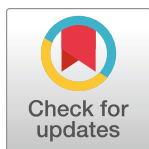# PLOS BIOLOGY

METHODS AND RESOURCES

Individual canopy tree species maps for the National Ecological Observatory Network

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, United States of America, **2** Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, United States of America, **3** School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, Florida, United States of America, **4** Department of Agricultural & Biological Engineering, University of Florida, Gainesville, Florida, United States of America, **5** Nelson Institute for Environmental Studies, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **6** Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, Maine, United States of America, **7** Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana, United States of America, **8** Department of Forest & Wildlife Ecology, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **9** Department of Geography, University of Colorado, Boulder, Colorado, United States of America, **10** School of Environment, Washington State University, Pullman, Washington, United States of America

* ben.weinstein@weecology.org

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 (NEON). 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 6 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 [1] 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 [2]. 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., [3]), and spatially continuous, but low-resolution data, from satellite or airborne sensors [4]. 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 [5,6]. 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 [7].

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., [8–10]), as well delineating the extent of tree crowns (e.g., [11]), is essential in many remote-sensing workflows and has been a rich area of algorithmic research (see reviews by [12,13]). 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 [14–16]. 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 [17]. Dozens of models have been proposed using classical image processing [18], feature-based machine learning [19,20], and deep learning [21–23] 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 [24]. 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 and create rigorous evaluation datasets [17].

Combining tree delineation and species classification to create broad scale tree maps is further complicated by the interaction between workflow components. Ref [25] reported that species classification decreased by more than 20% when moving from pixel-level to individual tree crown-level predictions. Changes in illumination during multiple days of remote-sensing data collection hampers generalization and species mapping has largely occurred at single flight line scales (e.g., [21]), or supported by terrestrial data in urban environments [26]. The changes in local species abundance over large areas contributes to further mismatch between training data and predicted landscapes at wide extents. Ref [27] 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
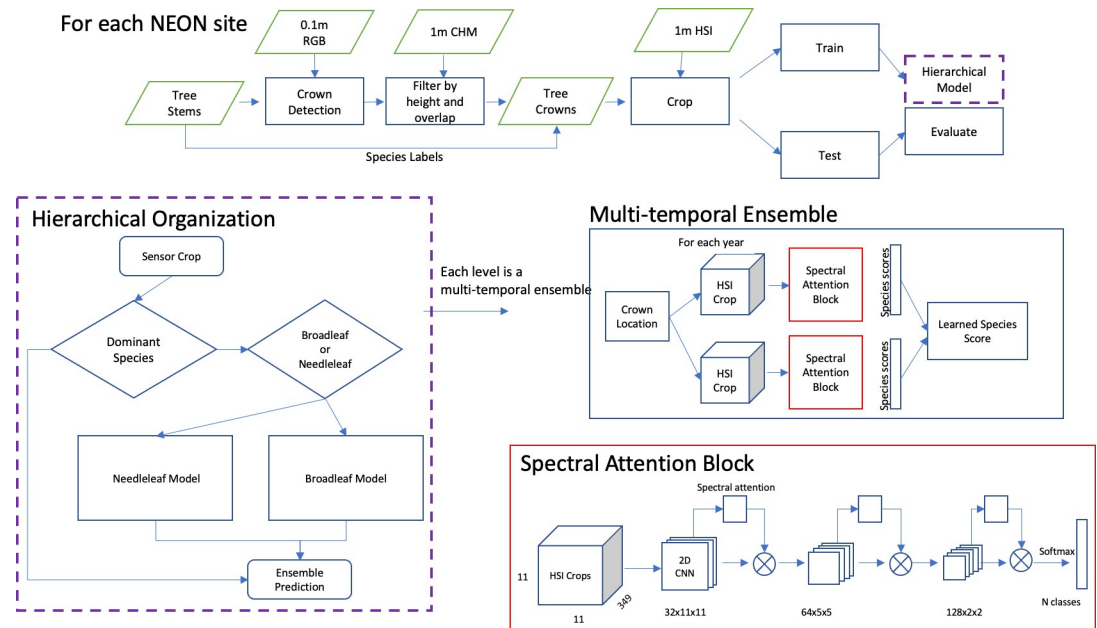
Fig 1. Conceptual workflow for species prediction at each NEON site. The fig is modified from [27].

https://doi.org/10.1371/journal.pbio.3002700.g001

be accurately classified at a single National Ecological Observatory Network (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 [27] to sites across the United States and assess its performance in order to provide data for ecological and computer vision research.

The 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 [28]. NEON collects standardized terrestrial and airborne data at dozens of sites across the US, 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 [29–32]. 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 US using machine learning models to predict crown position, species identity, health status, and height for individual trees visible in the canopy (Fig 1). Our work extends the crown location dataset published in [33] 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.

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 [28]. We used 4 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 [34]. 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 to 2,500 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 [35]. 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 [1] and public data [36]. We followed the taxonomic hierarchy used by NEON except for 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 (Fig 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 1 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 3 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. Ref [20] 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 [20] 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 [20]

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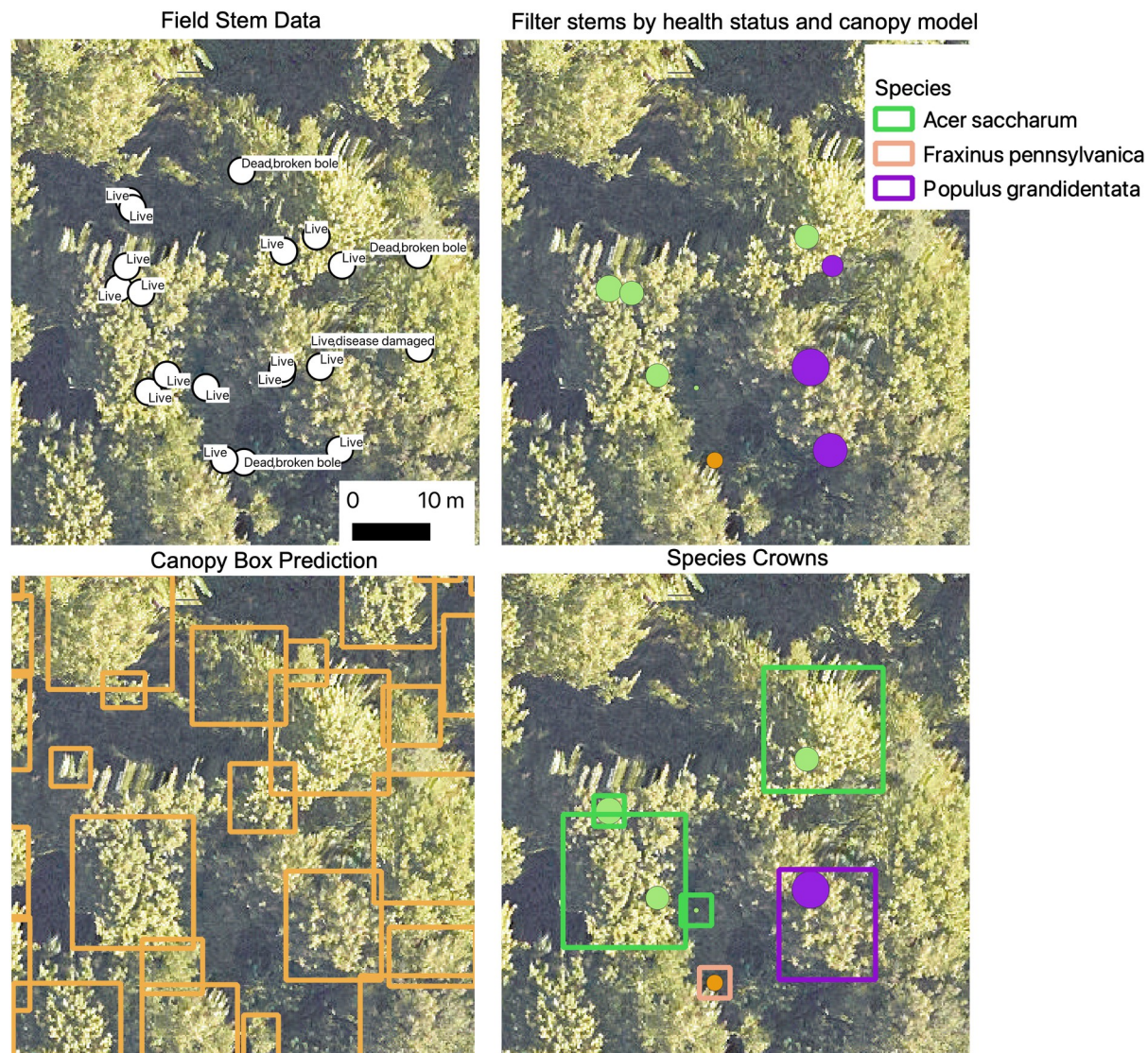
Fig 2. Example workflow for filtering stem data to associate with crown pixel area. Size of the dots in panels b and d are proportional to the individual tree DBH.

