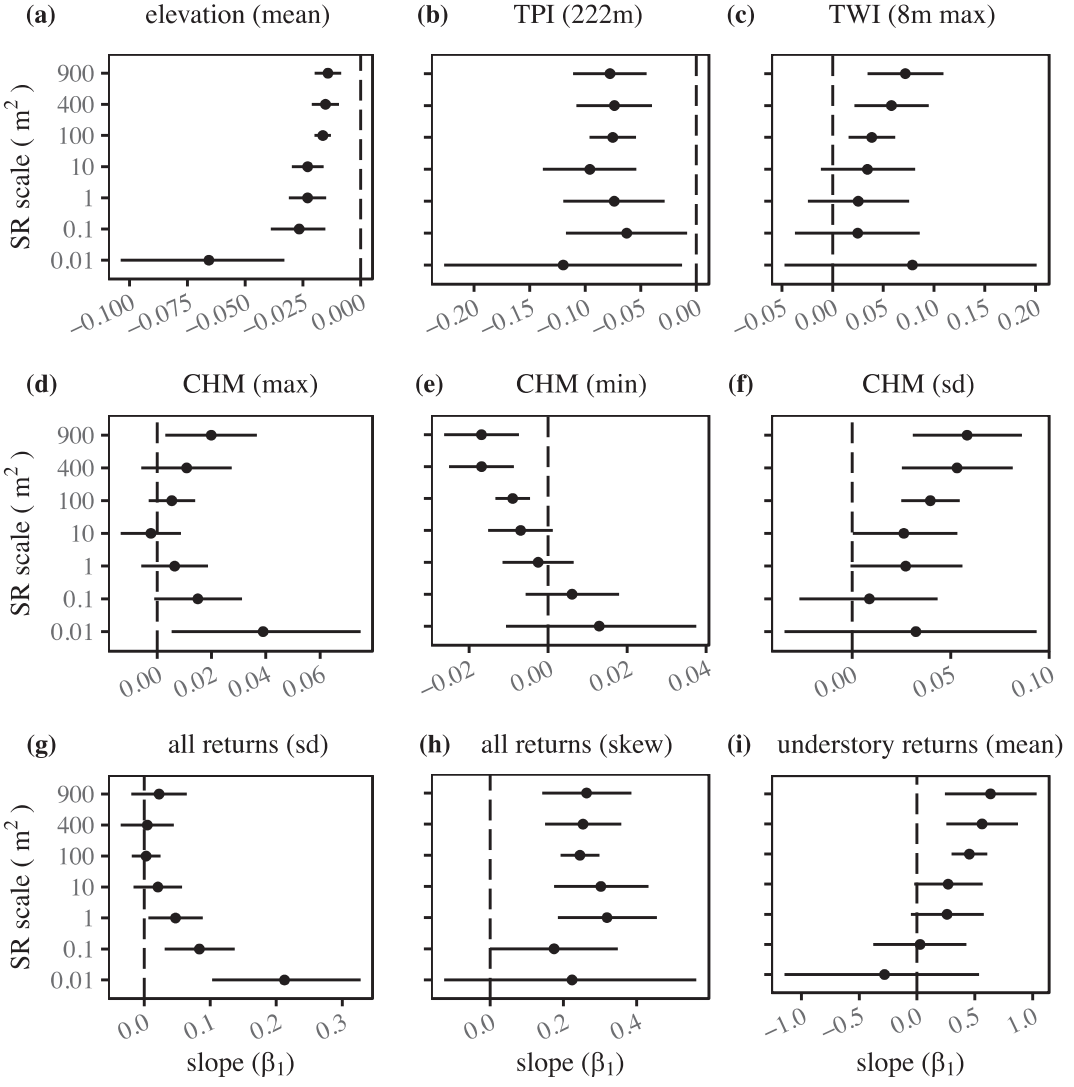

FIG. 3. Slope coefficients of derived LiDAR predictors vs. plant species richness across spatial scales. Estimated slope parameters (points) and 95% confidence intervals (bars) derived from univariate negative binomial GLMs of plant species richness across seven spatial scales. Predictor variables include LiDAR-derived measures of topography (a–c), canopy height (d–f), and canopy complexity (g–i). (TPI, topographic position index; TWI, topographic wetness index; CHM, canopy height model).

on the other hand, was not a significant predictor of plant richness across scales, especially after factoring zero-inflation and spatial autocorrelation (Fig. 3g; Appendices S3 and S4).

