

RESEARCH ARTICLE

# Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy

Claire A. Baldeck<sup>1\*</sup>, Gregory P. Asner<sup>1</sup>, Robin E. Martin<sup>1</sup>, Christopher B. Anderson<sup>1</sup>, David E. Knapp<sup>1</sup>, James R. Kellner<sup>2</sup>, S. Joseph Wright<sup>3</sup>

**1** Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, United States of America, **2** Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, United States of America, **3** Smithsonian Tropical Research Institute, Washington, DC, United States of America

\* [cbaldeck@carnegiescience.edu](mailto:cbaldeck@carnegiescience.edu)



**OPEN ACCESS**

**Citation:** Baldeck CA, Asner GP, Martin RE, Anderson CB, Knapp DE, Kellner JR, et al. (2015) Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLoS ONE* 10(7): e0118403. doi:10.1371/journal.pone.0118403

**Editor:** Lalit Kumar, University of New England, AUSTRALIA

**Received:** June 30, 2014

**Accepted:** January 14, 2015

**Published:** July 8, 2015

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** The spectral data are owned by The Carnegie Airborne Observatory. Requests to access this data should be directed to Gregory Asner ([gpa@stanford.edu](mailto:gpa@stanford.edu)).

**Funding:** This study has been supported by the Avatar Alliance Foundation, Gordon and Betty Moore Foundation, John D. and Catherine R. MacArthur Foundation, Grantham Foundation for the Protection of the Environment, W. M. Keck Foundation, Margaret A. Cargill Foundation, Mary Anne Nyburg Baker and G. Leonard Baker Jr., and William R. Hearst III. There are no grant numbers or URLs. The

## Abstract

Remote identification and mapping of canopy tree species can contribute valuable information towards our understanding of ecosystem biodiversity and function over large spatial scales. However, the extreme challenges posed by highly diverse, closed-canopy tropical forests have prevented automated remote species mapping of non-flowering tree crowns in these ecosystems. We set out to identify individuals of three focal canopy tree species amongst a diverse background of tree and liana species on Barro Colorado Island, Panama, using airborne imaging spectroscopy data. First, we compared two leading single-class classification methods—binary support vector machine (SVM) and biased SVM—for their performance in identifying pixels of a single focal species. From this comparison we determined that biased SVM was more precise and created a multi-species classification model by combining the three biased SVM models. This model was applied to the imagery to identify pixels belonging to the three focal species and the prediction results were then processed to create a map of focal species crown objects. Crown-level cross-validation of the training data indicated that the multi-species classification model had pixel-level producer's accuracies of 94–97% for the three focal species, and field validation of the predicted crown objects indicated that these had user's accuracies of 94–100%. Our results demonstrate the ability of high spatial and spectral resolution remote sensing to accurately detect non-flowering crowns of focal species within a diverse tropical forest. We attribute the success of our model to recent classification and mapping techniques adapted to species detection in diverse closed-canopy forests, which can pave the way for remote species mapping in a wider variety of ecosystems.

## Introduction

A comprehensive understanding of tropical forest biodiversity and function over large spatial scales is hindered by a lack of spatially extensive information on tree species composition, as

fundings had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

field measurements are often spatially localized and have difficulty resolving large geographic patterns [1]. Remote sensing can extend field-based knowledge over large areas by providing continuous spatial coverage of surface reflectance and structural properties [2,3]. In particular, imaging spectroscopy (or hyperspectral remote sensing) measures light reflected in hundreds of narrow, contiguous spectral bands, which may allow the discrimination of different plant species. However, the use of imaging spectroscopy for remote species mapping presents a major practical challenge in tropical forests where the species richness of trees  $\geq 10$  cm dbh may exceed 300 species per ha [4]. As automated tree species mapping has not yet been accomplished in a high diversity tropical forest, we currently have little understanding of the ability of imaging spectroscopy to map species in these ecosystems.

To assess the potential for remote species mapping in tropical forests, a number of studies have examined the spectral separability and classification performance of different tropical plant species and growth forms. Spectral differences have been found among leaf- and branch-level spectra of different species from the same ecosystem, as well as the spectra of different species detected from hyperspectral airborne and satellite sensors [5–11]. Important spectral differences have also been found between tree and liana species [12,13], as well as between groups of native and invasive species [14]. These results indicate the potential for classification models to differentiate species or groups of species based on their remotely-sensed spectral signatures.

In a foundational study, Clark et al. [7] attempted to classify seven canopy tree species from a diverse Costa Rican rainforest based on their remotely-sensed hyperspectral signatures collected with the airborne sensor HYDICE. Their most successful model classified the crowns of their study species with 92% accuracy. Despite these encouraging results, this classification model could not be used to map any of the seven study species across the study site. The primary reason for this was that the classification model differentiated the seven study species from one another, but not from the hundreds of other tree and liana species present at the site. In addition, the preferred model was based on crown-mean spectra, and thus operational species mapping would require automatic delineation of tree crowns throughout the image prior to applying the classification model to predict species identity. At the time of the study, the ability to accurately, automatically delineate individual tree crowns in a closed-canopy tropical forest did not exist, and nearly a decade later it still does not.

Since Clark et al. [7], others have performed successful classifications of tropical forest species based on their remotely-sensed spectral signatures, but these models were also non-operational in that they were unable to distinguish the study species from the other species present at the study site, e.g., [15]. Meanwhile, operational species mapping has been accomplished in several ecosystems with relatively low species diversity [16–20]. The inability of operational species mapping to move beyond lower diversity ecosystems may lie in the classification approach used. In the aforementioned classification and mapping studies, species classification models were based upon the traditional multi-class classification approach, which requires representative training data to be collected for all model classes (species) present in the scene. This requirement is not likely to be fulfilled in tropical forests that harbor hundreds to thousands of species. Even if representative training data were available for all canopy species in a high-diversity forest, the outlook for such a classification model would be quite dim as classification accuracy is known to decline as the number of classes increases [18,21]. In the face of these mounting challenges, it is not surprising that the remote sensing community has expressed ambivalence towards the prospect of remote species mapping in high diversity tropical forests [22].

Clearly, a different approach is needed to accomplish operational remote species mapping in high-diversity forests. A possible solution may come from single-class classification methods

that distinguish samples of a single focal class from those of all other classes, e.g., [23–25]. These methods are advantageous when the interest is in mapping a single class because they aim to maximize classification accuracy with respect to the focal class rather than the average classification accuracy over all model classes. They also reduce the overall amount of training data required compared to traditional multi-class classification, and some require training data from only the class of interest [23,24], making acquisition of an appropriate training dataset more feasible for tropical forests. Moreover, single-species classification models can be directly applied for species mapping in the area of study because they explicitly aim to distinguish the focal species from all other species. Thus far, single-class classification methods have been explored for species classification and mapping with promising results in a few ecosystems with relatively low species diversity [25–27]. If these methods are found to be effective in high-diversity ecosystems, single-species classification may solve many of the problems associated with remote species mapping in tropical forests.

Taking a single-species detection approach, we set out to map three important canopy tree species within the diverse, closed-canopy tropical forest of Barro Colorado Island, Panama, using high spatial resolution imaging spectrometer data. We evaluated two leading single-class classification methods (discussed in detail in Methods section) for their ability to distinguish the remotely-sensed spectra of the focal species from a diverse background of other tree and liana species based solely on their non-flowering crown spectral characteristics. We combined the selected single-species models into a multi-species model for mapping the three focal species across the island and evaluated the predictions by confirming the species identities in the field. Our species maps were highly reliable, demonstrating that tropical tree species can be effectively mapped with airborne imaging spectroscopy data. Together, the techniques presented here represent a high-performance, operational method for the identification of individual tree crown objects of one or more focal species that is adapted to the challenges encountered in tropical forests and greatly reduces the burden of training data collection.

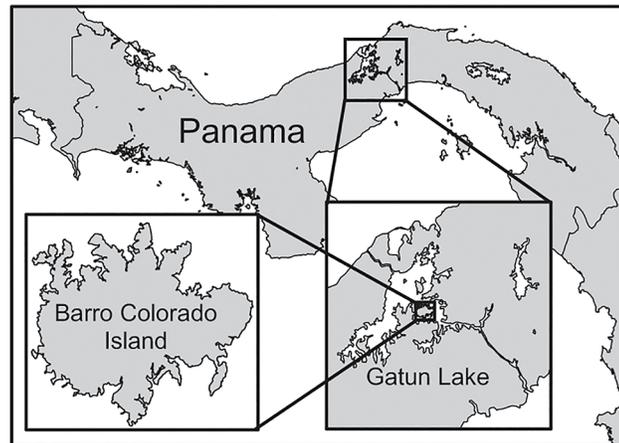
## Methods

### Study site

Barro Colorado Island (BCI) is a 1560-ha island located in Gatun Lake, which was formed in the early 20<sup>th</sup> century as part of the construction of the Panama Canal (Fig 1). The BCI forest is tropical moist, with a mean annual temperature of 26°C and mean annual rainfall of 2600 mm. There is a distinct dry season from January to April, and an estimated 6.3% of trees > 30 cm dbh are deciduous during the dry season [28]. There are approximately 500 tree species in the BCI forest [29,30], with nearly 60 tree species ha<sup>-1</sup> among trees ≥ 20 cm dbh [31], a size typically considered to occupy the canopy. In addition to the high species diversity of trees present in the canopy, there are many lianas present in this forest. Less is known about the diversity of lianas at the top of the canopy, but a census of liana stems on BCI found an average of 75 liana species ha<sup>-1</sup> for stems ≥ 1 cm in diameter [32]. Surveys of liana infestation of trees have estimated that ~78% of trees ≥ 20 cm dbh have at least one liana in their crown, though not necessarily overtopping the crown [33].

