ENVIRONMENTAL RESEARCH
ECOLOGY

LETTER • **OPEN ACCESS**

Evaluating individual tree species classification performance across diverse environments

To cite this article: M M Seeley *et al* 2024 *Environ. Res.: Ecology* **3** 011001

ENVIRONMENTAL RESEARCH
ECOLOGY

OPEN ACCESS

RECEIVED
25 August 2023

REVISED
29 November 2023

ACCEPTED FOR PUBLICATION
16 January 2024

PUBLISHED
23 January 2024

LETTER

Evaluating individual tree species classification performance across diverse environments

M M Seeley[1,2,*], N R Vaughn[1] and G A Asner[1,2]

[1] Center for Global Discovery and Conservation Science, Arizona State University, Tempe, AZ 85281, United States of America
[2] School of Geographic Sciences and Urban Planning, Arizona State University, Tempe, AZ 85281, United States of America
* Author to whom any correspondence should be addressed.

E-mail: mseeley1@asu.edu

Keywords: *Metrosideros polymorpha*, species classifications, support vector machine, random forest, ecosystems, vegetation mapping

Supplementary material for this article is available online

Abstract

Vegetation species mapping using airborne imaging spectroscopy yields accurate results and is important for advancing conservation objectives and biogeographic studies. As these data become more readily available owing to the upcoming launch of spaceborne imaging spectrometers, it is necessary to understand how these data can be used to consistently classify species across large geographic scales. However, few studies have attempted to map species across multiple ecosystems; therefore, little is known regarding the effect of intra-specific variation on the mapping of a single species across a wide range of environments and among varying backgrounds of other non-target species. To explore this effect, we developed and tested species classification models for *Metrosideros polymorpha*, a highly polymorphic canopy species endemic to Hawai'i, which is found in a diverse array of ecosystems. We compared the accuracies of support vector machine (SVM) and random forest models trained on canopy reflectance data from each of eight distinct ecosystems (ecosystem-specific) and a universal model trained on data from all ecosystems. When applied to ecosystem-specific test datasets, the ecosystem-specific models outperformed the universal model; however, the universal model retained high (>81%) accuracies across all ecosystems. Additionally, we found that models from ecosystems with broad variation in *M. polymorpha* canopy traits, as estimated using chemometric equations applied to canopy spectra, accurately predicted *M. polymorpha* in other ecosystems. While species classifications across ecosystems can yield accurate results, these results will require sampling procedures that capture the intra-specific variation of the target species.

1. Introduction

Spatial information on tree species distributions has many conservation applications including monitoring keystone or endangered species [1], invasive species tracking [2–4], and modeling climate change outcomes [5–7], among others. Remote sensing data are increasingly being used to circumvent the limitations of traditional field-based forest inventories. While species classifications have successfully used light detection and ranging (LiDAR) and multispectral images, complex cases often rely on airborne imaging spectroscopy to achieve sufficient accuracies [8–14]. Despite the potential of these data to operationally map plant species across large geographic regions, classification studies using imaging spectroscopy are often limited to small geographic regions [10, 12, 13], and thus the effects of intra-specific spectral variability, driven by variation in canopy traits, on classification accuracies remains underexplored [10].

By collecting solar-reflected radiance spectra from the visible (∼350–700 nm) through near-infrared (700–1100 nm) and into the shortwave infrared (SWIR; 1100–2500 nm) at narrow wavelength intervals (<10 nm) and over extensive geographic areas, imaging spectroscopy often yields higher tree classification accuracies than other remote sensing methods [15, 16]. Accurate classifications via imaging spectroscopy are largely a result of the ability of these data to capture chemical and morphological traits of canopies [17, 18], which are taxonomically driven [17, 19–24]. These data can even capture intra-specific variations at the canopy level [20, 25, 26]. Yet the effect of this intra-specific variability on species classifications when mapping a single species across ecosystems or environmental gradients is largely unknown [10].

Intra-specific leaf traits vary substantially across ecosystems [27, 28] and environmental gradients [26, 29–31]. For example, ecosystem of origin, as opposed to taxonomy, explained nearly 30% of leaf trait variation in species that crossed the Amazon-Cerrado transition [27]. *Metrosideros polymorpha,* a keystone canopy species endemic to Hawaii, demonstrates high polymorphism in that its four subspecies on Hawaii Island are environmentally filtered along elevation and soil substrate gradients and are characterized by morphological differences such as leaf pubescence, shape, and thickness [32–35]. Intra-specific trait variation helps drive community structure, often indirectly, such as through the microbial composition [36]. It also helps explain not only how a single species can exist in contrasting environmental conditions, but also how many species coexist in diverse environments [37, 38]. Because traits can vary within a species, leaf traits often converge within a given environment regardless of taxonomic association [39]. While a growing body of literature focuses on intra-specific variation, the relative effect of trait convergence versus taxonomy on species classifications across ecosystems remains poorly known.

To address how intra-specific variation affects species classifications across ecosystems, we employed the Hawaii Island model system where *M. polymorpha* exists along and across extreme environmental gradients. Hawaii Island is home to twelve natural and anthropogenic ecosystems, eight of which host *M. polymorpha*. The intra-specific variation that characterizes *M. polymorpha* includes morphological [32, 40], chemical, and spectral variation [26, 31] across environmental gradients and has led to its differentiation into four genotypes that self-sort along these gradients [34, 35]. The polymorphism of this species has challenged large-scale classification efforts [41] and makes it an ideal study organism to assess species classifications across ecosystems.

