Rethinking the fundamental unit of ecological remote sensing: Estimating individual level plant traits at scale

Sergio Marconi, Sarah J. Graves, Ben. G. Weinstein, Stephanie Bohlman, Ethan P. White

Abstract

Functional ecology has increasingly focused on describing ecological communities based on their traits (measurable features of individuals that affect their fitness and performance). Analyzing trait distributions within and among forests could significantly improve understanding of community composition and ecosystem function. Historically, data on trait distributions in ecosystems are generated by (1) collecting a few leaves from a small number of trees, or (2) using remote sensing images to infer traits. While the first method suffers from limited, and potentially biased, sampling, it produces information at the level of the fundamental ecological unit: the individual. On the contrary, remote sensing produces information in a continuum space whose unit is usually the plot, containing dozens of trees of different species.

The National Ecological Observatory Network (NEON) provides data collected both from the field and 1m^2 resolution remote sensing. Using these data we developed a method to scale up functional traits measured on 160 trees to the full spatial extent of NEON sites (millions of crowns). The first step of our pipeline involved the use of image segmentation methods to infer crown geometry. Then, we built models to infer Leaf Mass Area (LMA), nitrogen, carbon, and phosphorus content. We developed a naive multiple instance regression method based on an ensemble of Partial Least Squares Generalized Linear regressions (pls-GLR). Results were applied both at the pixel scale (1 m^2) and aggregated to the crown scale for comparison. Finally, we applied the models to two NEON sites, Ordway Swisher Biological Station (OSBS) and Talladega (TALL) to estimate the location and leaf traits of approximately 5 million trees.

Nitrogen, LMA and phosphorus models had R^2 values ranging from 0.50 to 0.75 on held out test data, comparable with what has been observed in studies conducted at the plot scale. The ensemble resulted in better predictability than single pls-GLM models. Models for carbon performed poorly compared to other traits (R^2 of 0.22). Crown delineation produced the highest uncertainty in the pipeline, generally due to over-segmentation of crowns. The intersection over union between segmented and ground-truth crowns ranged between 0.24% and 0.78%. Despite this uncertainty, crown scale predictions performed significantly better than pixel predictions, resulting in improvement of R^2 on test data of between 0.07 to 0.20 points.

We used the resulting models to produce individual trait data for ~2.5 million individual crowns for both OSBS and TALL, covering a total extent of ~360 km^2. In these “derived-data” LMA, %N, and %P were highly correlated as expected. Chemical traits were highly correlated to some physical traits like LAI, and to elevation. Structural traits like crown area and height were correlated to chemical traits only in OSBS.

Introduction

Functional traits are biochemical, physiological and structural characters that influence organism performance or fitness (Nock et al., 2016). They are central to how organisms perform under different environmental conditions, interact with other species, and influence the ecosystems to which they belong (McGill 2006, Dwyer et al., 2017). For individual organisms, traits influence core demographic parameters including survival and reproduction. At the species level, traits can influence species distributions and how species respond to changes in land use and climate. The traits of the organisms influence ecosystem biogeochemical cycles and habitat availability for other species (e.g., Fisichelli et al., 2015). Given their central importance across multiple levels of organization, understanding how traits vary within and among species, across environmental gradients, and through time is essential to understanding many areas of ecology and predicting how ecological systems will change in the future (McGill 2006, Lawler et al. 2010, Valladares et al., 2014).

In plants, the primary traits fall into two categories: chemical (e.g. nitrogen content, leaf mass per area, carotenoid, Fe, Zn…) and structural (height, canopy size). Historically, studying these plant functional traits has relied on collecting samples of a few leaves from a small number of trees to estimate the average trait values for that species. These average values are then used to explore how the ecosystem level traits vary biogeographically or through time by assuming that all individuals of a species in a region share the same trait value. This approach is necessary because it is expensive and time consuming to collect individual level trait data, but it ignores the fact that these traits can vary within species based on both gradual evolution and rapid plastic responses to the conditions an individual or population experiences (Callaway et al. 2003, Muller et al. 2010, Nicotra et al. 2010, Albert et al., 2010). These issues are magnified when studying changing environments (across space or time) because of bias in where the data for each species is collected. Data are typically collected in small subsets of the full range of conditions that species experience. Another issue is the biased choice of representative individual trees. Depending on the purpose of the original studies, sample design can be biased towards individuals of a particular health status, size or species. This means that the data may focus on environments with limited variation, or not accurately represent the broader landscape. As a result, it has been suggested that trait data contains less information than species data on the relationship between plant communities and their environment (Clark, 2016). Measuring traits systematically across geographic gradients would address this limitation, but this is challenging with traditional methods (Anderson-Teixeira 2015).

An alternative approach that allows continuous estimation of traits across the landscape is to use remote sensing data (Homolova et al. 2013, Houborg et al., 2015, Kerr & Ostrovsky, 2003) . Hyperspectral remote sensing imagery can be used to estimate chemical traits of sunlit leaves, and laser imaging, detection, and ranging (LiDAR) can infer vertical and horizontal vegetation structure. Together, these two data sources can be used to estimate many of the standard chemical and structural functional traits for trees (Serbin et al., 2014, Singh et al., 2015, Asner et al., 2017, Barbosa et al., 2017). However, remote sensing applications generally use either the pixel (the smallest resolution component of the image) or the plot (a region of space typically containing multiple individuals) as the fundamental unit. This is a natural result of the structure

of the remote sensing data combined with the difficulty of linking individual crowns between remote sensing and field data. However, it results in a disconnect with the fundamental biological unit: the individual (Liu et al., 2016, De Angelis, 2018, Marconi et al. 2018). Individuals reproduce, interact with their neighbors, and exhibit plastic responses to environmental conditions. Populations of individuals evolve in response to natural selection. As a result, our understanding of many biological processes is grounded in the individual and many field based survey methods focus on collecting data with individual trees as the primary unit of measurement. While the plot is a scale widely used in forestry (Newnham et al, 2015), it loses important information about the relative distribution of individuals and their traits within the plot and hence most of the embedded information about community structure. To fully understand how traits vary across space, change through time, and are determined by biological processes, it is important to develop approaches linking these fundamental characteristics to individual trees that can be applied at large scales.

Despite its importance for biological research, few studies have attempted to model leaf traits using remote sensing at the crown level (but see Chadwick & Asner 2016, Martin et al., 2018). Even when crown level models are developed, the resulting leaf trait predictions are made for pixels, not crowns, because of the challenges in aligning LiDAR and hyperspectral data. In contrast, studies modeling structural traits (like crown height and area) do generally make crown level predictions. Perhaps as a result of this difference, structural and chemical leaf traits are not predicted together at large scales. Consequently, ecology lacks the large scale individual level trait estimates that are necessary to fully understand relationships between traits, species, and the environment.

To address this gap, we developed a pipeline for making crown level trait predictions at scales of 500 km^2 with associated uncertainties on both crown delineation and trait estimation. Building on Chadwick & Asner (2016) and Martin et al. (2018), we: 1) identify individual crowns in remote sensing imagery that can be associated with field-based trait measurements; 2) build models relating the remote sensing measurements to field based traits; and 3) apply those models to estimate trait values and examine patterns of tree structural and chemical traits from individual to landscape scale. We advance the state of the art (Chadwick & Asner 2016, Martin et al. 2018) in all three steps in this pipeline. For identifying individual tree crowns, we adopted a digital field-mapping approach to accurately locate individual crowns in remote sensing images of closed canopy forests (Graves et al. 2018). The field-mapped individual tree crowns make it possible to calibrate, validate, and compare different crown delineation algorithms, and identify uncertainty in the first step of the pipeline. When modeling traits based on remote sensing data, we use new methods that allow uncertainty to be accurately estimated by producing likelihoods and assess whether crown-level modeling produces more accurate predictions than pixel level modeling. Most importantly we estimate traits at the crown level instead of at the pixel level. This combination of approaches allows us to produce and distribute derived data products on the location, size, shape, and leaf traits of millions of individual trees distributed over tens of thousands of hectares. Providing data as a downloadable data product will allow researchers to conduct individual level research at scales over 100 times greater than current field based datasets.

Methods

Predicting crown level leaf and structural traits from remote sensing requires: 1) field measurements of functional traits for building and evaluating models; 2) data on the shape and location of individual tree crowns (ITCs), for building accurate models and assessing uncertainty in crown delineation algorithms; and 3) high resolution remote sensing LiDAR (for crown delineation and estimation of structural traits) and hyperspectral data (for estimation of leaf chemical traits). To obtain these components, we built on the National Ecological Observatory Network's (NEON) pioneering infrastructure by combining NEON's LiDAR and hyperspectral data with field data we collected at NEON sites on traits and the location and shape of individual tree crowns.

Site descriptions

The study was conducted at two core terrestrial NEON sites; Ordway Swisher Biological Station in Florida (OSBS, NEON Domain 03) and the Oakmulgee Management District of Talladega National Forest in Alabama (TALL, NEON Doman 08). The two sites (Figure 1) have a mix of deciduous, evergreen, and mixed forest types (Homer et al., 2012). Upland areas at both sites are dominated by fire-tolerant oaks and pine species, primarily longleaf pine (*Pinus palustris*). Lowland areas near lakes or wetlands (OSBS), and riparian areas (TALL) are dominated by hardwood forests (Beckett and Golden 1982, Cox and Hart 2015).