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.

Crown prediction

The DeepForest algorithm used in this work was first proposed in [37] using a combination of hand-annotated tree crown delineations and large-scale synthetic pretraining data using

LiDAR-derived tree locations; [16,38] 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 [33], 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 1 prediction, so if 2 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 [10,26,39,40] with accuracies generally mirroring approximately 70% for fine-tuned models from independent analysis [41].

We follow the workflow described in [33] 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 [27], this Alive-Dead model is a 2 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).

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

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 5 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 4 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 2 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 4 letter code for the NEON site. See 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.

https://doi.org/10.1371/journal.pbio.3002700.t001

classify each predicted crown tree crown to species, we use the 1-m hyperspectral data and a multi-temporal hierarchical model. Ref [27] 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 [42]. 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 (Fig 1) with 3 convolutional layers and a max pooling spectral attention layer following [43]. 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 4 years of available hyperspectral data for a geographic location, we predicted 4 classification outputs and then combined them to create the final prediction. This assumes that canopy trees at each geographic location are unlikely to change species label among years at short time scales [44]. 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 [45,46] we did not find significant improvements when adding the 10 cm RGB data to species classification (Fig A in S1 File), 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 (Fig B in S1 File). 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 [17]. 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" versus "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). To visualize the predictions and overlapping RGB data, see visualize.idtrees.org. 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 dataset covered 47.5% of the total species richness for trees ≥10 cm dbh represented in the reference list at the sites (Fig 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 (Fig 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 [20]. Models performed more poorly in southern broadleaf forests, such Talladega National Forest, Alabama

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	9,782	1,194
OSBS, FL	Oak Savannah,	0.73	0.63	14	3,293	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
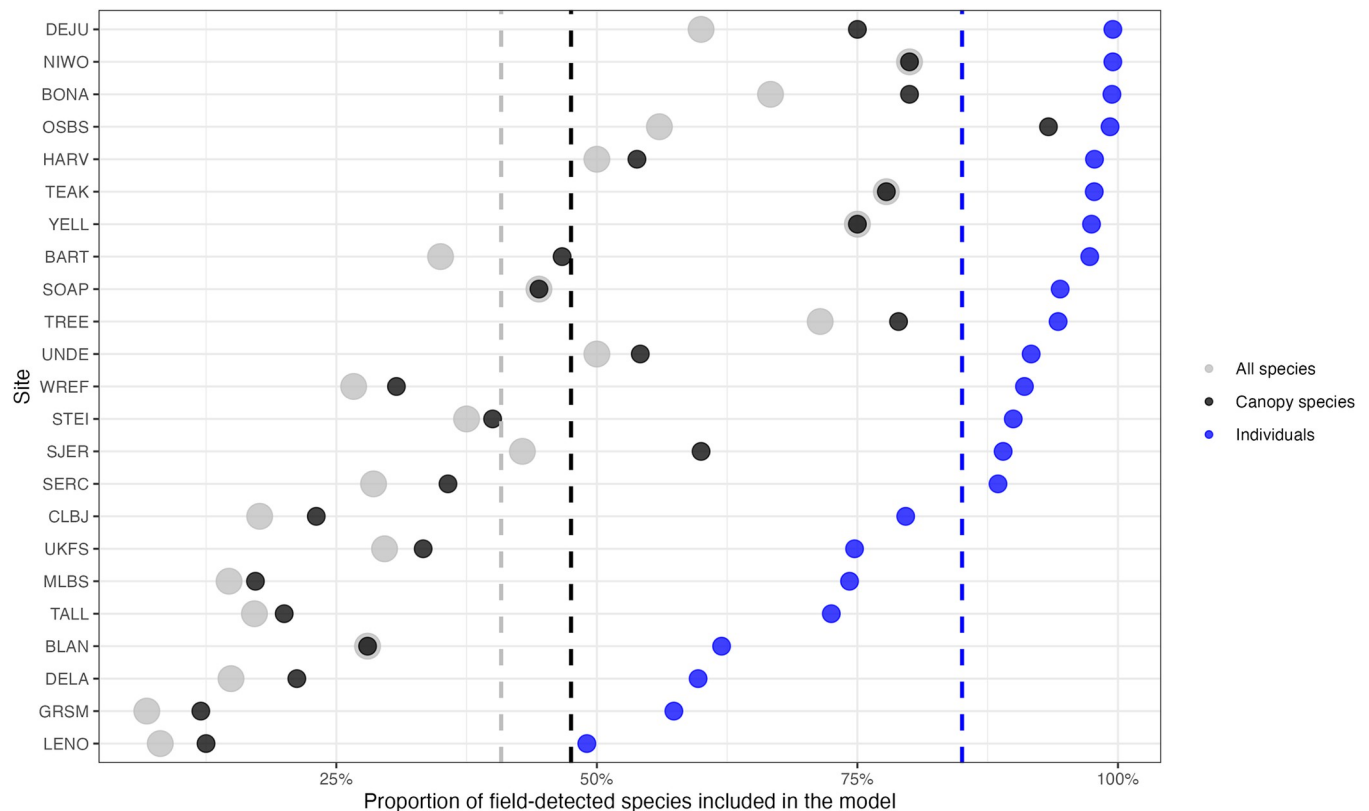
Fig 3. The proportion of species included in the model for each site compared to species with at least 2 records in the field-collected data. We calculated the proportion compared to: (1) all species—every record in the field collected data with at least 2 samples; 2) canopy species—the data filtered 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 A in S1 File. The dashed line is the mean number of species across sites for both species and individual proportions. The underlying data for this figure can be found in supplemental data "S1 Data."