2. Methods

2.1. Airborne imaging spectroscopy

Imaging spectroscopy data across Hawai'i Island were collected by the Arizona State University Global Airborne Observatory (GAO), which houses a high-fidelity imaging spectrometer (380–2510 nm) and a boresight aligned dual-laser LiDAR scanner. Collection and processing of the data followed Seeley *et al* [41], which are described briefly here. Data were collected in January 2019, with some regions filled in using previous campaign data from January 2016, July 2017, and January 2018 [42]. Spectroscopy data were orthorectified to 2 m × 2 m spatial resolution using accurate position and orientation data along with a ray tracing technique to match viewing geometry to LiDAR-derived surface maps. Observed radiance spectra were corrected for atmospheric effects and converted to reflectance using ACORN v6.0 (Atmospheric CORrection Now; AIG LLC; Boulder, CO) [43, 44]. After brightness-normalized cloud-free VSWIR surface reflectance mosaics across Hawaii Island were developed, a normalized difference vegetation index (NDVI) and height threshold as well as a shade mask were applied to the dataset. Only pixels with photosynthetic vegetation, as determined by a NDVI threshold of >0.7 were kept; NDVI was calculated using the 650 and 860 nm wavelengths of the VSWIR data. LiDAR-derived top of canopy height (TCH) was used to remove vegetation below one meter, and shaded portions of the canopy were removed using a shade mask created using a ray tracing technique applied to the LiDAR surface maps [42, 45]. Canopy traits (table S1) were estimated using the filtered canopy datasets by applying universal chemometric algorithms developed by Asner *et al* [46].

2.2. Crown data collection

Crown reflectance data was derived from the filtered canopy datasets using averaged spectra from within field-collected canopy polygons. A total of 3154 canopies in eight distinct Hawaii Island ecosystems were delineated and identified as either *M. polymorpha* or 'other vegetation' between August 2022 and April 2023. Crowns were delineated in the field using Garmin Glo GPS connected to PC tablets with GAO TCH, true and false color composites, and preliminary *M. polymorpha* classifications developed using a support vector machine (SVM). Additional delineations were made using Google Street View. As Balzotti *et al* [9] and Weingarten *et al* [47] employed similar crown delineation methods, their crown classifications were included in the training dataset. All pixels in the filtered reflectance dataset within a crown polygon were averaged to generate canopy spectra and leaf trait data. Crowns of other species were grouped together and represented the background vegetation spectra from which the classification models were trained to discriminate *M. polymorpha*.

