Individual canopy tree species maps for the National Ecological Observatory Network

Short title: Individual tree species maps at NEON

Ben. G. Weinstein[1*], Sergio Marconi[1], Alina Zare[2], Stephanie A. Bohlman[3], Aditya Singh[4], Sarah J. Graves[5], Lukas Magee[3], Daniel J. Johnson[3], Sydne Record[6], Vanessa E. Rubio[7], Nathan G. Swenson[7], Philip Townsend[8], Thomas T. Veblen[9], Robert A. Andrus[9, 10], Ethan P. White[1]

1 Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida, USA

2 Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA

3 School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, Florida, USA.

4 Department of Agricultural & Biological Engineering, University of Florida, Gainesville, Florida, USA

5 Nelson Institute for Environmental Studies, University of Wisconsin-Madison, Madison, Wisconsin, USA

6 Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, Maine, USA

7 Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556 USA

8 Department of Forest & Wildlife Ecology, University of Wisconsin-Madison, Madison, Wisconsin 53706 USA

9 Department of Geography, University of Colorado, Boulder, Colorado 80309

10 School of Environment, Washington State University, Pullman, WA

*Corresponding Author: ben.weinstein@weecology.org

Keywords: Forest Ecology; Biodiversity; Deep Learning; Biogeography; Hyperspectral Reflectance; Google Earth Engine

Abstract

The ecology of forest ecosystems depends on the composition of trees. Capturing fine-grained information on individual trees at broad scales provides a unique perspective on forest ecosystems, forest restoration and responses to disturbance. Individual tree data at wide extents promises to increase the scale of forest analysis, biogeographic research, and ecosystem monitoring without losing details on individual species composition and abundance. Computer vision using deep neural networks can convert raw sensor data into predictions of individual canopy tree species through labeled data collected by field researchers. Using over 40,000 individual tree stems as training data, we create landscape-level species predictions for over 100 million individual trees across 24 sites in the National Ecological Observatory Network. Using hierarchical multi-temporal models fine-tuned for each geographic area, we produce open-source data available as 1 km^2 shapefiles with individual tree species prediction, as well as crown location, crown area and height of 81 canopy tree species. Site-specific models had an average performance of 79% accuracy covering an average of six species per site, ranging from 3 to 15 species per site. All predictions are openly archived and have been uploaded to Google Earth Engine to benefit the ecology community and overlay with other remote sensing assets. We outline the potential utility and limitations of these data in ecology and computer vision research, as well as strategies for improving predictions using targeted data sampling.

Introduction

Broadscale tree taxonomic data is essential for forest management, conservation planning, ecosystem service modeling, and biodiversity research. Historically, collection of tree species data has largely relied on 1) field-censused plots ranging from dozens of individuals to several thousand trees (Davies et al. 2021) that provide high-quality data, but can only be monitored over small areas for each plot; and 2) satellite-based predictions of community-level taxonomic diversity, which can be made continuously over broad scales, but lack detailed information on individual trees (Schäfer et al. 2016). Individual tree predictions from high resolution airborne data complement these approaches by creating a bridge between high quality, but spatially restricted, field data (e.g. Jucker et al. 2022), and spatially continuous, but low resolution data, from satellite or airborne sensors (Wagner et al. 2023). The spatial coverage of high resolution airborne imagery from planes and UAVs allows a broader view of forest ecology over areas from dozens to 10,000s of hectares (Tucker et al. 2023, Li et al. 2023). Access to these data can complement field data and global satellite monitoring to facilitate the assessment of forest structure and dynamics and how they respond to ecological processes, human management and global change (Wallis et al. 2023, Tucker et al. 2023).

Individual tree detection is a long-standing task for remote sensing of the environment as it provides information on the densities of individual trees for large areas. Predicting the location of individual trees (e.g. Freudenberg et al. 2019, Zamboni et al. 2021, Velasquez-Camacho et al. 2023), as well delineating the extent of tree crowns (e.g. Aubry-Kientz et al. 2019), is essential in many remote sensing workflows and has been a rich area of algorithmic research (see reviews by Ke and Quackenbush 2011, Pulido et al. 2020). Deep learning algorithms using a combination of human-labeled imagery and field-based geospatial data have become the standard tool for tree detection for airborne RGB data (Bosch 2020, Schiefer et al. 2020, Weinstein et al. 2021a). The challenge for deep learning algorithms for tree detection is

collecting sufficient training data to capture the variation in tree crown shape when applied across land-use and forest types.

After individual tree crowns have been delineated, the next step towards airborne forest inventories is to assign each crown a taxonomic label (Fassnacht et al. 2016). Dozens of models have been proposed using classical image processing (Seeley et al. 2023), feature-based machine learning (Maschler et al. 2018, Marconi et al. 2022), and deep learning (Fricker et al. 2019, La Rosa et al. 2021, Veras et al. 2022), but it is unclear if they are successful when applied to a variety of ecosystems with differences in tree density, abundance distributions, and spectral backgrounds. Given the very low sample sizes of training data in most studies, it is difficult to capture the range of species present and the spectral representations for each species. One proposed solution to this is using an ensemble of multiple time points of airborne imagery to improve within-site performance (Onishi and Ise 2021, Veras et al. 2022). Sample size issues are magnified by class imbalance since the dominant taxa in many systems comprises more than 50% of training data and can thousands of times more common than the rarer species in the dataset. This imbalance makes it difficult to train large neural network models (La Rosa et al. 2021) and creates rigorous evaluation datasets (Fassnacht et al. 2016). Weinstein et al. (2023) proposed an approach to addressing these limitations by using a flexible hierarchical model structure that uses simple rules to define a series of models to create an ensemble species prediction. This approach uses both multiple views of the same crown across years, as well as a hierarchical structure to reduce the effect of species imbalance. It was effective at expanding the number of species that could be accurately classified at a single NEON site, but has yet to be tested and applied across sites with a diversity of forest types. Here we apply the tree delineation and species classification workflow proposed for a single site in Weinstein et al. (2023) to sites across the United States and assess its performance in order to provide data for ecological and computer vision research.

The National Ecological Observatory Network (NEON) provides an opportunity to advance our regional scale understanding of forests by collecting open-access, high resolution airborne remote sensing data over 10,000s of hectares (Musinsky et al. 2022). NEON collects standardized terrestrial and airborne data at dozens of sites across the United States, creating an ideal situation for constructing landscape scale maps of canopy tree species for ecological research. Our aim is to generate individual canopy tree crown maps to support the ongoing forest, ecosystem, natural history, community science, and wildlife research programs at NEON sites (Kampe et al. 2010, Egli et al. 2020, Ayres et al. 2021, Lombardozzi et al. 2023). Here, we combine airborne RGB, hyperspectral, and LiDAR data, to predict 100 million canopy tree locations for 81 species within 24 NEON sites across the United States using machine learning models to predict crown position, species identity, health status and height for individual trees visible in the canopy (Figure 1). Our work extends the crown location dataset published in Weinstein et al. (2021) by adding predictions of species identity and alive/dead classification. The addition of species labels significantly expands the utility of this dataset for biodiversity research and natural resource management.

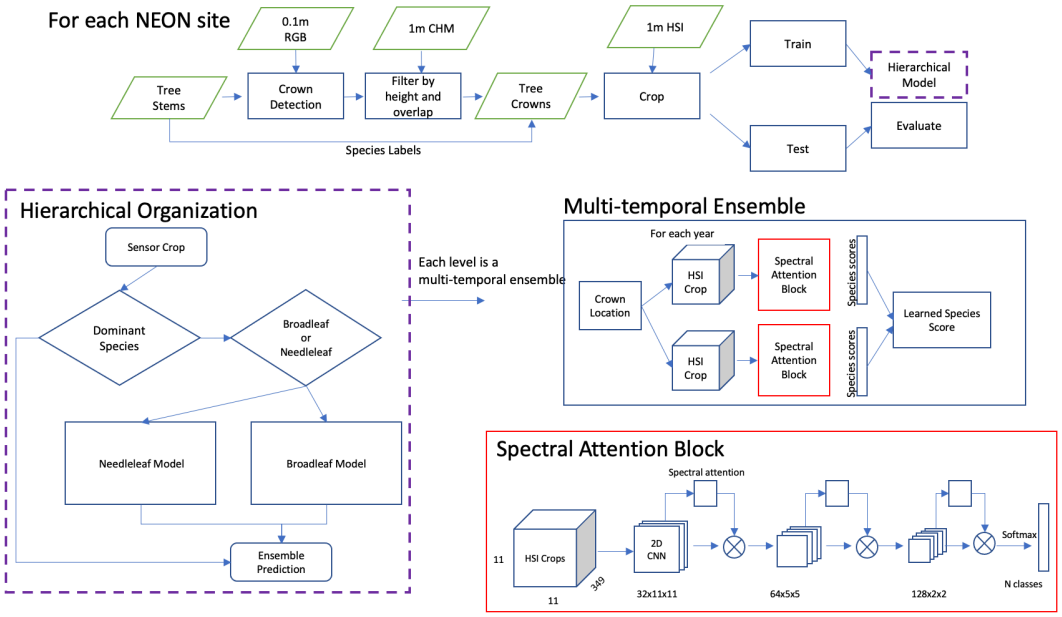


Figure 1. Conceptual workflow for species prediction at each NEON site. The figure is modified from Weinstein et al. (2023).

Materials and Methods

Airborne Sensor Data

The NEON airborne observation platform (AOP) collects remote sensing data on an annual basis during leaf-on conditions for all sites. For each site, data is collected at peak greenness to reduce variation due to phenological differences (Musinsky et al. 2022). We used four NEON data products 1) orthorectified camera mosaic (‘RGB’ NEON ID: DP3.30010.001), 2) ecosystem structure (‘Canopy Height Model’ NEON ID: DP3.30015.001), 3) hyperspectral surface reflectance (‘HSI’ NEON ID: DP1.30006.001), and 4) vegetation structure (NEON ID: DP1.10098.001). All data were downloaded in August 2022 and were the RELEASE form (NEON 2023). The 10 cm RGB data were used to predict tree crown locations necessary for

associating field labels and sensor data during model development. RGB data were also used to identify dead trees during our prediction workflow. The 1 m canopy-height model was used to determine which field collected data were likely to be visible from the air, as well as to define a 3 m minimum tree height threshold during the prediction workflow. The HSI data is used to differentiate tree species based on spectral reflectance. The HSI data spanned approximately 420-2500 nm with a spectral sampling interval of 5 nm producing a total of 426 bands. NEON provides orthorectified images with a pixel size of 1 m^2 in 1 km^2 tiles that are georectified and aligned with the RGB and Canopy-Height-Model. For more information on hyperspectral data processing and calibration see NEON technical document NEON.DOC.001288.

Field-based species labels

The NEON Vegetation Structure dataset is a collection of tree stem points within fixed-area field plots; plot locations are allocated across sites according to a stratified random, spatially balanced design (Barnett et al. 2019). All trees in sampled areas with a stem diameter > 10 cm are mapped and measured for diameter, height, health status, and species identity. Building on this NEON dataset, we contacted researchers at each NEON site to find as many mapped stems as possible outside the NEON woody vegetation sampling plots. We collected 22,072 additional canopy trees from a variety of sources, including several large ForestGEO plots co-located at NEON sites (Davies et al. 2021) and public data (Veblen et al. 2021). We followed the taxonomic hierarchy used by NEON with the exception of genus-only, subspecies and variety labels.

To connect species information from ground-based stem points with the airborne sensor data, we adopted a heuristic data filtering approach (Figure 2). We began with raw stem data for 41,036 individuals. We removed stems that were labeled as dead or broken, did not have a species label, or were less than 3 m in field-measured height. Whenever DBH was available, stems less than 10 cm were discarded. We then compared the field-measured height to the

height of the LiDAR-derived canopy model at the stem point for the closest available year. If the difference between the LiDAR-derived and field height was more than 4 m, we discarded the stem. We then overlaid these height-filtered points to crown bounding box predictions made from the DeepForest RGB algorithm. If more than one height-filtered point fell within the predicted canopy crown box, we selected the tallest point using the canopy height model since this was most likely to be the dominant tree in the canopy. The shorter tree stems that overlapped the bounding box were discarded. If a point did not overlap with any bounding box, we created a 1 m buffer around the point to serve as a crown box. We refer to these crowns as 'fixed boxes,' and these were only included in training data, but never in testing data due to lower confidence in associating species labels and sensor pixels. Finally, if there less than three matched stems per species at a site, the species and its stems were removed for that site. After these steps, there were 31,736 points remaining to be used for model training and validation. Marconi et al. (2022) used a portion of these training data to compare local versus global models for each site. Because of the differences in evaluation approaches, a precise comparison between Marconi et al. (2022) and this article is not possible. We emphasize that the focus of this article is on the publication of the crowns dataset rather than a comparison of a bounding box multi-temporal deep learning approach versus the pixel-based ensemble of machine learning classifiers presented in Marconi et al. (2022).

For predictions to be maximally useful, they should cover the dominant canopy tree species that occur within a site. There is a tradeoff between the filtering steps described above to strive for accurate matches with canopy trees versus a desire to include as many species as possible. We compared our final filtered data to all field-collected tree species to assess the proportion of field-estimated tree species richness at the site captured by the model. We calculated the proportion comparing image-predicted species data to 1) all species - every record in the field collected data with at least 2 samples, 2) canopy species - the data filtered to 3 m height, and labeled as visible in the canopy in NEON field-surveys, 3) individuals - the

proportion of individuals in the training data captured by the species in the model. For example, if we had 100 individuals in a geographic site in the original field data, with 97 individuals coming from species A and 3 individuals from species B, and the model only contained species A, the proportion of species covered would be 0.5, but the proportion of individuals would be 0.97.