### Spectral data

Imaging spectrometer data were collected over Barro Colorado Island, Panama, with the Carnegie Airborne Observatory Airborne Taxonomic Mapping System (CAO AToMS; [34]) in February of 2012. The CAO was flown at an altitude of 1 km providing data at 1.12 m spatial resolution, and was flown between 10 am and 1:30 pm under nearly cloud-free conditions to minimize shadow effects. The CAO AToMS system consists of three instrument subunits: i) a



**Fig 1. Map of Barro Colorado Island (BCI), Panama.**

doi:10.1371/journal.pone.0118403.g001

High-fidelity Imaging Spectrometer (HiFIS), ii) a Light Detection and Ranging (LiDAR) scanner, and iii) a Global Positioning System-Inertial Measurement Unit (GPS-IMU). The CAO HiFIS subsystem provided spectroscopic images spanning the visible to shortwave infrared portion of the spectrum between 380 and 2512 nm, which was resampled to 10.4 nm spectral resolution. The LiDAR subsystem was operated in discrete return mode with a density of four laser shots per square meter with up to four returns per shot, providing three-dimensional information on terrain and canopy structure [34]. The GPS-IMU subsystem provided positioning and attitude data to project the HiFIS and LiDAR data onto the land surface. Canopy height information provided by the LiDAR subsystem was used for accurate orthorectification of the spectral data. Radiance data were converted to surface reflectance using ACORN 5BatchLi (ImSpec LLC, Palmdale, CA) with a MODTRAN look-up table and reflectance was BRDF adjusted to correct for cross-track reflectance gradients [35]. Prior to analysis, we removed spectral bands that were heavily influenced by water absorption (1330–1460 nm and 1770–2022 nm) and noisy bands at the upper and lower portions of the spectrum (below 400 nm and above 2422 nm). This resulted in a total of 167 spectral bands used for analysis.

A field campaign was conducted to locate and identify individual tree crowns within the CAO image in January of 2013. Crown locations were recorded in the field with a handheld Global Positioning System (GS50 Leica Geosystems Inc., Norcross, GA, USA). These crowns were then manually delineated in the CAO imagery, taking care to ensure that the pixels extracted were as pure as possible, without contamination from lianas or neighboring trees. The crowns were collected along trails traversing the full east-west span of the island, sampling across all flight lines. The campaign was conducted to sample the tree diversity of the island as widely as possible, without sampling large numbers of any one species. This resulted in 340 crowns representing 84 tree species. Although not used in this study, leaves were collected from identified crowns with a permit from the Smithsonian Tropical Research Institution to Robin Martin.

To obtain a larger sample of crowns for a few key species, we used satellite imagery collected over the island from the Ikonos, QuickBird, and WorldView-2 sensors. Three common canopy tree species—*Dipteryx panamensis*, *Handroanthus guayacan* (formerly assigned to the genus *Tabebuia*), and *Jacaranda copaia*—flower at different times of the year. *D. panamensis* flowers in June–August, *H. guayacan* flowers after heavy rainfall events in the second half of the dry season, and *J. copaia* flowers in March–April before the onset of the wet season. In these events,

many mature canopy individuals will display colorful flowers that allow them to be identified in multispectral satellite images or aerial photographs. Individuals of these species were identified in a series of satellite images acquired between 2002 and 2011 during flowering events [36]. Locations of crowns identified from the satellite images were superimposed on the CAO image and 40 crowns of each species were manually delineated in the CAO image. Crowns transferred to the CAO image were distributed evenly across the island while avoiding proximity to trails (< 20 m). Crowns of *H. guayacan* and *J. copaia* were checked in the field at the time of the CAO flight and both species were found to have foliage but not flowers. Thus our efforts to identify these species using imaging spectroscopy are based solely on the optical reflectance properties of non-flowering tree crowns.

The field-collected crowns and those obtained from the satellite data were combined into a single dataset to compare methods for single species identification and to map the focal species across BCI. Each of the three focal species was well represented by over 40 training crowns, while the field-collected crowns provided a broad sampling of the tree species across the island. Prior to analysis, all of the spectral data were filtered to include only well-lit, leafy vegetation pixels with NDVI  $\geq 0.7$  and mean NIR (850–1050 nm) reflectance  $\geq 21\%$  (referred to here as “vegetation pixels”). A summary of the spectral data is presented in Table 1 and the mean and 95% confidence interval of the spectra for the three focal species is shown in Fig 2, superimposed on the 95% confidence interval for the spectra of all vegetation pixels on the island.

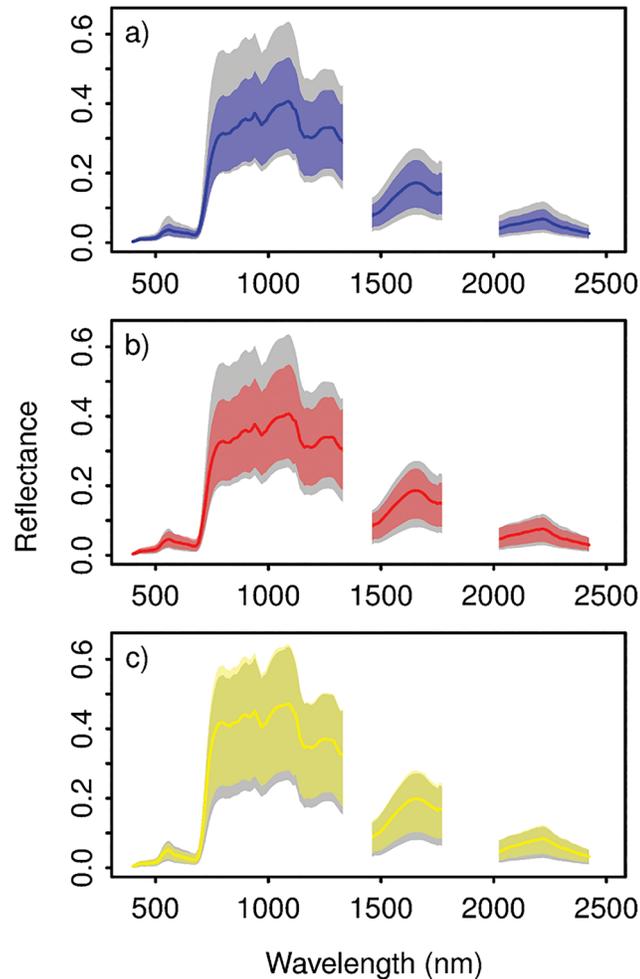
## Single-species classification methods

We investigated two support vector machine (SVM) techniques that have shown promise in distinguishing samples of a single focal species from those of all other species—binary SVM and biased SVM [26]. The basic SVM is a non-parametric classification method that is widely used in remote sensing because of its good classification performance and elegant handling of high-dimensional data [37–39]. It has also been shown to be among the top-performing classification methods in remote species identification [17,18,40]. SVM is a binary classifier that separates samples of two classes in feature space with a decision boundary [41,42]. This boundary is generated by maximizing the margins between the boundary and the closest training samples (the support vectors), while minimizing misclassification errors. A fitted cost parameter,  $C$ , controls the penalty associated with misclassification error and thus the complexity of the decision boundary; a higher misclassification penalty will result in a more complex model, but with more danger of overfitting. A kernel function may be used to perform an implicit mapping of the data into a higher dimensional feature space [43,44], allowing non-linear separation of the classes in the original feature space. The popular radial basis function (RBF) kernel has a single fitted parameter,  $\gamma$ , controlling the kernel width. Thus, constructing an SVM with an RBF kernel requires fitting of two parameters,  $\gamma$  and  $C$ . Optimal values of these parameters are found through cross-validation on the training dataset.

**Table 1. Summary of the spectral data from both the field-collection campaign and the crowns obtained from satellite data.**

Species	# Crowns	# Pixels
<i>D. panamensis</i>	50	10726
<i>H. guayacan</i>	44	3325
<i>J. copaia</i>	44	2091
Other	322	9350
Total	460	25492

doi:10.1371/journal.pone.0118403.t001



**Fig 2. Mean and 95% confidence interval of the reflectance spectra for each focal species.** (a) *Dipteryx panamensis*, (b) *Handroanthus guayacan*, and (c) *Jacaranda copaia*. The colored lines represent the mean reflectance, and the colored areas represent the 95% confidence interval for each spectral band. The gray area represents the 95% confidence interval for the well-lit leafy vegetation on BCI.

doi:10.1371/journal.pone.0118403.g002

For this analysis we used a specific type of binary SVM in which samples from a single class of interest, or the focal class, are contrasted against samples from all other classes, which together make up the non-focal, or outlier, class. This is also referred to as “one-against-all” binary SVM. The F-score on the focal class ( $F = 2pr/(p+r)$ , where  $r$  = sensitivity or recall accuracy and  $p$  = precision) is an appropriate choice of optimization criterion in single-class classification with highly imbalanced class sizes. Sensitivity is the proportion of focal class samples that are correctly assigned by the classifier to the focal class (i.e., the producer’s accuracy for the focal class) and precision is the proportion of samples assigned to the focal class that truly belong to the focal class (i.e., the user’s accuracy for the focal class). Maximization of the F-score selects a combination of parameter values that yields both high sensitivity and high precision, while keeping these two aspects of model performance relatively well balanced.