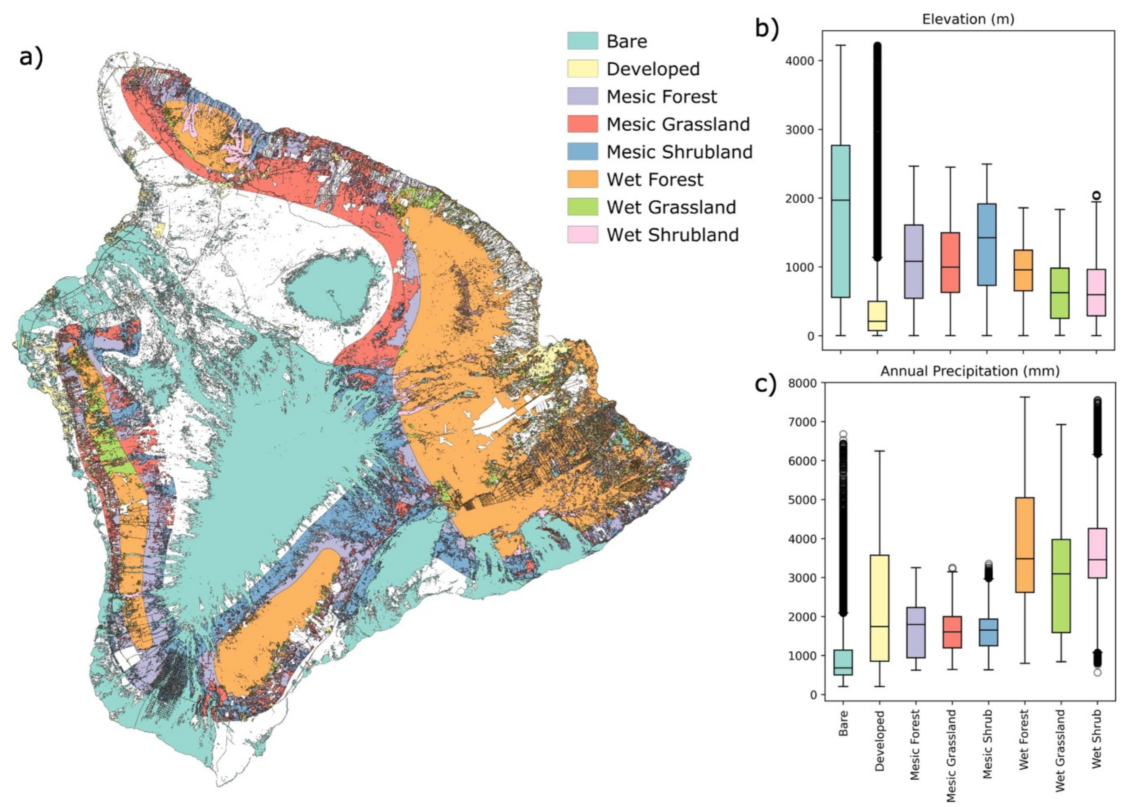

Figure 1. (a) Spatial arrangement of the eight ecosystems on Hawai'i Island inhabited by *Metrosideros polymorpha*. Ecosystems were designated by the U.S. Geological Survey Gap Analysis Program in 2001 [48]. Boxplots of (b) elevation (m) and (c) annual precipitation (mm) in each ecosystem are presented. Elevation data originated from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) digital elevation model (DEM), and the 30 year mean annual precipitation (1978–2007) was obtained from the University of Hawai'i-Manoa Hawaii Rainfall Atlas [49].

Using geospatial ecosystem delineations across Hawaii Island, canopy spectra were sorted into one of eight ecosystems based on their location. Ecosystem delineations were created using 2001 Landsat data by the U.S. Geological Survey Gap Analysis Program [48] and included grassland, shrubland, and forest in two precipitation regimes—wet and mesic (figure 1). Wet forests at low elevations often contain a mixture of invasive species, while wet and mesic forests at higher elevations are often comprised of native *Acacia koa* and *M. polymorpha.* Grasslands and shrublands on Hawai'i Island are commonly a consequence of abandoned cattle pastures, and these ecosystems include shrubby *M. polymorpha,* other native trees and shrubs, and opportunistic invasive species. Ecosystems designated as grasslands include savannas as large spreading trees like *Samanea saman* dominate these landscapes. The anthropogenic ecosystem included in this study, 'developed', included developed regions like cities and roads. Developed areas included non-native ornamental species, many of which are present in lowland wet forests, palm species, trees with edible fruits such as *Artocarpus altilis, Mangifera indica*, and *Persea americana*, and *M. polymorpha* cultivated from across the Hawaiian Islands. The bare ecosystem denoted regions with relatively recent lava flows where *M. polymorpha* and other shrubby vegetation grows out of rocky lava fields with no understory or vegetated ground cover. As *M. polymorpha* rarely grows in agricultural fields and not in any of the three dry ecosystems, these were excluded from the study.

2.3. Analysis

Several analyses were undertaken to explore the characteristics of *M. polymorpha* reflectance patterns across the eight ecosystems relative to background vegetation. The initial investigation included a visual assessment of canopy spectra. Mean spectra for *M. polymorpha* and background vegetation within each ecosystem were plotted together to understand patterns in the spectra. Reflectance across the VSWIR spectra is an expression of leaf traits and leaf and canopy structure, and different traits often have a stronger influence on different portions of the spectra. As a result, a qualitative assessment of the VSWIR spectra can provide insight into the variations of *M. polymorpha* across ecosystems relative to background vegetation. To further understand patterns in canopy spectra, canopy leaf traits for *M. polymorpha* and background vegetation across the ecosystems were compared. Within each ecosystem, leaf trait differences between *M. polymorpha* and other vegetation were assessed via a Mann-Whitney test using the scipy python package [50].

After assessing patterns in canopy spectra and traits, several SVM classification and random forest models were fit to the data using the scikit-learn python package (version 1.1.3). SVMs produce accurate species classifications and efficiently handle the highly dimensional imaging spectroscopy datasets by creating decision boundaries in feature space that maximize the distance between that boundary and the training data [8, 11, 51–55]. We used a radial kernel SVM with hyperparameter selection (kernel coefficient and regulation parameter) performed using a grid search available in the scikit-learn python package (version 0.24.1) [56]. Random forest is an ensemble classifier consisting of many decision trees and has been used for many remote sensing classification applications [57, 58]. The random forest model was likewise performed using a grid search to optimize classifier hyperparameters.

As the number of training crowns from each ecosystem was uneven (table S2), 88 crowns were selected from each ecosystem via stratified random sampling. As the ecosystem with the least number of crowns was ‘wet shrub’, which had 88 crowns, 88 crowns were selected from each ecosystem using random selection stratified based on species (*M. polymorpha* or other vegetation) to create datasets with an equivalent number of data points in each category, where possible. An additional category (‘Universal’) that included data from all the ecosystems was created using 11 crowns from each of the eight ecosystems. As before, half of the crowns in the Universal category were *M. polymorpha* (either five or six *M. polymorpha* from each ecosystem), and half were other vegetation. The 88 crowns from each ecosystem were split randomly into 70/30 training/test datasets. An SVM and random forest was fit to training crown reflectance data from each ecosystem category (ecosystem-specific models) and to data from all ecosystems (universal model). Accuracy scores were obtained by fitting the models trained on data from each ecosystem to all test datasets.

Lastly, the intra-specific variability of *M. polymorpha* was compared with the inter-specific variability across the ecosystems. For these investigations, we developed two SVM models—one that included only *M. polymorpha*, and the other included only other vegetation. The SVM was fit to each dataset and its performance in categorizing crowns based on their ecosystem of origin was assessed. For this investigation, a 60/40 training/test split via stratified random sampling was used to increase the test dataset size as the number of classification categories increased.