https://doi.org/10.1371/journal.pbio.3002700.g003

(TALL) and Smithsonian Environmental Research Center, Maryland (SERC), with higher biodiversity, closed canopy structure, and/or low data coverage per species. The most abundant species at a site typically had the highest accuracy, with lower accuracy for rarer species (Fig 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 (Fig 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 (Fig 6). Ranking the predicted species abundance for each site, the most predicted species represented approximately 60% of crown classifications (Fig 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 (Fig 8).
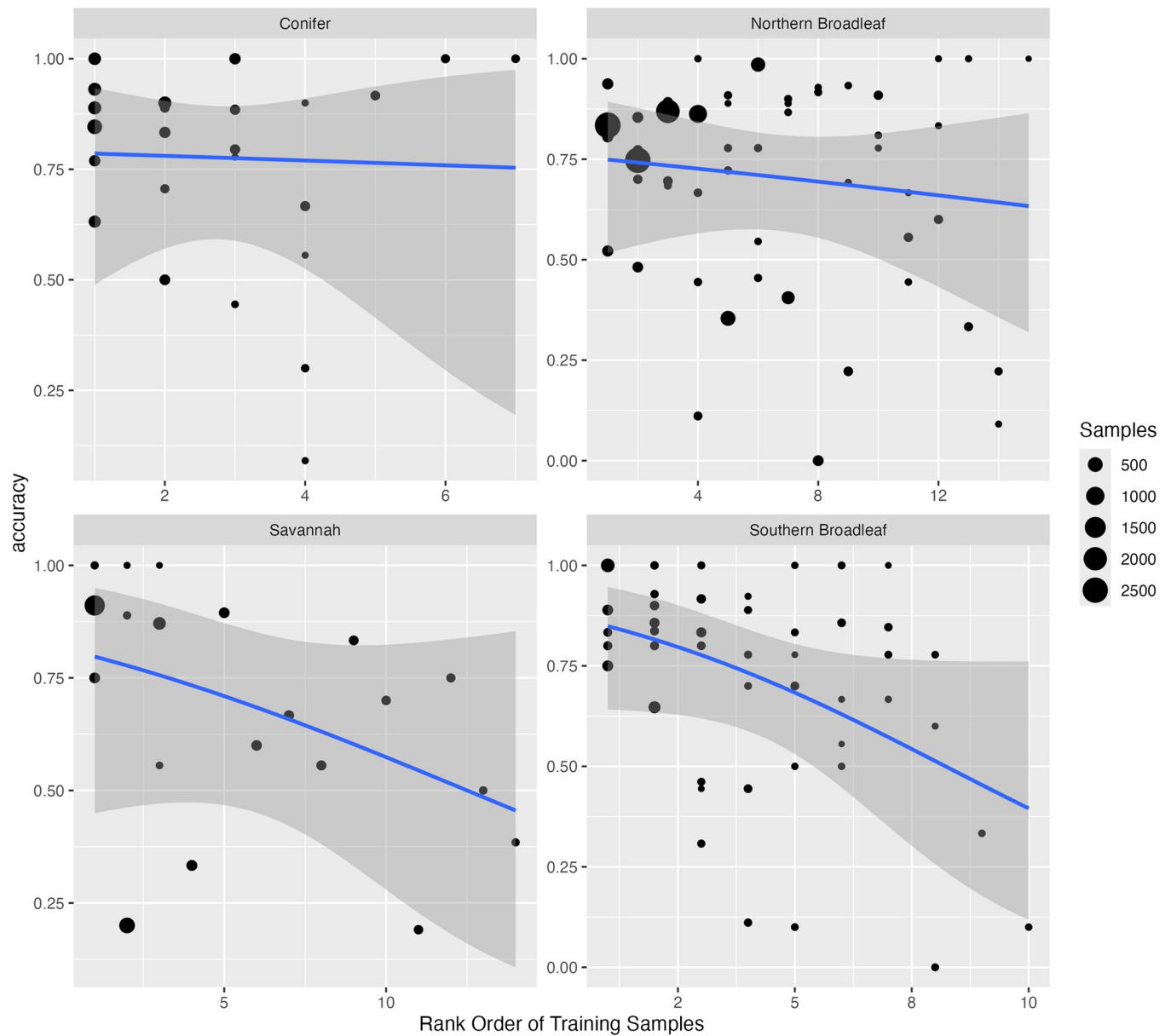
Fig 4. Rank order abundance and evaluation accuracy for each species for each NEON site aggregated by forest type. A binomial classification model was fit for each forest type to relate the rank order abundance of each species and evaluation accuracy. Each point is 1 species within 1 NEON site model. Point size is relative to the abundance of the species at the individual site in the training data. The underlying data for this figure can be found in supplemental data "S2 Data."

https://doi.org/10.1371/journal.pbio.3002700.g004

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 NEON. 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
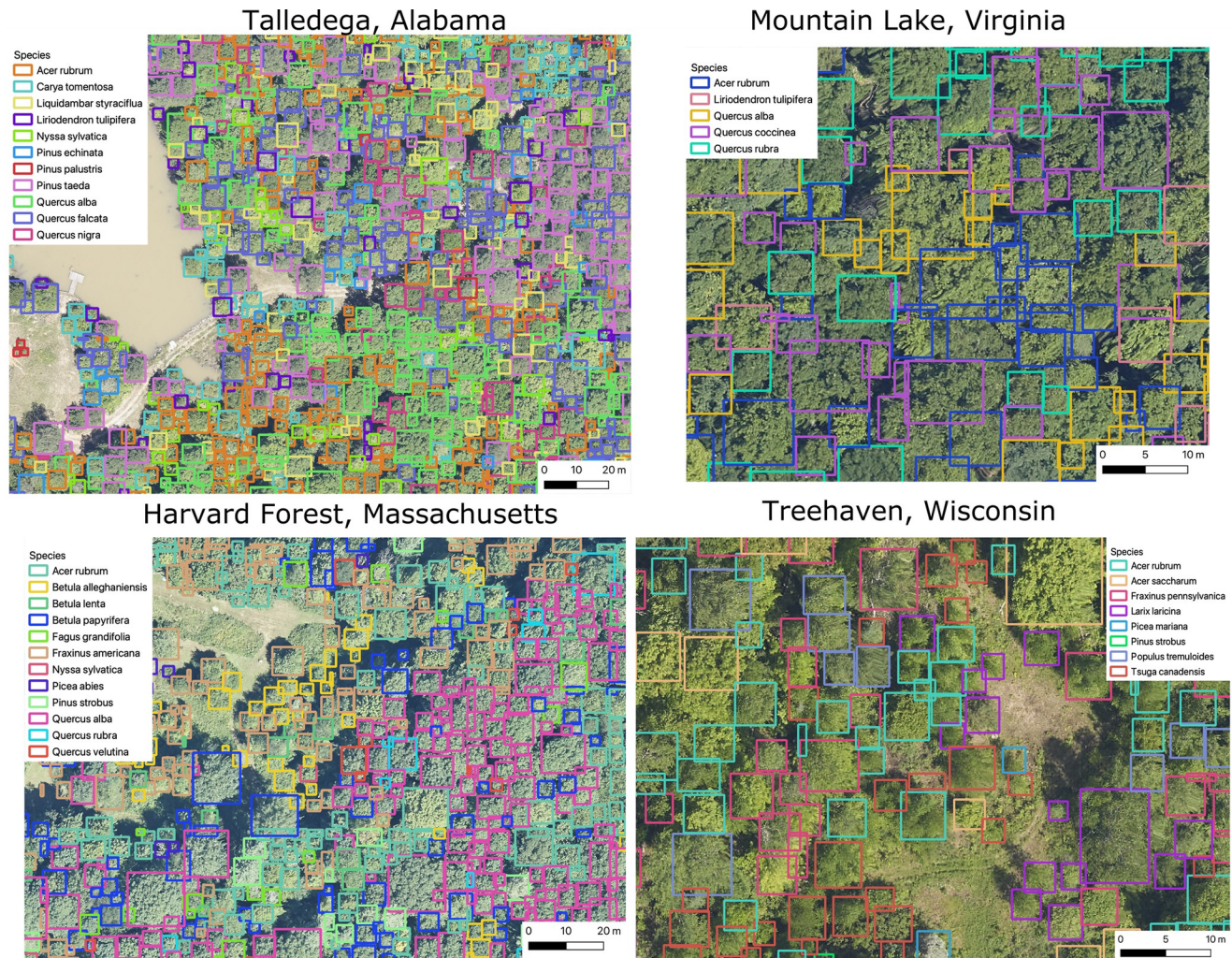
Fig 5. Example tree detections and species labels for 4 NEON sites with closed canopy deciduous forests.