Biased SVM [24] works similarly to the standard binary SVM in that it finds an optimal separation between two classes in feature space. However, the collected training data from the focal class are contrasted against samples that are randomly selected from the data pool (in this case, the vegetation pixels from the entire island), referred to here as “pseudo-outliers”. Because

the pseudo-outlier data are of unknown identity and will contain samples from the focal class, errors occurring within the pseudo-outlier class are given a lower penalty than errors occurring within the focal class. An extra parameter,  $w_c$ , is added to the model, which controls the relative cost of errors occurring in the two classes.

The parameterization of biased SVM is also performed through cross-validation on the training dataset. However, without known samples from the outlier class, precision cannot be calculated. A tuning criterion that works well for biased SVM is  $r^2/P[f(x) = 1]$ , where  $r$  is the sensitivity (recall accuracy) and  $P[f(x) = 1]$  is the probability that a sample is assigned to the focal class [26,45]. This criterion is based on the concept that model sensitivity should be maximized while simultaneously minimizing the number of samples that are assigned to the focal class, and is thought to work similarly to the F-score in that it will be large when both recall and precision are large [45]. In a study examining the optimization of biased SVM for remote tree species mapping, Baldeck and Asner [26] found that this criterion worked well for optimizing biased SVMs and generally outperformed an alternative optimization criterion. They also found that cross-validation generally worked best when performed at the crown level (i.e., by splitting crowns rather than pixels into the cross-validation groups).

One-against-all binary SVM and biased SVM have thus far been applied to species detection in only a few instances [26,27]. Binary SVM was initially used in a study exploring different methods for species detection in a Hawaiian forest [27]. Biased SVM is a relatively new single-class classification method that first showed promise in studies of web document and land-cover classification [24,46], and represented an attractive possibility for species detection because it could be trained with a fraction of the training data used in binary SVM. Only one study has investigated the potential of biased SVM for use in single-species detection [26]. That study compared biased SVM to binary SVM using crown spectral data from a savanna ecosystem and found that biased SVM performed well, though it did not achieve as high overall performance as binary SVM.

## Comparison of binary versus biased SVM

To create the binary SVM models, the training data were divided into the focal species class and the non-focal, or outlier, species class. The training dataset for each focal species consisted of all the focal species crowns plus all outlier species crowns from the field collection campaign. Only crowns from the field collection campaign were used to form the outlier class to prevent skewing this class heavily toward the other focal species. It has been shown that the balance in the amount of data from the focal and outlier classes can affect model performance in this type of one-against-all binary classification [26]; therefore, we conducted a preliminary analysis to assess model balance for these datasets, which is described in the Supporting Information. We found that the ratio of focal to outlier species crowns present in these datasets (approximately 0.15) yielded good results (S1 Fig), and so the full datasets were used in the tests of the binary SVM models.

For each focal species binary SVM model, the optimal values of  $\gamma$  and  $C$  were found by an exhaustive grid search over the values  $\gamma \in \{e^{-12}, e^{-11}, \dots, e^{-6}\}$  and  $C \in \{e^5, e^6, \dots, e^{15}\}$  using five-fold crown-level cross-validation [21]. In this procedure, the crowns of each class were split into five roughly equally sized groups. One group from each class was set aside to form the test dataset and a model was built from the remaining data using the trial parameter values. This was repeated five times such that the data from each group was predicted with a model built from a separate set of crowns. The F-score on the focal class was calculated based on the number of pixels that were correctly classified and this was repeated for every parameter

combination. The combination of  $\gamma$  and  $C$  resulting in the highest cross-validation F-score was chosen for the final model for each species.

To create the biased SVM models, the focal species crowns were used to form the focal class and 40,000 vegetation pixels were randomly sampled from the CAO imagery to form the pseudo-outlier class. The biased SVM models were optimized using the same crown-level cross-validation procedure described for binary SVM, except that the pseudo-outlier data were split into five groups randomly by pixel. The optimal values of  $\gamma$ ,  $C$ , and  $w_c$  were found through a grid search over the values  $\gamma \in \{e^{-8}, e^{-7}, \dots, e^{-2}\}$ ,  $C \in \{e^5, e^6, \dots, e^{11}\}$ , and  $w_c \in \{0.1, 0.2, \dots, 0.7\}$ . To save computation time, a coarse search was conducted over these ranges using every other value of each parameter, then a fine-scale search was conducted bracketing the optimal values identified from the coarse search. The biased SVMs were optimized using the  $r^2/P[f(x) = 1]$  criterion estimated using all focal species and pseudo-outlier pixels [45].

Once the optimal parameters were identified, we compared the performance of the binary and biased SVM models. For binary SVM, one-fifth of both the focal and outlier crowns were removed to form the test dataset and a model was built using the remaining data using the optimal parameter values. For biased SVM, one-fifth of the focal species crowns were removed and placed in the test dataset and a model was built using the remaining focal species crowns and 32,000 pseudo-outlier pixels. Focal and outlier crowns were randomly assigned to the training and test groups and the same group assignments were used for both the binary and biased models of the same species. This procedure was repeated 100 times. We were especially interested in comparing the sensitivity and specificity (the proportion of outlier class samples that are correctly assigned to the outlier class) of the two classifiers. Therefore, although the outlier crowns were not used to train the biased SVMs, the test outlier crowns were passed to the biased SVM models to estimate specificity. We also compared the performance of binary and biased SVM by using each to map the focal species across the island. For each focal species and classifier type, a classification model was built from the full dataset and this was applied to all vegetation pixels in the CAO imagery.

## Focal species mapping

Based on our comparison of binary and biased SVM methods for mapping the focal species (see [Results](#) section below), we chose to use the biased SVM method as the basis for creating the final maps of the three focal species. A final biased SVM model was created for each focal species using all of the data available for that species plus all of the pseudo-outlier data. Then the three final biased SVMs were combined into a multi-species classification to be used for mapping the three species simultaneously.

As the three biased SVM models are independent of one another, it is possible for a pixel to be assigned to the focal species of more than one model. To break such ties we trained an ordinary multi-class SVM with an RBF kernel to differentiate the three focal classes from one another. All crowns from the three focal species were used, but the data from *D. panamensis* and *H. guayacan* were randomly subset to 2091 pixels (the number available for *J. copaia*) to create evenly balanced species classes. The biased SVMs for the three focal species and the tie-breaking SVM were combined into the final, multi-species classification model with the following rules: (i) If a pixel was assigned to the outlier class by all of the biased SVM models, then that pixel was assigned to the outlier class. (ii) If a pixel was assigned to the focal species class by only one biased SVM model, then it was assigned to the corresponding focal species. (iii) If a pixel was assigned to the focal species class by more than one biased SVM model, the pixel was passed to the tie-breaking SVM to determine the species assignment. For the final multi-species classification model, we obtained estimates of the pixel-level producer's accuracies for

the three focal species (equivalent to the sensitivity, but for multi-class models) using five-fold crown-level cross-validation. We then applied this model to all vegetation pixels in the CAO imagery to map the three focal species.

The raster of the model prediction results contained some amount of undesirable noise (spatially random pixels assigned to the focal classes). To reduce this noise, we applied contextual filters to the model prediction results. First, we resized the raster grid by cutting each pixel into quarters and performed opening-closing on the raster of predicted occurrences for each species ( $3 \times 3$  pixel kernel; ENVI software, Exelis, Boulder, CO USA). The opening procedure removed small objects of two pixels in width and then the closing procedure filled small holes of two pixels in width (here, two pixels in width is equivalent to one pixel in width at the original resolution). The raster grid was then resized to the original resolution and each contiguous area belonging to the same focal class was designated as an individual object (contiguity was defined as pixels that were side-adjacent). Objects with fewer than ten pixels were eliminated and objects that remained were counted as crown objects. We refer to these areas as “crown objects” rather than “crowns” to acknowledge that a single crown object may contain the crowns of two or more individual conspecific trees, and that the crown of one individual tree may be represented by two or more smaller crown objects, separated perhaps by liana or shadow.