3. Results

3.1. Reflectance spectra

Overall, mean *M. polymorpha* canopy reflectance overlapped with that of other vegetation in the visible (450–700 nm). The mean reflectance of *M. polymorpha* and other vegetation converged before the green reflectance peak (∼550 nm) in the developed, mesic forest, bare, and all wet ecosystems, though the mean reflectance was slightly higher for *M. polymorpha* (figure 2). Where spectra diverged between *M. polymorpha* and other background vegetation, *M. polymorpha* had higher reflectance values in all ecosystems except the bare ecosystem (figures 2 and S1). Divergence between the vegetation categories was greatest in the mesic shrub ecosystem (figure S1). *M. polymorpha* spectra diverged from other vegetation in the near-infrared (750–1320 nm) across all ecosystems. Between 750 and 1000 nm, *M. polymorpha* spectra were lower than that of other vegetation, but the pattern reversed between 1170 and 1320 nm. While these patterns were consistent across all ecosystems in the visible and near-infrared, divergence between spectra in the SWIR (1450–2500 nm) of *M. polymorpha* and other species differed for each ecosystem (figure 2). The developed, bare, wet grassland, and to a lesser extent, wet shrub ecosystems, had similar reflectance in the SWIR. Reflectance of *M. polymorpha* and other vegetation in the wet shrub ecosystem diverged between 1450 and 1800 nm but nearly converged after 2000 nm. For all other ecosystems, *M. polymorpha* had higher reflectance in the SWIR than the background vegetation from any ecosystem (figure S1).

3.2. Canopy chemistry

Across all ecosystems, *M. polymorpha* had lower median values than background vegetation for all chemistry indices except leaf mass per area (LMA; figure 3). LMA indices of *M. polymorpha* were higher than background vegetation in all ecosystems except the bare ecosystem where the pattern reversed (figure 3). All differences between chemistry indices for *M. polymorpha* and background vegetation were significant according to the MannWhitney two-sided test ($p < 0.5$) except for nonstructural carbohydrates (NSC) in the mesic shrub ecosystem. Except for mesic shrub, variability of chlorophyll a and b (a + b), carotenoids, lignin, total carbon (C), and nitrogen (N) was greater in background vegetation than *M. polymorpha* across all ecosystems (figure 3). *M. polymorpha* in the mesic shrub ecosystem had greater variability for all chemistry indices than *M. polymorpha*, and often background vegetation, from the other ecosystems (figure 3). For background vegetation, there was high variability in chlorophyll a + b, carotenoids, and N from wet

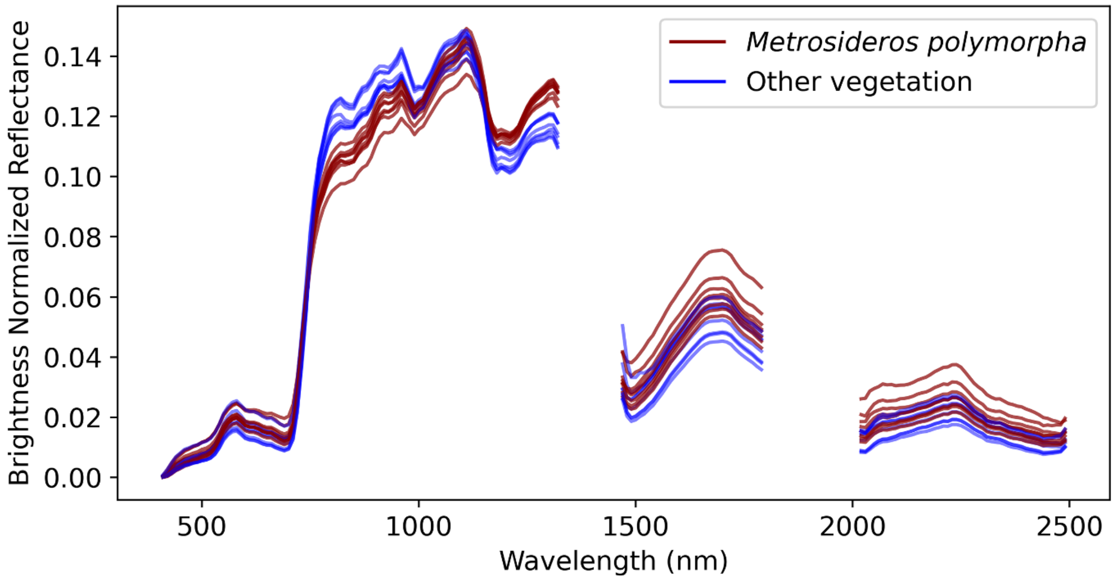

Figure 2. Mean brightness normalized reflectance of *Metrosideros polymorpha* (red) and other vegetation (blue). Each line represents a different ecosystem. Individual plots of spectra within each ecosystem can be seen in figure S1.

ecosystems (figure 3). *M. polymorpha* in wet shrub ecosystems likewise had higher N, phenols, and NSCs than that in other ecosystems. This pattern of higher values for cellulose, phenols, and total C in both wet forest and grasslands ecosystems as well as higher carotenoids, chlorophyll a + b, N, and NSC in wet grasslands, compared to other ecosystems was mirrored by background vegetation. Similarly, background vegetation in wet forest and grasslands had lower LMA than that in other ecosystems. All vegetation from the mesic shrub ecosystem had higher LMA than vegetation in other ecosystems. *M. polymorpha* from mesic shrub ecosystems had higher LMA and lower carotenoids, cellulose, chlorophyll a + b, lignin, and N than all but one ecosystem, most often the bare ecosystem. Background vegetation similarly had higher LMA than background vegetation from other ecosystems except for mesic grassland, but the other patterns did not hold.