Derived variables from hyperspectral imagery, including the mean and variance of full spectrum principal components and narrowband vegetation indices, were singled out in the feature-selection process at all seven scales. The slope coefficients of aggregated means like PCA2 exhibited a similar profile to those of narrowband vegetation indices like CRI1 and NDVI, growing increasingly negative at larger spatial scales (Fig. 4a–c). Variance metrics, on the other hand, which measure spectral heterogeneity within a focal window, were increasingly positively correlated with plant richness at larger spatial scales (Fig. 4d–f).

Discussion

Mapping vascular plant species richness across spatial scales

Multi-scale diversity maps provide a realistic representation of how landscape patterns in plant diversity vary with scale in relation to remotely-sensed environmental gradients like topography, canopy structure, and spectral variance. Among the remotely-derived environmental gradients used to predict plant richness, LiDAR-derived topography was the most consistent predictor across spatial scales (Fig. 2). At the largest scales, topography is most conspicuous as an elevational gradient segregating upland forests in the extreme eastern and western portions of the study site with the bottomlands in the center, while at the smallest spatial scales, the imprint of topography manifests in the discrete boundaries of riparian areas within the bounds of incised stream beds in the north of the study area (Fig. 2; and see TWI in Fig. 1 upper right). At intermediate scales, predictors related to canopy composition and structure are most prominent, as evidenced by textured variation in predictive maps that reflect spatial variation in the CHM (Fig. 1 lower left).

Ultimately, the predictive models used in this study are correlative and not mechanistic. Given technological limitations in detecting individual plants in the sub-canopy and at subpixel scales, correlative models predict by proxy - in this case, reflecting the net outcome of interactions among feature-selected predictors in the nonparametric model. Rather than being guided by ecological intuition per se, the model selection protocol was driven by an agnostic variable selection procedure designed to maximize explained variance when generalized to quasi-independent validation data. If models are explanatory, it is because they are sufficiently accurate and generalizable (Houlahan et al. 2016). In this way, we distinguish between prediction of pattern, where accuracy in the