## Field validation

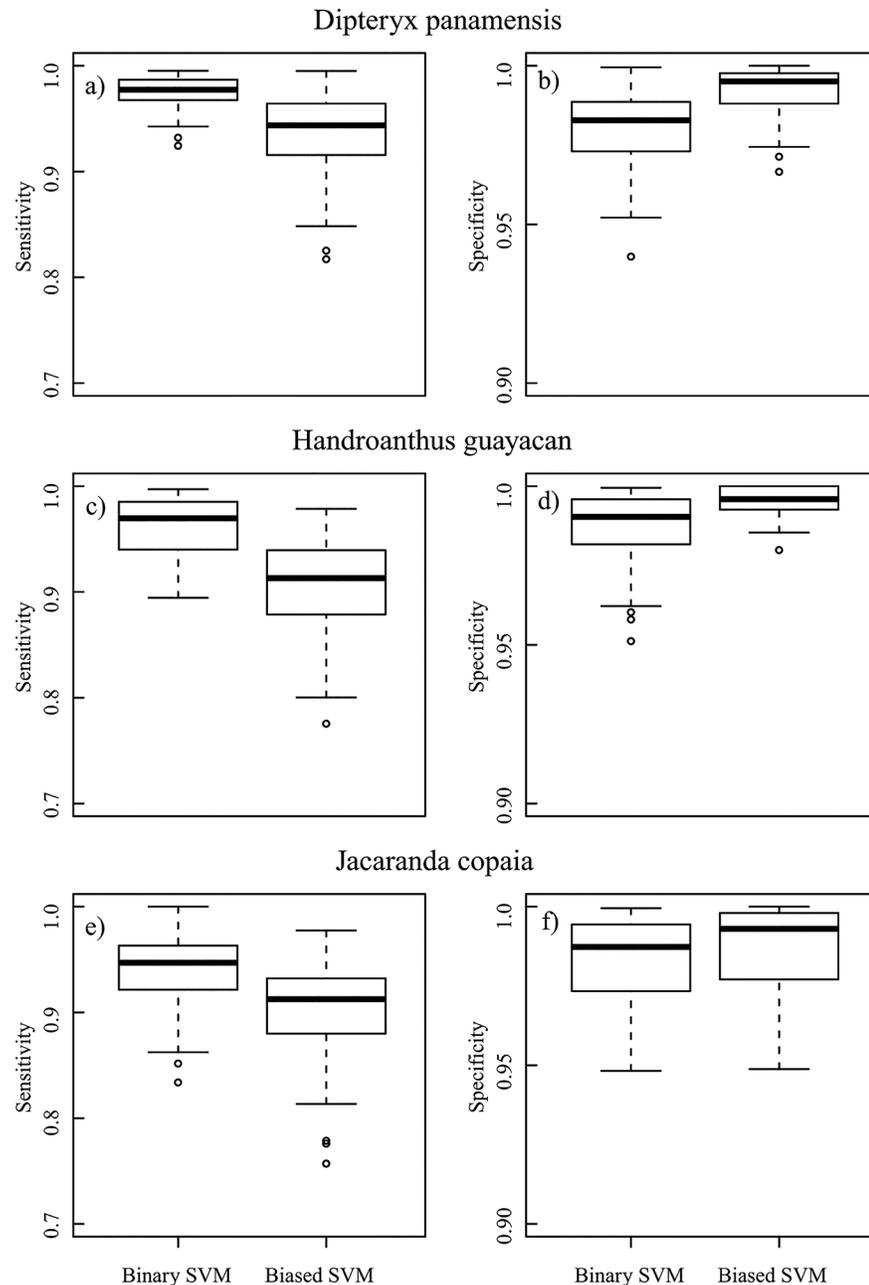
To obtain estimates of model precision for predicted crown objects, we returned to the island to verify their species identity. The prediction map was brought to the field on a tablet PC and predicted crown objects located near trails ( $< 20$  m away) were checked to verify their species identity. Only trail segments that were not visited in the first field campaign were used for model verification. No permits were needed to check the identities of the tree crowns on the island.

Additional field observations were recorded to help with the interpretation of the prediction map. We noted the presence of individuals of the focal species with crowns reaching the top of the canopy (at least 5 m in diameter) that were encountered in the field but not indicated by the prediction map, which might be considered as a failure of the model to detect the presence of that species. For all focal species trees encountered, the presence and severity of liana infestation of the tree crown was recorded, with special attention to whether the lianas were covering the top of the canopy. However, the field search for omitted focal species crowns was not systematic and there was ambiguity in which ones should be considered to be a genuine omission error (for example, the presence of lianas may indicate that the omission is correct from a remote sensing perspective). Therefore, we could not quantify the proportion of focal species canopy individuals that were not detected (i.e., we could not measure crown-object-level sensitivity). These observations were meant only to explore sources of error; we relied on the previously described accuracy analyses to estimate model sensitivity.

## Results

### Comparison of binary versus biased SVM

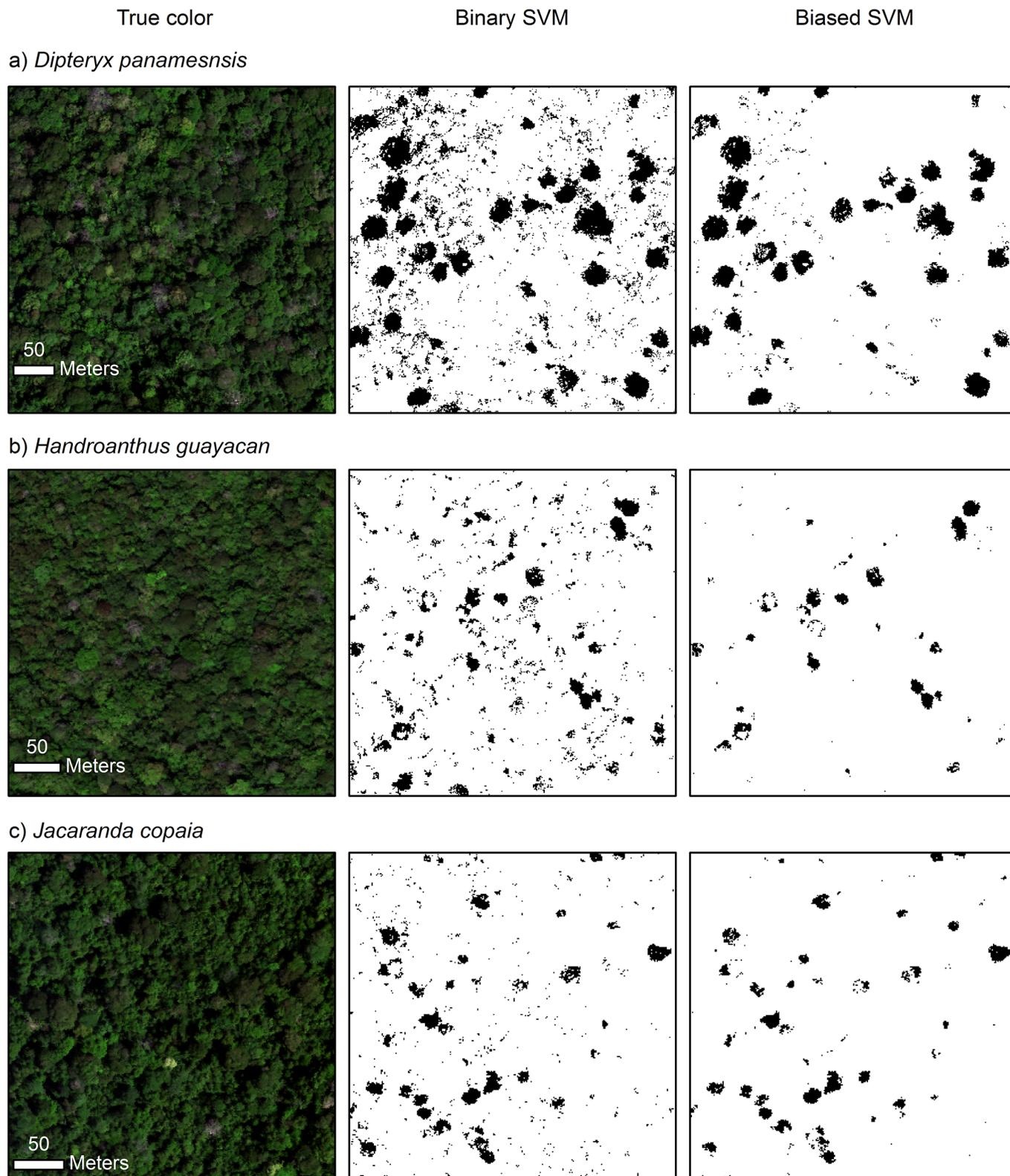
In general, binary SVM had higher sensitivity but lower specificity than biased SVM for the same focal species (Fig 3). The average model sensitivity for binary SVM was 0.036–0.057 higher across species than that of biased SVM. The difference was smaller for specificity, which was 0.004–0.012 higher across species for biased SVM compared to binary SVM. The maps produced from these models were consistent with this general result as more pixels were assigned to the focal class by the binary SVM models (Fig 4). However, the contrast in the



**Fig 3. Performance of binary and biased SVM for the three focal species.** Results are given for both model sensitivity (a, c, e) and specificity (b, d, f) measured on the spectra from a separate set of crowns. Sensitivity is the proportion of focal class samples that are correctly assigned to the focal class and specificity is the proportion of outlier class samples that are correctly assigned to the outlier class. Note the difference in the y-axis between sensitivity and specificity.