3.3. Classification models

M. polymorpha classification across all Hawaii Island ecosystems was most accurately performed by the universal SVM model (81.5% accuracy; table 1). The best model for each ecosystem classified *M. polymorpha* canopies with 81.5%–100% accuracy (table 1). As *M. polymorpha* from all ecosystems except for mesic shrub and mesic forest was most accurately classified by SVM, as opposed to random forest, the following results refer to SVM unless stated otherwise (tables 1 and S3). Mesic grassland was best classified by data trained from mesic forests as well as the developed ecosystem (96.3% accuracy for both; table 1). Wet grasslands were most accurately classified by the bare ecosystem model (85.2%; table 1). Mesic forest was most accurately classified by the random forest model trained on data from developed ecosystems (92.6%; table S3), though the mesic forest SVM was the most accurate SVM model (88.9%; table 1). For all other ecosystems, *M. polymorpha* classification accuracies were highest for models trained on data from the same ecosystem (table 1). Models tested on data from developed ecosystems had the same high accuracy when trained using data from three other ecosystems (table 1). Models trained using data from mesic and wet grasslands had some of the lowest classification accuracies across all ecosystems (table 1). The mesic forest SVM model was the most universal, followed by the bare model, in that the highest accuracies for four ecosystems were achieved with this model (table 1).

To address whether vegetation from the same environment is more similar spectrally than individuals of the same species from different environments, we trained an SVM to classify *M. polymorpha* and other vegetation spectra into their respective ecosystems (table 2). *M. polymorpha* in mesic grasslands was most accurately classified into the correct ecosystem (66.7%), while that from the other ecosystems had 22.2% (developed) to 55.6% (mesic forest) accuracy (table 2). *M. polymorpha* from wet forests was most often misclassified as wet shrub (table 2). *M. polymorpha* canopies from developed and bare areas were misclassified with similar rates into all ecosystems, though they were most often misclassified as being from the mesic shrub ecosystem (table 2). Each ecosystem had at least one misclassification in the developed ecosystem (table 2). *M. polymorpha* spectra were more accurately classified based on ecosystem than other