https://doi.org/10.1371/journal.pbio.3002700.g005

environmental and biotic interactions [47,48], the phylogenetic structure of tree assemblages [49], and scale dependance of species plant communities [50]; ecosystem scientists can improve estimates of biomass using species-specific allometry [7,51], and foresters can measure impacts of habitat disturbance and landscape history [52,53]. 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 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
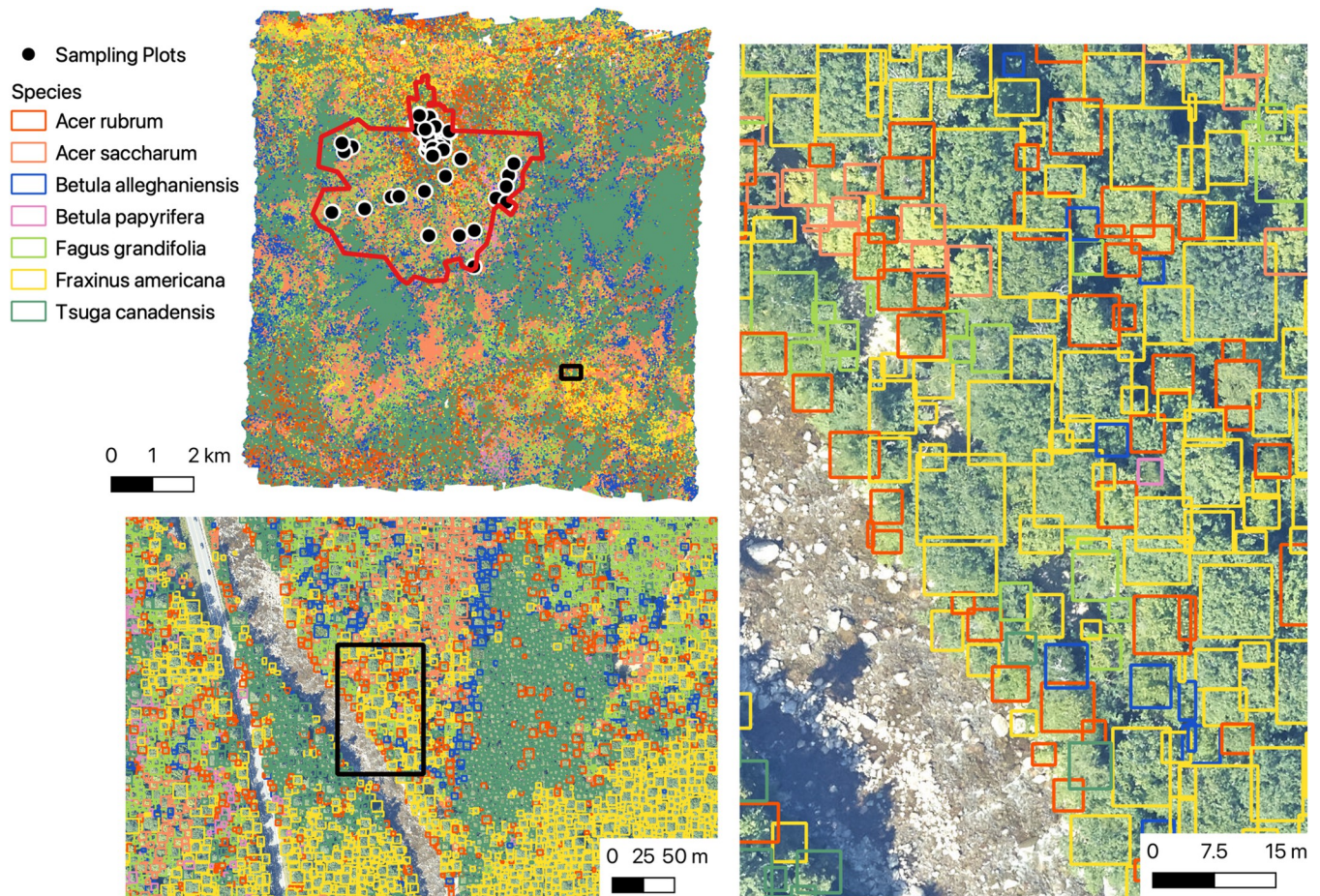
Fig 6. Overview from Bartlett Experimental Forest, New Hampshire (BART) showing 4,352,930 tree predictions for 7 species at 3 spatial scales. The location of NEON sampling plots and the NEON boundary are shown in the top left image.