Remote sensing data

All aerial remote sensing data products were provided by NEON Airborne Observation Platform (NEON-AOP) (Table 1). We used data from the May 2014 flight for OSBS, and the June 2015 for TALL. We used the raw L1 data products: (1) "classified LiDAR point cloud", and (2) "hyperspectral surface reflectance" data, orthorectified and atmospherically corrected. The LiDAR data consist of 3D spatial point coordinates (4-6 points/m^2) which provides high resolution information about crown shape and height. These data are released in 1 km x 1 km tiles. Hyperspectral data consist of 1m^2 spatial resolution images with 426 channels (or bands), each one collecting reflectance data in a 5 nm wide interval of wavelengths, ranging from visible to near infrared light (from 350 to 2500 nm). Hyperspectral data provide information about the chemical composition of sunlit leaves. These images were provided as multiple flight lines covering about 15 km x 0.8 km each, for a total area of about 215 km^2 in OSBS, and 145 km^2 in TALL. The hyperspectral images were provided as "prototype" data, pre-processed differently than post 2017 data, and delivered on hard drives (Table 1).

During this project, leaf traits data collected by the NEON Terrestrial Observation System (TOS) were not available. For this reason we collected leaf samples for 226 trees near NEON inventory plots, which are randomly located across the study site and stratified by land cover type. The sample of individual trees were located away from major roads (since roads can affect the spectral reflectance of the remote sensing data), had a crown that was visible from above and could be identified in the image, and had sunlit branches that were accessible for leaf

collection (Graves et al., 2018). Trees were not sampled within NEON plots to avoid disturbing NEON's long-term monitoring efforts.

Field Data

Field data were collected in the early part of the peak growing season in 2015 and 2016. In particular, 87 individual trees (for 16 species) were sampled from OSBS in May-June 2015, and 137 individuals (29 species) in July 2015 from TALL. Sunlit leaves were sampled with a shotgun. Immediately after collection, the leaves were placed in a labeled plastic bag and stored in a cooler. Leaf samples were further processed within 2-4 hours of collection. Samples were analyzed of C, N, and P concentrations and LMA according
to standard protocols (Murphy and Riley 1962; Cornelissen et al. 2003).
For LMA, a random sample of leaves was weighed and scanned on a flatbed scanner to determine leaf area. The leaves were then dried at 65 degrees C for at least 72 hours and reweighed. A separate sample of leaves was dried and used for measurements of nitrogen (%N), carbon (%C), and phosphorus (%P).

The year following leaf and airborne data collection, individual trees were mapped in the remote sensing images using a field tablet and GIS software. This process involved mapping individual tree crowns on the hyperspectral image directly in the field to ensure the sampled trees matched directly with image pixels (Graves et al., 2018). This individual tree crown (ITC) data is necessary for accurately linking field measurements with pixels from remote sensing spectral data and for quantifying uncertainty in crown the delineation algorithms.

The algorithm pipeline

We developed a modular pipeline based on three steps: (1) develop an ensemble of statistical models to infer leaf mass per area (LMA, g/m^2), nitrogen (%N), phosphorus (%P), and carbon (%C) at an individual tree scale from hyperspectral data; (2) build and evaluate crown delineations from LiDAR data; (3) scale predictions for each individual tree crown in each NEON site (extent of more than 200 km^2 per site). For each crown we also extracted values of height, elevation, slope, aspect and the LAI Determining Index (Delialieux 2008), provided as NEON AOP data products (Table 2). The structure of the pipeline presented in this paper is summarized in Figure 2.

Leaf chemistry model

We extracted all of the hyperspectral pixels from within each crown for sampled trees, whose crown boundaries were delineated in the field (ITC). Data were normalized and preprocessed to reduce noisy, highly mixed, or unvegetated pixels (Figure S.1). We filtered bad pixels by using both near infrared band (reflectance in 860 nm < 0.3) and Normalized Difference Vegetation Index (NDVI < 0.7) thresholds (Colgan et al., 2012, Graves et al., 2016). We did not use a height threshold (Asner et al., 2017), since height was indirectly used to select pixels belonging to each crown in the field mapping process. We normalized the spectral values for each pixel by

dividing each spectral vector by its root sum of squares. We used this method to reduce the effect of peripheral light and shadows within each crown (Singh et al., 2015).

Field data were split at the tree level, and stratified by species, into training (n = 168), validation (n = 42), and test sets (n = 15). Since the two sites have similar species composition, we aggregated the two datasets to build a joint model. As is common for trait studies, our field data on foliar traits was a single value for an individual tree. However, individual crowns contain multiple pixels, and crowns vary in the number and quality of these pixels. In each crown, some pixels are better than others for linking traits to hyperspectral signatures because some register light reflected primarily by leaves, others from branches, understory, or ground. To capture the range of possible models from different pixels in each crown, we randomly sampled one pixel from each training crown 1000 times, and used the resulting 1000 vectors of pixels (one single pixel for each crown in the dataset) to build 1000 independent partial least squares generalized linear regression (PLS-GLR) models (Bastien et al., 2005, Bertrand et al, 2014). Hereafter we we refer to these models as “Single Pixel Models” (SPM). We used PLS-GLR (despite the higher computational demand than regular PLS regressions) because it calculates components vectors by using maximum likelihood estimation which allows the calculation of AIC values to compare model goodness of fit, and provides a measure of uncertainty for each model prediction. We treated site (OSBS versus TALL) as a fixed effect to account for site specific ancillary conditions. The number of components included in each model was determined using 10-fold cross-validation (CV) using the PRESS statistic on the training set. We tested how different transformations of leaf spectra affected goodness of fit. Following Feilhauer et al. (2015), we used the log transformation of the first derivative of the spectra for Phosphorus and Carbon.

To pixel and crown based approaches to modeling leaf traits, we compared three modeling strategies: 1) A traditional pixel based approach with no information on tree crown identity. We selected the 100 best SPMs; 2) A pixel based approach that included information on crown identity that we refer to as the “Ensemble Pixel Based Model” (EPBM). This model is an ensemble of the 100 best SPM models with individual models weighted based on their performance on validation data using AIC based multi-model averaging (Burnham & Anderson 2002).; 3) A crown based ensemble model that focuses on the crown as the fundamental unit of remote sensing that we refer to as the “Crown Ensemble Aggregation Model” (CEAM). This model calculates an average trait value across all pixels in a crown using the EPBM calculated for each pixel. We built these models from two different pixels datasets: we used either (1) pixels belonging to ground delineated crowns (Graves et al., 2018), or (2) pixels belonging to algorithmically delineated crowns (Silva et al., 2016). We finally compared SPBs, EPBMs, and CEAMs for the two datasets to quantify how uncertainty in crown delineation would affect our ability to predict leaf traits.

All models were tested on each pixel of the 15 crowns withheld in the test dataset. The test data were not used at any phase of the fitting or the ensemble process. Accuracy was evaluated using the predictive coefficient of determination (R^2) and the root-mean-square error (RMSE). The coefficient of determination produces values between 1 and negative infinity, where negative values indicate that the model predictability is lower than the sample average. As such,

negative R^2 values indicate that the statistical model did not learn any meaningful information from the data. A value of 1 indicates that predicted values perfectly match observations. We evaluated the uncertainty of predictions for each hypotheses modeled using the coverage of the 95% prediction interval (95PI). The prediction interval is the range values that is expected to contain 95% of the observed data points, and therefore a model with good estimates of uncertainty should have approximately 95% of the test data falling in within this range. Since the CEAM was generated by the ensemble of the 100 best SPMs, we estimated the 95PI for CEAM predictions by averaging the error functions for the best 100 SPMs. We used the same data split, data transformation and PLS-GLR parameterization for all models. For pixel based estimations we compared ground measures of LMA, N, C, and P with predictions from each pixel in the test dataset. For crown based estimations, we averaged pixel based predictions belonging to all of the pixels in the crown. The same rationale was used for comparing pixel and crown based uncertainty.

Crown Segmentation

We used pre-classified point cloud LiDAR data to create a 0.5 m^2 resolution canopy height model (CHM). Building a higher resolution CHM than the one provided by NEON data products allowed us to better capture the canopy structure, and produce better predictions of individual tree crowns. We used the CHM to determine the number of trees in the scene, then applied three unsupervised crown segmentation algorithms (Dalponte & Coomes 2016, Silva et al., 2016, and a watershed algorithm as in Barnes et al., 2014)**.** These algorithms were chosen because they were fast enough to scale up to hundreds of km^2, and they performed well when applied on the same data in a similar context (Marconi et al, 2018). We used a single parametrization (Table S. 1) for each method at both sites to facilitate cross-model and cross-sites comparison. Data were processed using the lidR R package (Roussel & Auty, 2017). We used the mean pairwise Jaccard index coefficient (Real & Vargas, 1996) to choose the most suitable algorithm, and quantify individual tree crowns (ITCs) related uncertainty. The Jaccard index was calculated by comparing ITCs collected in the field with the single most overlapping predicted crown. Field delineated crowns not overlapping with any crown segmented by the algorithm were labelled as undetected. The Jaccard index was chosen to evaluate the segmentation accuracy because it is fully automatable, has good continuity, and can be used when only a subset of ITCs in the picture has been collected on the ground (Zhen et al., 2016, Marconi et al. 2018). The Silva et al. (2016) segmentation algorithm was determined as the best method for our study (Table S1, see results section), consistent with results from a multi-group data science competition (Marconi et al. 2018). Therefore, this algorithm was used to extract ITCs from the CHM data for the full remote sensing foot print for both NEON sites.