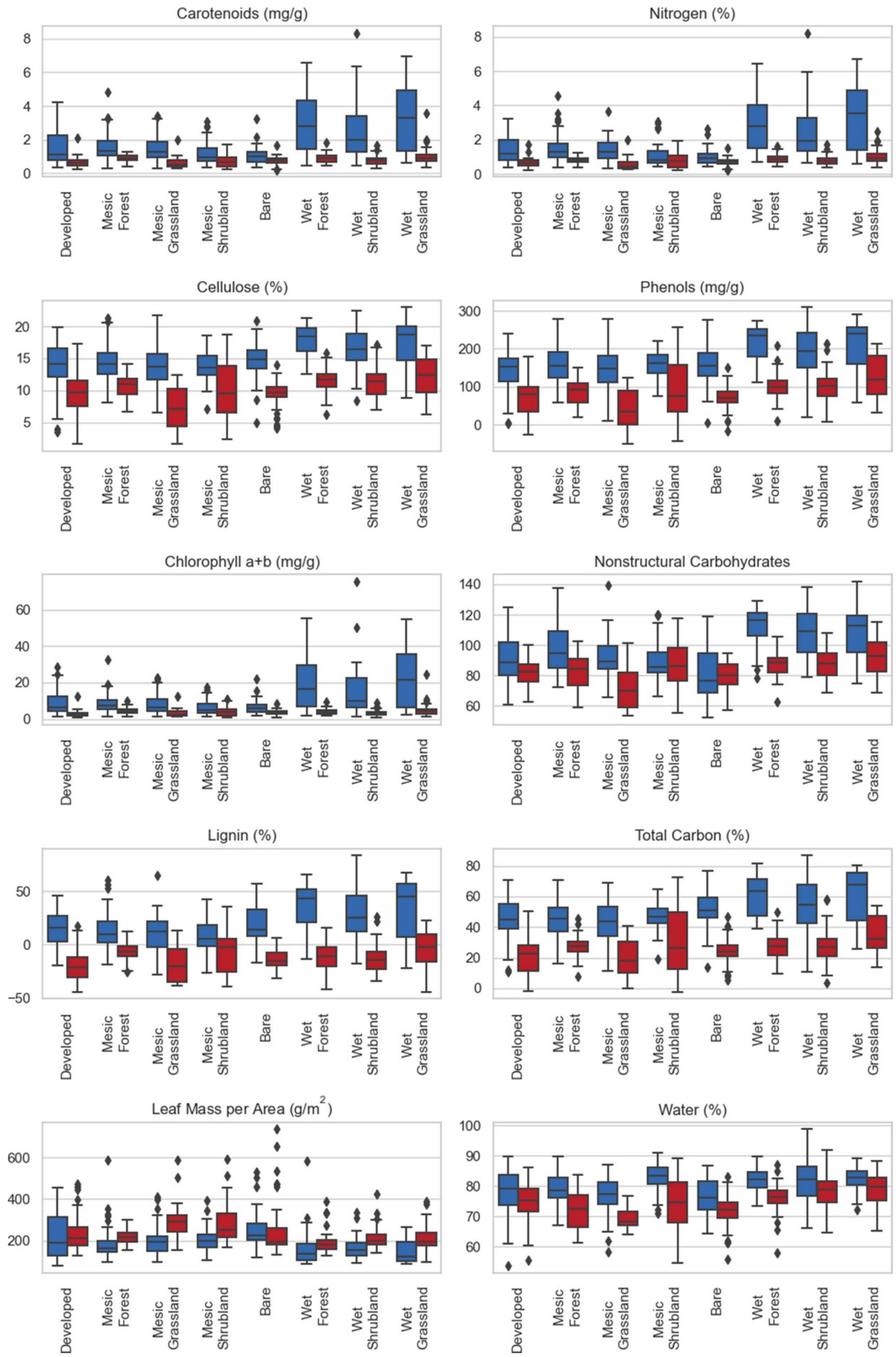

Figure 3. Boxplots of ten canopy traits as estimated by chemometric equations created by Asner *et al* [46]. Red boxes represent the canopy trait variation of the *Metrosideros polymorpha* training datasets for each ecosystem while blue represents that of background vegetation.

vegetation for all ecosystems except wet shrub, wet forest, and bare (table 2). For other vegetation, the canopy spectra from the bare ecosystem were the most accurately classified (55.6%) while that from the wet grassland and developed ecosystems were the least accurately classified (16.7%; table 2). Other vegetation canopy spectra from mesic and wet ecosystems were most often misclassified as another ecosystem in a similar moisture regime.

Table 1. Accuracy of support vector machine (SVM) when trained on one ecosystem (columns) and applied to another (rows). The 'Universal' category represents data from all ecosystems combined. Bold values indicate the best scores in each row. See table S3 for random forest accuracy results.

	SVM								
	Mesic grassland	Mesic shrub	Mesic forest	Wet grassland	Wet shrub	Wet forest	Developed	Bare	Universal
Mesic Grassland	85.2	88.9	**96.3**	14.8	66.7	74.1	**96.3**	59.3	85.2
Mesic Shrub	70.4	**88.9**	66.7	44.4	59.3	70.4	66.7	55.6	85.2
Mesic Forest	51.9	74.1	**88.9**	59.3	85.2	85.2	81.5	85.2	81.5
Wet Grassland	66.7	81.5	77.8	81.5	77.8	77.8	70.4	**85.2**	81.5
Wet Shrub	55.6	81.5	**96.3**	74.1	**96.3**	88.9	92.6	88.9	92.6
Wet Forest	44.4	92.6	92.6	88.9	88.9	**100.0**	92.6	96.3	92.6
Developed	44.4	**85.2**	**85.2**	63.0	81.5	77.8	**85.2**	**85.2**	81.5
Bare	63.0	88.9	96.3	63.0	63.0	81.5	81.5	**100.0**	92.6
Universal	56.1	77.4	78.8	66.5	73.6	81.1	79.2	80.7	**81.5**

Table 2. Confusion matrix of SVM classifying (top) *Metrosideros polymorpha* and (bottom) other vegetation crowns into their respective ecosystems.

		Predicted								
		Metrosideros polymorpha								
		Mesic grassland	Mesic shrub	Mesic forest	Wet grassland	Wet shrub	Wet forest	Developed	Bare	Accuracy
Observed	Mesic Grassland	6	1	0	0	0	0	2	0	66.7
	Mesic Shrub	2	8	1	0	0	2	3	1	47.1
	Mesic Forest	0	1	10	0	5	1	1	0	55.6
	Wet Grassland	1	0	0	7	4	0	2	4	38.9
	Wet Shrub	1	1	2	1	6	3	3	1	33.3
	Wet Forest	0	0	1	1	7	6	1	1	35.3
	Developed	0	4	2	2	3	2	4	1	22.2
	Bare	1	4	1	2	1	1	2	6	33.3
		Other Vegetation								
	Mesic Grassland	12	6	7	0	0	0	1	0	46.2
	Mesic Shrub	2	8	5	0	0	0	2	0	47.1
	Mesic Forest	3	3	6	0	2	1	2	1	33.3
	Wet Grassland	1	0	2	3	7	3	2	0	16.7
	Wet Shrub	0	0	1	2	8	3	3	0	47.1
	Wet Forest	1	0	0	3	3	7	3	0	41.2
	Developed	3	3	1	1	1	1	3	5	16.7
	Bare	1	1	1	0	3	0	2	10	55.6

4. Discussion

Intra-specific spectral variation of dominant canopy species has been demonstrated at the leaf [21, 24] and canopy [20, 25, 26, 59, 60] levels, yet the effect of intra-specific variation on species classifications in remote sensing, specifically imaging spectroscopy, data has been largely unexplored. Here, we addressed this question using *M. polymorpha,* a highly polymorphic canopy species that varies morphologically and chemically across environmental gradients [26, 31, 32, 40]. For example, *M. polymorpha* had a four-fold increase in leaf mass per area and three-fold change in foliar nitrogen across an elevation gradient [32]. Here, we first demonstrated that *M. polymorpha* varies spectrally across eight Hawaii Island ecosystems by visually comparing reflectance spectra and by building a classifier that predicted the ecosystem of origin for *M. polymorpha* canopy spectra. We then assessed how the intra-specific variation of *M. polymorpha* across these ecosystems affected species classification accuracies. We found that, while *M. polymorpha* classifiers performed best when trained and tested on data from the same ecosystem, most models performed with reasonably high accuracies (<70%) on data from other ecosystems.