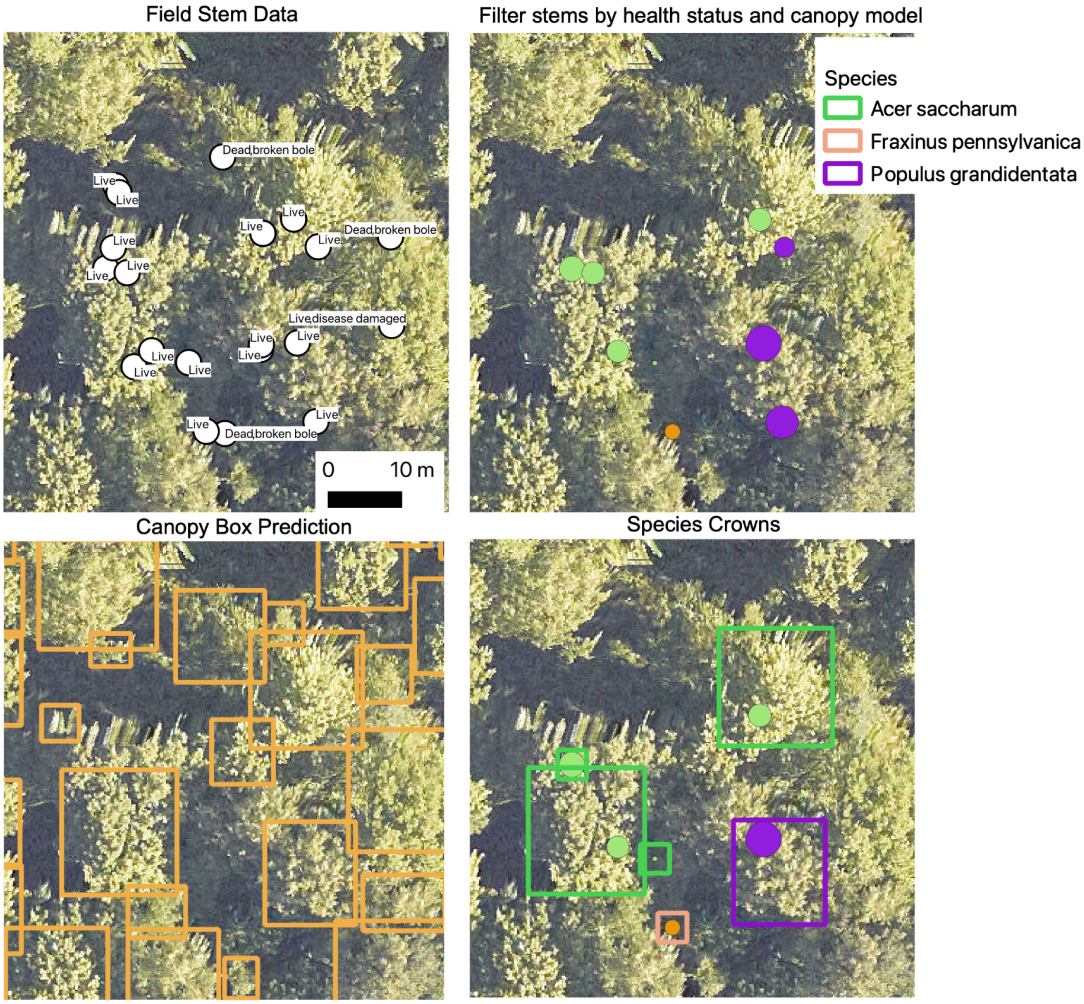


Figure 2. Example workflow for filtering stem data to associate with crown pixel area. Size of the dots in panels b and d are proportional to the individual tree DBH.

Crown Prediction

The DeepForest algorithm used in this work was first proposed in Weinstein et al. (2019) using a combination of hand-annotated tree crown delineations and large-scale synthetic pretraining data using LiDAR-derived tree locations. Weinstein et al. (2020a) and Weinstein et al. (2020b) compared the performance of tree detection algorithms across NEON sites and released the DeepForest model as an open-source python package with an average recall of 72%. Recall was measured using intersection-over-union, a common object detection metric, with a threshold for overlap of 0.4 for a positive match between predicted crown box and hand-annotation. In Weinstein et al. (2021), we released a dataset of 100 million crowns and calculated the performance of our workflow in matching crown predictions to individual trees by scoring the proportion of field stems that fall within a prediction. Field stems can only be applied to one prediction, so if two predictions overlap over a field stem, only one is considered a positive match. The average stem recall was 69.4%, with better performance in well-spaced western forests, and weaker performance in alpine conifer forests. DeepForest has been used widely outside of NEON sites (Reiersen et al. 2022, Kapil et al. 2022, Velasquez-Camacho et al. 2023, Kwon et al. 2023), with accuracies generally mirroring ~70% for fine-tuned models from independent analysis (Gan et al. 2023).

We follow the workflow described in Weinstein et al. (2021), with tree crowns less than 3 m maximum height in the LiDAR-derived canopy height model removed. Each predicted crown in the RGB imagery had a unique ID, predicted crown location, crown area and confidence score from the DeepForest tree detection model. Following tree detection, we classified each predicted crown as ‘Alive’ or ‘Dead’ based on the RGB data. Presented in Weinstein et al (2023), this Alive-Dead model is a two class resnet-50 deep learning neural network trained on hand-annotated images from across all NEON sites. During prediction, the location of each predicted crown was cropped and passed to the Alive-Dead model for labeling as Alive (0) or

Dead (1) with a confidence score for each class. Combining the information from the crown prediction, alive/dead prediction and species classification, we release shapefiles for each 1 km NEON HSI tile that has overlapping RGB and LiDAR data (Table 1).

Table 1. Data available for each predicted crown. Crowns are organized into 1 km shapefiles with UTM projection and follow the naming scheme from NEON's AOP data, with a geographic index at the top left corner. For sites with fewer than five species, the broadleaf and conifer labels are not available, as they are largely redundant with the species present and were all modeled jointly.

Column Name	Definition
Geometry	A four pointed bounding box location in UTM coordinates.
indiv_id	A unique crown identifier that combines the year, site and geoindex of the NEON airborne tile (e.g. 732000_4707000). The UTM coordinate is the northwest corner of the tile.
sci_name	The full latin name of predicted species aligned with NEON's taxonomic nomenclature.
ens_score	The confidence score of the species prediction. This score is the output of the multi-temporal model for the ensemble hierarchical model.
bleaf_taxa	Highest predicted category for the broadleaf model
bleaf_score	The confidence score for the broadleaf taxa submodel
oak_taxa	Highest predicted category for the oak model
dead_label	A two class alive/dead classification based on the RGB data. 0=Alive/1=Dead.
dead_score	The confidence score of the Alive/Dead prediction.
site_id	The four letter code for the NEON site. See https://www.neonscience.org/field-sites/explore-field-sites for site locations.
conif_taxa	Highest predicted category for the conifer model
conif_score	The confidence score for the conifer taxa submodel
dom_taxa	Highest predicted category for the dominant taxa mode submodel
dom_score	The confidence score for the dominant taxa submodel

Species Prediction

To train species classification models, we opted to build a different model for each NEON site to create the best possible set of species predictions for downstream ecological analysis. To classify each predicted crown tree crown to species, we use the 1-m hyperspectral data and a multi-temporal hierarchical model. Weinstein et al. (2023) found that a hierarchical model outperforms a flat model by improving rare species accuracy. The hierarchical model organizes tree species into submodels, allowing each model to learn better features related to distinguishing similar classes. The submodels also allow species that are well sampled to be separated from poorly sampled species, thereby reducing the effect of class imbalance in favoring common species (Liu et al. 2019). Within each submodel, we combine predictions for each year of available sensor data to reduce the potential overfitting and bias due to georectification of ground-truth trees and image acquisition conditions. The top model predicts ‘Broadleaf’, ‘Conifer’ and optionally the dominant tree species class at that site based on its frequency in the training data. A species was considered ‘dominant’ if it consisted of more than 40% of the training samples. Without this, common machine learning approaches will predict most samples as the dominant class regardless of spectral signal. After prediction in the first subgroup, samples that are predicted as ‘Broadleaf’ are then passed to the Broadleaf submodule, and samples that are predicted as ‘Conifer’ are then passed to the Conifer submodule. This structure was maintained for the majority of sites, but we did allow some site specific customization. For example at the Ordway Swisher Biological Station, Florida (OSBS) site, the many similar oak congenerics were split off into their own oak submodule within the broadleaf submodule.

Each submodule consists of a 2D spectral attention block (Figure 1) with three convolutional layers and a max pooling spectral attention layer following Hang et al. (2021). Batch normalization is used to normalize layer weights after each convolution. This spectral

attention block was repeated for each year of airborne sensor data to create an ensemble model. For example, if there are four years of available hyperspectral data for a geographic location, we predicted four classification outputs and then combined them to create the final prediction. This is based on the assumption that canopy trees at each geographic location are unlikely to change species label among years at short time scales (Busing 2005). A weighted average among all years was used to create the sample prediction for each crown. This relative weight among years was a learned parameter for each submodel. Despite multiple publications that highlight performance gains through multi-modal data fusion in remote sensing classification (Liao et al. 2018, Ghamisi et al. 2018, Sumbul et al. 2019), we did not find significant improvements when adding the 10 cm RGB data to species classification (Figure S1), but continue to believe it will have a role in distinguishing similar species.

For each site, we pretrained the hierarchical model using data from all sites, but only including the species at the focal site. We then fine-tuned this model using samples only at the target site. We experimented with a single NEON-wide model across all sites, but found consistently worse performance, especially for rare species (Figure S2). For each site we pretrained for 200 epochs, decreasing the learning rate of each submodel based on performance on the focal site test data. We then fine-tuned this model with the available annotations at the target site for 200 epochs. Learning rates differed among submodules, with the dominant class and conifer submodules having an initial learning rate of 10e-5, and the broadleaf model starting at 10e-4. We allowed batch size to vary between 12 and 24 depending on the site to account for differences in class imbalance and dataset size.

To determine the evaluation accuracy of species predictions, we developed a train-test split with a minimum of 10 samples per class. To minimize the potential effect of spatial autocorrelation in hyperspectral signature between training and test datasets, we adopted a spatial block approach (Fassnacht et al. 2016). All samples within a NEON plot, or within a 40-m grid for the non-NEON contributed data were assigned to training or test. We performed this

assignment iteratively until the minimum number of samples per class were in the test dataset. The remaining samples were used to train the model. For each site we evaluated the accuracy and precision of each species. To get the site-level score, we used both micro-averaged accuracy and macro-averaged accuracy. Micro-averaging weights all samples the same, and therefore is largely driven by the performance of the common species. Macro-average weights all species the same, giving greater importance to the rare species as compared to their frequency in the dataset. We also computed the accuracy of the higher order taxonomic labels (e.g. 'Broadleaf' vs 'Conifer'), which may be useful to downstream applications in which coarser categories are sufficient.

Results

We developed individual canopy tree species predictions for 81 species at 24 NEON sites (Table 2). There was an average of 6.56 species per site, with a maximum of 15 species (Harvard Forest, Massachuesetts) and minimum of 3 (Delta Junction, Alaska and San Joaquin Experimental Range, California). Compared to reference species lists filtered for canopy species, the crown data set covered 47.5% of the total species richness for trees ≥ 10 cm dbh represented in the reference list at the sites (Figure 3). These species account for an average of 85.0% of the stems ≥ 10 cm dbh from the forest plot data at the NEON sites. The average model had a micro-averaged accuracy of 78.8% and a macro-accuracy of 75.8% (Table 2). Sites with more data generally performed well, with a general pattern of decreasing species-level accuracy with fewer data (Figure 4). Consistent with previous work, the highest performing sites, including Teakettle Canyon, CA (TEAK), Niwot Ridge Colorado (NIWO), and Yellowstone National Park, Wyoming (YELL), were dominated by conifers and had relatively low species diversity (Marconi et al 2022). Models performed more poorly in southern broadleaf forests, such Talladega National Forest, Alabama (TALL) and Smithsonian Environmental

Research Center, Maryland (SERC), with higher biodiversity, closed canopy structure, and/or low data coverage per species (Marconi et al 2022). The most abundant species at a site typically had the highest accuracy, with lower accuracy for rarer species (Figure 4).

Applying the best model for each site to all available airborne tiles, we predicted 103,441,970 trees with an average of 4.31 million trees per site. Of the 24 sites, 17 are heavily forested with near continuous canopy cover. Sites vary in both area and forest density, with the smallest size in San Joaquin Valley, CA (SJER) with 0.85 million trees predicted, and the largest site in TreeHaven, Wisconsin with 7.1 million trees predicted. The sites with the most predicted trees tend to have high species diversity at local scales with complex, overlapping crown boundaries (Figure 5). Patterns of biodiversity are highly scale dependent with grouping of similar species in local areas and complex patterns of species patches at broader scales within the same site (Figure 6). Ranking the predicted species abundance for each site, the most commonly predicted species represented approximately 60% of crown classifications (Figure 7). The dominant species was slightly less abundant in the southern broadleaf sites with 30% to 40% of crowns belonging to the most commonly predicted species. Viewing the predictions at the largest spatial extents, there is a broad range of species presence patterns, from sites showing highly mixed species to sites with distinct autocorrelation and species patterns at all spatial scales (Figure 8).

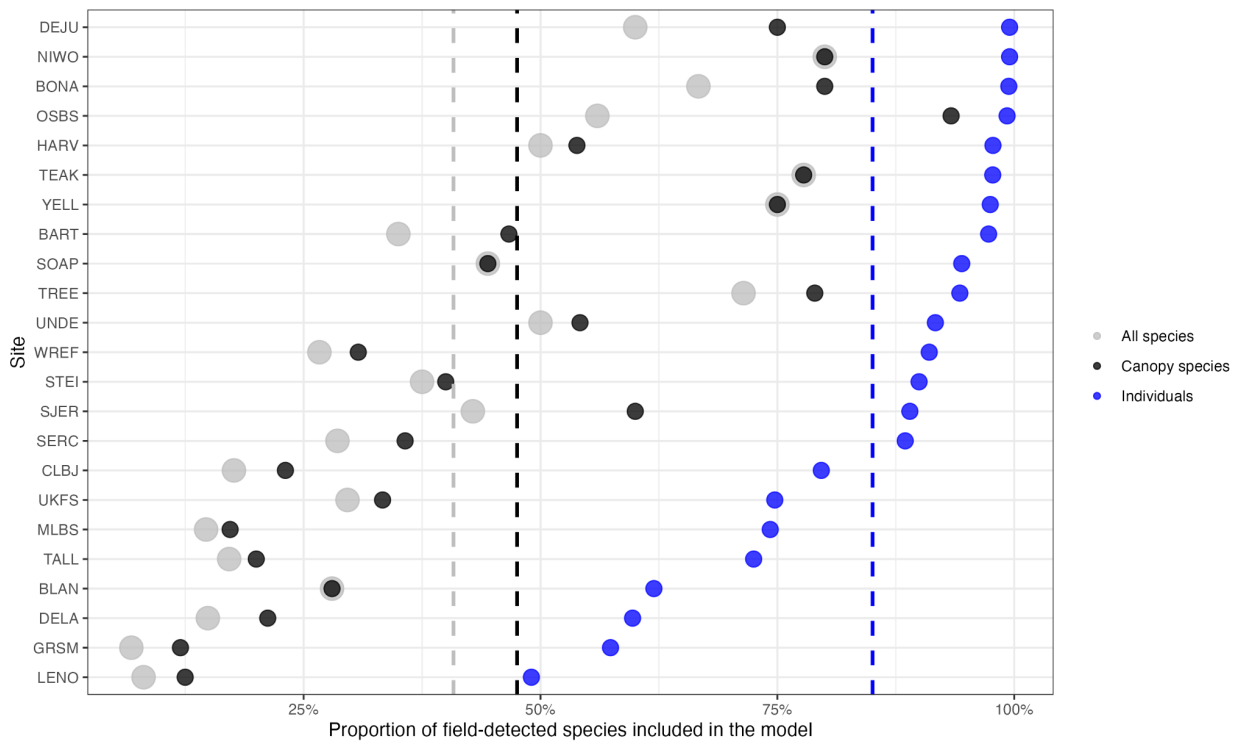


Figure 3. The proportion of species included in the model for each site compared to species with at least two records in the field-collected data. We calculated the proportion compared to 1) all species - every record in the field collected data with at least 2 samples, 2) canopy species - the data filtered to 3 m height, and labeled as visible in the canopy in NEON field-surveys, 3) individuals - the proportion of individuals in the training data captured by the species in the model. For example, the BART model has 35% of species found during field-surveys, 46% of the species judged to be in the canopy, but these species represent over 97% of the sampled individuals at the site. For a complete list of each species in the model and the canopy-filtered data, see Table S1. The dashed line is the mean number of species across sites for both species and individual proportions.