Building individual-level derived data for full flight paths

Each hyperspectral image was subset into single 1 km^2 tiles. We pre-processed each tile using the same filters as used for developing models. We ensembled the best 20 models for prediction to reduce the computational processing for each tile to <48 core hours (~ 76,800 hours across all tiles). We predicted trait average values with associated uncertainty for each

suitable pixel in the tile independently. We also calculated the standard deviation of each PLS-GLR prediction as a measure of uncertainty. Crown-based predictions were achieved by averaging the values of all suitable pixels lying within the corresponding ITC boundaries. For those areas where the ITC overlapped with more than one flight path (flight paths overlap by ~30%), we averaged the crown-based predictions results from overlapping data.

The resulting dataset was then compiled as a comma delimited file containing all the geometry information to rebuild polygons shape and location. The data is distributed as both raw data files in a Zenodo archive, and through a shiny app that lets the user observe and download subsets of the data (LINK).The resulting files can be imported into GIS systems and spatial analysis packages in languages like R and Python for analysis. To demonstrate the potential uses of this data product we analyzed the correlation between terrain, leaf, and structural traits to summarize the general patterns observed at the scale of an entire NEON footprint.

Results

The crown delineation algorithm described in Silva et al. (2016) yielded the highest overlap with the field-delineated ITC crowns (Table S.1). The Jaccard Index ranged between 0 (undetected trees) and 0.81, with average of 0.35 and median of 0.34. Crowns identified by the algorithm were generally larger than those delineated in the field. Tree detection was relatively good, with the algorithmic approach producing overlapping crowns for 88% of the ground delineated crowns. Visual analysis of similarity showed that some algorithmically delineated crowns were shifted by a couple of meters relative to the field-delineated crowns as a result of misalignment between LiDAR and Hyperspectral data (Figure S.2; Marconi et al. 2018).

Figure 3 summarizes the performance of the CEAM, EPBM and SPM methods. CEAM explained 72% of the variance in LMA, 67% of the variance in %N, 44% of the variance in %P, and 23% of the variance in %C (Table 2). Except for %C, these results were comparable to those obtained by Martin et al. (2018) for trees in Borneo (71%, 46%, 44%, 0.48% respectively for LMA, %N, %P, %C). Predictions of %P showed lower accuracy despite being highly correlated to observed values (0.80 p-correlation), due tight relationship that deviated from the 1:1 line. This suggests that there is potential to improve predictions for %P since more information about the observations is available in the predicted values. Except for %C, RMSE was always between 10 and 15% of the range of observed values, meeting the quality threshold recommended by Singh et al. (2015).

These crown level results (CEAM) were better in almost every case compared to the methods using pixels as the fundamental unit, and the pixel based methods that considered crown identity (EPBM) performed better than the simple pixel based approach (SPM) (Table 2; Figure 3). Few individual PBMs performed better than the EPBM ensembles (Figure 3) and when individual pixel models did outperform the ensemble, they usually were not among the best models (i.e., the models with the lowest dAIC in the validation). As such, these pixel models could not have been selected and used to provide better fits. This shows that EPBM provide the best method for making out-of-sample predictions at the pixel level. CEAM models outperformed EPBM with higher predicted R^2 values for all traits except %P, where we found no

significant difference between crown and pixel based ensemble. The main improvements in predictability were for for LMA (from 0.60 to 0.73) and %N (0.60 to 0.67) (Table 2; Figure 3). While predictability of %C also increased (from 0.13 to 0.23), we considered the model's performance too low to provide useful individual-level information, and we discourage the use of %C products to address meaningful ecological questions. Crown based prediction also produced the best estimates of uncertainty. CEAM 95PI showed an average coverage of 83% of held out observations, EPSM 81%, and SBM 74%, with the ideal value being 95% (Figure 4).

Extracting pixels algorithmically, using the automatically delineated crowns from the LIDAR rather than ground delineated crowns, did not affect negatively the models. R^2 from held out predictions using CEAMs improved by 0.05 for both LMA and %N and the RMSE was reduced by 3.37 g m-2 and 0.1 % respectively. While SPMs and EPBMs performed significantly worse for %P and %C (0.44 to -0.21 and -0.17 to -0.51 respectively), predictions for the CEAM were comparable to those built from ground delineated crowns (0.44 to 0.37 for %P and 0.23 to 0.22 for %C). This suggested that scaling remote sensing predictions to tree objects can reduce the effect of uncertain crown delineation on final models' performance (Table 2).

Scaling algorithmic crown-delineation and trait estimation to the full extent of the NEON remote sensing data (Figure 5) yielded trait predictions for approximately 5 million ITCs for the two sites combined (Figure 6). Despite some striping patterns associated with the NEON flight lines, landscape patterns in traits are evident, including east-west gradients in LMA, %N, and %P at OSBS. At TALL, lower LMA and higher %N and %P are found in a dendritic pattern association with the stream network. On average, OSBS showed higher LMA and lower %N than TALL (Figure 7). Assessing correlations between estimated structural traits, leaf traits, and abiotic environmental conditions showed strong correlations between LMA, %N, and %P, consistent with the leaf economic spectrum (Wright et al., 2004). Results showed correlations of leaf traits with elevation and LAI (Figure 8). There were also correlations between leaf traits and tree structure (tree height and crown area), but only for trees at OSBS where the forest stature is about 10 m lower (Figures S.4, S.8), the canopy is more open, and the water table more superficial.

Discussion

We were interested in testing whether predicting leaf functional traits at the level of tree crown objects is a better option than the more traditional pixel based approaches, and explore how leaf and structural crown traits are affected by each other and by the environment in south east US mixed forests. A key challenge to do so is the delineation of the crown objects themselves. We needed to to identify a method to delineate crown objects at scale, calculate the associated uncertainty, and address if poor crown delineation would affect traits models significantly. With this perspective we built a fully automated modular pipeline to link NEON remote sensing products to data collected in the field, convert the NEON products into ecological data at landscape scale, and distribute information about millions of derived tree data in an accessible (Cassey et al., 2006) open (Hampton et al., 2016) format for use by the scientific community.

In general, we found many advantages in predicting traits at the crown level, including improved accuracy and reduced uncertainty in the predictions. Our method produced models with predictive power comparable to other methods in literature (Martin et al., 2018; Barbosa and Asner, 2017). First, by linking pixels to crowns we could ensemble models built from different pixel combinations sampled from training tree crowns. Each combination of pixels likely grasped different elements of the underlying chemometric relationship between foliar traits and their spectral signature. Ensembling these slightly different models resulted in more generalizable and accurate relationships (Figure 3, 4). Second, aggregating pixel predictions to crowns reduced the relative weight of outlier pixels. In fact, comparison between models built on field-delineated crowns and algorithmically-delineated crowns showed that uncertain crown delineation generally did not negatively affect model predictability (Table 2). Using algorithmically-delineated crowns for leaf trait modeling resulted in no change in accuracy for %C, improved accuracy for LMA and %N and a decrease in accuracy only for %P. These results suggest that the uncertainty in using automated crown delineation is somehow comparable to the uncertainty in drawing ITCs from the field, at least for mixed forests in southeast United States. We think that overestimating crown boundaries may reduce uncertainty by including representative pixels such that the algorithm can discard shaded or particularly mixed pixels from the external boundaries. On the other hand, including few representative pixels from crowns (underestimating crown boundaries) potentially affects accuracy in predictions by decreasing the sample from which leaf trait relationships can be drawn. Hence, this does not mean that unsupervised crown delineation is generally better than ground data collection, but suggests that a good integration of the two could further improve trait predictions.

A crown based approach is also potentially favorable to align structural and leaf trait products across years. The same area covered by a pixel in two consecutive years could significantly vary in a trait because of poor alignment in pixels through time, where the same pixel contains different portions of the crown and surrounding vegetation on different dates. Predicting traits at the crown level reduces this source of uncertainty because ITCs are more flexible than gridded data, and do not require perfect spatial match, as long the same crown is identified in all images. Finally, this approach allowed us to deliver data products as spatial polygons, resulting in lighter and more and manageable data, hundreds of times computational less demanding than raster gridded data.