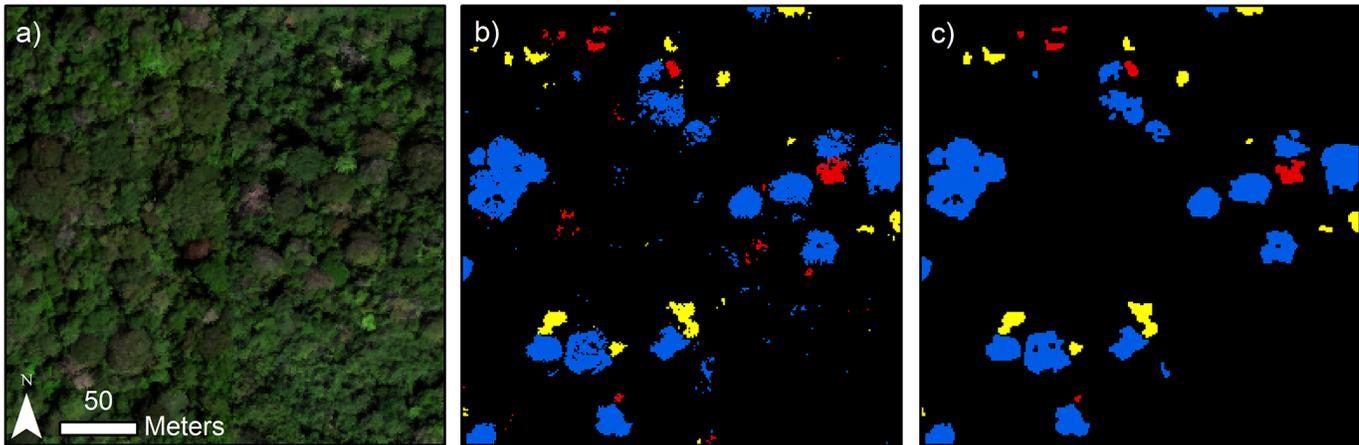
doi:10.1371/journal.pone.0118403.g003

amount of pixels assigned to the focal classes was quite strong, as 11.9%, 3.9%, and 4.3% of the vegetation pixels were classified as *D. panamensis*, *H. guayacan*, and *J. copaia*, respectively, using the binary SVM models versus only 4.1%, 0.7%, and 1.0%, respectively, using the biased SVM models. Accordingly, the binary SVM maps were noticeably noisier than those produced from biased SVM (Fig 4). We found the higher precision provided by the biased SVM models



**Fig 4. Mapped prediction results of binary and biased SVM models for the three focal species.** The true-color representation of the raw imagery is on the left-hand side, results from the binary SVM models are in the center, and results from the biased SVM models are on the right-hand side. The three spectral bands used to display the true-color image are R = 640 nm, G = 550 nm, and B = 460 nm.

doi:10.1371/journal.pone.0118403.g004



**Fig 5. Mapped results of the multi-species model before and after applying the contextual filters.** The panels are (a) true-color representation of the raw imagery, (b) prediction results for the multi-species classification model, and (c) prediction results after undergoing crown smoothing and size filtering. The three spectral bands used to display the true-color image in (a) are R = 640 nm, G = 550 nm, and B = 460 nm. The colors used to represent the species in (b) and (c) are blue = *D. panamensis*, red = *H. guayacan*, and yellow = *J. copaia*.

doi:10.1371/journal.pone.0118403.g005

to be desirable for our mapping application; therefore, we chose biased SVM as the basis for the final species classification model.

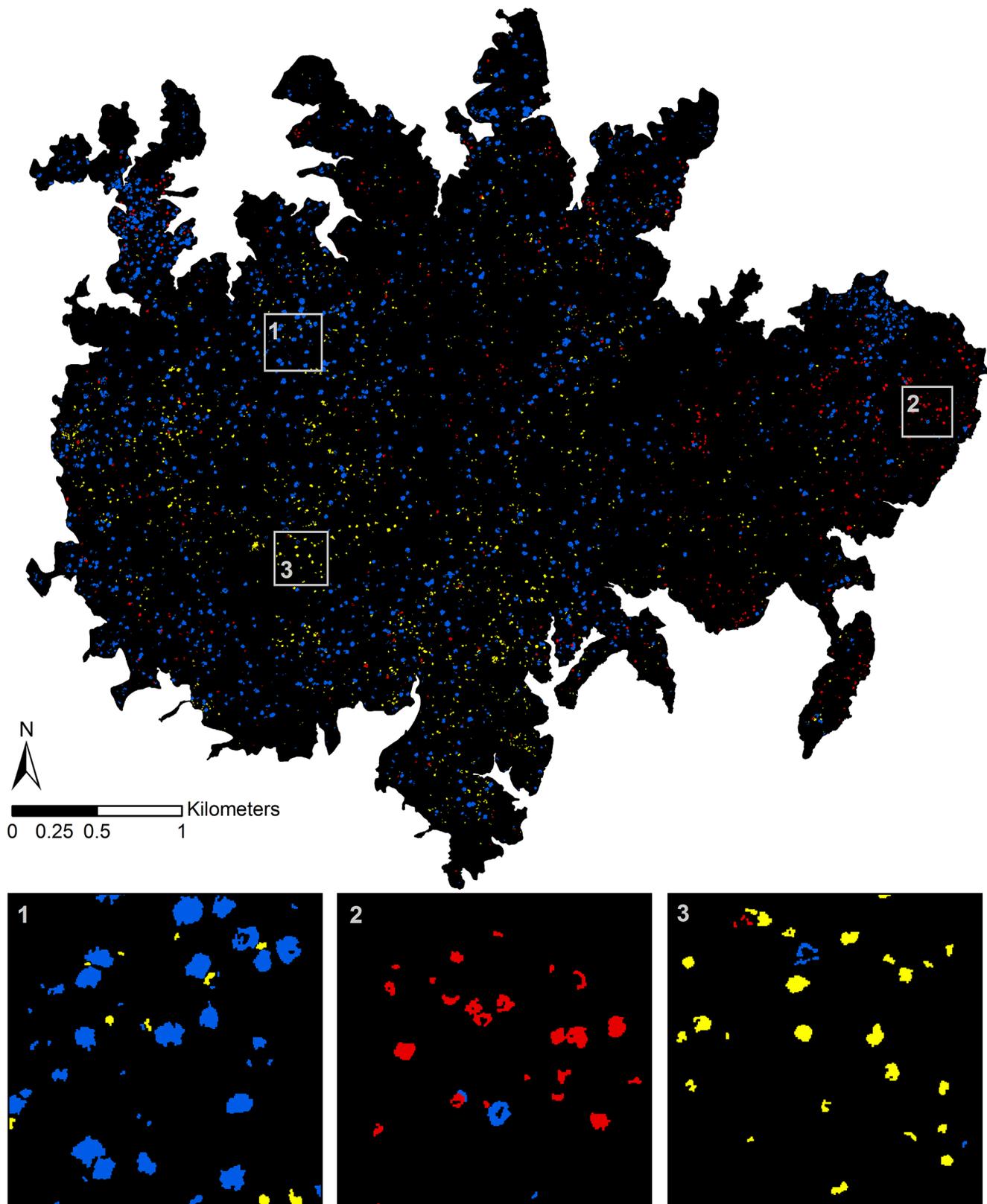
### Focal species mapping

The tie-breaking three-species classification model showed good classification performance, with an overall accuracy of 97.9% based on five-fold crown-level cross-validation. However, the tie-breaking model was rarely needed to break prediction ties. When the biased SVM models were applied to the imagery, 5.6% of the vegetation pixels were assigned to at least one of the focal species classes and of these, only 2.1% were assigned to more than one focal species class. Thus, the tie-breaking model was used to predict the identity of only 0.1% of the vegetation pixels. We estimated that the final classification model, which included the three individual biased SVM models plus the tie-breaking model, had pixel-level producer's accuracies (sensitivity for the focal class  $\times$  100) of 97.4%, 94.3%, and 93.9% for *D. panamensis*, *H. guayacan*, and *J. copaia*, respectively, based on five-fold crown-level cross-validation. A portion of the mapped results of the final multi-species classification is shown in Fig 5B.

The contextual filters reduced the number of pixels assigned to the focal species classes by 12.8%, 26.4%, and 23.1% for *D. panamensis*, *H. guayacan*, and *J. copaia*, respectively. The number of objects (contiguous areas) assigned to the three species was reduced much more drastically by 90.4%, 88.1%, and 85.1%, respectively (Fig 5C). This occurred because the filters removed a large quantity of small objects, often only one pixel in size. After filtering, there were a total of 2,107 predicted crown objects of *D. panamensis*, 837 of *H. guayacan*, and 1,405 of *J. copaia* across the entire island (Fig 6).

### Field validation

Field validation indicated that our map of predicted crown objects was very precise. We visited 39 crown objects assigned to *D. panamensis*, 38 of *H. guayacan*, and 32 of *J. copaia*. Of these, only one crown object assigned to *D. panamensis* and two assigned to *H. guayacan* were found to be incorrect. This resulted in an estimated false-positive error rate of 2.6% for *D. panamensis*, 5.3% for *H. guayacan*, and 0% for *J. copaia*.



**Fig 6. Prediction results for the three focal species across BCI.** Blue = *D. panamensis*, red = *H. guayacan*, and yellow = *J. copaia*. Insets show close-up of results for high-density areas of each focal species.

doi:10.1371/journal.pone.0118403.g006

We observed some individuals of each species in the field that were not detected by our model. Six *D. panamensis* crowns were encountered that were not predicted by our maps. Of these, one was found to have been cut by the NDVI filter (through inspection of the CAO imagery), one had interference by lianas, and two were small (~ 5 m in diameter). We encountered nine crowns of *H. guayacan* which were not predicted, of which one had low NDVI, one had interference by lianas, and one was small. We also encountered eight *J. copaia* crowns in the field that were not predicted, seven of which were small. Most of the small non-detected *J. copaia* crowns had been swept out of the prediction map by the image processing steps designed to remove noise and small objects. Both the false-positive rates and the relative lack of plausible reasons for crown omission errors indicate that *H. guayacan* was the most troublesome of the three focal species to detect.