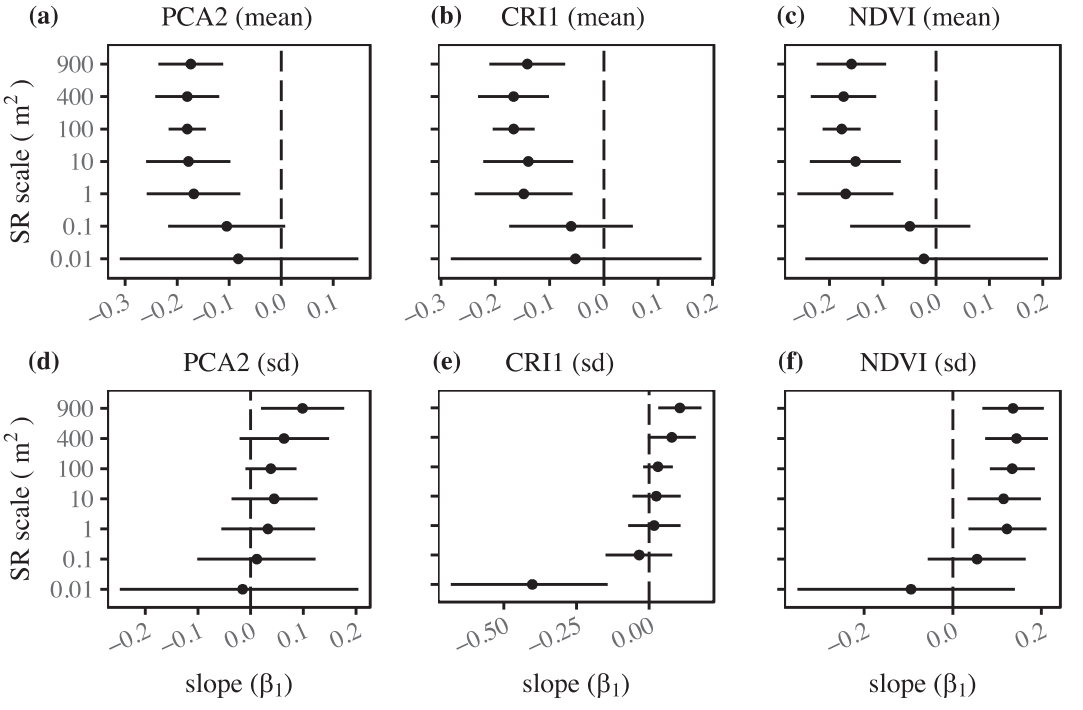

Fig. 4. Slope coefficients of derived hyperspectral predictors vs. plant species richness across spatial scales. Estimated slope parameters (points) and 95% confidence intervals (bars) from univariate negative binomial GLMs of plant species richness across seven spatial scales. Predictor variables are multi-pixel aggregates of hyperspectral imagery, including full spectrum principal components (a, d), as well as narrowband vegetation indices (b, c, e, f). (PCA, principal component axis; CRI1, Carotenoid Reflectance Index 1; NDVI, Normalized Difference Vegetation Index).

nonparametric model is the ultimate concern, and inference of process, for which parametric tests of specific predictors are best suited (Lo et al. 2015).

In general, cross-validated model accuracies of predictive models compare favorably with those in similar studies, mainly due to the added value of including disparate environmental data domains enabled by combined LiDAR-hyperspectral datasets. 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. Fricker et al. (2015) accounted for 35% (adj$R^2 = 0.35$) in tree species richness, and up to 52% (adj$R^2 = 0.52$) for understory trees and shrubs alone using high resolution optical imagery fused with LiDAR on Barro Colorado Island, Panama. However, caution is advised with direct comparison to other studies, especially when they differ widely in the specific response variable (e.g. all plant richness vs. woody species only), and statistics used to quantify model fit or predictive power. That predictive models explain an increasing proportion of variance in plant richness at larger spatial scales reflects a shift towards optimal convergence of model constraints (such as spatial resolution and precision of remotely-sensed data), and ecological factors of the study site such as characteristic stand structure (e.g. plant traits, size, and density) in combination with local resource dynamics (e.g. gap size, insolation, and edaphic heterogeneity).

Remotely-sensed predictors of vascular plant richness

Topography - LiDAR last returns.—The preponderance of terrain variables derived from LiDAR ground returns among all feature-selected variables indirectly confirms the predominant role of topography in driving landscape patterns of plant diversity in the North Carolina Piedmont. Elevation and TPI, both proxies for soil moisture and nutrients along hillslope catenal sequences, were significant predictors of plant richness across all spatial scales, such that higher plant richness totals are expected at lower elevations and within concave topographic features (Fig. 3a, b). These results reinforce those from other studies where remotely-derived abiotic factors related to topographic and edaphic properties were found to be significant predictors of plant species richness (Wolf et al. 2012, Fricker et al. 2015, Zellweger et al. 2016). Similar patterns have likewise been observed in empirical studies in Piedmont forests, where bottomland and riparian communities tend towards higher plant species richness values compared with their upland counterparts (Matthews et al. 2011, Peet et al. 2014).

Canopy height - LiDAR first returns.—Canopy height models (CHM) are derived products that interpret the three-dimensional coordinates of LiDAR first returns as a continuous surface representing the uppermost canopy. Though first returns 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 (Kane et al. 2010). In this study, maximum canopy height was significantly correlated with plant richness at the 900 m^2 scale – at a scale large enough to adequately detect stand height without being confounded by height variance (and local minima) within the extent of the aggregated pixel (Fig. 3d; Appendix S3). This correlational relationship between stand height and plant diversity likely reflects stand age – plant diversity relationships observed in Piedmont forests (Peet and Christensen 1988), as well as more proximate drivers of richness, such as environmental conditions and productivity (Marks et al. 2016).