Table 2. Evaluation scores for each NEON site included in the dataset. Sites are ranked from highest to lowest micro accuracy

Site, State	Forest Description	Micro Accuracy	Macro Accuracy	Species	Train Samples	Test Samples
SJER, CA	Oak Savannah	1.00	1.00	3	47	27
GRSM, NC	Southern Hardwoods	0.90	0.89	3	200	29
TEAK, CA	Western Conifer	0.82	0.83	7	713	67
BONA, AK	Riparian and Taiga	0.82	0.74	4	584	103
STEI, MI	Northern Hardwoods	0.80	0.83	6	283	82
NIWO, CO	Alpine Conifer	0.80	0.77	4	852	46
YELL, WY	Western Conifer	0.80	0.83	3	390	10
SERC, MD	Southern Hardwood	0.80	0.68	11	816	287
DELA, AL	Southern Hardwood	0.79	0.79	7	166	72
DEJU, AK	Taiga	0.79	0.78	3	571	52
UNDE, WI	Northern Hardwood	0.79	0.79	13	547	178
SOAP, CA	Western Conifer	0.78	0.78	4	223	37
MLBS, VA	Southern Hardwood	0.78	0.75	5	363	54
TREE, MI	Northern Hardwood	0.78	0.72	15	643	168
WREF, WA	Western Conifer	0.76	0.66	4	598	97
TALL, AL	Southern Hardwood	0.76	0.72	6	250	125
HARV, MA	Northern Hardwood	0.76	0.57	15	9782	1194

OSBS, FL	Oak Savannah,	0.73	0.63	14	3293	240
CLBJ, TX	Oak Savannah	0.73	0.73	3	187	30
BLAN, VA	Riparian	0.72	0.73	8	271	79
LENO, AL	Southern Hardwood	0.71	0.71	3	74	28
RMNP, CO	Alpine Conifer	0.70	0.70	7	671	99
BART, VT	Northern Hardwood	0.68	0.66	7	514	125
UKFS, KT	Southern Hardwood, Riparian	0.60	0.60	8	204	85

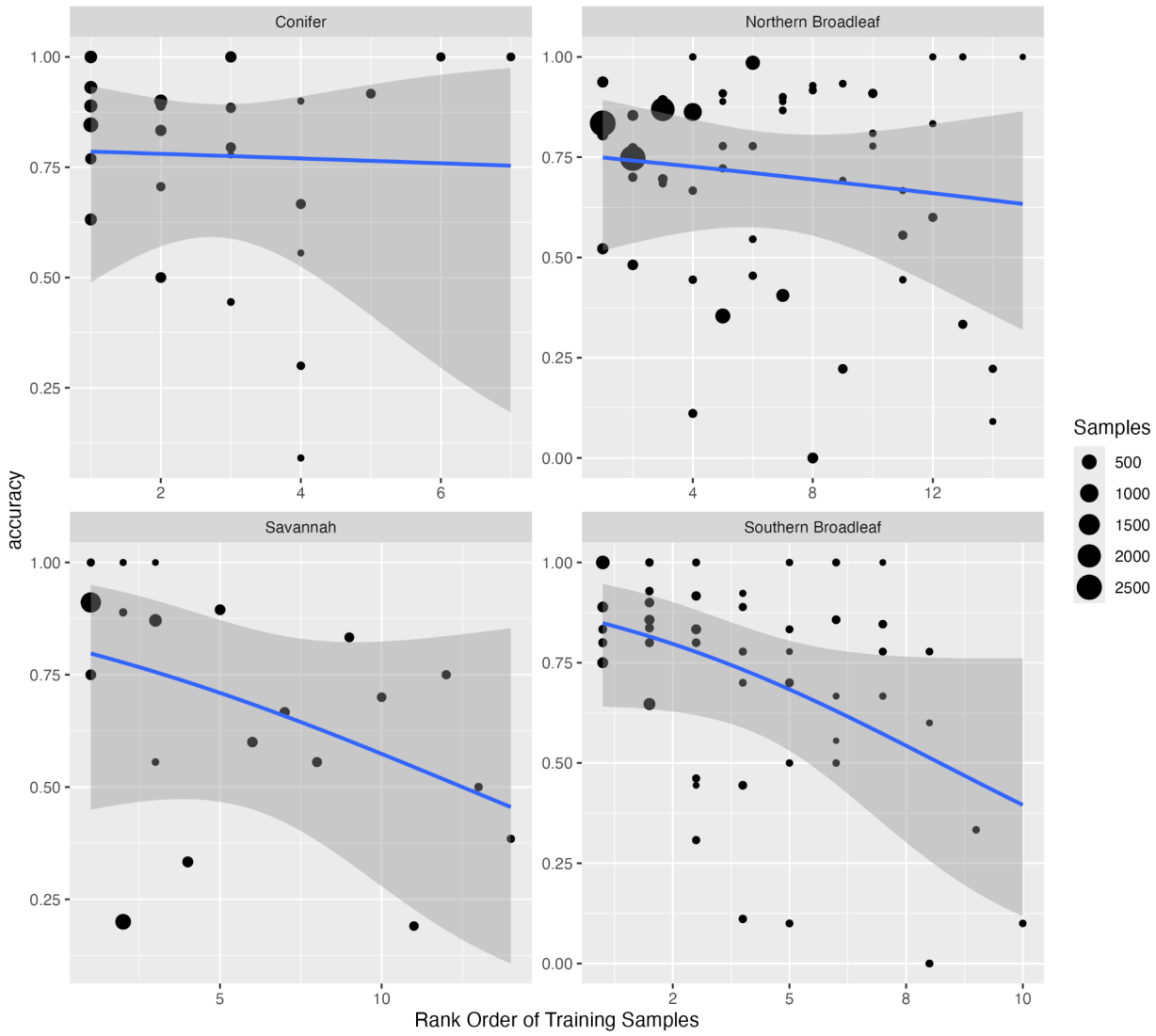


Figure 4. Rank order abundance and evaluation accuracy for each species for each NEON site aggregated by forest type. A binomial classification model was fit for each forest type to relate the rank order abundance of each species and evaluation accuracy. Each point is one species within one NEON site model. Point size is relative to the abundance of the species at the individual site in the training data.

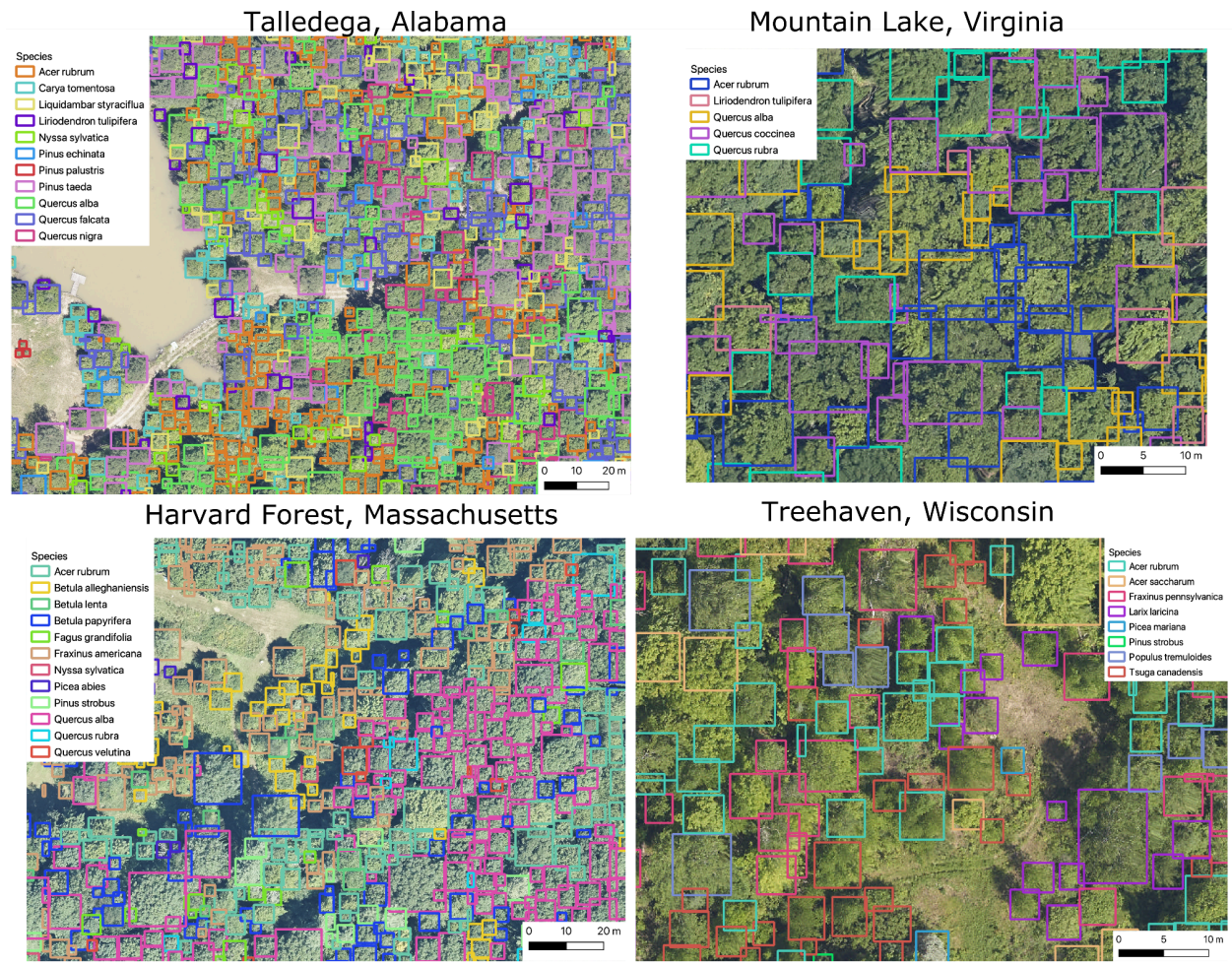


Figure 5. Example tree detections and species labels for four NEON sites with closed canopy deciduous forests.

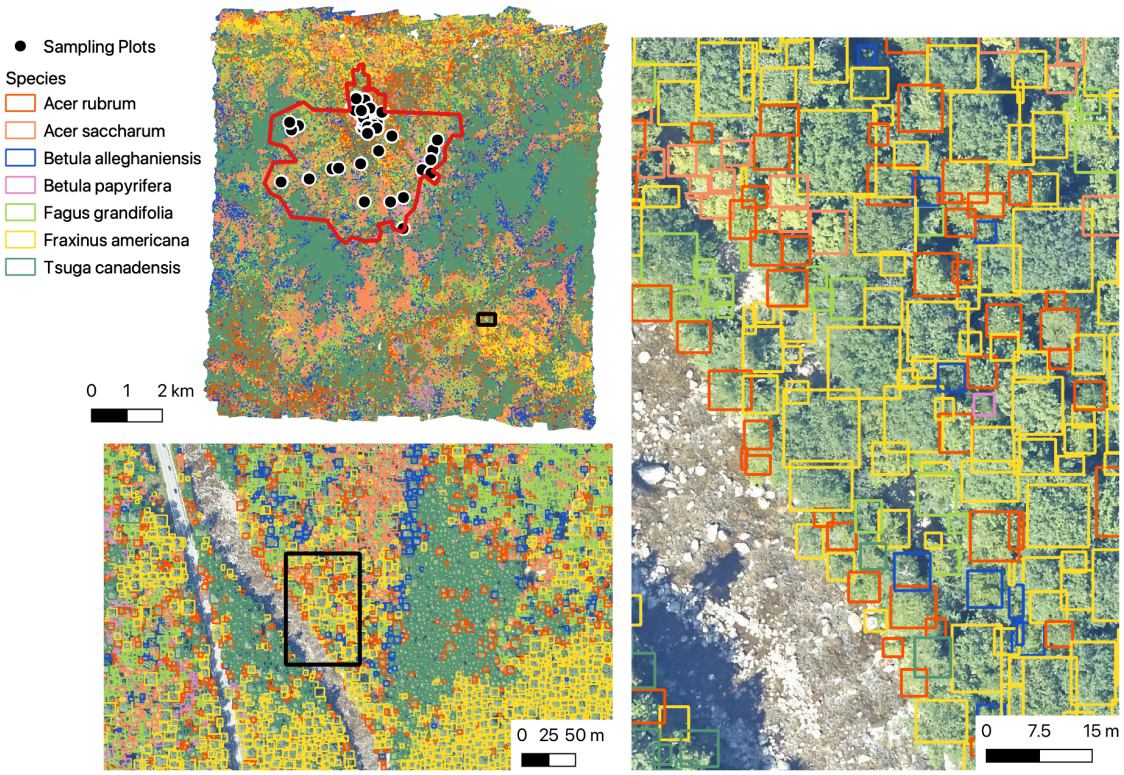


Figure 6. Overview from Bartlett Experimental Forest, New Hampshire (BART) showing 4,352,930 tree predictions for seven species at three spatial scales. The location of NEON sampling plots and the NEON boundary are shown in the top left image.