Uncertainty in crown delineation was high for all methods, potentially affecting the ability to scale predictions, especially for structural traits like crown area and shape. The availability of a number of ITCs collected in the field had a key role in calibrating and quantifying uncertainty in the crown delineation module. The use of ground data allowed comparison of different algorithms, including methods already applied on the same data (Marconi et al., 2018). We found slightly higher performance when using the method described in Silva et al. (2016). However, low similarity with ground truth ITCs is partly attributable to the discrete nature of raster data and the resolution of the remote sensing data (1m^2 pixels) relative to the scale of most tree crowns (average tree crown area in this study = 35 m^2). For example, if the extent of a 4x4 m crown (area = 16 m^2) is overestimated by a single pixel in each direction (6x6m = 36 m^2), the Jaccard index is only 0.44. Similarly, the Jaccard index of a 4x4 m crown that is

misaligned by a single pixel on the diagonal would be only 0.39. The prototype NEON AOP data used in this study may have misalignments between LiDAR and Hyperspectral products on the scale of 1-2 meters (Marconi et al. 2018), which could reduce the Jaccard indices to roughly this level even for highly accurate delineations. Misalignment of this extent should only affect pixels at crown borders in most cases; while we expect little influence on predictions of partly isolated crowns (highly mixed pixels are discarded by pre-filtering), it potentially alters accuracy in predicting traits from overlapping crowns.

Scaling to thousands of hectares required the development of a pipeline that can be fully automated and is computationally sustainable. Processing each NEON site required over 10,000 core hours. The main computational limitations came from maximum likelihood estimation of PLS-GLR hyperparameters, the step that allowed us to calculate AIC and uncertainty. Reporting uncertainties is a key component in providing estimations of ecological information (citation), especially if these estimates are to be used as derived data, taken indirectly from remote sensing data. A good grasp of what is the actual range of likely results of a specific model not only provides information about how strong a prediction is, but allows propagation of uncertainty when using model predictions to test scientific hypotheses (Miller-Gulland and Shea, 2017). To our knowledge, this is one of the few works pairing remote sensing predictions with associated uncertainties at scale (Li et al., 2017, Sexton et al., 2015, Singh et al., 2015).

Landscape scale predictions showed multimodal distributions of LMA, %N, and %P. This is expected for a mixed forest that contains both coniferous and deciduous trees and is consistent with field observations at the sites. Interestingly, despite the similar species composition, ITCs in TALL showed higher %N and lower LMA than those in OSBS (Figure 7). LMA, %N and %P showed similar patterns to the Leaf Economic Spectrum (LES), despite the large difference in the spatial extent from which these datasets were derived and the different ecological units (ITCs and species, respectively) (Wright et al. 2004). Compared to the field data, our method showed that for the full AOP footprint the distribution of LMA and %P exhibited fewer large values. This difference could result from insufficiencies in either the estimation methods or the field data used to build them.The PLS method can produce biased predictions (Sarstedt et al., 2017) that potentially lead to underestimation of large values. Averaging the pixel information at crown-level shrinks extreme values to their mean, reducing their variance. On the other hand, small sample design could determine oversampling of the tails of traits distributions. For example, field data used in this study were collected to investigate other ecological questions and focused on particular species, sampled in a subset of the broader region. Values of %P larger than 0.2% collected on the field belong to trees of the Magnolia genus, which, according to NEON TOS vegetation census, represent only 0.2% of the trees in our study area: excluding magnolia from field traits data, distribution of observed and predicted %P match. The inherently unbalanced and small sampling design for field traits data is a general challenge for any attempt to address the relationship among the environment, leaf and structural traits at broader scales (Swenson et al., 2018). This is even more challenging when comparing field datasets collected for different studies or by different investigators, whose choice of sampled crowns is partly biased towards the most attractive (e.g. wider crowns, more healthy …) or accessible ones. Using the field data to build relationships that translate remote sensing in derived ecological

data for each individual tree crown at scale of hundreds of km2 can bridge this gap and provide a better insight to explore how traits are distributed across scales.

Quantities like crown area, stem density and size are properties of individual trees, and do not necessarily scale well at pixel level. In general, individual tree objects are better suited for merging and comparing biological, physical, and chemometric information. For example, our results suggest that relationship between functional and structural traits for individual trees differ across systems, even in case of forests with relatively similar species and trait composition. At OSBS there were significant relationships between crown area, height, and the predicted chemical traits (LMA, %N, %P), while no strong correlation to chemical traits was found in TALL (Figure S.8). This difference could be attributable to differences in edaphic conditions (Walter & Gerlach, 2013). OSBS soil is thin and mainly sandy, with a superficial water table (Varner et al., 2009). These conditions may potentially affect nutrients availability, determining the lower %N showed in OSBS (Figure 7), and have a potentially negative effect on tree vertical growth, on average 10 m lower than TALL (Figure S.4). It in also noteworthy that ontogeny strongly affects allometric strategies and correlation between structural and leaf traits (Prado-Junior et al., 2016). Observed patterns may be affected by different developmental stage and forest structure in different patches of the landscape, which suggests that estimation of ITCs age may be a fundamental feature for further analyses.

We showed that scaling pixels to individual tree crowns is a valuable and viable option to produce more reliable traits predictions, reduce the related uncertainty, and merge ecological data from different sources. Crown objects represent a more suitable unit to merge image, structural traits, leaf chemistry, and environment, and are more flexible in linking objects partly misaligned or changing in time. The data we produced could be extended by including predictions for other leaf and structural traits, environmental variables, management, or disturbance. Moreover, our pipeline could potentially be used to extract ecological information from each single tree in any NEON AOP site for 30 years. This will produce a publicly available, spatially explicit database of very detailed ecological information for hundreds of millions of trees across US (link). Such data could be used to further understand the biology behind traits tradeoffs and investigate cross scale ecological processes and patterns from individual to landscape to continental scale.

References

1. Albert, C.H., Thuiller, W., Yoccoz, N.G., Soudant, A., Boucher, F., Saccone, P. and Lavorel, S., 2010. Intraspecific functional variability: extent, structure and sources of variation. Journal of Ecology, 98(3), pp.604-613.

2. Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Joseph Wright, S., Abu Salim, K., Almeyda Zambrano, A.M., Alonso, A., Baltzer, J.L. and Basset, Y., 2015. CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. Global change biology, 21(2), pp.528-549.

3. Asner, G.P., Martin, R.E., Knapp, D.E., Tupayachi, R., Anderson, C.B., Sinca, F., Vaughn, N.R. and Llactayo, W., 2017. Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. Science, 355(6323), pp.385-389.

4. Barbosa, J.M. and Asner, G.P., 2017. Prioritizing landscapes for restoration based on spatial patterns of ecosystem controls and plant–plant interactions. Journal of Applied Ecology, 54(5), pp.1459-1468.

5. Barnes, R., Lehman, C., Mulla, D., 2014. Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. Computers & Geosciences 62, 117–127. doi:10.1016/j.cageo.2013.04.024

6. Bastien, P., Vinzi, V.E. and Tenenhaus, M., 2005. PLS generalised linear regression. Computational Statistics & data analysis, 48(1), pp.17-46.

7. Callaway, R.M., Pennings, S.C. and Richards, C.L., 2003. Phenotypic plasticity and interactions among plants. Ecology, 84(5), pp.1115-1128.

8. Cassey, P. and Blackburn, T.M., 2006. Reproducibility and repeatability in ecology. BioScience, 56(12), pp.958-959.

9. Chadwick, K.D. and Asner, G.P., 2016. Organismic-scale remote sensing of canopy foliar traits in lowland tropical forests. Remote Sensing, 8(2), p.87.

10. Clark, J.S., Gelfand, A.E., Woodall, C.W. and Zhu, K., 2014. More than the sum of the parts: forest climate response from joint species distribution models. Ecological Applications, 24(5), pp.990-999.

11. Clark, J.S., 2016. Why species tell more about traits than traits about species: predictive analysis. Ecology, 97(8), pp.1979-1993.

12. Colgan, M.S., Baldeck, C.A., Féret, J.B. and Asner, G.P., 2012. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. Remote Sensing, 4(11), pp.3462-3480.

13. Cornelissen, J. H. C., S. Lavorel, E. Garnier, S. Diaz, N. Buchmann, D. E. Gurvich, P. B. Reich, H. ter Steege, H. D. Morgan, M. G. A. van der Heijden, J. G. Pausas, and H. Poorter. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Australian Journal of Botany 51:335 – 380.

14. Dalponte, M. and Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods in ecology and evolution, 7(10), pp.1236-1245.

15. De Angelis, D.L., 2018. Individual-based models and approaches in ecology: populations, communities and ecosystems. CRC Press.

16. Dwyer, J.M. and Laughlin, D.C., 2017. Constraints on trait combinations explain climatic drivers of biodiversity: The importance of trait covariance in community assembly. Ecology letters, 20(7), pp.872-882.

17. Feilhauer, H., Asner, G.P. and Martin, R.E., 2015. Multi-method ensemble selection of spectral bands related to leaf biochemistry. Remote Sensing of Environment, 164, pp.57-65.

18. Frederic Bertrand, Nicolas Meyer and Myriam Maumy-Bertrand (2014). Partial Least Squares Regression for Generalized Linear Models, R package version 1.1.1.