## Discussion

### Model performance

Our final classification model was found to perform very well, which was especially remarkable considering that it was produced using 138 training crowns from only the species that were mapped. The crown object map that we produced from the classification results was highly reliable in the sense that it had a very low false-positive error rate for predicted crown objects. The classification model had high pixel-level producer's accuracies for the three focal species, though some of the model's sensitivity in detecting the focal species was certainly lost after the contextual filters were applied to the model results to create the map of crown objects. This was evidenced by the fact that omission errors were observed for crowns of each species while we were validating the maps in the field.

When observing focal species crowns in the field, it could be quite unclear as to what constituted an omission error. With examples from each of the focal species, we observed crowns in the field that were not predicted or only partially predicted by our model that were highly contaminated or partially or fully covered by lianas. Additionally, it was not uncommon that a crown object appearing to represent a single crown both in the CAO imagery and in the prediction map was found to be two or more adjacent individuals of the same species. Both of these occurrences are important from an ecological perspective because they contribute to an underestimation of the number of individuals of a species, though we do not consider them to be errors from a remote species detection perspective. Some of the omitted crowns, such as many of those of *J. copaia*, were simply small, and may have been captured by our model if we had used a less conservative size filter (though this may have introduced false-positives). There were other instances from each species where crowns were missed by the prediction model for no apparent reason of interference by other species, insufficient size, or lower NDVI, which occurred most often for *H. guayacan*.

The three species examined here have previously been mapped across the island either by their flowering characteristics or through visual interpretation of aerial photographs. Flowering crowns of *H. guayacan* were mapped across BCI using two Quickbird images from different years with an automatic approach combining spectral angle mapping and one-class SVM [47]. Using visual interpretation of aerial photographs, *D. panamensis* crowns were identified based on canopy structure [48], and *H. guayacan* and *J. copaia* crowns were identified through their flowering traits [49,50]. Additionally, J. R. Kellner produced island-wide maps of all three focal species (the same used to obtain focal species training data used in this study) based on crown flowering using multispectral satellite imagery with an automatic detection method [36]. The island-wide distributions of the three focal species produced by our species detection model resemble those found in these previous studies [36,47,48,50]. However, our model detected

**Table 2. The number of crown objects detected in the present study compared to previous studies.**

<i>D. panamensis</i>	<i>H. guayacan</i>	<i>J. copaia</i>	Study
2107	837	1405	Present study
	706		Sanchez-Azofeifa et al. [46]
	688	977	Garzon-Lopez et al. [49]
904	804	432	J. R. Kellner [50]

doi:10.1371/journal.pone.0118403.t002

many more crown objects than the number of crowns reported in these studies (Table 2). This suggests that our model, which uses imaging spectroscopy and is not limited to detecting flowering individuals, was more powerful. However, caution is warranted with these comparisons as our results pertain to crown objects and not individuals *per se* (though the same logic likely applies to other studies as well).

The forest of Barro Colorado Island posed several challenges to remote species mapping; namely, high species diversity of the canopy, high prevalence of lianas, and a continuous canopy layer in which individual tree crowns are difficult to distinguish. We attribute the success of remote species mapping in this forest to two main factors: i) a shift from multi-species to single-species classification, and ii) performing species classification at the pixel-level, with crown objects identified after classification.

### Single-species detection

A single-species classification approach has considerable advantages over the traditional multi-species classification framework when the interest is in mapping one or a few focal species. Single-species classification can increase model performance with respect to the focal species and decrease the overall amount of data required to train the species that are not of interest (or in the case of biased SVM, it eliminates this requirement altogether) [25,26]. However, a focus on single-species detection need not restrict mapping to only one species; single-class classification models can be combined in different ways to make a multi-class classification [51]. For example, Muñoz-Marí et al. [52] combined several support vector domain description classifiers (very similar to one-class SVM) into a multi-class model to map land cover classes. One-against-all binary SVM and biased SVM were found to outperform one-class SVM for remote species identification [26], and for this analysis, we combined biased SVMs to create a multi-species identification model. Thus, improving single-species classification models can be one important route towards building more effective multi-species classification models.

Both binary and biased SVM models yielded reasonable performance with this dataset. Tested on a separate crown dataset, we found that binary SVM provided greater sensitivity but slightly lower specificity than biased SVM. However, comparing the performance of these methods in this way may give an incomplete view of their relative performances. When the models were applied to the imagery, binary SVM assigned approximately three to six times more pixels to the focal species compared to biased SVM. We may expect that small differences in model specificity have a large impact when the model is applied to the image as the non-focal species make up the majority of the landscape. However, it is also possible that noisy results from the binary SVM model could be an artifact of less-than-complete representation of the non-focal vegetation by our training dataset of over 300 crowns. This phenomenon was suggested by the results of a previous study comparing binary and biased SVM for remote species identification of five focal species in a savanna ecosystem [26]. In that study, binary SVM generally performed better than biased SVM, except that binary SVM produced noticeably

noisier maps than biased SVM for one of the focal species. This was hypothesized to be caused by less representative training of the non-focal class for the higher-diversity riparian areas in which that particular focal species was found. In the current study we found that biased SVM consistently produced more precise maps than binary SVM, which supports the hypothesis that biased SVM has an advantage in high diversity landscapes.

Thus, although our preliminary experiments using cross-validation of the training dataset indicated that the binary SVM models had very high specificity, maps produced by these models were not as precise as we wished. The precision of binary SVM would likely improve if more training data were gathered for the non-focal species, but pursuing this track may be impractical in such an ecosystem. With high tree species diversity it becomes very costly to comprehensively sample the non-focal tree species, and in the BCI forest there is also a high diversity of other vegetation types, mostly lianas, that contributes to the non-focal class in the canopy. Biased SVM offers a solution to this problem by comprehensively sampling the spectra of all vegetation present and using this information to constrain the focal species class. Whereas the details of optimizing and implementing biased SVM for single-species detection were previously unclear, recent work exploring possible alternatives has made this method more straightforward to use for remote species mapping [26]. It now appears that biased SVM, through its ability to constrain the focal class independently of the quality of the outlier training data available, is not only a viable alternative but likely a preferred alternative to binary SVM when performing remote species mapping in very high diversity ecosystems.

## Delineation of tree crowns

The automatic delineation of individual tree crowns, or crown segmentation, has long been an obstacle for operational species mapping. Segmentation of individual crowns has generally been viewed as the first step before applying species classification to an image [7,53–56]. This is not without reason, as some studies have found that accuracy of species classification models was higher when classification was performed on crowns instead of pixels; for example, by using the mean spectra from crowns for classification or majority vote of the pixel class assignments [7,35]. Additionally, a map of crown identities may be more ecologically meaningful than a map of pixel identities.

Crown segmentation may rely on structural differences among crowns detected by LiDAR or indicated by shadow, spatial separation of crowns, or spectral differences among crowns. Segmentation of individual tree crowns has been successfully accomplished in many places where structural differences among crowns facilitate crown separation, such as coniferous forest, savannas, or urban areas, e.g., [20,35,53,56–58], and this is especially facilitated by the use of LiDAR. In closed-canopy, non-conifer forests where there are fewer structural clues, spectral differences have been emphasized as a way to differentiate crowns, e.g., [18,59]. Segmentation of individual tree crowns based on spectral data alone is extremely difficult, and to date there has been very limited success delineating individual tree crowns in tropical forests (though progress is being made; see [59,60]). This may partially explain why some crown-level species identification models have not made it to the mapping step despite promising classification results, e.g., [7,15].

The closed-canopy and high prevalence of lianas in the BCI forest meant that individual crowns were difficult to observe in the image. However, during the model validation field campaign we found many instances where the pixel-level species predictions correctly determined the shape and size of individual tree crowns (more specifically, the shape and size of the crown area uncovered by lianas), even where *a priori* visual interpretation had failed. Therefore, we found that it was preferable to train a pixel-level species classifier and use the prediction results

to dictate the spatial coverage of the focal species crowns. This approach had the advantage that errors due to incorrect delineation of tree crowns were not fed into a species identification model. It also suited our objectives because we were only interested in delineating crowns of the focal species, rather than delineating all crowns on the island. Although we could not eliminate cases where a single crown object identified by our procedure was actually two or more adjacent conspecific individuals (as observed in the field), this limitation is unlikely to be overcome using existing methods with spectral data alone. Our mapped crown results show that automatic crown segmentation can be accomplished for selected species in a closed-canopy, high diversity tropical forest when paired with species identification as a first step, and that this strategy may be a preferable alternative to attempting to perform individual tree crown segmentation based on raw spectral data followed by object-based classification of crowns.

## Remote species mapping in the tropics

In their review of high resolution remote sensing in tropical forests, Nagendra and Rocchini [22] argued that high spatial resolution remote sensing data are best suited to accurately locating features within an image and less well suited for species identification. They concluded that high-spatial resolution data may make it more difficult to characterize the spectral signatures of different species to perform species classification, citing the problem of variability among pixels within the same crown when pixels are smaller than the crown being identified. However, we found that the high spatial and spectral resolution data are suitable for both accurate feature location and species identification, and that within-crown spectral variability does not preclude species classification of pixels. A possible explanation for this disparity could be the recent shift from endmember-based classification methods such as spectral angle mapper, which use one or a few exemplar spectra for each class, towards classification methods such as linear discriminant analysis, random forest, and SVM, which can accommodate greater within-class variability. The latter two methods are even able to accommodate complex, non-normal, multi-modal within-class variation.