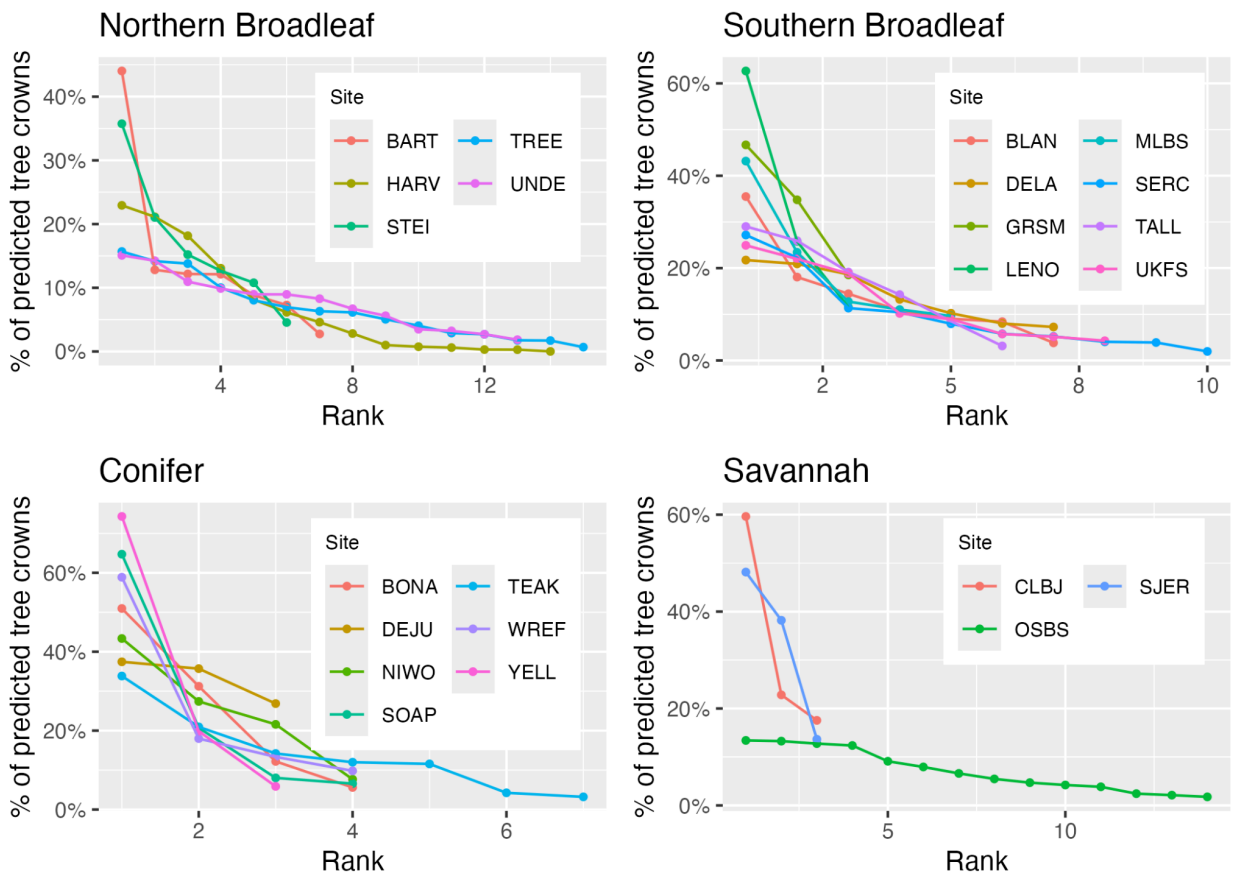


Figure 7. Rank order abundance for the predicted crown species labels for each site. The most commonly predicted species is rank 1, the second most commonly predicted species is rank 2, etc. Each point represents a species predicted at a site. For species identity and totals per site see Table S1.

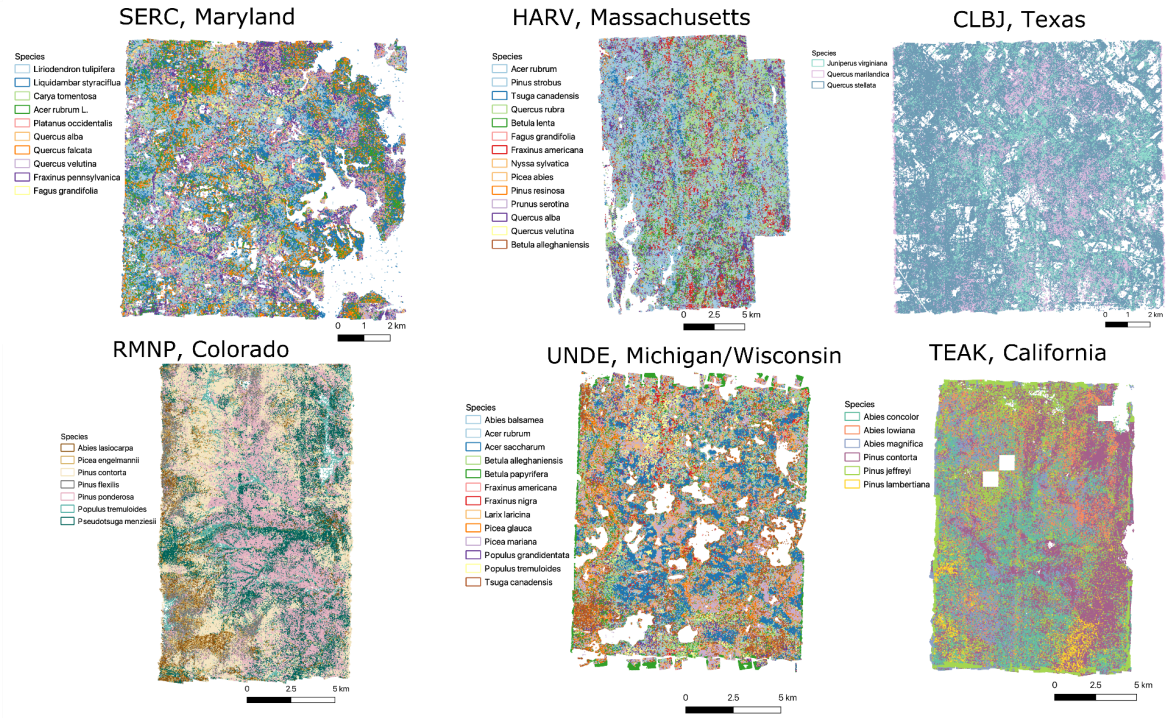


Figure 8. Overview of multiple sites spanning a broad range of forest types. Site names, from top left to bottom right, Smithsonian Environmental Research Center (SERC), Harvard Forest (HARV), Lyndon B. Johnson National Grassland (CLBJ), Rocky Mountain National Park (RMNP), University of Notre Dame Environmental Research Center (UNDE), Teakettle Canyon (TEAK).

Discussion

We used a multi-step deep learning workflow to generate individual level canopy tree species predictions continuously across large landscapes in a diverse array of forest types at sites within the National Ecological Observatory Network. The result is an extensive dataset on individual canopy tree species distribution that can be used for studying large scale forest ecology, used as a baseline dataset for guiding field sampling, and integrated into larger scale remote sensing tasks as training data for satellite based models. These data will inform a broad array of

research programs, for example community ecologists can study the patterns of species distributions as a function of environmental and biotic interactions (Peterson et al. 2011, Condit et al. 2013), the phylogenetic structure of tree assemblages (Cavender-Bares et al. 2004), and scale dependance of species plant communities (Freestone and Inouye 2006); ecosystem scientists can improve estimates of biomass using species-specific allometry (Martínez-Sánchez et al. 2020, Wallis et al. 2023), and foresters can measure impacts of habitat disturbance and landscape history (Duncanson et al. 2015, Hemming-Schroeder et al. 2023). To facilitate the broad use of this dataset we have uploaded the dataset to Google Earth Engine, which provides tools and computational resources that facilitate very large scale data analysis integrating numerous remote sensing assets that are collectively stored in the Earth Engine catalog.

The species classification models used to generate this dataset generally performed well with the accuracy for most common species ranging from 75 to 85% at well-sampled, diverse sites. Repeating a general model architecture for tree species prediction across a broad array of sites, revealed several general tendencies in the accuracy of predicted tree crowns including: 1) decreased accuracy with an increasing number of species, 2) higher accuracy at sites with more open canopy structure, and 3) a general tendency of higher performance for conifer over broadleaf species. This led to geographic patterns in accuracy even among sites in similar ecosystems, with northern broadleaf sites in general having better accuracy than the more diverse southern broadleaf sites. As local species diversity increases, classification errors are more likely due to increased numbers of model parameters (leading to potential overfitting), greater complexity in splitting similar species, and increased frequency of neighboring trees being from different species resulting in pollution of crown edge pixels. High local turnover may also decrease accuracy because it makes training data taken from a subset of the predicted region less representative of the total biodiversity and spectral background. For example, unique habitats in the remote sensing footprint appear to be more well sampled by NEON's

terrestrial plot design (Barnett et al. 2019) in ‘Northern Broadleaf’ forests than in ‘Southern Broadleaf’ forests, likely due to the northern forests being more admixed.

Data derived from airborne remote sensing should be seen as a complement to, not a replacement for, field data. While the dataset will facilitate capturing dynamics at scales infeasible for ground-based surveys, we stress that the data are imperfect predictions that can, and should, be improved with increased data collection and model exploration. Because of the nature of the airborne data, the dataset only includes crowns in the top layer of the canopy (sunlit tree crowns), and users should be careful when calculating stand-level metrics such as abundance, crown area, or DBH and comparing them to ground-data that includes smaller subcanopy trees. We expect that, compared to field surveys, this canopy dataset will include fewer, larger, trees (due to the lack of subcanopy trees; Figure S3; Weinstein 2021). There will also be fewer species represented in the dataset than observed in the field, in part because subcanopy only species are explicitly excluded from the model (Figure 3).

In addition to the restriction to canopy trees, each part of the workflow has associated uncertainty and tradeoffs in defining fixed labels. DeepForest, the crown detection algorithms, has been evaluated against hand-annotated imagery (Weinstein et al. 2021a), field-stem recall (Weinstein et al. 2021b) and images-drawn by observers on tablets directly in the field (Weinstein et al. 2020a), and consistently found to have roughly 70 to 75% accuracy for crown delineation. Errors occur due to over segmentation (one tree is identified as multiple trees), under segmentation (two or more trees are identified as a single tree), and imprecisely defined crown edges (Weinstein 2022). In general, counts of canopy trees on a landscape are often more accurate (because over and under segmentation errors cancel out), but detailed boundaries and crown area are less accurate. Beyond tree detection, the alive/dead label should be interpreted as provisional since trees can lose leaves due to a variety of causes such as insect defoliation in one year, but ultimately recover over time (Atkinson et al. 2014). Species predictions are also uncertain, and while they include the most common species at each site,

they still fail to include a number of species that do occur in the canopy (Figure 3). The discrepancy between canopy species in the filtered field data set and species predicted in our model is a result of several factors. Some canopy species are rare and thus have too few samples in our data set to be included. This may be due to species that are common but only in rare habitats, or are rare throughout a broad area of each site. On the other hand, some species may be common, but are shorter statured species that tend to mostly be in the subcanopy or only rarely reach in the canopy. When they do reach the canopy, the crowns are very small, providing poor spectral signature. Some canopy species are rare (either throughout the entire region or only occurring in rare habitats) and thus have too few samples to be modeled.

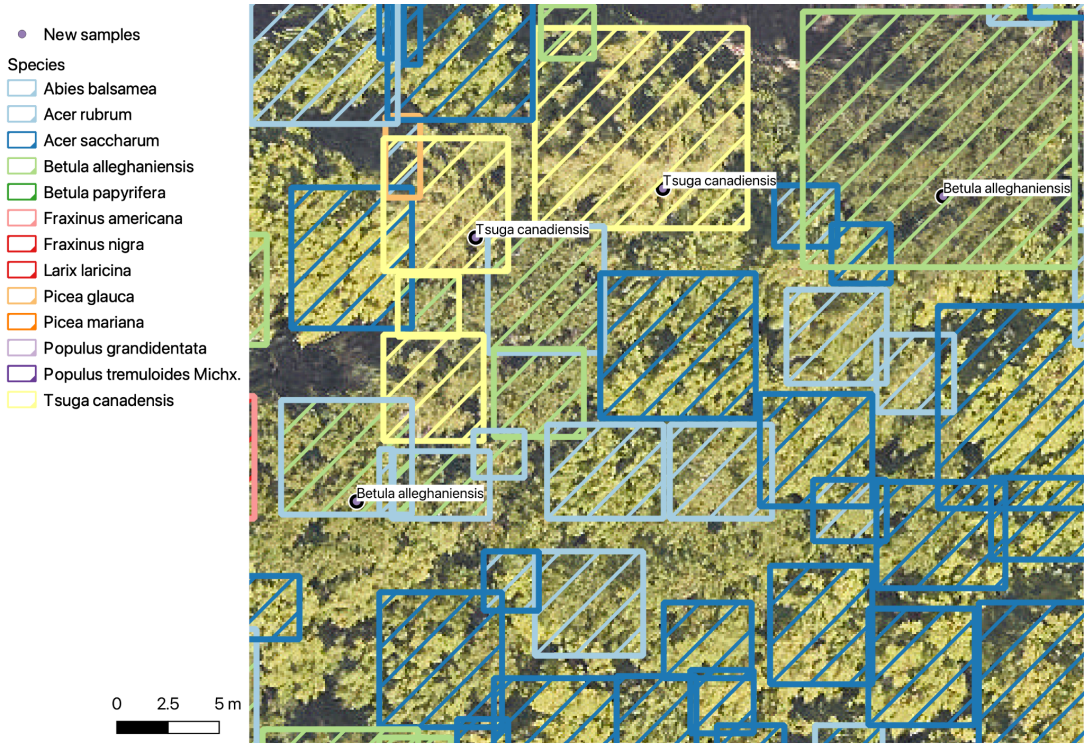
Given the uncertainties inherent in creating large scale species maps it is important to consider potential approaches for incorporating this uncertainty in analyses involving this and similar datasets. Weinstein et al. (2023) outlined multiple options for incorporating model uncertainty when using the data in downstream analysis. We compared data uncertainty through multiple training and test splits, model uncertainty by repeatedly training the model from the same training data, and prediction uncertainty using a multinomial draw of the confusion matrix to generate predicted counts for each species within a single site. While this is a useful first step, ultimately hierarchical models that can directly incorporate model uncertainty should be developed to improve downstream ecological analyses of remote sensing based data (e.g. Augustine et al. 2023). Calibrating confidence scores using held-out data from training or test is an important step in this direction (Guo et al. 2017), but there was insufficient data to set aside for this purpose while maintaining less common species in the model. This will be a common limitation in ecological studies where the limited data can be crucial for improving model accuracy and incorporating rarer species. Post Hoc corrections of predicted counts (e.g. Orenstein et al. 2020) or models that account for multiple types of uncertainty will be crucial in making robust predictions at larger spatial extents going forward.