19. Graves, S.J., Asner, G.P., Martin, R.E., Anderson, C.B., Colgan, M.S., Kalantari, L. and Bohlman, S.A., 2016. Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. Remote Sensing, 8(2), p.161.

20. Graves S, Gearhart J, Caughlin TT, Bohlman S. 2018. A digital mapping method for linking high-resolution remote sensing images to individual tree crowns. PeerJ Preprints6:e27182v1 https://doi.org/10.7287/peerj.preprints.27182v1

21. Hampton, S.E., Anderson, S.S., Bagby, S.C., Gries, C., Han, X., Hart, E.M., Jones, M.B., Lenhardt, W.C., MacDonald, A., Michener, W.K. and Mudge, J., 2015. The Tao of open science for ecology. Ecosphere, 6(7), pp.1-13.

22. Houborg, R., Fisher, J.B. and Skidmore, A.K., 2015. Advances in remote sensing of vegetation function and traits.

23. Kerr, J.T. and Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. Trends in ecology & evolution, 18(6), pp.299-305.

24. Lawler, J.J., Tear, T.H., Pyke, C., Shaw, M.R., Gonzalez, P., Kareiva, P., Hansen, L., Hannah, L., Klausmeyer, K., Aldous, A. and Bienz, C., 2010. Resource management in a changing and uncertain climate. *Frontiers in Ecology and the Environment*, *8*(1), pp.35-43.

25. Li, X., Messina, J.P., Moore, N.J., Fan, P. and Shortridge, A.M., 2017. MODIS land cover uncertainty in regional climate simulations. Climate Dynamics, 49(11-12), pp.4047-4059.

26. Liu, X., Swenson, N.G., Lin, D., Mi, X., Umaña, M.N., Schmid, B. and Ma, K., 2016. Linking individual-level functional traits to tree growth in a subtropical forest. Ecology, 97(9), pp.2396-2405.

27. Marconi, S., Graves, S.J., Gong, D., Nia, M.S., Le Bras, M., Dorr, B.J., Fontana, P., Gearhart, J., Greenberg, C., Harris, D.J. and Kumar, S.A., 2018. A data science challenge for converting airborne remote sensing data into ecological information (No. e26966v1). PeerJ Preprints.

28. Martin, M. E., L. C. Plourde, S. V. Ollinger, M. L. Smith, and B. E. McNeil. 2008. A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. Remote Sensing of Environment 112:3511–3519.

29. Martin, R.E., Chadwick, K.D., Brodrick, P.G., Carranza-Jimenez, L., Vaughn, N.R. and Asner, G.P., 2018. An approach for foliar trait retrieval from airborne imaging spectroscopy of tropical Forests. Remote Sensing, 10(2), p.199.

30. Milner-Gulland, E.J. and Shea, K., 2017. Embracing uncertainty in applied ecology. Journal of applied ecology, 54(6), pp.2063-2068.

31. Müller, I., Schmid, B. and Weiner, J., 2000. The effect of nutrient availability on biomass allocation patterns in 27 species of herbaceous plants. Perspectives in plant ecology, evolution and systematics, 3(2), pp.115-127.

32. Murphy, J., and J. P. Riley. 1962. A modified single solution method for the determination of phosphate in natural waters. Analytica Chimica Acta 27:31 – 36.

33. National Ecological Observatory Network. 2014. Data Products NEON.DP1.10098, NEON.DP1.30010, NEON.DP3.30015, NEON.DP1.30008. Provisional data downloaded from http://data.neonscience.org on 26 Jan 2016. Battelle, Boulder, CO, USA

34. Newnham, G.J., Armston, J.D., Calders, K., Disney, M.I., Lovell, J.L., Schaaf, C.B., Strahler, A.H. and Danson, F.M., 2015. Terrestrial laser scanning for plot-scale forest measurement. Current Forestry Reports, 1(4), pp.239-251.

35. Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F. and van Kleunen, M., 2010. Plant phenotypic plasticity in a changing climate. Trends in plant science, 15(12), pp.684-692.

36. Nock, C. A., Vogt, R. J. and Beisner, B. E. (2016). Functional Traits. In eLS, John Wiley & Sons, Ltd (Ed.). doi:10.1002/9780470015902.a0026282

37. Prado-Junior, J.A., Schiavini, I., Vale, V.S., Raymundo, D., Lopes, S.F. and Poorter, L., 2016. Functional traits shape size-dependent growth and mortality rates of dry forest tree species. Journal of Plant Ecology, 10(6), pp.895-906.

38. Real, R. and Vargas, J.M., 1996. The probabilistic basis of Jaccard's index of similarity. Systematic biology, 45(3), pp.380-385.

39. Roussel, J.R. and Auty, D., 2017. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications.

40. Sarstedt, M., Hair, J.F., Ringle, C.M., Thiele, K.O. and Gudergan, S.P., 2016. Estimation issues with PLS and CBSEM: Where the bias lies!. *Journal of Business Research*, *69*(10), pp.3998-4010.

41. Sexton, J.O., Noojipady, P., Anand, A., Song, X.P., McMahon, S., Huang, C., Feng, M., Channan, S. and Townshend, J.R., 2015. A model for the propagation of uncertainty from continuous estimates of tree cover to categorical forest cover and change. Remote Sensing of Environment, 156, pp.418-425.

42. Silva, C.A., Hudak, A.T., Vierling, L.A., Loudermilk, E.L., O'Brien, J.J., Hiers, J.K., Jack, S.B., Gonzalez-Benecke, C., Lee, H., Falkowski, M.J. and Khosravipour, A., 2016. Imputation of individual Longleaf Pine (Pinus palustris Mill.) Tree attributes from field and LiDAR data. Canadian journal of remote sensing, 42(5), pp.554-573.

43. Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M.B., Balaguer, L., Benito-Garzón, M., Cornwell, W., Gianoli, E., Kleunen, M., Naya, D.E. and Nicotra, A.B., 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecology letters, 17(11), pp.1351-1364.

44. Varner, J. Morgan, et al. "Post-fire tree stress and growth following smoldering duff fires." *Forest Ecology and Management* 258.11 (2009): 2467-2474.

45. Walters, Michael B., and John P. Gerlach. "Intraspecific growth and functional leaf trait responses to natural soil resource gradients for conifer species with contrasting leaf habit." *Tree physiology* 33.3 (2013): 297-310.

46. Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H., Diemer, M. and Flexas, J., 2004. The worldwide leaf economics spectrum. Nature, 428(6985), p.821.

47. Zhen, Z., Quackenbush, L.J. and Zhang, L., 2016. Trends in automatic individual tree crown detection and delineation—evolution of lidar data. Remote Sensing, 8(4), p.333.

Table 1. Data products and sources (National Ecological Observatory Network, 2016). Information about data products can be found on the NEON data products catalogue (http://data.neonscience.org/data-product-catalog).

Name	NEON data product ID	Data date	How it was used
Spectrometer orthorectified at-sensor radiance	NEON.DP1.30008	2014 and 2015	Hyperspectral images used to infer foliar chemical and physical properties
Discrete return LiDAR point cloud	NEON.DP1.30003.001	2014 and 2015	Crown delineation and calculation of tree height
AOP L2 and L3 data products (LAI, Albedo, Elevation, Slope, Aspect)	NEON.DP2.30012.001, DP3.30024.001, DP3.30025.001	2014 and 2015	Link inferred tree crown properties with other NEON-AOP data products
Field ITC	Internal	2017	Validate crown delineation, define tree objects.

Table 2 Models performance on 15 held-out crowns (predicted R2 and RMSE) for the three modeling strategies CEAM (Crown ensemble average model), EPBM (Ensemble pixel based model) and SPM (single pixel model). Higher performance in bold. (a) Train, validation and test pixels extracted from ground delineated crowns (least uncertainty in pixel labeling). (b) Train, validation and test pixels extracted from automatically delineated crowns (highest uncertainty in pixel labeling).

a) Field data aligned with ground delineated crowns						
	R2	RMSE	R2	RMSE	R2	RMSE
	CEAM		EPBM		SPM	
%P	**0.44**	0.19	0.43	0.19	0.44	0.22
LMA (g m-2)	**0.73**	**32.33**	0.60	33.49	0.46	38.90
%N	**0.67**	**0.23**	0.60	0.24	0.46	0.28
%C	**0.23**	**1.15**	0.13	1.23	-0.17	1.34
b) Field data aligned with automatically delineated crowns						
	R2	RMSE	R2	RMSE	R2	RMSE
	CEAM		EPBM		SPM	
%P	**0.37**	**0.021**	0.12	0.022	-0.21	0.026
LMA (g m-2)	**0.78**	**28.86**	0.65	28.24	0.46	35.03
%N	**0.72**	**0.22**	0.58	0.24	0.44	0.28
%C	**0.22**	**1.19**	-0.11	1.33	-0.51	1.56