Unlike canopy height, the minimum and standard deviation of the canopy height model indicates trait diversity in tree crown size and shape as well as the presence and frequency of canopy gaps – each factors found to influence diversity levels in temperate forests (Franklin 1988, Peet and Christensen 1988, Canham et al. 1994). At scales above 10 m^2, lower canopy minimums (e.g. large gaps) and greater height variance (e.g. structural heterogeneity) are each significantly correlated with increasing plant richness (Fig. 3e, f). At these larger scales, aggregated values sufficiently encompass variance in plant size and density, and by extension, heterogeneity in the resource environment that can affect understory recruitment and diversity levels (Canham et al. 1994). At the same time, aggregate predictors at larger spatial scales subsume geolocational misalignment between the vertical projection of fine-resolution LiDAR-determined canopy gaps and the ground location of gap regrowth due to interacting factors like gap shape and sun illumination angles (Kane et al. 2010).

Structural heterogeneity - LiDAR all returns.—LiDAR returns offer relatively precise estimates of canopy and sub-canopy structural elements (Dubayah and Drake 2000). These atemporal (structural) signatures reflect the net outcome of temporal (successional) processes including biotic competition, environmental conditions, and stochastic events (Franklin 1988, Peet and Christensen 1988). Measures of forest structural heterogeneity, such as variance in tree heights and the distribution of biomass among vertical strata, have been found to be among the more promising predictors of overall tree species diversity in Piedmont forests (Hakkenberg et al. 2016) and in temperate forests more generally (Chiarucci and Bonini 2005, Wolf et al. 2012). Our results corroborate these findings and roughly conform to the expectation of a correlative relationship between biodiversity and structural complexity described by the "diversity begets diversity" hypothesis (MacArthur and MacArthur 1961, Palmer and Maurer 1997). According to this line of thought, 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 (Canham et al. 1994). Concurrently, an alternative and equally plausible explanation for this phenomenon is that structural complexity is the result, rather than cause, of increased variance in plant traits like growth form, size, and crown shape in communities with higher species richness totals.

Unlike the CHM, which is restricted to the uppermost canopy surface, the distribution of all LiDAR returns characterizes the vertical distribution of biomass throughout the entire canopy. Although the relationship between the standard deviation of LiDAR returns and plant richness was largely inconclusive, the skewness of all returns was a consistently significant predictor of plant richness at scales above 0.1 m^2 (Fig. 3g, h). 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) vs. the upper canopy (negative skewness). That LiDAR skew was significantly (positively) correlated with plant richness indicates that, in the study area, top-heavy canopies (e.g. the single layer canopies of younger, single-cohort stands) tend towards lower diversity – a result corroborated by other studies in the NC Piedmont (Peet and Christensen 1988). Stands with more bottom-heavy distributions of ground herbs and woody shrubs, on the other hand, tend towards higher richness levels (Matthews et al. 2011). The positive correlation between understory return height and plant richness adds support to this conclusion, in part attributable to sampling effect and a statistical expectation of higher richness totals among the highly-abundant stems in the smaller size classes of understory vegetation (Knox et al. 1989). That the strength of the correlation between plant richness and LiDAR skewness or understory height increases in magnitude at larger spatial scales reaffirms the utility of modeling at ground extents sufficiently large enough to encompass variance in plant size and density (Canham et al. 1994).

Spectral reflectance - hyperspectral imagery.—Hyperspectral imagery excels in the remote detection of foliar biochemicals, water content, and leaf morphology, and when coupled with field and lab-measured spectral libraries can guide species identification of exposed canopy crowns (Curran 1989, Asner and Martin 2009). However, in structurally heterogeneous forests, the contingencies of interspersed taxa, intra-specific trait variation, as well as varying environmental and illumination conditions may hinder species identification (Price 1994, Ollinger 2011). In these cases, aggregate hyperspectral data representing subpixel means and variance, can be quite valuable for predicting large-scale patterns in the spatial distribution of biodiversity (Rocchini et al. 2015). While subpixel means suffer from information loss by filtering out subpixel variation, they can be highly effective in detecting large-scale reflectance patterns due to landscape gradients in compositional and structural attributes. Spectral variance, on the other hand, specifically utilizes spectral entropy as a proxy for foliar biochemical diversity, and by extension, higher relative species richness.