The process of making predictions for 100 millions trees across a broad range of habitat types helped identify areas for improvement in computer vision needed to address obstacles in assembling tree maps at massive scales. The main obstacle to improving model accuracy is the availability of training data. We have found that targeted sampling can yield 10 to 20% improvements in accuracy, and significantly broaden the number of species included in the model predictions, with only a few days or weeks of field work (Box 1). The simplest form of data needed is a geospatial point of a tree stem (precise enough to ensure it falls within a predicted crown box) and its species label. Data collection should focus on less common species, since more data on common species will have limited impact on model performance. Strategies for prioritizing new data collection include; 1) Using expert knowledge to identify areas containing underrepresented species; 2) Using the model confusion matrix and predictions from the initial model to select species with unexpected confusion patterns, such as underrepresented species that are not visually similar that are confused by the model (a possible indicator of spectra being polluted by neighboring trees); and 3) Sampling individuals with low confidence scores for their species predictions indicating either poor model performance or a species not included in the model.

There are also areas for improvement in associating tree stems with crown pixels. Our models perform better in open forests with low diversity, where spacing among trees improves crown delineation and fewer species reduces the chance of neighboring tree species polluting the spectral signature. This can be partially overcome by using crown polygons drawn on a tablet in the field, rather than relying on stem points taken by a GPS (Graves et al. 2022). Even a limited number of these crown polygons could allow the adoption of ‘weak labeling’ approaches common in computer vision that rely on access to a small number of confident samples and a larger set of less confident samples (Northcutt et al. 2020).

Box 1. In-depth examination of new data collection to improve models.

To increase the species coverage and accuracy of these models, we need additional data collection at each NEON site. Here, we outline one effort by N.G. Swenson and V.E. Rubio to improve the model at the University of Notre Dame Environmental Research Center (UNDE) site through targeted data collection (Box 1; Figure 1). The original model had 67.8% micro-accuracy, 61.6% macro-accuracy, and included 12 species. Overlaying the predictions over a recently mapped forestry plot, three areas of need were identified, 1) several key species were missing from current predictions, 2) there was overprediction of *Fraxinus nigra* compared to the abundance expected by field researchers, 3) there was high confusion between two closely related *Populus* species. Using these goals to target trees, data on an additional 157 stems of twelve species were collected along easy to access roads and forest edges. After training the model on the additional stems, the micro-averaged accuracy increased from 67.8% to 77.7% and the macro-averaged accuracy increased from 61.6% to 79.1% while adding an additional species to the test dataset. The accuracies of the two closely related *Populus* species increased from 66% and 54% to 72% and 82%, respectively.



Box 1 Figure 1. Original versus re-trained model predictions for UNDE. New sample trees were collected in the field without guidance from the predictions. The outline color is the original label, the filled shade is the revised label. The two *Tsuga canadensis* (top center) and the field samples were correctly predicted in the original model. The *Betula allenghensis* field samples were split. The tree on the right was correctly predicted in both models. The tree on the left was originally predicted as *Acer rubrum*, but was correctly predicted in the revised model. Overall, most labels do not change among models, with only a small number of trees changing labels. For example, several trees that were originally predicted as *Acer rubrum* have been revised, and a single *Picea glauca* was revised to *A. rubrum* in the top left.

One of the reasons additional data collection can be so beneficial is that compared to the typical computer vision application, the data sample sizes of the classes used in these models are extremely low. Therefore the emerging area of research on 'few shot learning', in which foundation models are used to predict new classes with only 1 to 5 samples, may be a useful avenue for further improving tree species predictions (e.g. Sohn et al. 2020). In the extreme, the task of zero-shot learning (Sumbul et al. 2018, Stork et al. 2021), or unknown class detection, in which the model can identify classes not included in training, will help address the challenge of identifying individuals not included in the models and have utility in rapid applying models trained on NEON data to new areas. This approach is limited by our current modeling design since the site-level model approach limits portability, and the hierarchical organization can be cumbersome to apply in new regions and as new species are added. While we chose this approach because it currently produces the most accurate predictions and therefore the best resulting dataset, a single NEON-wide model that is robust to class imbalance, but maintains good separability among co-occurring species, would be a major step forward.

Extending the models used in our workflow to non-NEON sites will be important in broadening access to high quality tree species prediction. There is considerable interest in developing species predictions for large areas using high resolution satellites and UAVs with low

cost hyperspectral sensors. Using NEON data as a source for training data to project into these coarser resolution data has large benefits since the NEON data is both high spectral and spatial resolution. This kind of ‘’Domain adaptation’ is an open challenge in computer vision, with many proposed approaches to try to align either the input data or learned features among disparate sensors or geographic areas (Koh et al. 2021). The ample unlabeled airborne data at NEON opens the possibility of a combination of supervised and unsupervised learning to increase transferability among geographic sites, spectral resolutions and spatial scales. In conjunction with automated methods for data collection, these approaches will move the community towards airborne classification models for tree species that can generalize across sampling events, geography, and acquisition hardware.

As the number of researchers working at NEON sites increases, the diversity of overlapping datasets will foster richer areas of understanding for forest ecology and ecosystem functioning. The goal of this work was to provide initial predictions for canopy trees at the landscape scale to document the broad pattern of tree species distributions, which in turn influence ecological communities and nutrient cycling. Combining these data with organismal surveys, fine-scaled environmental data, and landscape history will bring greater insights into the mechanisms underlying forest distribution and function. NEON’s on-going data collection will allow these maps to be updated both in terms of geographic coverage, as well as temporal change in species abundance and individual traits.

Acknowledgements

We would like to thank NEON staff and in particular Tristan Goulden and Courtney Meier for their assistance and support. We thank Natalie Heaton, Nicollete Lyons, Matthew Raulerson, Alex Seeley, Camille Sicangco, Luis Tirado and Stuart Wilkin and for field data collection efforts. This research was supported by the Gordon and Betty Moore Foundation’s Data-Driven

Discovery Initiative (GBMF4563) to EP White, by the USDA National Institute of Food and Agriculture McIntire Stennis project 1024612 and the Forest Systems Jumpstart program administered by the Florida Agricultural Experiment Station to SA Bohlman, and by the National Science Foundation (1926542) to EP White, SA Bohlman, A Zare, DZ Wang, and A Singh. This work was supported by the USDA National Institute of Food and Agriculture, Hatch project FLA-WEC-005944. PT acknowledges funding support from NSF Macrosystems Biology and NEON-Enabled Science (MSB-NES) award DEB 1638720 and NSF ASCEND Biology Integration Institute (BII) through DBI award 2021898. SR acknowledges funding support from NASA award 80NSSC23K0421 P00001 and Hatch project number ME022425. NGS and VER were supported by funding from NASA (80NSSC22K1625) and NSF Dimensions of Biodiversity (DEB-2124466). RAA was supported by the NWT LTER (NSF DEB-2224439), USDA NIFA McIntire Stennis project (1019284), and USDA NIFA postdoctoral award (2022-67012-37200).

Data Availability

The predictions, training data crops and shapefiles with predicted training crowns are available at https://zenodo.org/record/10067302. A web visualization is available to preview predictions over RGB imagery: https://visualize.idtrees.org/. A csv file per site was uploaded to Google Earth engine and a public link is available as a FeatureCollection. For example, 'https://code.earthengine.google.com/?asset=users/benweinstein2010/RMNP' is the RMNP, Rocky Mountain National Park, predictions. For more on using NEON data and earth engine, see https://www.neonscience.org/resources/learning-hub/tutorials/intro-aop-gee-image-collections. The code used in this manuscript is available both as an archive resource on zenodo (https://zenodo.org/records/10689811) and as a github repository (https://github.com/weecology/DeepTreeAttention).

References

Atkinson, R. R. L., M. M. Burrell, K. E. Rose, C. P. Osborne, and M. Rees. 2014. The dynamics of recovery and growth: how defoliation affects stored resources. Proceedings of the Royal Society B: Biological Sciences 281:20133355.

Aubry-Kientz, M., R. Dutrieux, A. Ferraz, S. Saatchi, H. Hamraz, J. Williams, D. Coomes, A. Piboule, and G. Vincent. 2019. A Comparative Assessment of the Performance of Individual Tree Crowns Delineation Algorithms from ALS Data in Tropical Forests. Remote Sensing 11:1086.

Audebert, N., B. Le Saux, and S. Lefevre. 2019. Deep Learning for Classification of Hyperspectral Data: A Comparative Review. IEEE Geoscience and Remote Sensing Magazine 7:159–173.

Augustine, B. C., M. D. Koneff, B. A. Pickens, and J. A. Royle. 2023, February 21. Towards estimating marine wildlife abundance using aerial surveys and deep learning with hierarchical classifications subject to error. bioRxiv.

Ayres, E., A. Colliander, M. H. Cosh, J. A. Roberti, S. Simkin, and M. A. Genazzio. 2021. Validation of SMAP Soil Moisture at Terrestrial National Ecological Observatory Network (NEON) Sites Show Potential for Soil Moisture Retrieval in Forested Areas. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14:10903–10918.

Barnett, D. T., P. A. Duffy, D. S. Schimel, R. E. Krauss, K. M. Irvine, F. W. Davis, J. E. Gross, E. I. Azuaje, A. S. Thorpe, D. Gudex-Cross, M. Patterson, J. M. McKay, J. T. McCorkel, and C. L. Meier. 2019. The terrestrial organism and biogeochemistry spatial sampling design for the National Ecological Observatory Network. Ecosphere 10:e02540.

Bosch, M. 2020. DetecTree: Tree detection from aerial imagery in Python. Journal of Open Source Software 5:2172.

Busing, R. T. 2005. Tree mortality, canopy turnover, and woody detritus in old cove forests of the southern Appalachians. Ecology 86:73–84.

Cavender-Bares, J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz. 2004. Phylogenetic Overdispersion in Floridian Oak Communities. The American Naturalist 163:823–843.

Condit, R., B. M. J. Engelbrecht, D. Pino, R. Pérez, and B. L. Turner. 2013. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. Proceedings of the National Academy of Sciences 110:5064–5068.

Davies, S. J., I. Abiem, K. Abu Salim, S. Aguilar, D. Allen, A. Alonso, K. Anderson-Teixeira, A. Andrade, G. Arellano, P. S. Ashton, P. J. Baker, M. E. Baker, J. L. Baltzer, Y. Basset, P. Bissiengou, S. Bohlman, N. A. Bourg, W. Y. Brockelman, S. Bunyavejchewin, D. F. R. P. Burslem, M. Cao, D. Cárdenas, L.-W. Chang, C.-H. Chang-Yang, K.-J. Chao, W.-C. Chao, H. Chapman, Y.-Y. Chen, R. A. Chisholm, C. Chu, G. Chuyong, K. Clay, L. S. Comita, R. Condit, S. Cordell, H. S. Dattaraja, A. A. de Oliveira, J. den Ouden, M. Detto, C. Dick, X. Du, Á. Duque, S. Ediriweera, E. C. Ellis, N. L. E. Obiang, S. Esufali, C. E. N. Ewango, E. S. Fernando, J. Filip, G. A. Fischer, R. Foster, T. Giambelluca, C. Giardina, G. S. Gilbert, E. Gonzalez-Akre, I. A. U. N. Gunatilleke, C. V. S. Gunatilleke, Z. Hao, B. C. H. Hau, F. He, H. Ni, R. W. Howe, S. P. Hubbell, A. Huth, F. Inman-Narahari, A. Itoh, D. Janík, P. A. Jansen, M. Jiang, D. J. Johnson, F. A. Jones, M. Kanzaki, D. Kenfack, S. Kiratiprayoon, K. Král, L. Krizel, S. Lao, A. J. Larson, Y. Li, X. Li, C. M. Litton, Y. Liu, S. Liu, S. K. Y. Lum, M. S. Luskin, J. A. Lutz, H. T. Luu, K. Ma, J.-R. Makana, Y. Malhi, A. Martin, C. McCarthy, S. M. McMahon, W. J. McShea, H. Memiaghe, X. Mi, D. Mitre, M. Mohamad, L. Monks, H. C. Muller-Landau, P. M. Musili, J. A. Myers, A. Nathalang, K. M. Ngo, N. Norden, V. Novotny, M. J. O'Brien, D. Orwig, R. Ostertag, K. Papathanassiou, G. G. Parker, R. Pérez, I. Perfecto, R. P. Phillips, N. Pongpattananurak, H. Pretzsch, H. Ren, G. Reynolds, L. J. Rodriguez, S. E. Russo, L. Sack, W. Sang, J. Shue, A. Singh, G.-Z. M. Song, R. Sukumar, I.-F. Sun, H. S. Suresh, N. G. Swenson, S. Tan, S. C.

Thomas, D. Thomas, J. Thompson, B. L. Turner, A. Uowolo, M. Uriarte, R. Valencia, J. Vandermeer, A. Vicentini, M. Visser, T. Vrska, X. Wang, X. Wang, G. D. Weiblen, T. J. S. Whitfeld, A. Wolf, S. J. Wright, H. Xu, T. L. Yao, S. L. Yap, W. Ye, M. Yu, M. Zhang, D. Zhu, L. Zhu, J. K. Zimmerman, and D. Zuleta. 2021. ForestGEO: Understanding forest diversity and dynamics through a global observatory network. Biological Conservation 253:108907.

Duncanson, L., R. Dubayah, and B. Enquist. 2015. Assessing the general patterns of forest structure: Quantifying tree and forest allometric scaling relationships in the United States. Global Ecology and Biogeography 24.