https://doi.org/10.1371/journal.pbio.3002700.g006

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 [35] 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. Compared to field surveys, the canopy dataset will include fewer trees, with a bias towards large trees. Comparing the predicted canopy count and ground counts for the NEON field plots, the average undercount at each site was 8.51 individuals (range -2.45, 22.85)
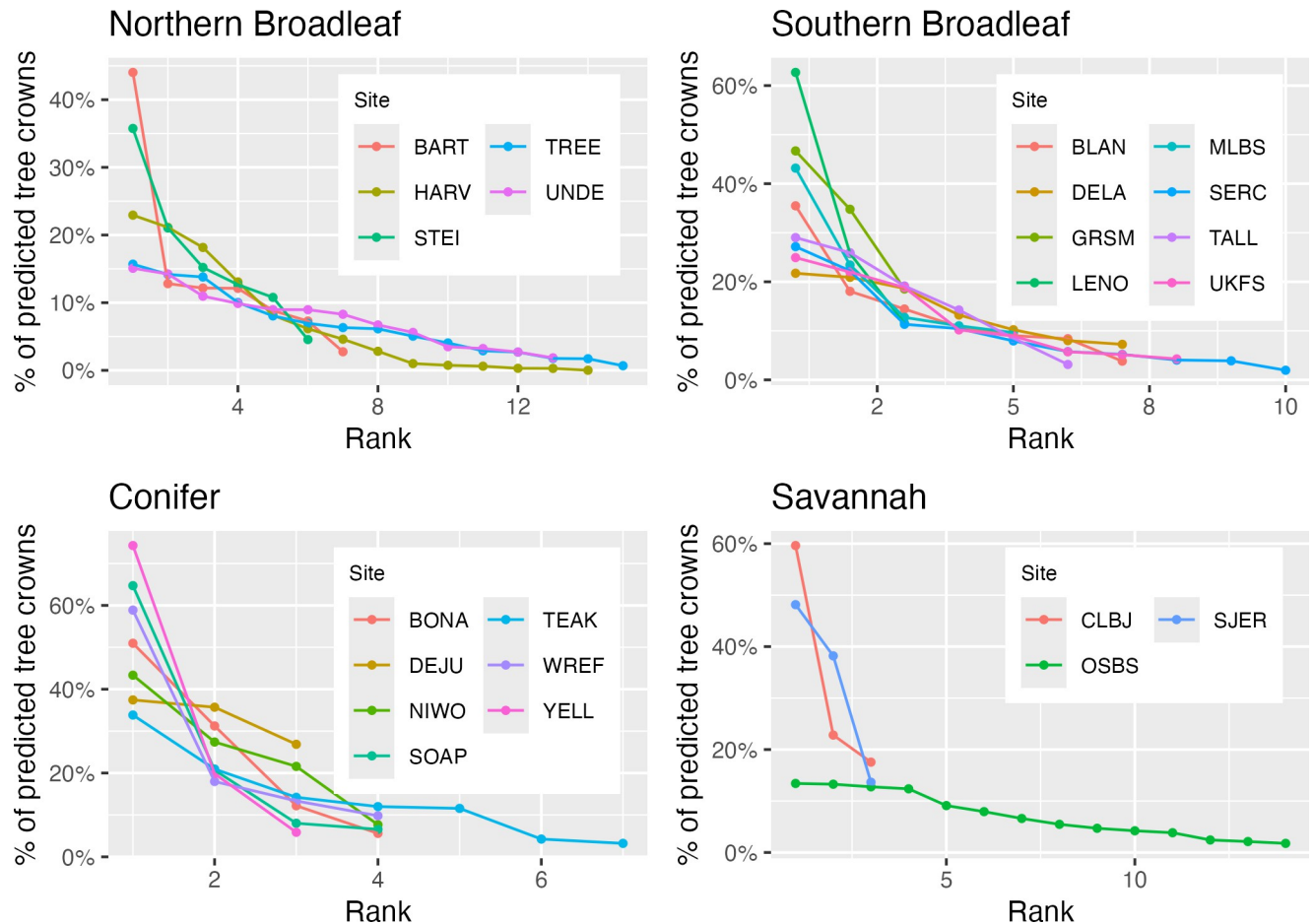
Fig 7. Rank order abundance for the predicted crown species labels for each site. The most commonly predicted species is rank 1, the second most commonly predicted species is rank 2, etc. Each point represents a species predicted at a site. For species identity and totals per site, see Table A in S1 File. The underlying data for this figure can be found in supplemental data "S3 Data."

https://doi.org/10.1371/journal.pbio.3002700.g007

(Fig C and Table C in S1 File). 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 (Fig 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 [16], field-stem recall [33], and images-drawn by observers on tablets directly in the field [54], and consistently found to have roughly 70% to 75% accuracy for crown delineation. Errors occur due to over segmentation (1 tree is identified as multiple trees), under segmentation (2 or more trees are identified as a single tree), and imprecisely defined crown edges. 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 1 year, but ultimately recover over time [55]. Species predictions are also uncertain, and while they include the most common species at each site, they still fail to include several species that do occur in the canopy (Fig 3). The discrepancy between canopy species in the filtered field dataset and species predicted in our model is a result of several factors. Some
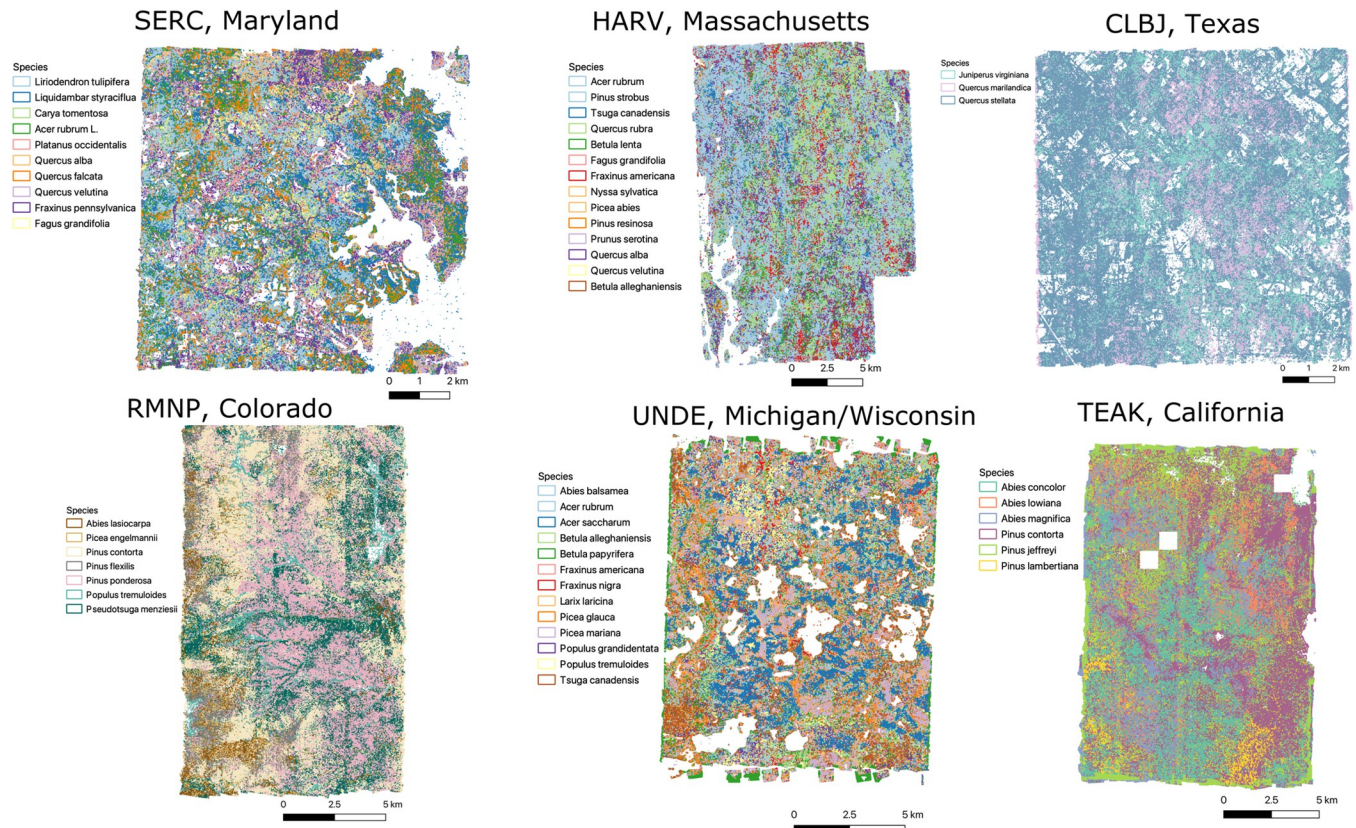
Fig 8. Overview of multiple sites spanning a broad range of forest types. Site names, from top left to bottom right, Smithsonian Environmental Research Center (SERC), Harvard Forest (HARV), Lyndon B. Johnson National Grassland (CLBJ), Rocky Mountain National Park (RMNP), University of Notre Dame Environmental Research Center (UNDE), Teakettle Canyon (TEAK).

https://doi.org/10.1371/journal.pbio.3002700.g008

canopy species are rare and thus have too few samples in our dataset 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. Ref [27] 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., [56]). Calibrating confidence scores using held-out data from training or test is an important step in this direction [57], 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., [58]) 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 million 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. 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.