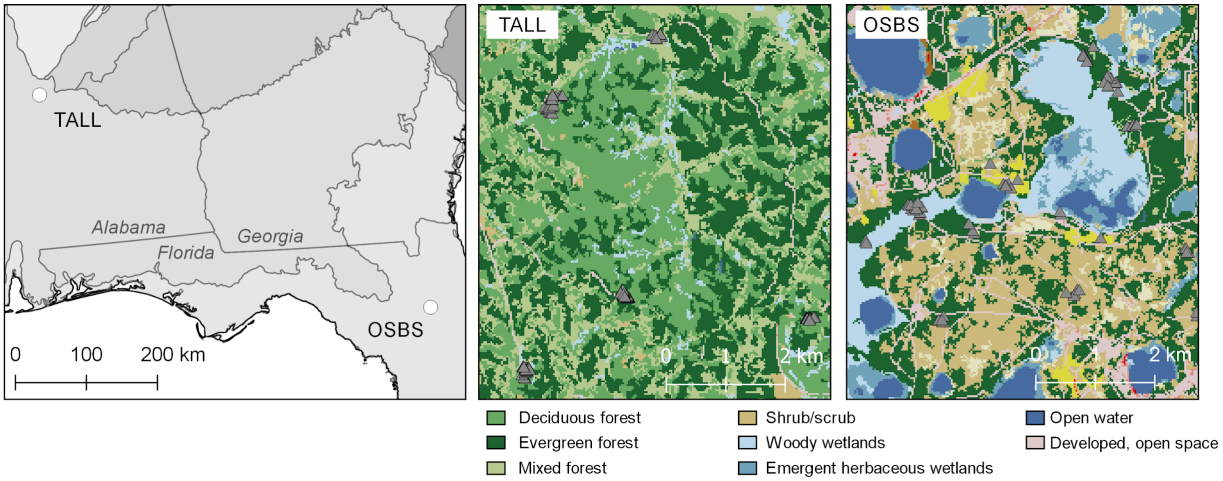


Figure 1. Regional map showing OSBS and TALL locations in the southeastern US ecoregions. Insets show TALL and OSBS locations of individual tree crowns (gray triangles) over the 2011 National Land Cover Database.

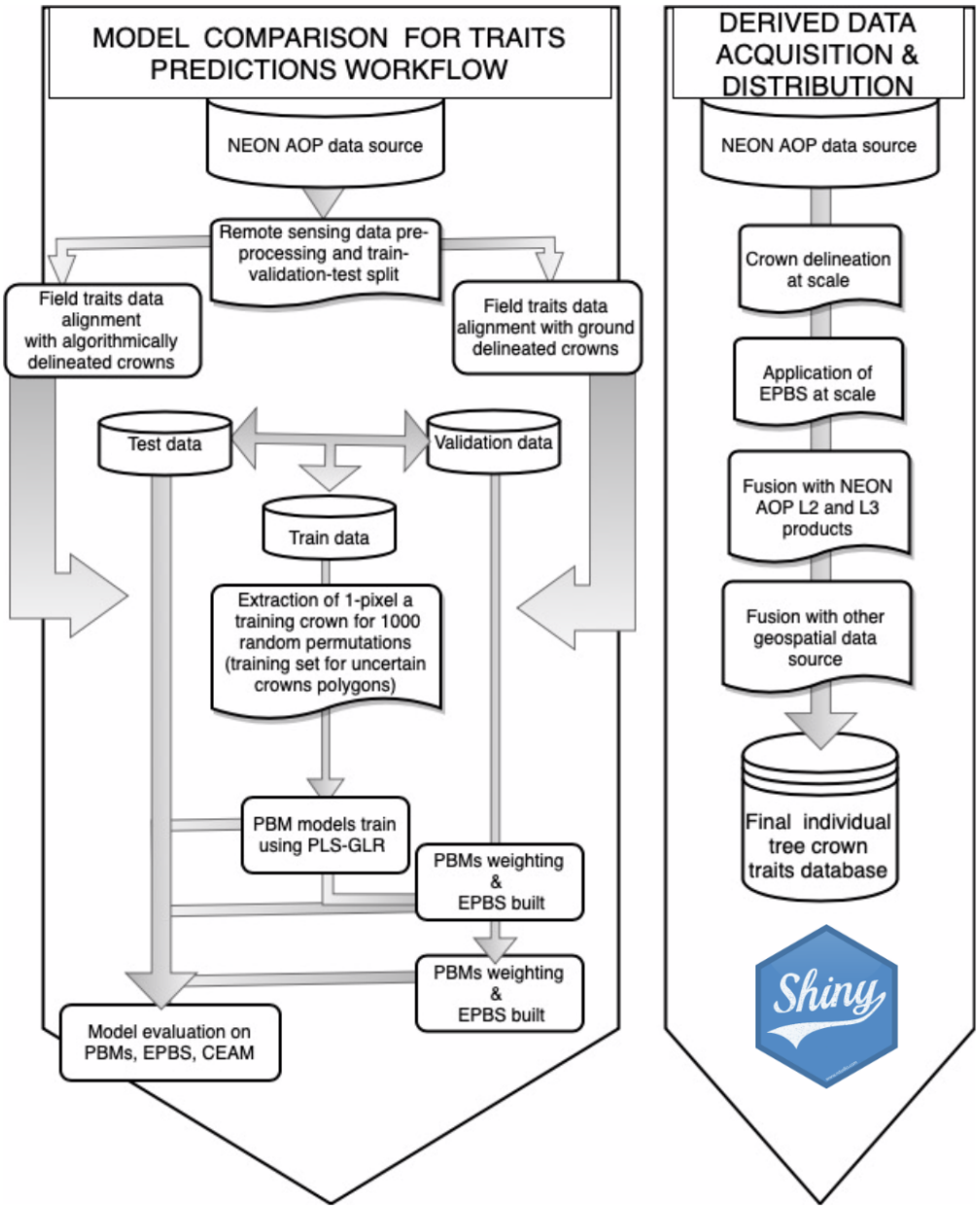


Figure 2. Workflow of the pipeline. Left panel shows the method to build and test the three modeling approaches: PBM (pixel based model), EPBS (ensemble of pixel based models), CEAM (crown ensemble averaged model). Cylinders represent raw or derived data. Rounded rectangles represent computational steps. Wavy rounded rectangles represent processes repeated for a number of times (e.g. 1000 extractions of 1 pixel a training crown permutations). Right panel represents the part of the pipeline dealing with scaling, fusion, and public distribution of derived data.

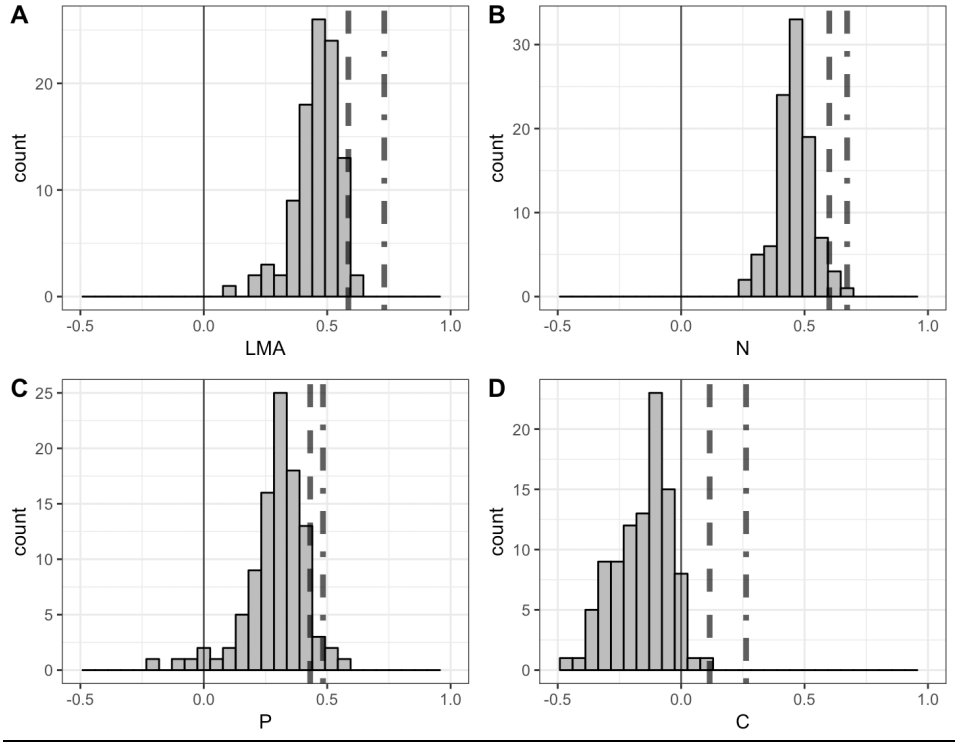


Figure 3. Comparison of test R^2 using the 100 best SPMs (histogram), the EPBM (dashed line), and the CEAM (dot-dashed lines). The vertical line on 0 represents the threshold below which the residual sum of squares (RSS) is higher than the total sum of squares (TSS). A negative R2 value means that the observed sample mean is a better predictor than the model, and so that specific model is meaningless.