Egli, L., K. E. LeVan, and T. T. Work. 2020. Taxonomic error rates affect interpretations of a national-scale ground beetle monitoring program at National Ecological Observatory Network. Ecosphere 11:e03035.

Fassnacht, F. E., H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, C. Straub, and A. Ghosh. 2016. Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment 186:64–87.

Freestone, A. L., and B. D. Inouye. 2006. Dispersal Limitation and Environmental Heterogeneity Shape Scale-Dependent Diversity Patterns in Plant Communities. Ecology 87:2425–2432.

Freudenberg, M., N. Nölke, A. Agostini, K. Urban, F. Wörgötter, and C. Kleinn. 2019. Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net. Remote Sensing 11:312.

Fricker, G. A., J. D. Ventura, J. A. Wolf, M. P. North, F. W. Davis, and J. Franklin. 2019. A Convolutional Neural Network Classifier Identifies Tree Species in Mixed-Conifer Forest from Hyperspectral Imagery. Remote Sensing 11:2326.

Gan, Y., Q. Wang, and A. Iio. 2023. Tree Crown Detection and Delineation in a Temperate Deciduous Forest from UAV RGB Imagery Using Deep Learning Approaches: Effects of

Spatial Resolution and Species Characteristics. Remote Sensing 15:778.

Ghamisi, P., B. Rasti, N. Yokoya, Q. Wang, B. Hofle, L. Bruzzone, F. Bovolo, M. Chi, K. Anders, R. Gloaguen, P. M. Atkinson, and J. A. Benediktsson. 2018. Multisource and Multitemporal Data Fusion in Remote Sensing. arXiv:1812.08287 [cs, eess, stat].

Guo, C., G. Pleiss, Y. Sun, and K. Q. Weinberger. 2017. On Calibration of Modern Neural Networks. Pages 1321–1330 Proceedings of the 34th International Conference on Machine Learning. PMLR.

Hang, R., Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya. 2021. Hyperspectral Image Classification With Attention-Aided CNNs. IEEE Transactions on Geoscience and Remote Sensing 59:2281–2293.

Hemming-Schroeder, N. M., A. A. Gutierrez, S. D. Allison, and J. T. Randerson. 2023. Estimating Individual Tree Mortality in the Sierra Nevada Using Lidar and Multispectral Reflectance Data. Journal of Geophysical Research: Biogeosciences 128:e2022JG007234.

Jucker, T., F. J. Fischer, J. Chave, D. A. Coomes, J. Caspersen, A. Ali, G. J. Loubota Panzou, T. R. Feldpausch, D. Falster, V. A. Usoltsev, S. Adu-Bredu, L. F. Alves, M. Aminpour, I. B. Angoboy, N. P. R. Anten, C. Antin, Y. Askari, R. Muñoz, N. Ayyappan, P. Balvanera, L. Banin, N. Barbier, J. J. Battles, H. Beeckman, Y. E. Bocko, B. Bond-Lamberty, F. Bongers, S. Bowers, T. Brade, M. van Breugel, A. Chantrain, R. Chaudhary, J. Dai, M. Dalponte, K. Dimobe, J.-C. Domec, J.-L. Doucet, R. A. Duursma, M. Enríquez, K. Y. van Ewijk, W. Farfán-Rios, A. Fayolle, E. Forni, D. I. Forrester, H. Gilani, J. L. Godlee, S. Gourlet-Fleury, M. Haeni, J. S. Hall, J.-K. He, A. Hemp, J. L. Hernández-Stefanoni, S. I. Higgins, R. J. Holdaway, K. Hussain, L. B. Hutley, T. Ichie, Y. Iida, H. Jiang, P. R. Joshi, H. Kaboli, M. K. Larsary, T. Kenzo, B. D. Kloeppel, T. Kohyama, S. Kunwar, S. Kuyah, J. Kvasnica, S. Lin, E. R. Lines, H. Liu, C. Lorimer, J.-J. Loumeto, Y. Malhi, P. L. Marshall, E. Mattsson, R. Matula, J. A. Meave, S. Mensah, X. Mi, S. Momo, G. R. Moncrieff, F.

Mora, S. P. Nissanka, K. L. O'Hara, S. Pearce, R. Pelissier, P. L. Peri, P. Ploton, L. Poorter, M. J. Pour, H. Pourbabaei, J. M. Dupuy-Rada, S. C. Ribeiro, C. Ryan, A. Sanaei, J. Sanger, M. Schlund, G. Sellan, A. Shenkin, B. Sonké, F. J. Sterck, M. Svátek, K. Takagi, A. T. Trugman, F. Ullah, M. A. Vadeboncoeur, A. Valipour, M. C. Vanderwel, A. G. Vovides, W. Wang, L.-Q. Wang, C. Wirth, M. Woods, W. Xiang, F. de A. Ximenes, Y. Xu, T. Yamada, and M. A. Zavala. 2022. Tallo: A global tree allometry and crown architecture database. Global Change Biology 28:5254–5268.

Jung, C., K. Zhou, and J. Feng. 2020. Fusionnet: Multispectral Fusion of RGB and NIR Images Using Two Stage Convolutional Neural Networks. IEEE Access 8:23912–23919.

Kampe, T. U., B. R. Johnson, M. A. 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 4:043510.

Kapil, R., S. M. Marvasti-Zadeh, D. Goodsman, N. Ray, and N. Erbilgin. 2022, August 21. Classification of Bark Beetle-Induced Forest Tree Mortality using Deep Learning. arXiv.

Ke, Y., and L. J. Quackenbush. 2011. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. International Journal of Remote Sensing 32:4725–4747.

Koh, P. W., S. Sagawa, H. Marklund, S. M. Xie, M. Zhang, A. Balsubramani, W. Hu, M. Yasunaga, R. L. Phillips, I. Gao, T. Lee, E. David, I. Stavness, W. Guo, B. Earnshaw, I. Haque, S. M. Beery, J. Leskovec, A. Kundaje, E. Pierson, S. Levine, C. Finn, and P. Liang. 2021. WILDS: A Benchmark of in-the-Wild Distribution Shifts. Pages 5637–5664 Proceedings of the 38th International Conference on Machine Learning. PMLR.

Kwon, R., Y. Ryu, T. Yang, Z. Zhong, and J. Im. 2023. Merging multiple sensing platforms and deep learning empowers individual tree mapping and species detection at the city scale. ISPRS Journal of Photogrammetry and Remote Sensing 206:201–221.

La Rosa, L. E. C., C. Sothe, R. Q. Feitosa, C. M. de Almeida, M. B. Schimalski, and D. A. B.

Oliveira. 2021. Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. ISPRS Journal of Photogrammetry and Remote Sensing 179:35–49.

Li, S., M. Brandt, X. Tong, S. Oehmcke, C. Igel, F. Gieseke, T. Nord-Larsen, R. Fensholt, J. Chave, and P. Ciais. 2023. Deep learning tree and forest biomass from sub-meter resolution images. preprint, In Review.

Liao, W., F. Van Coillie, L. Gao, L. Li, B. Zhang, and J. Chanussot. 2018. Deep Learning for Fusion of APEX Hyperspectral and Full-Waveform LiDAR Remote Sensing Data for Tree Species Mapping. IEEE Access 6:68716–68729.

Liu, Z., Z. Miao, X. Zhan, J. Wang, B. Gong, and S. X. Yu. 2019. Large-Scale Long-Tailed Recognition in an Open World. Pages 2532–2541 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA.

Lombardozzi, D. L., W. R. Wieder, N. Sobhani, G. B. Bonan, D. Durden, D. Lenz, M. SanClements, S. Weintraub-Leff, E. Ayres, C. R. Florian, K. Dahlin, S. Kumar, A. L. S. Swann, C. M. Zarakas, C. Vardeman, and V. Pascucci. 2023. Overcoming barriers to enable convergence research by integrating ecological and climate sciences: the NCAR–NEON system Version 1. Geoscientific Model Development 16:5979–6000.

Marconi, S., B. G. Weinstein, S. Zou, S. A. Bohlman, A. Zare, A. Singh, D. Stewart, I. Harmon, A. Steinkraus, and E. P. White. 2022. Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network. Remote Sensing of Environment 282:113264.

Martínez-Sánchez, J. L., C. Martínez-Garza, L. Cámara, and O. Castillo. 2020. Species-specific or generic allometric equations: which option is better when estimating the biomass of Mexican tropical humid forests? Carbon Management 11:241–249.

Maschler, J., C. Atzberger, and M. Immitzer. 2018. Individual Tree Crown Segmentation and Classification of 13 Tree Species Using Airborne Hyperspectral Data. Remote Sensing

10:1218.

Musinsky, J., T. Goulden, G. Wirth, N. Leisso, K. Krause, M. Haynes, and C. Chapman. 2022. Spanning scales: The airborne spatial and temporal sampling design of the National Ecological Observatory Network. Methods in Ecology and Evolution 13:1866–1884.

Northcutt, C. G., L. Jiang, and I. L. Chuang. 2020. Confident Learning: Estimating Uncertainty in Dataset Labels. arXiv:1911.00068 [cs, stat].

Onishi, M., and T. Ise. 2021. Explainable identification and mapping of trees using UAV RGB image and deep learning. Scientific Reports 11:903.

Orenstein, E. C., K. M. Kenitz, P. L. D. Roberts, P. J. S. Franks, J. S. Jaffe, and A. D. Barton. 2020. Semi- and fully supervised quantification techniques to improve population estimates from machine classifiers. Limnology and Oceanography: Methods 18:739–753.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo. 2011. Ecological Niches and Geographic Distributions (MPB-49). Page Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press.

Pulido, D., J. Salas, M. Rös, K. Puettmann, and S. Karaman. 2020. Assessment of Tree Detection Methods in Multispectral Aerial Images. Remote Sensing 12:2379.

Reiersen, G., D. Dao, B. Lütjens, K. Klemmer, K. Amara, A. Steinegger, C. Zhang, and X. Zhu. 2022, January 26. ReforesTree: A Dataset for Estimating Tropical Forest Carbon Stock with Deep Learning and Aerial Imagery. arXiv.

Schäfer, E., J. Heiskanen, V. Heikinheimo, and P. Pellikka. 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. Ecological Indicators 64:49–58.

Schiefer, F., T. Kattenborn, A. Frick, J. Frey, P. Schall, B. Koch, and S. Schmidtlein. 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing

170:205–215.

Seeley, M. M., N. R. Vaughn, B. L. Shanks, R. E. Martin, M. König, and G. P. Asner. 2023, July 24. Classifying a Highly Polymorphic Tree Species across Landscapes Using Airborne Imaging Spectroscopy. Preprints.

Sohn, K., D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li. 2020. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. Pages 596–608 Advances in Neural Information Processing Systems. Curran Associates, Inc.

Stork, L., A. Weber, J. van den Herik, A. Plaat, F. Verbeek, and K. Wolstencroft. 2021. Large-scale zero-shot learning in the wild: Classifying zoological illustrations. Ecological Informatics 62:101222.

Sumbul, G., R. G. Cinbis, and S. Aksoy. 2018. Fine-Grained Object Recognition and Zero-Shot Learning in Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing 56:770–779.

Sumbul, G., R. G. Cinbis, and S. Aksoy. 2019. Multisource Region Attention Network for Fine-Grained Object Recognition in Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing 57:4929–4937.

Tucker, C., M. Brandt, P. Hiernaux, A. Kariryaa, K. Rasmussen, J. Small, C. Igel, F. Reiner, K. Melocik, J. Meyer, S. Sinno, E. Romero, E. Glennie, Y. Fitts, A. Morin, J. Pinzon, D. McClain, P. Morin, C. Porter, S. Loeffler, L. Kergoat, B.-A. Issoufou, P. Savadogo, J.-P. Wigneron, B. Poulter, P. Ciais, R. Kaufmann, R. Myneni, S. Saatchi, and R. Fensholt. 2023. Sub-continental-scale carbon stocks of individual trees in African drylands. Nature 615:80–86.

Veblen, T., R. Andrus, and R. Chai. 2021. Permanent forest plot data from 1982-2019 at Niwot Ridge. Environmental Data Initiative.

Velasquez-Camacho, L., M. Etxegarai, and S. de-Miguel. 2023. Implementing Deep Learning

algorithms for urban tree detection and geolocation with high-resolution aerial, satellite, and ground-level images. Computers, Environment and Urban Systems 105:102025.

Veras, H. F. P., M. P. Ferreira, E. M. da Cunha Neto, E. O. Figueiredo, A. P. D. Corte, and C. R. Sanquetta. 2022. Fusing multi-season UAS images with convolutional neural networks to map tree species in Amazonian forests. Ecological Informatics 71:101815.

Wagner, F. H., R. Dalagnol, C. H. L. Silva-Junior, G. Carter, A. L. Ritz, M. C. M. Hirye, J. P. H. B. Ometto, and S. Saatchi. 2023. Mapping Tropical Forest Cover and Deforestation with Planet NICFI Satellite Images and Deep Learning in Mato Grosso State (Brazil) from 2015 to 2021. Remote Sensing 15:521.

Wallis, C. I. B., A. L. Crofts, D. Inamdar, J. P. Arroyo-Mora, M. Kalacska, É. Laliberté, and M. Vellend. 2023. Remotely sensed carbon content: The role of tree composition and tree diversity. Remote Sensing of Environment 284:113333.

Weinstein, B. G., S. J. Graves, S. Marconi, A. Singh, A. Zare, D. Stewart, S. A. Bohlman, and E. P. White. 2021a. A benchmark dataset for canopy crown detection and delineation in co-registered airborne RGB, LiDAR and hyperspectral imagery from the National Ecological Observation Network. PLOS Computational Biology 17:e1009180.

Weinstein, B. G., S. Marconi, M. Aubry-Kientz, G. Vincent, H. Senyondo, and E. P. White. 2020. DeepForest: A Python package for RGB deep learning tree crown delineation. Methods in Ecology and Evolution 11:1743–1751.

Weinstein, B. G., S. Marconi, S. A. Bohlman, A. Zare, A. Singh, S. J. Graves, and E. P. White. 2021b. A remote sensing derived data set of 100 million individual tree crowns for the National Ecological Observatory Network. eLife 10:e62922.

Weinstein, B. G., S. Marconi, S. Bohlman, A. Zare, and E. White. 2019. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sensing 11:1309.