One of the reasons additional data collections can be 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

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 (Fig 9). 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, 3 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 2 closely related *Populus* species. Using these goals to target trees, data on an additional 157 stems of 12 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 2 closely related *Populus* species increased from 66% and 54% to 72% and 82%, respectively.
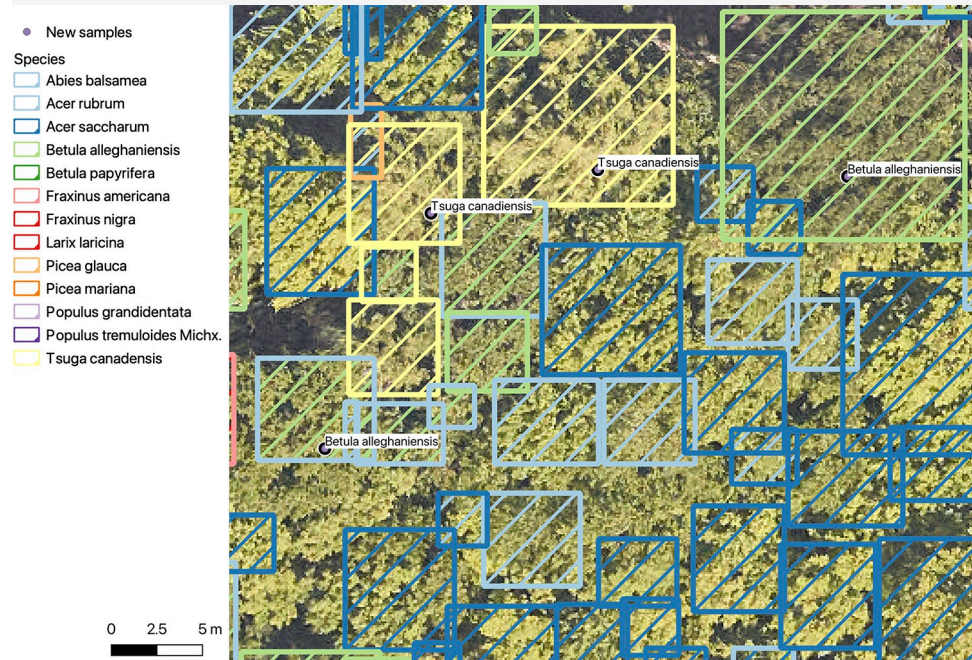

Fig 9. Original versus re-trained model predictions for UNDE. New sample trees were collected in the field without guidance from the predictions. The outline color is the original label, the filled shade is the revised label. The 2 *Tsuga canadensis* (top center) and the field samples were correctly 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.

https://doi.org/10.1371/journal.pbio.3002700.g009

for further improving tree species predictions (e.g., [59]). In the extreme, the task of zero-shot learning [59,60], 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 [61]. 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.

Supporting information

S1 Data. The underlying data for Fig 3.
(CSV)

S2 Data. The underlying data for Fig 4.
(CSV)

S3 Data. The underlying data for Fig 7.
(CSV)

S4 Data. The underlying data for Fig B in S1 File.
(CSV)

S5 Data. The underlying data for Fig C in S1 File.
(CSV)

S1 File. Supplemental materials. Table A. 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. Fig A. An example model architecture for data fusion between 1 m HSI data and 10 cm RGB for tree species classification. In this example, a batch of crowns (*n* = 20), each with an 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 2 features were min-max normalized separately before combined and a joint classifier was used to predict tree species classes. Table B. 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. Fig B. 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. For the underlying data, see S4 Data. Fig C. 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. For the underlying data, see S5 Data. Table C. Mean differences between predicted and observed counts, and RMSE for a generalized linear model with Poisson link function between field-measured counts of all trees and predicted canopy tree count (Fig C in S1 File).
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Ben G. Weinstein, Sergio Marconi, Alina Zare, Stephanie A. Bohlman, Aditya Singh, Sarah J. Graves, Lukas Magee, Daniel J. Johnson, Nathan G. Swenson, Philip Townsend, Thomas T. Veblen, Ethan P. White.

Data curation: Ben G. Weinstein, Sergio Marconi, Stephanie A. Bohlman, Sarah J. Graves, Lukas Magee, Daniel J. Johnson, Sydne Record, Vanessa E. Rubio, Nathan G. Swenson, Philip Townsend, Thomas T. Veblen, Robert A. Andrus, Ethan P. White.

Formal analysis: Ben G. Weinstein.

Funding acquisition: Ben G. Weinstein, Alina Zare.

Investigation: Ben G. Weinstein, Alina Zare, Aditya Singh, Lukas Magee, Ethan P. White.

Methodology: Ben G. Weinstein, Stephanie A. Bohlman, Aditya Singh, Sarah J. Graves, Ethan P. White.

Project administration: Ben G. Weinstein.

Resources: Ben G. Weinstein.

Software: Ben G. Weinstein.

Validation: Ben G. Weinstein.

Writing – original draft: Ben G. Weinstein, Sergio Marconi.

Writing – review & editing: Ben G. Weinstein.

References

1. Davies SJ, Abiem I, Abu Salim K, Aguilar S, Allen D, Alonso A, et al. ForestGEO: Understanding forest diversity and dynamics through a global observatory network. Biol Conserv. 2021; 253:108907. https://doi.org/10.1016/j.biocon.2020.108907

2. Schäfer E, Heiskanen J, Heikinheimo V, Pellikka P. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. Ecol Indic. 2016; 64:49–58. https://doi.org/10.1016/j.ecolind.2015.12.026

3. Jucker T, Fischer FJ, Chave J, Coomes DA, Caspersen J, Ali A, et al. Tallo: A global tree allometry and crown architecture database. Glob Change Biol. 2022; 28:5254–5268. https://doi.org/10.1111/gcb.16302 PMID: 35703577

4. Wagner FH, Dalagnol R, Silva-Junior CHL, Carter G, Ritz AL, Hirye MCM, et al. Mapping Tropical Forest Cover and Deforestation with Planet NICFI Satellite Images and Deep Learning in Mato Grosso State (Brazil) from 2015 to 2021. Remote Sens. 2023; 15:521. https://doi.org/10.3390/rs15020521

5. Tucker C, Brandt M, Hiernaux P, Kariryaa A, Rasmussen K, Small J, et al. Sub-continental-scale carbon stocks of individual trees in African drylands. Nature. 2023; 615:80–86. https://doi.org/10.1038/s41586-022-05653-6 PMID: 36859581

6. Liu S, Brandt M, Nord-Larsen T, Chave J, Reiner F, Lang N, et al. The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. Sci Adv. 2023; 9:eadh4097. https://doi.org/10.1126/sciadv.adh4097 PMID: 37713489