Here we showed that remote species identification can be successfully achieved in a diverse, closed-canopy tropical forest using high spatial resolution imaging spectroscopy data. We produced an accurate crown object map of three focal species that was found to be very precise. The recently developed biased SVM performed well as the basis for our final multi-species classification model, and its performance was judged superior to the current standard method of binary SVM for our dataset. Thus our final multi-species classification model had the considerable advantage of only requiring representative training data from the species of interest, which can greatly ease the burden associated with producing species classification models, especially in high diversity ecosystems. Our method of producing crown objects from the pixel-level classification results also has implications for automatic tree crown delineation, indicating that individual crowns in can be delineated in high diversity, closed-canopy tropical forests when coupled with species identification. Further experimentation is needed to determine whether these methods will prove useful for remote species mapping in other ecosystems, especially those with high species diversity and a continuous closed canopy.

## Supporting Information

**S1 Fig. Overall performance (F-score) of binary SVM models for the three focal species as the number of training crowns was varied for both the focal species and the outlier species classes.**  $n_f$  indicates the number of crowns of the focal species class and  $n_o$  indicates the number of crowns for the outlier species class.

(EPS)

## S1 Methods. Methods and results for the analysis of the training data amount and balance for binary SVM.

(DOC)

## Acknowledgments

We thank Raul Tupayachi, Felipe Sinca, Nestor Jaramillo, Katie Kryston, Byron Tsang, and Kelly McManus for their outstanding support during field collection of tree crowns. We also thank Rufino Gonzalez for his botanical expertise during the field crown collection and model validation. The Carnegie Airborne Observatory has been supported by the Avatar Alliance Foundation, Gordon and Betty Moore Foundation, John D. and Catherine T. MacArthur Foundation, Grantham Foundation for the Protection of the Environment, W. M. Keck Foundation, Margaret A. Cargill Foundation, Mary Anne Nyburg Baker and G. Leonard Baker Jr., and William R. Hearst III. We also thank Stephanie Bohlman and one anonymous reviewer for helpful comments on the manuscript.

## Author Contributions

Conceived and designed the experiments: CAB GPA. Performed the experiments: CAB. Analyzed the data: CAB. Contributed reagents/materials/analysis tools: CAB GPA REM CBA DEK JRK SJW. Wrote the paper: CAB GPA REM CBA JRK SJW.

## References

1. Schimel DS, Asner GP, Moorcroft P (2013) Observing changing ecological diversity in the Anthropocene. *Front Ecol Environ* 11: 129–137. doi: [10.1890/120111](https://doi.org/10.1890/120111)
2. Turner W, Spector S, Gardiner N, Fladeland M, Sterling E, Steininger M. (2003) Remote sensing for biodiversity science and conservation. *Trends Ecol Evol* 18: 306–314. doi: [10.1016/S0169-5347\(03\)00070-3](https://doi.org/10.1016/S0169-5347(03)00070-3)
3. Gillespie TW, Foody GM, Rocchini D, Giorgi AP, Saatchi S (2008) Measuring and modelling biodiversity from space. *Prog Phys Geogr* 32: 203–221. doi: [10.1177/0309133308093606](https://doi.org/10.1177/0309133308093606)
4. Valencia R, Balslev H, C GPYM (1994) High tree alpha-diversity in Amazonian Ecuador. *Biodivers Conserv* 3: 21–28. doi: [10.1007/BF00115330](https://doi.org/10.1007/BF00115330)
5. Fung T, Ma FY, Siu WL (1998) Hyperspectral data analysis for subtropical tree species recognition. *Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International. Vol. 3. pp. 1298–1300 vol.3.* doi: [10.1109/IGARSS.1998.691383](https://doi.org/10.1109/IGARSS.1998.691383)
6. Cochrane MA (2000) Using vegetation reflectance variability for species level classification of hyperspectral data. *Int J Remote Sens* 21: 2075–2087. doi: [10.1080/01431160050021303](https://doi.org/10.1080/01431160050021303)
7. Clark ML, Roberts DA, Clark DB (2005) Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens Environ* 96: 375–398. doi: [10.1016/j.rse.2005.03.009](https://doi.org/10.1016/j.rse.2005.03.009)
8. Castro-Esau KL, Sánchez-Azofeifa GA, Rivard B, Wright SJ, Quesada M (2006) Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. *Am J Bot* 93: 517–530. doi: [10.3732/ajb.93.4.517](https://doi.org/10.3732/ajb.93.4.517) PMID: [21646212](https://pubmed.ncbi.nlm.nih.gov/21646212/)
9. Zhang J, Rivard B, Sánchez-Azofeifa A, Castro-Esau K (2006) Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: implications for species identification using HYDICE imagery. *Remote Sens Environ* 105: 129–141. doi: [10.1016/j.rse.2006.06.010](https://doi.org/10.1016/j.rse.2006.06.010)
10. Asner GP, Martin RE (2011) Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian forest. *New Phytol* 189: 999–1012. doi: [10.1111/j.1469-8137.2010.03549.x](https://doi.org/10.1111/j.1469-8137.2010.03549.x) PMID: [21118261](https://pubmed.ncbi.nlm.nih.gov/21118261/)
11. Asner GP, Martin RE, Carranza-Jiménez L, Sinca F, Tupayachi R, Anderson CB, et al. (2014) Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. *New Phytol* 204: 127–139. doi: [10.1111/nph.12895](https://doi.org/10.1111/nph.12895) PMID: [24942328](https://pubmed.ncbi.nlm.nih.gov/24942328/)
12. Castro-Esau KL, Sánchez-Azofeifa GA, Caelli T (2004) Discrimination of lianas and trees with leaf-level hyperspectral data. *Remote Sens Environ* 90: 353–372. doi: [10.1016/j.rse.2004.01.013](https://doi.org/10.1016/j.rse.2004.01.013)
13. Kalacska M, Bohlman S, Sanchez-Azofeifa GA, Castro-Esau K, Caelli T (2007) Hyperspectral discrimination of tropical dry forest lianas and trees: comparative data reduction approaches at the leaf and canopy levels. *Remote Sens Environ* 109: 406–415. doi: [10.1016/j.rse.2007.01.012](https://doi.org/10.1016/j.rse.2007.01.012)

14. Somers B, Asner GP (2012) Hyperspectral time series analysis of native and invasive species in Hawaiian rainforests. *Remote Sens* 4: 2510–2529. doi: [10.3390/rs4092510](https://doi.org/10.3390/rs4092510)
15. Papeş M, Tupayachi R, Martínez P, Peterson AT, Asner GP, et al. (2013) Seasonal variation in spectral signatures of five genera of rainforest trees. *IEEE J Sel Top Appl Earth Obs Remote Sens* 6: 339–350. doi: [10.1109/JSTARS.2012.2228468](https://doi.org/10.1109/JSTARS.2012.2228468)
16. Lucas R, Bunting P, Paterson M, Chisholm L (2008) Classification of Australian forest communities using aerial photography, CASI and HyMap data. *Remote Sens Environ* 112: 2088–2103. doi: [10.1016/j.rse.2007.10.011](https://doi.org/10.1016/j.rse.2007.10.011)
17. Dalponte M, Bruzzone L, Gianelle D (2012) Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sens Environ* 123: 258–270. doi: [10.1016/j.rse.2012.03.013](https://doi.org/10.1016/j.rse.2012.03.013)
18. Féret J, Asner GP (2013) Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans Geosci Remote Sens* 51: 73–84. doi: [10.1109/TGRS.2012.2199323](https://doi.org/10.1109/TGRS.2012.2199323)
19. Baldeck CA, Colgan MS, Féret J-B, Levick SR, Martin RE, et al. (2014) Landscape-scale variation in plant community composition of an African savanna from airborne species mapping. *Ecol Appl* 24: 84–93. doi: [10.1890/13-0307.1](https://doi.org/10.1890/13-0307.1) PMID: [24640536](https://pubmed.ncbi.nlm.nih.gov/24640536/)
20. Alonzo M, Bookhagen B, Roberts DA (2014) Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sens Environ* 148: 70–83. doi: [10.1016/j.rse.2014.03.018](https://doi.org/10.1016/j.rse.2014.03.018)
21. Baldeck CA, Asner GP (2014) Improving remote species identification through efficient training data collection. *Remote Sens* 6: 2682–2698. doi: [10.3390/rs6042682](https://doi.org/10.3390/rs6042682)
22. Nagendra H, Rocchini D (2008) High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. *Biodivers Conserv* 17: 3431–3442. doi: [10.1007/s10531-008-9479-0](https://doi.org/10.1007/s10531-008-9479-0)
23. Tax DMJ, Duin RPW (1999) Support vector domain description. *Pattern Recognit Lett* 20: 1191–1199. doi: [10.1016/S0167-8655\(99\)00087-2](https://doi.org/10.1016/S0167-8655(99)00087-2)
24. Liu B, Dai Y, Li X, Lee WS, Yu PS (2003) Building text classifiers using positive and unlabeled examples. *Third IEEE International Conference on Data Mining, 2003. ICDM 2003*. pp. 179–186. doi: [10.1109/ICDM.2003.1250918](https://doi.org/10.1109/ICDM.2003.1250918)
25. Foody GM, Atkinson PM, Gething PW, Ravenhill NA, Kelly CK (2005) Identification of specific tree species in ancient semi-natural woodland from digital aerial sensor imagery. *Ecol Appl* 15: 1233–1244. doi: [10.1890/04-1061](https://doi.org/10.1890/04-1061)
26. Baldeck CA, Asner GP (in press) Single species detection with airborne imaging spectroscopy data: a comparison of support vector techniques. *IEEE J Sel Top Appl Earth Obs Remote Sens*. doi: [10.1109/JSTARS.2014.2346475](https://doi.org/10.1109/JSTARS.2014.2346475)
27. Féret J-B, Asner GP (2012) Semi-supervised methods to identify individual crowns of lowland tropical canopy species using imaging spectroscopy and LiDAR. *Remote Sens* 4: 2457–2476.
28. Condit R, Watts K, Bohlman SA, Pérez R, Foster RB, Hubbell SP. (2000) Quantifying the deciduousness of tropical forest canopies under varying climates. *J Veg Sci* 11: 649–658. doi: [10.2307/3236572](https://doi.org/10.2307/3236572)
29. Croat TB (1978) *Flora of Barro Colorado Island*. Stanford, CA: Stanford University Press.
30. Garwood NC (2009) *Seedlings of Barro Colorado Island and the Neotropics*. Ithaca, NY: Comstock Publishing Associates.
31. Thorington RW Jr, Tannenbaum B, Tarak A, Rudran R (1982) Distribution of trees on Barro Colorado Island: a five hectare sample. In: Leigh EG Jr, Rand AS, Windsor DM, editors. *The ecology of a tropical forest: seasonal rhythms and long-term changes*. Smithsonian Institution Press.
32. Schnitzer SA, Mangan SA, Dalling JW, Baldeck CA, Hubbell SP, Ledo A, et al. (2012) Liana abundance, diversity, and distribution on Barro Colorado Island, Panama. *PLoS ONE* 7: e52114. doi: [10.1371/journal.pone.0052114](https://doi.org/10.1371/journal.pone.0052114) PMID: [23284889](https://pubmed.ncbi.nlm.nih.gov/23284889/)
33. Ingwell LL, Joseph Wright S, Becklund KK, Hubbell SP, Schnitzer SA (2010) The impact of lianas on 10 years of tree growth and mortality on Barro Colorado Island, Panama. *J Ecol* 98: 879–887. doi: [10.1111/j.1365-2745.2010.01676.x](https://doi.org/10.1111/j.1365-2745.2010.01676.x)
34. Asner GP, Knapp DE, Boardman J, Green RO, Kennedy-Bowdoin T, et al. (2012) Carnegie Airborne Observatory-2: increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sens Environ* 124: 454–465. doi: [10.1016/j.rse.2012.06.012](https://doi.org/10.1016/j.rse.2012.06.012)
35. Colgan M, Baldeck C, Féret J-B, Asner G (2012) Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. *Remote Sens* 4: 3462–3480. doi: [10.3390/rs4113462](https://doi.org/10.3390/rs4113462)
36. Kellner JR (2008) *Population and community dynamics of tropical rain forest canopy trees* Athens, GA: The University of Georgia.