Zamboni, P., J. M. Junior, J. de A. Silva, G. T. Miyoshi, E. T. Matsubara, K. Nogueira, and W. N.

Gonçalves. 2021. Benchmarking Anchor-Based and Anchor-Free State-of-the-Art Deep Learning Methods for Individual Tree Detection in RGB High-Resolution Images. Remote Sensing 13:2482.

Table S1: Species included in each model for each NEON site. The number of samples (n) for each species in the canopy filtered data. To be included in the model, a species needs to have at least 10 training samples and 10 test samples at a site in the final filtered data. The number of predicted trees at each site, the proportion of total predictions at the site, and the rank abundance of each species is shown.

sci_name	count	site	proportion	rank	Habitat
Tsuga canadensis (L.) Carrire	1916656	BART	0.44	1	Northern Broadleaf
Fagus grandifolia Ehrh.	557790	BART	0.13	2	Northern Broadleaf
Acer saccharum Marshall	529581	BART	0.12	3	Northern Broadleaf
Betula alleghaniensis Britton	527648	BART	0.12	4	Northern Broadleaf
Acer rubrum L.	384249	BART	0.09	5	Northern Broadleaf
Fraxinus americana L.	317331	BART	0.07	6	Northern Broadleaf
Betula papyrifera Marshall	119675	BART	0.03	7	Northern Broadleaf
Juglans nigra L.	856379	BLAN	0.35	1	Southern Broadleaf
Cornus florida L.	435672	BLAN	0.18	2	Southern Broadleaf
Celtis occidentalis L.	348581	BLAN	0.14	3	Southern Broadleaf
Quercus rubra L.	260101	BLAN	0.11	4	Southern Broadleaf
Liriodendron tulipifera L.	217721	BLAN	0.09	5	Southern Broadleaf
Pinus strobus L.	202535	BLAN	0.08	6	Southern Broadleaf
Platanus occidentalis L.	92087	BLAN	0.04	7	Southern Broadleaf
Picea mariana (Mill.) Britton, Sterns & Poggenb.	3858957	BONA	0.51	1	Conifer
Betula neoalaskana Sarg.	2365676	BONA	0.31	2	Conifer

Populus tremuloides Michx.	923889	BONA	0.12	3	Conifer
Picea glauca (Moench) Voss	424178	BONA	0.06	4	Conifer
Quercus stellata Wangenh.	1276070	CLBJ	0.60	1	Savannah
Juniperus virginiana L.	488138	CLBJ	0.23	2	Savannah
Quercus marilandica M√ºnchh.	375595	CLBJ	0.18	3	Savannah
Picea glauca (Moench) Voss	1282877	DEJU	0.37	1	Conifer
Picea mariana (Mill.) Britton, Sterns & Poggenb.	1223419	DEJU	0.36	2	Conifer
Populus tremuloides Michx.	920182	DEJU	0.27	3	Conifer
Pinus taeda L.	786987	DELA	0.22	1	Southern Broadleaf
Quercus nigra L.	756838	DELA	0.21	2	Southern Broadleaf
Acer rubrum L.	675465	DELA	0.19	3	Southern Broadleaf
Liquidambar styraciflua L.	478992	DELA	0.13	4	Southern Broadleaf
Fraxinus pennsylvanica Marshall	370191	DELA	0.10	5	Southern Broadleaf
Celtis laevigata Willd.	289826	DELA	0.08	6	Southern Broadleaf
Carya tomentosa (Lam.) Nutt.	262398	DELA	0.07	7	Southern Broadleaf
Acer rubrum L.	1142770	GRSM	0.47	1	Southern Broadleaf
Quercus montana Willd.	851493	GRSM	0.35	2	Southern Broadleaf
Liriodendron tulipifera L.	453621	GRSM	0.19	3	Southern Broadleaf
Acer rubrum L.	2092750	HARV	0.23	1	Northern Broadleaf
Pinus strobus L.	1928207	HARV	0.21	2	Northern Broadleaf
Quercus rubra L.	1656996	HARV	0.18	3	Northern Broadleaf
Quercus alba L.	1192530	HARV	0.13	4	Northern Broadleaf
Betula lenta L.	743261	HARV	0.08	5	Northern Broadleaf
Fraxinus americana L.	561297	HARV	0.06	6	Northern Broadleaf
Tsuga canadensis (L.) Carri√®re	419849	HARV	0.05	7	Northern Broadleaf

Betula alleghaniensis Britton	257040	HARV	0.03	8	Northern Broadleaf
Pinus resinosa Aiton	91640	HARV	0.01	9	Northern Broadleaf
Fagus grandifolia Ehrh.	68370	HARV	0.01	10	Northern Broadleaf
Prunus serotina Ehrh.	55833	HARV	0.01	11	Northern Broadleaf
Picea abies (L.) Karst.	27501	HARV	0.00	12	Northern Broadleaf
Nyssa sylvatica Marshall	26159	HARV	0.00	13	Northern Broadleaf
Quercus velutina Lam.	1259	HARV	0.00	14	Northern Broadleaf
Pinus palustris Mill.	2159896	JERC	0.58	1	Savannah
Quercus hemisphaerica W. Bartram ex Willd.	1280439	JERC	0.35	2	Savannah
Quercus margaretta	257687	JERC	0.07	3	Savannah
Liquidambar styraciflua L.	2153931	LENO	0.63	1	Southern Broadleaf
Quercus pagoda Raf.	886055	LENO	0.26	2	Southern Broadleaf
Quercus nigra L.	396110	LENO	0.12	3	Southern Broadleaf
Quercus rubra L.	1610301	MLBS	0.43	1	Southern Broadleaf
Acer rubrum L.	873498	MLBS	0.23	2	Southern Broadleaf
Quercus alba L.	475352	MLBS	0.13	3	Southern Broadleaf
Liriodendron tulipifera L.	411167	MLBS	0.11	4	Southern Broadleaf
Quercus coccinea M√ºnchh.	359546	MLBS	0.10	5	Southern Broadleaf
Pinus contorta Douglas ex Loudon	2005920	NIWO	0.43	1	Conifer
Picea engelmannii Parry ex Engelm.	1268215	NIWO	0.27	2	Conifer
Abies lasiocarpa (Hook.) Nutt.	999052	NIWO	0.22	3	Conifer
Pinus flexilis James	354697	NIWO	0.08	4	Conifer
Quercus hemisphaerica W. Bartram ex Willd.	464452	OSBS	0.13	1	Savannah
Pinus palustris Mill.	459380	OSBS	0.13	2	Savannah

Pinus taeda L.	441837	OSBS	0.13	3	Savannah
Pinus elliottii Engelm.	428306	OSBS	0.12	4	Savannah
Quercus laevis Walter	315498	OSBS	0.09	5	Savannah
Quercus geminata Small	274916	OSBS	0.08	6	Savannah
Quercus virginiana Mill.	228768	OSBS	0.07	7	Savannah
Pinus clausa (Chapm. ex Engelm.) Vasey ex Sarg.	189431	OSBS	0.05	8	Savannah
Liquidambar styraciflua L.	162751	OSBS	0.05	9	Savannah
Nyssa sylvatica Marshall	145971	OSBS	0.04	10	Savannah
Magnolia sp.	132960	OSBS	0.04	11	Savannah
Quercus nigra L.	84128	OSBS	0.02	12	Savannah
Acer rubrum L.	73579	OSBS	0.02	13	Savannah
Carya glabra (Mill.) Sweet	61204	OSBS	0.02	14	Savannah
Pinus contorta Douglas ex Loudon	2656483	RMNP	0.35	1	Conifer
Pinus ponderosa Lawson & C. Lawson	1491041	RMNP	0.20	2	Conifer
Pseudotsuga menziesii (Mirb.) Franco	1185424	RMNP	0.16	3	Conifer
Pinus flexilis James	812500	RMNP	0.11	4	Conifer
Populus tremuloides Michx.	513952	RMNP	0.07	5	Conifer
Picea engelmannii Parry ex Engelm.	419298	RMNP	0.06	6	Conifer
Abies lasiocarpa (Hook.) Nutt.	408857	RMNP	0.05	7	Conifer
Liriodendron tulipifera L.	614569	SERC	0.27	1	Southern Broadleaf
Liquidambar styraciflua L.	503363	SERC	0.22	2	Southern Broadleaf
Acer rubrum L.	256490	SERC	0.11	3	Southern Broadleaf
Quercus falcata Michx.	236338	SERC	0.10	4	Southern Broadleaf

Fraxinus pennsylvanica Marshall	179711	SERC	0.08	5	Southern Broadleaf
Platanus occidentalis L.	130393	SERC	0.06	6	Southern Broadleaf
Quercus velutina Lam.	117590	SERC	0.05	7	Southern Broadleaf
Quercus alba L.	91438	SERC	0.04	8	Southern Broadleaf
Fagus grandifolia Ehrh.	87899	SERC	0.04	9	Southern Broadleaf
Carya tomentosa (Lam.) Nutt.	44351	SERC	0.02	10	Southern Broadleaf
Quercus wislizeni A. DC.	410332	SJER	0.48	1	Savannah
Quercus douglasii Hook. & Arn.	325524	SJER	0.38	2	Savannah
Pinus sabiniana Douglas ex Douglas	116312	SJER	0.14	3	Savannah
Calocedrus decurrens (Torr.) Florin	1957264	SOAP	0.65	1	Conifer
Pinus ponderosa Lawson & C. Lawson	626354	SOAP	0.21	2	Conifer
Quercus kelloggii Newberry	242541	SOAP	0.08	3	Conifer
Quercus chrysolepis Liebm.	198656	SOAP	0.07	4	Conifer
Abies balsamea (L.) Mill.	2390776	STEI	0.36	1	Northern Broadleaf
Populus tremuloides Michx.	1407230	STEI	0.21	2	Northern Broadleaf
Tilia americana L.	1017129	STEI	0.15	3	Northern Broadleaf
Acer rubrum L.	846965	STEI	0.13	4	Northern Broadleaf
Quercus rubra L.	721269	STEI	0.11	5	Northern Broadleaf
Acer saccharum Marshall	304397	STEI	0.05	6	Northern Broadleaf
Quercus alba L.	1128147	TALL	0.29	1	Southern Broadleaf
Pinus taeda L.	1007905	TALL	0.26	2	Southern Broadleaf
Pinus palustris Mill.	744862	TALL	0.19	3	Southern Broadleaf
Liquidambar styraciflua L.	554307	TALL	0.14	4	Southern Broadleaf

Pinus echinata Mill.	330868	TALL	0.09	5	Southern Broadleaf
Liriodendron tulipifera L.	122261	TALL	0.03	6	Southern Broadleaf
Pinus contorta Douglas ex Loudon	1238741	TEAK	0.34	1	Conifer
Abies concolor (Gord. & Glend.) Lindl. ex Hildebr.	766721	TEAK	0.21	2	Conifer
Pinus jeffreyi Balf.	520066	TEAK	0.14	3	Conifer
Abies lowiana (Gordon & Glend.) A. Murray bis	439012	TEAK	0.12	4	Conifer
Abies magnifica A. Murray bis	423231	TEAK	0.12	5	Conifer
Calocedrus decurrens (Torr.) Florin	155247	TEAK	0.04	6	Conifer
Pinus lambertiana Douglas	117623	TEAK	0.03	7	Conifer
Quercus rubra L.	1117038	TREE	0.16	1	Northern Broadleaf
Abies balsamea (L.) Mill.	1009185	TREE	0.14	2	Northern Broadleaf
Populus tremuloides Michx.	982204	TREE	0.14	3	Northern Broadleaf
Betula papyrifera Marshall	713772	TREE	0.10	4	Northern Broadleaf
Picea mariana (Mill.) Britton, Sterns & Poggenb.	572793	TREE	0.08	5	Northern Broadleaf
Acer saccharum Marshall	494667	TREE	0.07	6	Northern Broadleaf
Tilia americana L.	449669	TREE	0.06	7	Northern Broadleaf
Larix laricina (Du Roi) K. Koch	437108	TREE	0.06	8	Northern Broadleaf
Pinus resinosa Aiton	359730	TREE	0.05	9	Northern Broadleaf
Thuja occidentalis L.	288164	TREE	0.04	10	Northern Broadleaf
Picea glauca (Moench) Voss	206034	TREE	0.03	11	Northern Broadleaf
Acer rubrum L.	191871	TREE	0.03	12	Northern Broadleaf
Fraxinus pennsylvanica Marshall	125480	TREE	0.02	13	Northern Broadleaf
Pinus strobus L.	122127	TREE	0.02	14	Northern Broadleaf

Tsuga canadensis (L.) Carri√®re	48679	TREE	0.01	15	Northern Broadleaf
Maclura pomifera (Raf.) C.K. Schneid.	399538	UKFS	0.25	1	Southern Broadleaf
Juniperus virginiana L.	352218	UKFS	0.22	2	Southern Broadleaf
Celtis occidentalis L.	301882	UKFS	0.19	3	Southern Broadleaf
Ulmus americana L.	163371	UKFS	0.10	4	Southern Broadleaf
Quercus muehlenbergii Engelm.	143089	UKFS	0.09	5	Southern Broadleaf
Carya ovata (Mill.) K. Koch	91871	UKFS	0.06	6	Southern Broadleaf
Gleditsia triacanthos L.	82320	UKFS	0.05	7	Southern Broadleaf
Juglans nigra L.	68406	UKFS	0.04	8	Southern Broadleaf
Acer saccharum Marshall	727435	UNDE	0.15	1	Northern Broadleaf
Picea mariana (Mill.) Britton, Sterns & Poggenb.	686836	UNDE	0.14	2	Northern Broadleaf
Picea glauca (Moench) Voss	528496	UNDE	0.11	3	Northern Broadleaf
Tsuga canadensis (L.) Carri√®re	476682	UNDE	0.10	4	Northern Broadleaf
Abies balsamea (L.) Mill.	432414	UNDE	0.09	5	Northern Broadleaf
Betula alleghaniensis Britton	432101	UNDE	0.09	6	Northern Broadleaf
Acer rubrum L.	399475	UNDE	0.08	7	Northern Broadleaf
Betula papyrifera Marshall	323397	UNDE	0.07	8	Northern Broadleaf
Larix laricina (Du Roi) K. Koch	270436	UNDE	0.06	9	Northern Broadleaf
Populus tremuloides Michx.	168410	UNDE	0.03	10	Northern Broadleaf
Populus grandidentata Michx.	155654	UNDE	0.03	11	Northern Broadleaf
Fraxinus nigra Marshall	130453	UNDE	0.03	12	Northern Broadleaf
Fraxinus americana L.	89307	UNDE	0.02	13	Northern Broadleaf
Pseudotsuga menziesii (Mirb.) Franco	3987937	WREF	0.59	1	Conifer
Tsuga heterophylla (Raf.) Sarg.	1218052	WREF	0.18	2	Conifer

Abies amabilis (Douglas ex Loudon) Douglas ex Forbes	903627	WREF	0.13	3	Conifer
Thuja plicata Donn ex D. Don	664059	WREF	0.10	4	Conifer
Pinus contorta Douglas ex Loudon	3326783	YELL	0.74	1	Conifer
Pseudotsuga menziesii (Mirb.) Franco	889082	YELL	0.20	2	Conifer
Populus tremuloides Michx.	262641	YELL	0.06	3	Conifer

RGB Data Experimentation

We tested several model architectures that fuse RGB and HSI information for tree species classification. We expected that the 10cm RGB data would provide additional spatial information compared to the 1m HSI data. Data fusion is a common strategy in deep learning for remote sensing classification (e.g. Jung et al. 2020). The challenge is understanding at what level the data should be fused, with several reasonable options; 1) “Early”, as a stack of input data, with the HSI and RGB data resampled to the same higher resolution, 2) “Middle”, as two separate arms of a convolutional neural network with concatenation of deeply learned features during prediction, 3) “Late”, ensemble of predicted classes, either as a weighted average or a learned combination of predicted score for each tree class (Audebert et al. 2019). We opted for the “Middle” strategy (Figure S1), since fusing the “Early” layers would create enormous input data and slow models to resample the HSI data to 10cm, and the “Late” strategy was unlikely to produce high quality perfection results for the RGB data on its own.