7. Wallis CIB, Crofts AL, Inamdar D, Arroyo-Mora JP, Kalacska M, Laliberté É, et al. Remotely sensed carbon content: The role of tree composition and tree diversity. Remote Sens Environ. 2023; 284:113333. https://doi.org/10.1016/j.rse.2022.113333

8. Freudenberg M, Nölke N, Agostini A, Urban K, Wörgötter F, Kleinn C. Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net. Remote Sens. 2019; 11:312. https://doi.org/10.3390/rs11030312

9. Zamboni P, Junior JM, Silva J de A, Miyoshi GT, Matsubara ET, Nogueira K, et al. Benchmarking Anchor-Based and Anchor-Free State-of-the-Art Deep Learning Methods for Individual Tree Detection in RGB High-Resolution Images. Remote Sens. 2021; 13:2482. https://doi.org/10.3390/rs13132482

10. Velasquez-Camacho L, Etxegarai M, de-Miguel S. Implementing Deep Learning algorithms for urban tree detection and geolocation with high-resolution aerial, satellite, and ground-level images. Comput Environ Urban Syst. 2023; 105:102025. https://doi.org/10.1016/j.compenvurbsys.2023.102025

11. Aubry-Kientz M, Dutrieux R, Ferraz A, Saatchi S, Hamraz H, Williams J, et al. A Comparative Assessment of the Performance of Individual Tree Crowns Delineation Algorithms from ALS Data in Tropical Forests. Remote Sens. 2019; 11:1086. https://doi.org/10.3390/rs11091086

12. Pulido D, Salas J, Rös M, Puettmann K, Karaman S. Assessment of Tree Detection Methods in Multispectral Aerial Images. Remote Sens. 2020; 12:2379. https://doi.org/10.3390/rs12152379

13. Ke Y, Quackenbush LJ. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. Int J Remote Sens. 2011; 32:4725–4747. https://doi.org/10.1080/01431161.2010.494184

14. Bosch M. DetecTree: Tree detection from aerial imagery in Python. JOSS. 2020; 5:2172. https://doi.org/10.21105/joss.02172

15. Schiefer F, Kattenborn T, Frick A, Frey J, Schall P, Koch B, et al. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. ISPRS J Photogramm Remote Sens. 2020; 170:205–215. https://doi.org/10.1016/j.isprsjprs.2020.10.015

16. Weinstein BG, Marconi S, Aubry-Kientz M, Vincent G, Senyondo H, White EP. DeepForest: A Python package for RGB deep learning tree crown delineation. Methods Ecol Evol. 2020; 11:1743–1751. https://doi.org/10.1111/2041-210X.13472

17. Fassnacht FE, Latifi H, Stereńczak K, Modzelewska A, Lefsky M, Waser LT, et al. Review of studies on tree species classification from remotely sensed data. Remote Sens Environ. 2016; 186:64–87. https://doi.org/10.1016/j.rse.2016.08.013

18. Seeley MM, Vaughn NR, Shanks BL, Martin RE, König M, Asner GP. Classifying a Highly Polymorphic Tree Species across Landscapes Using Airborne Imaging Spectroscopy. Preprints. 2023. https://doi.org/10.20944/preprints202307.1570.v1

19. Maschler J, Atzberger C, Immitzer M. Individual Tree Crown Segmentation and Classification of 13 Tree Species Using Airborne Hyperspectral Data. Remote Sens. 2018; 10:1218. https://doi.org/10.3390/rs10081218

20. Marconi S, Weinstein BG, Zou S, Bohlman SA, Zare A, Singh A, et al. Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network. Remote Sens Environ. 2022; 282:113264. https://doi.org/10.1016/j.rse.2022.113264

21. Fricker GA, Ventura JD, Wolf JA, North MP, Davis FW, Franklin J. A Convolutional Neural Network Classifier Identifies Tree Species in Mixed-Conifer Forest from Hyperspectral Imagery. Remote Sens. 2019; 11:2326. https://doi.org/10.3390/rs11192326

22. La Rosa LEC, Sothe C, Feitosa RQ, de Almeida CM, Schimalski MB, Oliveira DAB. Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. ISPRS J Photogramm Remote Sens. 2021; 179:35–49. https://doi.org/10.1016/j.isprsjprs.2021.07.001

23. Veras HFP, Ferreira MP, da Cunha Neto EM, Figueiredo EO, Corte APD, Sanquetta CR. Fusing multi-season UAS images with convolutional neural networks to map tree species in Amazonian forests. Ecol Inform. 2022; 71:101815. https://doi.org/10.1016/j.ecoinf.2022.101815

24. Onishi M, Ise T. Explainable identification and mapping of trees using UAV RGB image and deep learning. Sci Rep. 2021; 11:903. https://doi.org/10.1038/s41598-020-79653-9 PMID: 33441689

25. Lee J, Cai X, Lellmann J, Dalponte M, Malhi Y, Butt N, et al. Individual Tree Species Classification From Airborne Multisensor Imagery Using Robust PCA. IEEE J Sel Top Appl Earth Obs Remote Sens. 2016; 9:2554–2567. https://doi.org/10.1109/JSTARS.2016.2569408

26. Kwon R, Ryu Y, Yang T, Zhong Z, Im J. Merging multiple sensing platforms and deep learning empowers individual tree mapping and species detection at the city scale. ISPRS J Photogramm Remote Sens. 2023; 206:201–221. https://doi.org/10.1016/j.isprsjprs.2023.11.011

27. Weinstein BG, Marconi S, Graves SJ, Zare A, Singh A, Bohlman SA, et al. Capturing long-tailed individual tree diversity using an airborne imaging and a multi-temporal hierarchical model. Remote Sens Ecol Conserv. 2023; 9:656–670. https://doi.org/10.1002/rse2.335

28. Musinsky J, Goulden T, Wirth G, Leisso N, Krause K, Haynes M, et al. Spanning scales: The airborne spatial and temporal sampling design of the National Ecological Observatory Network. Methods Ecol Evol. 2022; 13:1866–1884. https://doi.org/10.1111/2041-210X.13942

29. Kampe TU, Johnson BR, Kuester MA, Keller M. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. JARS. 2010; 4:043510. https://doi.org/10.1117/1.3361375

30. Egli L, LeVan KE, Work TT. Taxonomic error rates affect interpretations of a national-scale ground beetle monitoring program at National Ecological Observatory Network. Ecosphere. 2020; 11:e03035. https://doi.org/10.1002/ecs2.3035

31. Ayres E, Colliander A, Cosh MH, Roberti JA, Simkin S, Genazzio MA. Validation of SMAP Soil Moisture at Terrestrial National Ecological Observatory Network (NEON) Sites Show Potential for Soil Moisture Retrieval in Forested Areas. IEEE J Sel Top Appl Earth Obs Remote Sens. 2021; 14:10903–10918. https://doi.org/10.1109/JSTARS.2021.3121206

32. Lombardozzi DL, Wieder WR, Sobhani N, Bonan GB, Durden D, Lenz D, et al. Overcoming barriers to enable convergence research by integrating ecological and climate sciences: the NCAR–NEON system Version 1. Geosci Model Dev. 2023; 16:5979–6000. https://doi.org/10.5194/gmd-16-5979-2023

33. Weinstein BG, Marconi S, Bohlman SA, Zare A, Singh A, Graves SJ, et al. A remote sensing derived data set of 100 million individual tree crowns for the National Ecological Observatory Network. eLife. 2021; 10:e62922. https://doi.org/10.7554/eLife.62922 PMID: 33605211