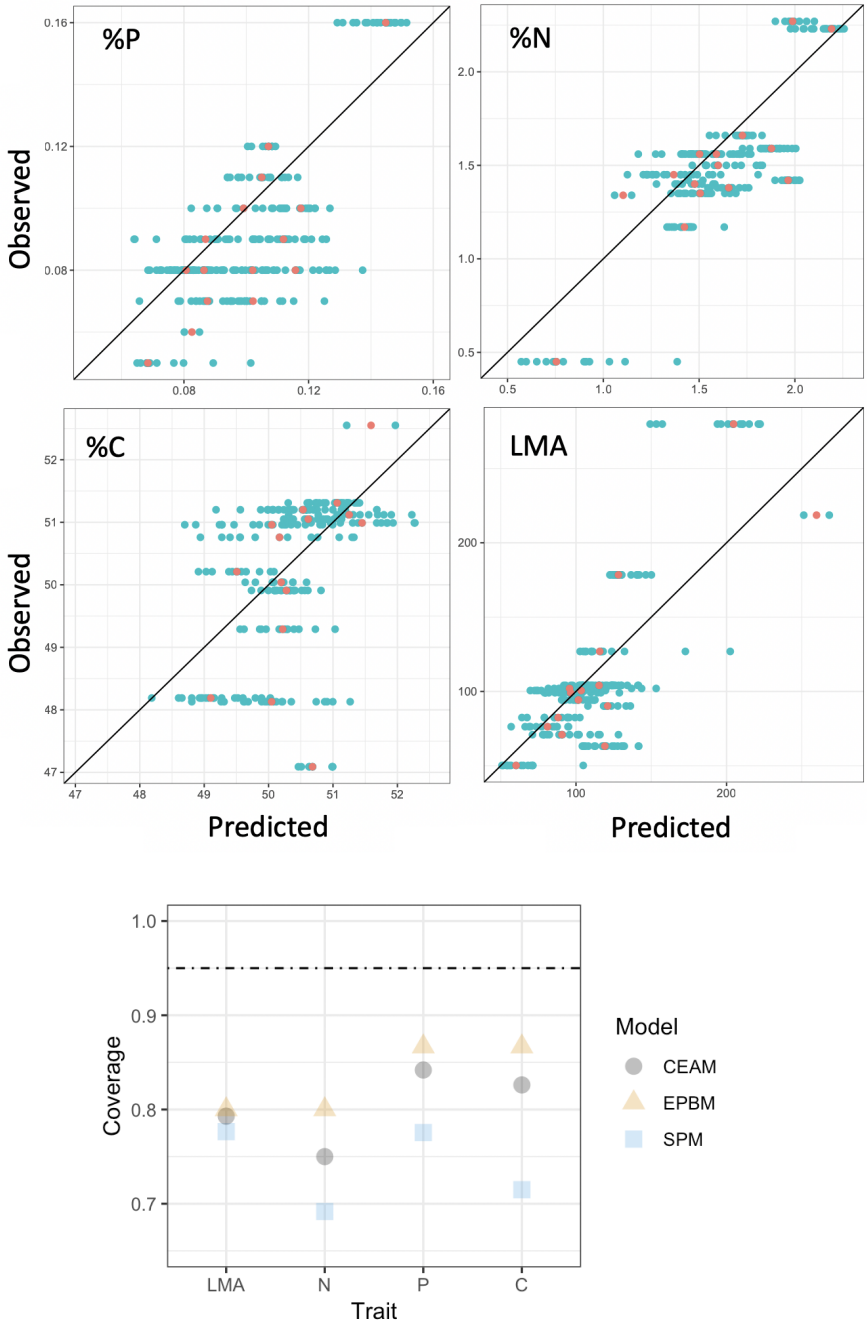


Figure 4. (A-D) Comparison of pixel (azure) and crown aggregated (red) predictions on an independent test set of 15 observed crowns for %P (A), %N (B), %C (C) and LMA (D) respectively. Black diagonal is the 1:1 line. (E) Coverage of the 95 predictions intervals on held out data for the four independent models (LMA, N, P and C). Grey circle represents the coverage for the crown ensemble averaged model (CEAM), yellow triangle for the ensemble pixel based model (EPBM), azure square for the single pixel model (SPM). Dash-dotted line represent the ideal coverage value of 95%.

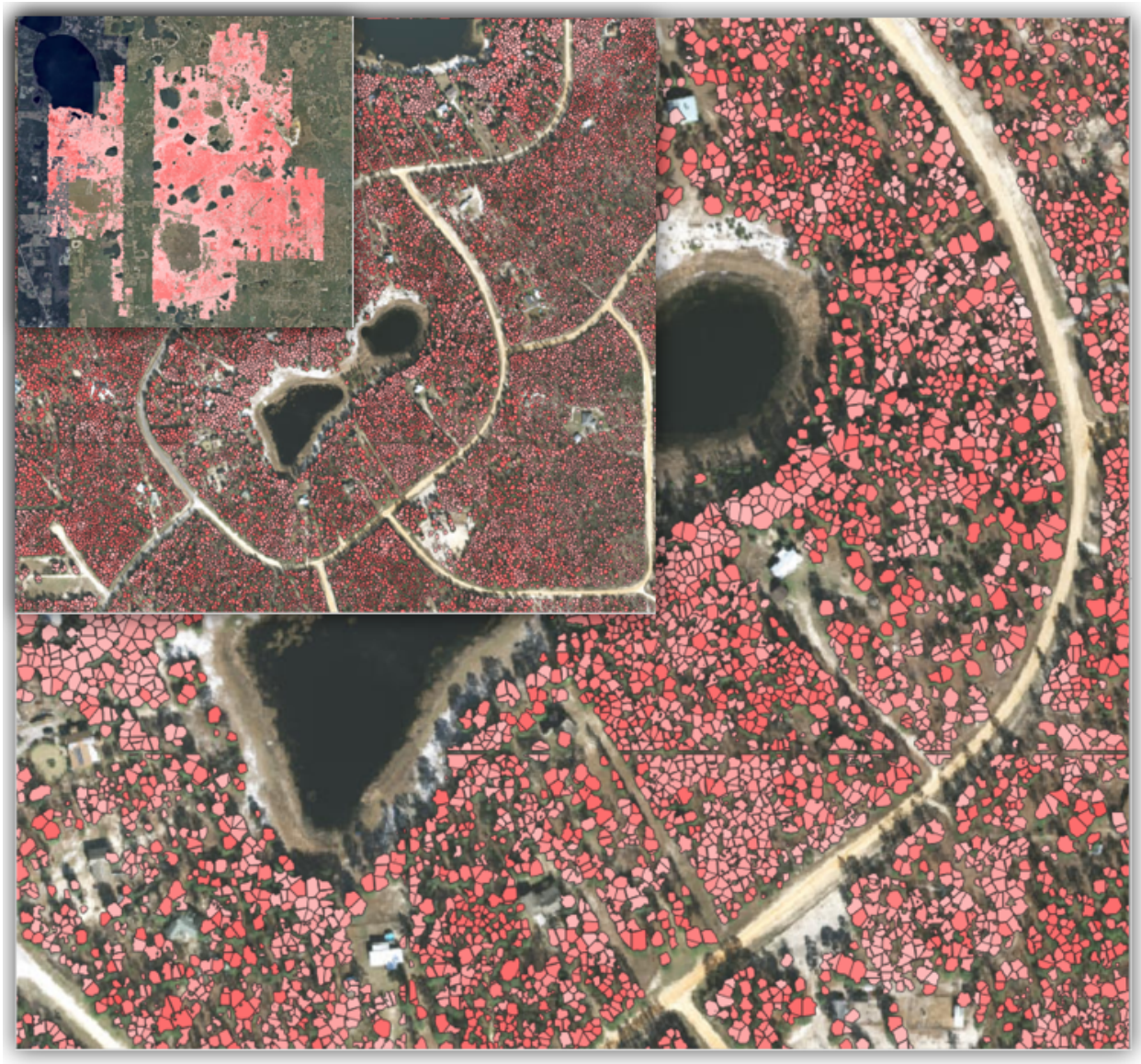
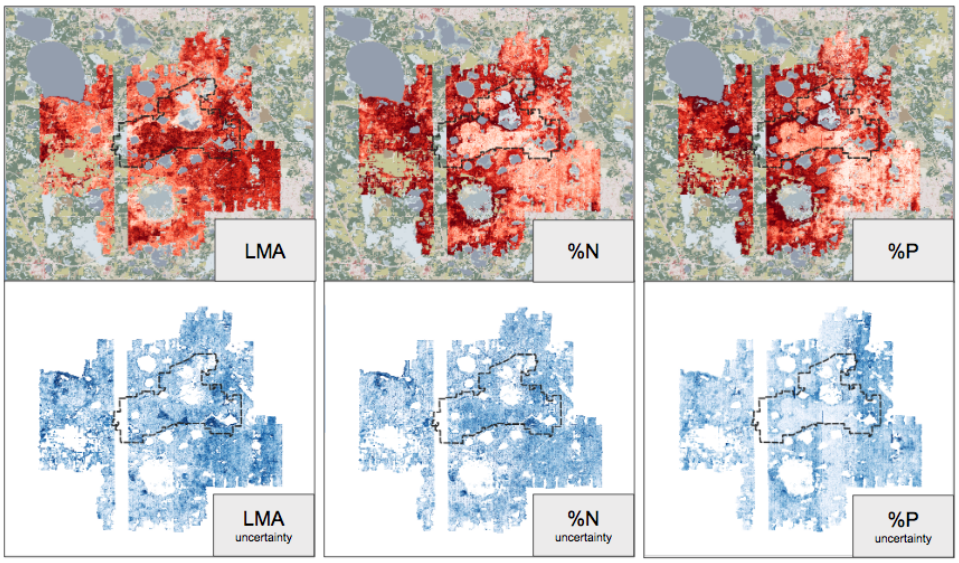


Figure 5. Graphic representation of site level pipeline products ranging from 1:2000 to 1:125,000 scale.