37. Camps-Valls G, Gómez-Chova L, Calpe-Maravilla J, Martín-Guerrero JD, Soria-Olivas E, et al. (2004) Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Trans Geosci Remote Sens* 42: 1530–1542. doi: [10.1109/TGRS.2004.827262](https://doi.org/10.1109/TGRS.2004.827262)
38. Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans Geosci Remote Sens* 42: 1778–1790. doi: [10.1109/TGRS.2004.831865](https://doi.org/10.1109/TGRS.2004.831865)
39. Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. *ISPRS J Photogramm Remote Sens* 66: 247–259. doi: [10.1016/j.isprsjprs.2010.11.001](https://doi.org/10.1016/j.isprsjprs.2010.11.001)
40. Dalponte M, Orka HO, Gobakken T, Gianelle D, Naesset E (2013) Tree species classification in boreal forests with hyperspectral data. *IEEE Trans Geosci Remote Sens* 51: 2632–2645. doi: [10.1109/TGRS.2012.2216272](https://doi.org/10.1109/TGRS.2012.2216272)
41. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20: 273–297. doi: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
42. Vapnik VN (1998) *Statistical Learning Theory*. Wiley. 736 p.
43. Aizerman A, Braverman E, Rozoner L (1964) Theoretical foundations of the potential function method in pattern recognition learning. *Autom Remote Control* 25: 821–837.
44. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory. COLT '92*. New York, NY, USA: ACM. pp. 144–152. Available: <http://doi.acm.org/10.1145/130385.130401>. Accessed 5 March 2014.
45. Lee WS, Liu B (2003) Learning with positive and unlabeled examples using weighted logistic regression. *Twentieth international conference on machine learning, ICML-2003*. Vol. 3. pp. 448–455.
46. Muñoz-Marí J, Bovolo F, Gómez-Chova L, Bruzzone L, Camp-Valls G (2010) Semisupervised one-class support vector machines for classification of remote sensing data. *IEEE Trans Geosci Remote Sens* 48: 3188–3197. doi: [10.1109/TGRS.2010.2045764](https://doi.org/10.1109/TGRS.2010.2045764)
47. Sánchez-Azofeifa A, Rivard B, Wright J, Feng J-L, Li P, Chong MM, et al. (2011) Estimation of the distribution of *Tabebuia guayacan* (Bignoniaceae) using high-resolution remote sensing imagery. *Sensors* 11: 3831–3851. doi: [10.3390/s110403831](https://doi.org/10.3390/s110403831) PMID: [22163825](https://pubmed.ncbi.nlm.nih.gov/22163825/)
48. Caillaud D, Crofoot MC, Scarpino SV, Jansen PA, Garzon-Lopez CX, Winkelhagen AJ, et al. (2010) Modeling the spatial distribution and fruiting pattern of a key tree species in a neotropical forest: methodology and potential applications. *PLoS ONE* 5: e15002. doi: [10.1371/journal.pone.0015002](https://doi.org/10.1371/journal.pone.0015002) PMID: [21124927](https://pubmed.ncbi.nlm.nih.gov/21124927/)
49. Garzon-Lopez CX, Bohlman SA, Olf H, Jansen PA (2013) Mapping tropical forest trees using high-resolution aerial digital photographs. *Biotropica* 45: 308–316. doi: [10.1111/btp.12009](https://doi.org/10.1111/btp.12009)
50. Garzon-Lopez CX, Jansen PA, Bohlman SA, Ordonez A, Olf H (2014) Effects of sampling scale on patterns of habitat association in tropical trees. *J Veg Sci* 25: 349–362. doi: [10.1111/jvs.12090](https://doi.org/10.1111/jvs.12090)
51. Duin RPW, Tax DMJ (2000) Experiments with classifier combining rules. *Multiple Classifier Systems. Lecture Notes in Computer Science*. Springer Berlin Heidelberg. pp. 16–29. Available: [http://link.springer.com/chapter/10.1007/3-540-45014-9\\_2](http://link.springer.com/chapter/10.1007/3-540-45014-9_2). Accessed 21 May 2014.
52. Muñoz-Marí J, Bruzzone L, Camps-Valls G (2007) A support vector domain description approach to supervised classification of remote sensing images. *IEEE Trans Geosci Remote Sens* 45: 2683–2692. doi: [10.1109/TGRS.2007.897425](https://doi.org/10.1109/TGRS.2007.897425)
53. Gougeon FA (1995) A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Can J Remote Sens* 21: 274–284.
54. Meyer P, Staenz K, Ittena KI (1996) Semi-automated procedures for tree species identification in high spatial resolution data from digitized colour infrared-aerial photography. *ISPRS J Photogramm Remote Sens* 51: 5–16. doi: [10.1016/0924-2716\(96\)00003-2](https://doi.org/10.1016/0924-2716(96)00003-2)
55. McGraw JB, Warner TA, Key TL, Lamar WR (1998) High spatial resolution remote sensing of forest trees. *Trends Ecol Evol* 13: 300–301. PMID: [21238315](https://pubmed.ncbi.nlm.nih.gov/21238315/)
56. Leckie DG, Gougeon FA, Walsworth N, Paradine D (2003) Stand delineation and composition estimation using semi-automated individual tree crown analysis. *Remote Sens Environ* 85: 355–369. doi: [10.1016/S0034-4257\(03\)00013-0](https://doi.org/10.1016/S0034-4257(03)00013-0)
57. Wang L, Gong P, Biging GS (2004) Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogramm Eng Remote Sens* 70: 351–357. doi: [10.14358/PERS.70.3.351](https://doi.org/10.14358/PERS.70.3.351)
58. Bunting P, Lucas R (2006) The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data. *Remote Sens Environ* 101: 230–248. doi: [10.1016/j.rse.2005.12.015](https://doi.org/10.1016/j.rse.2005.12.015)

59. Tochon G, Feret J, Martin RE, Tupayachi R, Chanussot J, Asner GP. (2012) Binary partition tree as a hyperspectral segmentation tool for tropical rainforests. *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. pp. 6368–6371. doi: [10.1109/IGARSS.2012.6352716](https://doi.org/10.1109/IGARSS.2012.6352716)
60. Ferreira MP, Zanotta DC, Zortea M, Korting TS, Fonseca LMG, Shimabukuro YE. (2014) Automatic tree crown delineation in tropical forest using hyperspectral data. *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. pp. 784–787. doi: [10.1109/IGARSS.2014.6946541](https://doi.org/10.1109/IGARSS.2014.6946541)