The challenge of the “Middle” strategy is how to normalize and concat the two data streams. Since the 10cm RGB data is much higher resolution than the 1m HSI data, applying 2D convolutions to each input data will lead to a mismatch in feature lengths, with many more RGB features compared to HSI features. We experimented with several normalization and data reduction techniques to make these features more similar. Unfortunately, we found no evidence that the RGB data made any improvement in model accuracy (Table S2). This does not preclude that other architectures, strategies or normalization approaches won’t lead to improvement, especially when creating a unified model across sites.

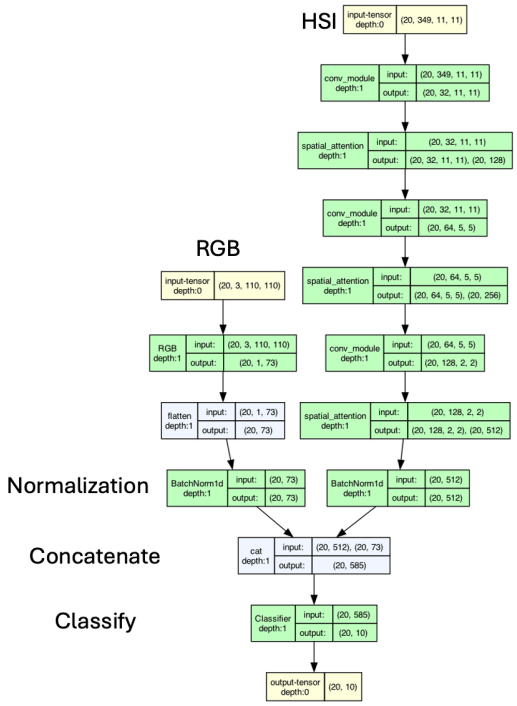


Figure S1. An example model architecture for data fusion between 1m HSI data and 10cm RGB for tree species classification. In this example, a batch of crowns (n=20), each with a HSI and RGB pair, is run through the network to jointly predict tree classes (n=10). The RGB model was a resnet-50 pretrained backbone, a common RGB architecture for image-classification. The HSI architecture was the same spectral attention network used throughout the rest of the paper. The two features were min-max normalized separately before combined and a joint classifier was used to predict tree species classes.

Table S2. Experiments comparing RGB, HSI and joint model for a single NEON site (OSBS). The experiments were done without the hierarchical model, or multi-temporal ensemble approaches to highlight the difference solely from source data type.

Model Description	Micro	Macro
HSI	0.40	0.41
RGB	0.37	0.30
RGB + HSI (Figure S1)	0.34	0.40

NEON-wide versus per-site models

To generate the species crown predictions, we used a separate model for each NEON site. While this yielded the best score, it greatly increased the amount of training time and reduced the portability of the workflow to new sites. Per-site models performed both better on average across all datasets, as well as at the local site level for the majority of sites (Figure S2). The majority of the improvement is in the rare species detection, as shown by the dramatic increase in the macro-averaged recall. For example, at OSBS, a well sampled site that we used for model validation in Weinstein et al. (2023), a per-site model had a micro/macro accuracy of 0.73, 0.63 versus the test samples at OSBS in the full site scale had a micro/macro accuracy of 0.61, 0.21. This comparison should be seen as approximate, since there are 3 species not included in the per-site model that are known to occur at OSBS, but which did not meet the minimum number of test data when considering OSBS only. While we cannot know without significant further exploration, we believe that the poorer performance using a single NEON-wide model is due to three main factors: 1) combining data across sites eliminates the utility of the hierarchical approach, since it removes the logic of the initial step of modeling of the dominant class in a particular area. It is very easy to differentiate between the dominant class at OSBS (*Pinus palustris*) and the dominant species at other well-sampled sites like Harvard Forest (*Acer Saccharum*), 2) generalization across geography is complicated by environmental differences, as well as potential leaf trait variation in response to local conditions. Pooling individuals from the same species across sites may make a model that muddies the features for all sites, leading to poor performance even in well sampled areas. 3) a NEON-wide model exacerbates the class imbalance challenge within the dataset. This has a disproportionately large impact on rare species. For example, at OSBS, *Quercus hemisphaerica* co-occurs with several other difficult to identify *Quercus* species. It does not occur widely, just at two sites in the entire dataset. Despite this, the performance from the OSBS-only model drops from 76% in Weinstein et al. 2023 to 37% in the single model for all NEON sites.

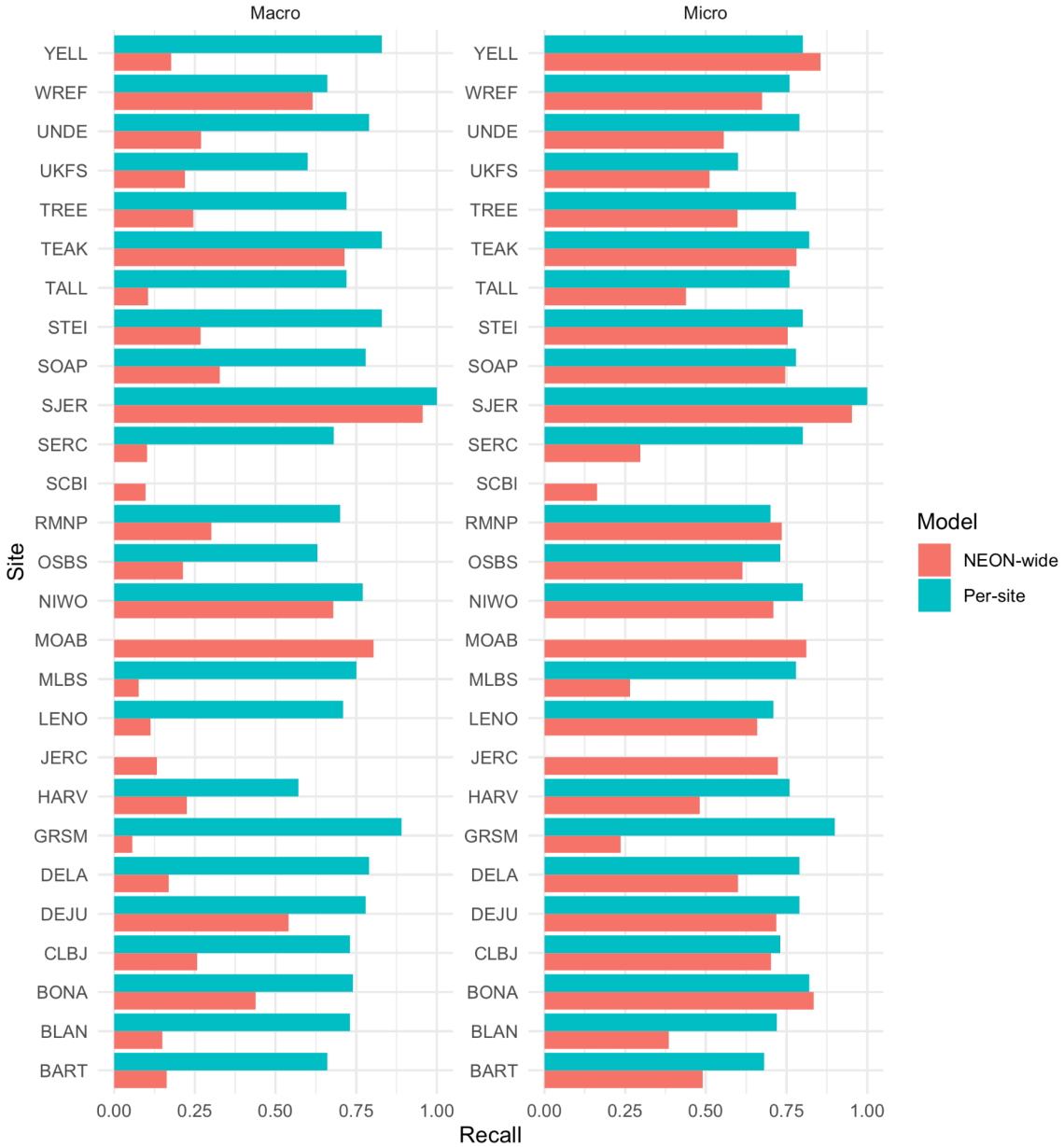


Figure S2. Comparison of site-level performance for modeling workflows that use training data solely from a single site ('per-site') and pool training data across all sites 'NEON-wide'. Micro averaged recall is the proportion of correctly predicted ground truth stems. Macro-averaged recall is the average recall per species, thereby weighing all species equally regardless of abundance. Several sites (JERC, MOAB, SCBI) lacked site-level predictions because the sample size per species at the individual site was too low.

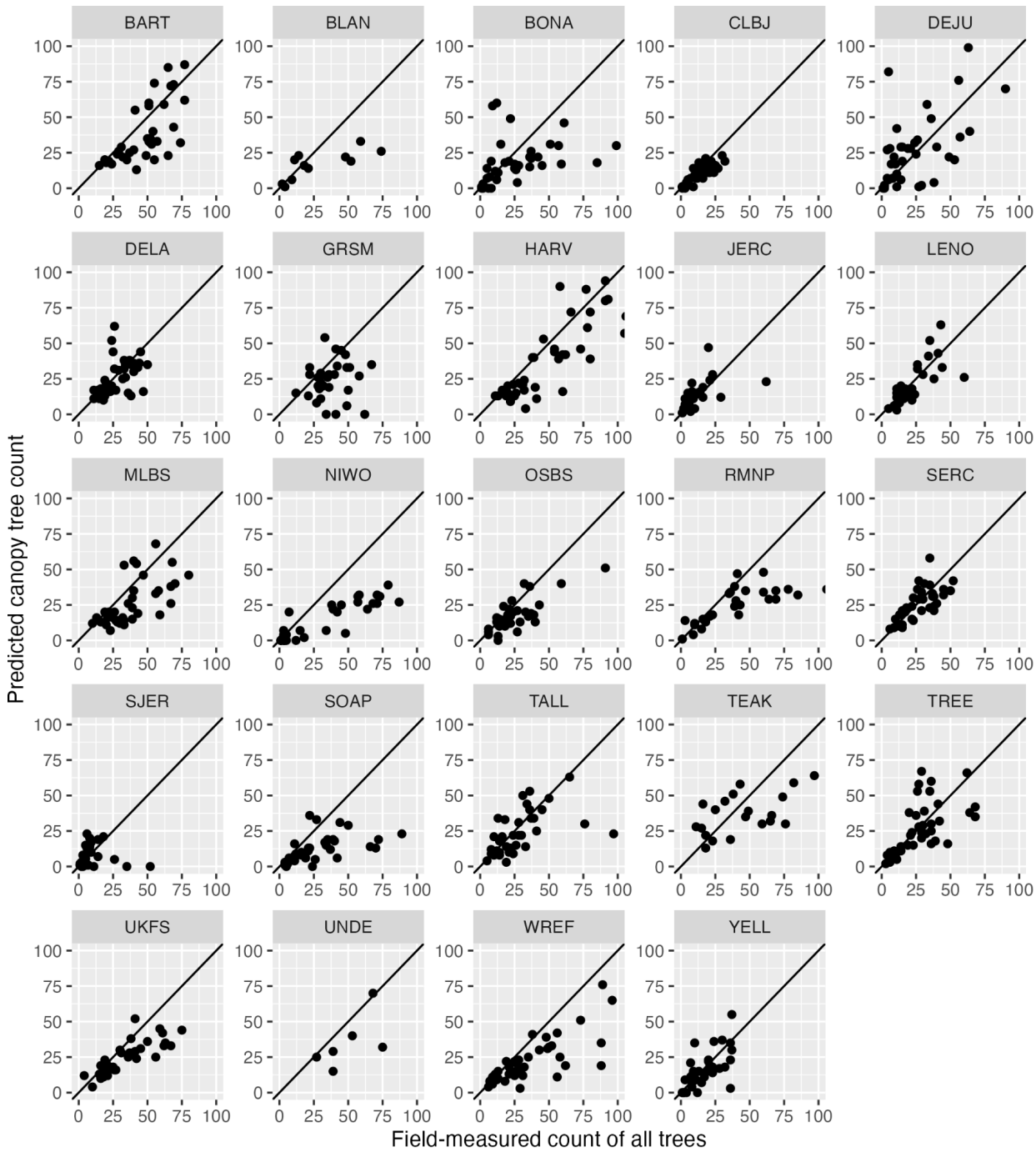


Figure S3. Predicted canopy trees versus the count of all field measured trees in the NEON Woody Vegetation Structure plots. For each NEON site, the number of tree detections in the prediction data is compared to the number of field-measured detections for that NEON subplot. The 1:1 line is shown in solid black, and a linear model is shown in blue with 95% CI intervals.