OSBS



TALL

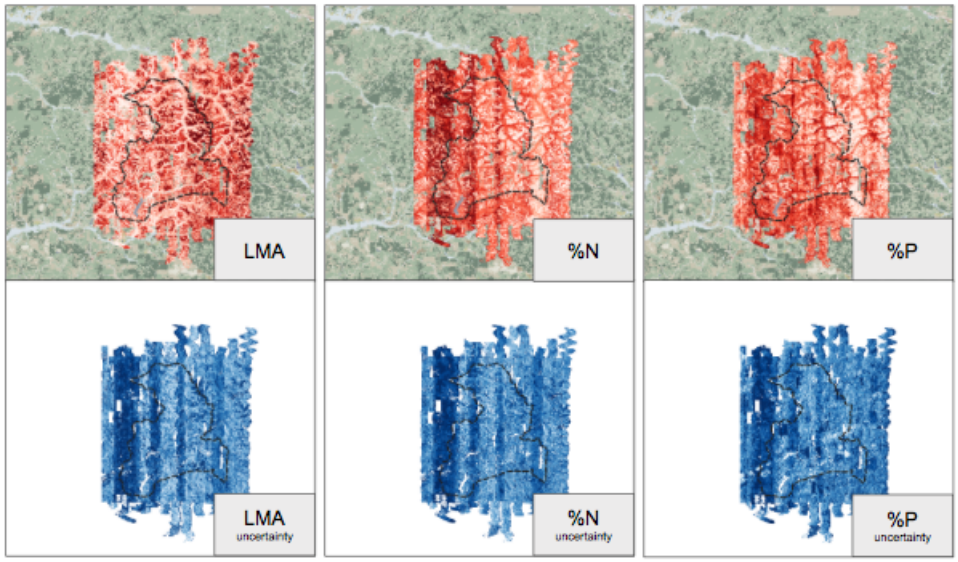


*Figure 6. Thematic maps for LMA, %N and %P for the two NEON sites; The top panel is OSBS (215 km^2), bottom panel TALL (145 km^2). The maps show the centroid of around 2.5 million trees per site, colored by gradient of intensity. The top panel (red) represents individual predictions for LMA, %N and %P ITCs respectively (n = 5 * 10$^{\wedge 6}$). The lower panel (blue) represents the uncertainty for each tree based prediction as the width of the 95%PI. Black lines represent the borders of NEON sites. Maps of structural traits and environment in appendix.*

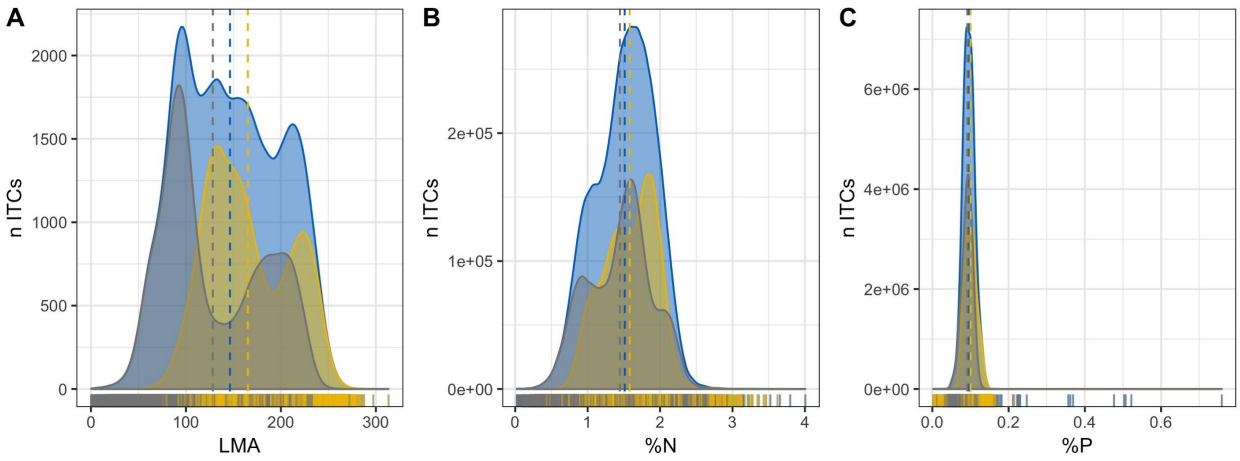


Figure 7. Comparison between distributions of LMA (panel A), %N (panel B), and %P (panel C) for the whole derived dataset (blue), and per individual sites (gray and yellow for TALL and OSBS respectively). Vertical dotted lines represent the mean value of the full (blue), TALL (grey) and OSBS (grey) datasets. Rug plots on the x-axis represent the marginal distribution of full, TALL, and OSBS data between the minimum and maximum range of derived observations.

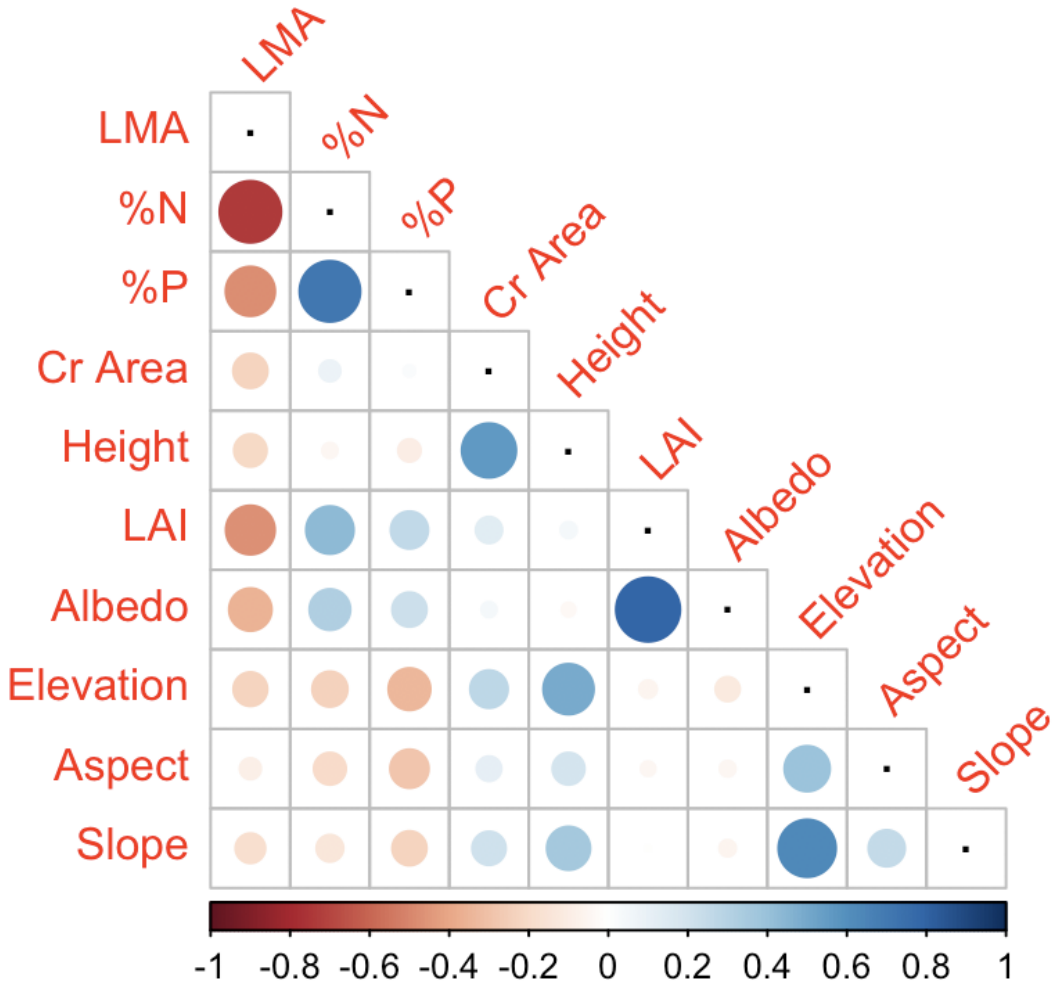


Figure 8. Correlation between terrain, leaf and structural traits for both OSBS and TALL predicted crowns (5M ITCs). Circles size scaled by Pearson's correlation coefficient. Site specific relationships in appendix.