Although the sign of a PCA scores is arbitrary and relative to any particular loading, those for narrowband vegetation indices like NDVI and CRI1 are not, and reflect absolute change in foliar chlorophyll and carotenoid levels, respectively. Interestingly, because vegetation indices tend to saturate at low levels of canopy density, our finding that mean vegetation index values were negatively correlated with plant richness implies that in areas of continuous canopy cover, it can be the absence, rather than presence, of foliar pigmentation that best explains diversity patterns (Haboudane et al. 2004). In this case, lower vegetation index values indicate a larger proportion of canopy shadow as a component of the entire remotely-sensed signal, and thus a high degree of mutual leaf shading in more structurally-complex canopies (Asner 1998). Provided structural complexity is positively correlated with richness (see Fig. 3d–h), remotely-sensed estimates of canopy complexity can be useful for predicting plant diversity (White et al. 2010).

Metrics of spectral entropy quantify total information content in an aggregate pixel, reducing subpixel variance to a single variable reflective of canopy biochemical diversity that is agnostic to the identity of individual taxa (Rocchini et al. 2015). In this regard, the spectral variation hypothesis (SVH) posits a positive relationship between spectral variation due to foliar biochemical diversity and phylogenetically-conserved trait variation among taxa that can aid remotely-sensed biodiversity modeling efforts (Palmer et al. 2002, Cavender-Bares et al. 2016). Our results generally support this conclusion, though only at pixel resolutions coarse enough to adequately incorporate sub-pixel heterogeneity and at ground scales large enough to subsume local diversity (Rocchini 2007, Schmidtlein and Fassnacht 2017). These findings likewise suggest the role of phenological status in determining the strength of the SVH. That inter-specific variance in chlorophyll and carotenoid reflectance was enhanced due to partial leaf senescence in the mid-Fall imagery date supports the potential of seasonally strategic retrievals to exploit plant phenological status in temperate broadleaf forests (Gitelson et al. 2002, Haboudane et al. 2004). On this point, it is important to note that with the majority of total vascular plant richness lying below the canopy surface, this relationship between spectral variation and species richness only exists because canopy diversity was concurrently highly correlated with total vascular plant diversity in the study area (Appendix S7).

Conclusion

This study employs airborne LiDAR-hyperspectral data to predict vascular plant species richness at seven spatial scales in a structurally- and compositionally-complex

NC Piedmont forest landscape. Multi-scale predictive mapping of this sort accounts for the role of spatial scale in constraining landscape patterns in plant richness and species-area relationships, as well as the remotely-sensible essential biodiversity variables (EBVs) used to predict it. Our findings identify LiDAR-derived estimates of topography and forest structural complexity as well as spectral variance from hyperspectral imagery as particularly effective EBVs to parsimoniously describe and predict landscape variation in vascular plant species richness in closed canopy, multi-cohort forest stands. While the finding that remotely-derived EBVs like topography, forest structure, and spectral variance are significant predictors of vascular plant species richness should best be interpreted in the context of the extent of the Piedmont NC study site, that these results are corroborated, to various extents, by other studies in disparate environments invites further study into the limits and potential of this set of EBVs for cross-biome transferability.

Importantly, our results likewise suggest scale dependence among predictors of plant richness, such that predictive models are optimized at those scales that best ensure coherence in the spatial attributes of field plot data (sampling size, design, and intensity), remotely-sensed data (spatial resolution and geolocational precision), species diversity patterns (species-area relationships), and target biota (size, density, and coverage extent). However, coherence in scale need not mean equivalence in grain or resolution. Predictive models can effectively predict diversity at spatial scales finer than the nominal resolution of remotely-sensed imagery, or conversely, with high accuracy at coarser resolutions by exploiting the increased informational content of aggregated EBVs like spectral variance. With the recent and upcoming deployment of a series of high precision LiDAR and hyperspectral airborne and space-borne sensors, remotely-sensed EBVs will increasingly enable ecologists to test hypotheses regarding patterns and drivers of floristic diversity from plot, to landscape, to global scales.