34. NEON (National Ecological Observatory Network). High-resolution orthorectified camera imagery mosaic (DP3.30010.001). RELEASE-2023. 2023. https://doi.org/10.48443/67by-mq58

35. Barnett DT, Duffy PA, Schimel DS, Krauss RE, Irvine KM, Davis FW, et al. The terrestrial organism and biogeochemistry spatial sampling design for the National Ecological Observatory Network. Ecosphere. 2019; 10:e02540. https://doi.org/10.1002/ecs2.2540

36. Veblen T, Andrus R, Chai R. Permanent forest plot data from 1982–2019 at Niwot Ridge. Environmental Data Initiative. 2021. https://doi.org/10.6073/pasta/48fa11a8f5bc6541b0472bc3fd4c0c71

37. Weinstein BG, Marconi S, Bohlman S, Zare A, White E. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sens. 2019; 11:1309. https://doi.org/10.3390/rs11111309

38. Weinstein BG, Marconi S, Bohlman SA, Zare A, White EP. Cross-site learning in deep learning RGB tree crown detection. Ecol Inform. 2020; 56:101061. https://doi.org/10.1016/j.ecoinf.2020.101061

39. Reiersen G, Dao D, Lütjens B, Klemmer K, Amara K, Steinegger A, et al. ReforesTree: A Dataset for Estimating Tropical Forest Carbon Stock with Deep Learning and Aerial Imagery. arXiv. 2022. Available from: http://arxiv.org/abs/2201.11192.

40. Kapil R, Marvasti-Zadeh SM, Goodsman D, Ray N, Erbilgin N. Classification of Bark Beetle-Induced Forest Tree Mortality using Deep Learning. arXiv. 2022. Available from: http://arxiv.org/abs/2207.07241.

41. Gan Y, Wang Q, Iio A. 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 Sens. 2023; 15:778. https://doi.org/10.3390/rs15030778

42. Liu Z, Miao Z, Zhan X, Wang J, Gong B, Yu SX. Large-Scale Long-Tailed Recognition in an Open World. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE; 2019. p. 2532–2541. https://doi.org/10.1109/CVPR.2019.00264

43. Hang R, Li Z, Liu Q, Ghamisi P, Bhattacharyya SS. Hyperspectral Image Classification With Attention-Aided CNNs. IEEE Trans Geosci Remote Sens. 2021; 59:2281–2293. https://doi.org/10.1109/TGRS.2020.3007921

44. Busing RT. Tree mortality, canopy turnover, and woody detritus in old cove forests of the southern Appalachians. Ecology. 2005; 86:73–84. https://doi.org/10.1890/04-0410

45. Liao W, Van Coillie F, Gao L, Li L, Zhang B, Chanussot J. Deep Learning for Fusion of APEX Hyper-spectral and Full-Waveform LiDAR Remote Sensing Data for Tree Species Mapping. IEEE Access. 2018; 6:68716–68729. https://doi.org/10.1109/ACCESS.2018.2880083

46. Sumbul G, Cinbis RG, Aksoy S. Multisource Region Attention Network for Fine-Grained Object Recognition in Remote Sensing Imagery. IEEE Trans Geosci Remote Sens. 2019; 57:4929–4937. https://doi.org/10.1109/TGRS.2019.2894425

47. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, et al. Ecological Niches and Geographic Distributions (MPB-49). Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press; 2011. https://doi.org/10.1515/9781400840670

48. Condit R, Engelbrecht BMJ, Pino D, Pérez R, Turner BL. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. Proc Natl Acad Sci U S A. 2013; 110:5064–5068. https://doi.org/10.1073/pnas.1218042110 PMID: 23440213

49. Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA. Phylogenetic Overdispersion in Floridian Oak Communities. Am Nat. 2004; 163:823–843. https://doi.org/10.1086/386375 PMID: 15266381

50. Freestone AL, Inouye BD. Dispersal Limitation and Environmental Heterogeneity Shape Scale-Dependent Diversity Patterns in Plant Communities. Ecology. 2006; 87:2425–2432. https://doi.org/10.1890/0012-9658(2006)87[2425:dlaehs]2.0.co;2 PMID: 17089651

51. Martínez-Sánchez JL, Martínez-Garza C, Cámara L, Castillo O. Species-specific or generic allometric equations: which option is better when estimating the biomass of Mexican tropical humid forests? Carbon Management. 2020; 11:241–249. https://doi.org/10.1080/17583004.2020.1738823

52. Duncanson L, Dubayah R, Enquist B. Assessing the general patterns of forest structure: Quantifying tree and forest allometric scaling relationships in the United States. Glob Ecol Biogeogr. 2015; 24:1465–1475. https://doi.org/10.1111/geb.12371

53. Hemming-Schroeder NM, Gutierrez AA, Allison SD, Randerson JT. Estimating Individual Tree Mortality in the Sierra Nevada Using Lidar and Multispectral Reflectance Data. JGR. Biogeosciences. 2023; 128:e2022JG007234. https://doi.org/10.1029/2022JG007234

54. Weinstein BG, Graves SJ, Marconi S, Singh A, Zare A, Stewart D, et al. 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 Comput Biol. 2021; 17:e1009180. https://doi.org/10.1371/journal.pcbi.1009180 PMID: 34214077

55. Atkinson RRL, Burrell MM, Rose KE, Osborne CP, Rees M. The dynamics of recovery and growth: how defoliation affects stored resources. Proc R Soc B Biol Sci. 2014; 281:20133355. https://doi.org/10.1098/rspb.2013.3355 PMID: 24671974

56. Augustine BC, Koneff MD, Pickens BA, Royle JA. Towards estimating marine wildlife abundance using aerial surveys and deep learning with hierarchical classifications subject to error. bioRxiv. 2023:2023.02.20.529272. https://doi.org/10.1101/2023.02.20.529272

57. Guo C, Pleiss G, Sun Y, Weinberger KQ. On Calibration of Modern Neural Networks. Proceedings of the 34th International Conference on Machine Learning. PMLR. 2017:1321–1330. Available from: https://proceedings.mlr.press/v70/guo17a.html.

58. Orenstein EC, Kenitz KM, Roberts PLD, Franks PJS, Jaffe JS, Barton AD. Semi- and fully supervised quantification techniques to improve population estimates from machine classifiers. Limnol Oceanogr Methods. 2020; 18:739–753. https://doi.org/10.1002/lom3.10399

59. Sumbul G, Cinbis RG, Aksoy S. Fine-Grained Object Recognition and Zero-Shot Learning in Remote Sensing Imagery. IEEE Trans Geosci Remote Sens. 2018; 56:770–779. https://doi.org/10.1109/TGRS.2017.2754648

60. Stork L, Weber A, van den Herik J, Plaat A, Verbeek F, Wolstencroft K. Large-scale zero-shot learning in the wild: Classifying zoological illustrations. Ecol Inform. 2021; 62:101222. https://doi.org/10.1016/j.ecoinf.2021.101222

61. Koh PW, Sagawa S, Marklund H, Xie SM, Zhang M, Balsubramani A, et al. WILDS: A Benchmark of in-the-Wild Distribution Shifts. Proceedings of the 38th International Conference on Machine Learning. PMLR. 2021:5637–5664. Available from: https://proceedings.mlr.press/v139/koh21a.html.