When comparing ecosystem-specific models with a universal model trained on data from all ecosystems, we found that the universal model only outperformed ecosystem-specific models when tested on combined data from all ecosystems. In contrast, Marconi *et al* (2022) found that a generalized model trained using data from 27 National Ecological Observatory Network sites across the contiguous US, Alaska, and Guam outperformed site-specific tree crown classifiers. Marconi *et al* (2022) included the site centroid in the model to reduce site-level effects, which may have resulted in improved classification. Further, they did not specify whether their site-specific models have the same number of training samples as their general model, which would also affect classification accuracies as a larger training dataset in the general model would result in higher accuracies. Another study achieved high (>70%) accuracies when classifying species and habitats in five diverse sites across the contiguous United States (US), though they did not include site-specific models [12]. Both our results and these studies demonstrate that species classifications with high accuracies (>70%) across large geographic regions is possible, but accurate species mapping will require training data that represents the intra-specific variation of target species across the region.

In remote sensing classifications, selecting training datasets that best represent the spectral variation of the data results in models with higher accuracies [11, 61, 62]. In this study, ecosystem-specific models trained on data from mesic grasslands had the lowest accuracies (min 44.4%) when classifying data from other ecosystems. When looking at the range of the first and third quantiles, data from mesic grasslands often had *M. polymorpha* canopy chemical indices outside that of other ecosystems. Though this difference in canopy chemistries may be a result of a sampling bias and not representative of the intra-specific variation of *M. polymorpha* in mesic grasslands, it likely resulted in lower classifications accuracies as *M. polymorpha* canopy characteristics present in datasets from other ecosystems was not represented in the mesic grasslands data. In contrast, the SVM trained on mesic shrubs, which had the largest *M. polymorpha* canopy chemistry variation, had consistently high accuracies across all ecosystems (>74.1%). This was not necessarily the case for ecosystems with high variation in background vegetation canopy traits as models trained on data from wet ecosystems yielded low accuracies (min 14.8%) when tested on other ecosystems. Training data selection is an important consideration when developing classification models across large geographic scales, though, due to the hyperplanes SVMs create as decision boundaries, simply increasing dataset size without consideration for the target species variation may not necessarily result in more accurate classifications [63, 64].

M. polymorpha is a unique case study for this investigation due to its high intra-specific variation [24, 26, 32, 34, 35, 40]. Here, we found that a SVM was able to predict the ecosystem of origin more accurately for *M. polymorpha* canopy spectra than those of other species. This may be a result of other species growing in ecosystems of the same moisture regime and having less intra-specific variation across grassland, shrub, and forest designations than *M. polymorpha*, especially as many of these species are non-native invasives and likely experienced genetic bottlenecks [65]. The developed ecosystem had the least unique spectra for both *M. polymorpha* and other vegetation. This is likely a result of developed areas on Hawai'i Island existing in regions with high and low rainfall regimes as well as people cultivating plants from multiple origins across the Hawaiian Islands.

5. Conclusions

In the upcoming era of spaceborne imaging spectrometers [66–69], large-scale vegetation classifications will become more feasible. Yet despite the imminent data availability, few studies have attempted to classify trees species across ecosystems using imaging spectroscopy data [12, 13]. This investigation addressed how intraspecific variation of tree species across ecosystems affected species classifications from imaging spectroscopy data, which is an important consideration in large-scale vegetation mapping. Implications of accurate vegetation species classifications across large geographic areas include developing baseline species distributions for tracking future range shifts [5–7], identifying locations where rare species exist [1], and monitoring invasive species [2–4]. Here we demonstrated that despite the high intra-specific variability of *M. polymorpha* across ecosystems, accurate classifications of this species were possible. While ecosystem-specific models tended to outperform the universal model, the universal model achieved high (>80%) accuracies. When classifying species across environmental gradients, developing a dataset that represents the intra-specific variation of the target species will be key to developing accurate models.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.6084/m9.figshare.23949099.

Acknowledgments

This research was funded by the National Science Foundation Doctoral Dissertation Improvement Grant (Award Number 2218932) awarded to Gregory Asner and Megan Seeley. The Global Airborne Observatory

(GAO) is managed by the Center for Global Discovery and Conservation Science at Arizona State University. The GAO is made possible by support from private foundations, visionary individuals, and Arizona State University. Author contributions are as follows: Conceptualization, M M S; methodology, M M S; data curation, N R V, G P A; fieldwork, MMS; code development, M M S; formal analysis, M M S; writing—original draft preparation, M M S; writing—review and editing, all authors; funding acquisition, M M S and G P A. All authors have read and agreed to the published version of the manuscript. The authors declare no conflicts of interest. Crown data used for model training will be made openly available in Figshare upon acceptance of the manuscript.