Acknowledgments

This work was supported under NASA Earth and Space Science Fellowship Grant NNX14AK92H and a grant from the Center for the Study of the American South. In addition, we would like to thank Bruce D. Cook for provision of G-LiHT data, the Duke Forest field office, Michael Lee for assistance with data entry and management protocols, Samantha Tessel and Bianca Lopez for help with taxonomic identification, as well as field assistants: Jeffery Berger, Joshua Bradley, Per Frisk, Alec Keiper, Nicole Murnighan, Matthew Seibert, and Stephanie Trillo-ordonez. Finally, we would also like to thank the subject matter editor and two anonymous reviewers for providing insightful suggestions that greatly improved the quality of this article.

Literature Cited

Anderson-Teixeira, K. J., et al. 2015. CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. Global Change Biology 21:528–549.

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

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

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

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

Canham, C. D., A. C. Finzi, S. W. Pacala, and D. H. Burbank. 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.

Cavender-Bares, J., et al. 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., J. M. R. Benayas, A. Justel, and J. Salas-Rey. 2006. Modelling tree diversity in a highly fragmented tropical montane landscape. Global Ecology and Biogeography 15: 602–613.

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

Civitello, D. J., et al. 2015. Biodiversity inhibits parasites: broad evidence for the dilution effect. PNAS 112:8667–8671.

Condit, R. 1995. Research in large, long-term tropical forest plots.

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

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

Crawley, M. J., and J. E. Harral. 2001. Scale dependence in plant biodiversity. Science 291:864–868.

Curran, P. J. 1989. Remote sensing of foliar chemistry.

De Reu, J., et al. 2013. Application of the topographic position index to heterogeneous landscapes. Geomorphology 186:39–49.

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

Dubayah, R. O., and J. B. Drake. 2000. Lidar remote sensing for forestry applications. Journal of Forestry 98:44–46.

ESRI. 2016. ArcGIS Desktop: Release 10.4. Environmental Systems Research Institute, Redlands, CA.

Evans, J. S. 2011. Modeling species distribution and change using random forest. Pages 139–159 *in* C. A. Drew, Y. F. Wiersma, and F. Huettmann, editors. Predictive species and habitat modeling in landscape ecology: concepts and applications. Springer, New York, New York, USA.

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

Féret, J. B., and G. P. Asner. 2014. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. Ecological Applications 24:1289–1296.

Franklin, J. 1988. Structural and functional diversity in temperate forests. Pages 166–175 *in* E. O. Wilson, editor. Biodiversity. National Academy Press, Washington, D.C.

Fricker, G. A., J. A. Wolf, S. S. Saatchi, and T. W. Gillespie. 2015. Predicting spatial variations of tree species richness in tropical forests from high resolution remote sensing. Ecological Applications 25:1776–1789.

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

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

Haboudane, D., J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan. 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., C. Song, R. K. Peet, and P. S. White. 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., R. K. Peet, D. L. Urban, and C. Song. 2018. Modeling plant composition as community-continua in a forest landscape with LiDAR and hyperspectral remote sensing. Ecological Applications 28:177–190.

Hooper, D. U., E. C. Adair, B. J. Cardinale, J. E. K. Byrnes, B. A Hungate, K. L. Matulich, A. Gonzalez, J. E. Duffy, L. Gamfeldt, and M. I. O'Connor. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature 486:105–108.

Houlahan, J. E., S. T. Mckinney, T. M. Anderson, and B. J. Mcgill. 2016. The priority of prediction in ecological understanding. Oikos 126:1–7.

Kampe, T., B. R. Johnson, M. Kuester, and M. Keller. 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., R. J. McGaughey, J. D. Bakker, R. F. Gersonde, J. A. Lutz, and J. F. Franklin. 2010. Comparisons between field- and LiDAR-based measures of stand structural complexity. Canadian Journal of Forest Research 40:761–773.

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

Levin, S. A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. Ecology 73:1943–1967.

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

Lo, A., H. Chernoff, T. Zheng, and S. H. Lo. 2015. Why significant variables aren't automatically good predictors. Proceedings of the National Academy of Sciences 112:13892–13897.

MacArthur, R., and J. MacArthur. 1961. On bird species diversity. Ecology 42:594–598.

Marks, C. O., H. C. Muller-Landau, and D. Tilman. 2016. Tree diversity, tree height and environmental harshness in eastern and western North America. Ecology Letters 19:743–751.

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

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

NC OneMap Geospatial Portal. 2016. NC OneMap Geospatial Portal. North Carolina Department of Information Technology, Data Division, Center for Geographic Information and Analysis.

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

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

Palmer, M. W., P. Earls, B. W. Hoagland, P. S. White, and T. Wohlgemuth. 2002. Quantitative tools for perfecting species lists. Environmetrics 13:121–37.

Peet, R. K., and N. L. Christensen. 1988. Changes in species diversity during secondary forest succession on the North Carolina piedmont. Pages 233–245 *in* J. J. During, M. J. A. Werger, and J. H. Willems, editors. Diversity and Pattern in Plant Communities. SPB Academic Publishing, The Hague, The Netherlands.

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

Peet, R. K., M. T. Lee, M. F. Boyle, T. R. Wentworth, M. P. Schafale, and A. S. Weakley. 2012*a*. Vegetation plot database of the Carolina Vegetation Survey database. Biodiversity and Ecology 4:243–253.

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

Peet, R. K., K. A. Palmquist, and S. M. Tessel. 2014. Herbaceous layer species richness of southeastern forests and woodlands: patterns and causes. Pages 255–276 *in* M. F. S. Gilliam, and M. R. Roberts, editors. The Herbaceous Layer in Forests of Eastern North America, Chapter 10. Second edition. Oxford University Press, Oxford, UK.

Pettorelli, N., et al. 2016. Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. Remote Sensing in Ecology and Conservation 2:122–131.

Prasad, A. M., L. R. Iverson, and A. Liaw. 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.

Pu, R., and S. Landry. 2012. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. Remote Sensing of Environment 124:516–533.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rocchini, D. 2007. Effects of spatial and spectral resolution in estimating ecosystem alpha-diversity by satellite imagery. Remote Sensing of Environment 111:423–434.

Rocchini, D., J. L. Hernández-Stefanoni, and K. S. He. 2015. Advancing species diversity estimate by remotely sensed proxies: a conceptual review. Ecological Informatics 25: 22–28.

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

Schmidtlein, S., and F. E. Fassnacht. 2017. The spectral variability hypothesis does not hold across landscapes. Remote Sensing of Environment 192:114–125.

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

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

Townshend, J. R., et al. 2012. Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. International Journal of Digital Earth 5:373–397.

Ustin, S. L., A. A. Gitelson, S. Jacquemoud, M. Schaepman, G. P. Asner, J. A. Gamon, and P. Zarco-Tejada. 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sensing of Environment 113:S67–S77.

Venables, W., and B. Ripley. 2002. Modern applied statistics with S.

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

White, J. C., C. Gómez, M. A. Wulder, and N. C. Coops. 2010. Characterizing temperate forest structural and spectral diversity with Hyperion EO-1 data. Remote Sensing of Environment 114:1576–1589.

Wolf, J., G. Fricker, V. Meyer, S. Hubbell, T. Gillespie, and S. Saatchi. 2012. Plant species richness is associated with canopy height and topography in a neotropical forest. Remote Sensing 4:4010–4021.

Yu, L., Y. Shi, and P. Gong. 2015. Land cover mapping and data availability in critical terrestrial ecoregions: a global perspective with Landsat thematic mapper and enhanced thematic mapper plus data. Biological Conservation 190: 34–42.

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

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

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

Zhu, K., C. W. Woodall, J. V. D. Monteiro, and J. S. Clark. 2015. Prevalence and strength of density-dependent tree recruitment. Ecology 96:2319–2327.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.2109/suppinfo