ORCID iDs

M M Seeley https://orcid.org/0000-0003-1945-1162
N R Vaughn https://orcid.org/0000-0003-0428-2909
G A Asner https://orcid.org/0000-0001-7893-6421

References

[1] Kivinen S *et al* 2020 A keystone species, European aspen (Populus tremula L.), in boreal forests: ecological role, knowledge needs and mapping using remote sensing *For. Ecol. Manage.* **462** 118008

[2] Hawthorne T L, Elmore V, Strong A, Bennett-Martin P, Finnie J, Parkman J, Harris T, Singh J, Edwards L and Reed J 2015 Mapping non-native invasive species and accessibility in an urban forest: a case study of participatory mapping and citizen science in Atlanta, Georgia *Appl. Geogr.* **56** 187–98

[3] Joshi C, de Leeuw J and van Duren I C 2004 Remote sensing and GIS applications for mapping and spatial modelling of invasive species *ISPRS 2004: Proc. 20th ISPRS Congress: Geo-imagery Bridging Continents* (*Istanbul, Turkey Comm VII*, *12–23 July 2004*) pp 669–77 (available at: http://www.itc.nl/library/Papers_2004/peer_conf/joshi.pdf)

[4] Randall J, Inglis N C, Smart L and Vukomanovic J 2022 From meadow to map: integrating field surveys and interactive visualizations for invasive species management in a National Park *ISPRS Int. J. Geo-Inf.* **11** 525

[5] Cursach J, Far A J and Ruiz M 2020 Geospatial analysis to assess distribution patterns and predictive models for endangered plant species to support management decisions: a case study in the Balearic Islands *Biodivers. Conserv.* **29** 3393–410

[6] Rathore P, Roy A and Karnatak H 2022 Predicting the future of species assemblages under climate and land use land cover changes in Himalaya: a geospatial modelling approach *Clim. Change Ecol.* **3** 100048

[7] Saran S, Joshi R, Sharma S, Padalia H and Dadhwal V K 2010 Geospatial modeling of Brown oak (Quercus semecarpifolia) habitats in the Kumaun Himalaya under climate change scenario *J. Indian Soc. Remote Sens.* **38** 535–47

[8] Baldeck C A, Asner G P, Martin R E, Anderson C B, Knapp D E, Kellner J R and Wright S J 2015 Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy *PLoS One* **10** e0118403

[9] Balzotti C S, Asner G P, Adkins E D, Parsons E W and Magrach A 2020 Spatial drivers of composition and connectivity across endangered tropical dry forests *J. Appl. Ecol.* **57** 1593–604

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

[11] Feret J-B and Asner G P 2013 Tree species discrimination in tropical forests using airborne imaging spectroscopy *IEEE Trans. Geosci. Remote Sens.* **51** 73–84

[12] Roth K L, Roberts D A, Dennison P E, Alonzo M, Peterson S H and Beland M 2015 Differentiating plant species within and across diverse ecosystems with imaging spectroscopy *Remote Sens. Environ.* **167** 135–51

[13] Marconi S, Weinstein B G, Zou S, Bohlman S A, Zare A, Singh A, Stewart D, Harmon I, Steinkraus A and White E P 2022 Continental-scale hyperspectral tree species classification in the United States national ecological observatory network *Remote Sens. Environ.* **282** 113264

[14] Mäyrä J *et al* 2021 Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks *Remote Sens. Environ.* **256** 112322

[15] Dalponte M, Bruzzone L and 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** 258–70

[16] Ferreira M P, Zortea M, Zanotta D C, Shimabukuro Y E and de Souza Filho C R 2016 Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data *Remote Sens. Environ.* **179** 66–78

[17] Asner G P and Martin R E 2009 Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests *Front. Ecol. Environ.* **7** 269–76

[18] Asner G P and Martin R E 2016 Spectranomics: emerging science and conservation opportunities at the interface of biodiversity and remote sensing *Glob. Ecol. Conserv.* **8** 212–9

[19] Asner G P, Knapp D E, Anderson C B, Martin R E and Vaughn N 2016 Large-scale climatic and geophysical controls on the leaf economics spectrum *Proc. Natl Acad. Sci.* **113** E4043–51

[20] Blonder B *et al* 2020 Remote sensing of ploidy level in quaking aspen (Populus tremuloides Michx.) *J. Ecol.* **108** 175–88

[21] Cavender-Bares J *et al* 2016 Associations of leaf spectra with genetic and phylogenetic variation in oaks: prospects for remote detection of biodiversity *Remote Sens.* **8** 221

[22] Féret J-B and Asner G P 2014 Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy *Ecol. Appl.* **24** 1289–96

[23] McManus K M, Asner G P, Martin R E, Dexter K G, Kress W J and Field C B 2016 Phylogenetic structure of foliar spectral traits in tropical forest canopies *Remote Sens.* **8** 196

[24] Seeley M M, Stacy E A, Martin R E and Asner G P 2023 Foliar functional and genetic variation in a keystone Hawaiian tree species estimated through spectroscopy *Oecologia* **202** 15–28

[25] Czyż E A, Guillén Escribà C, Wulf H, Tedder A, Schuman M C, Schneider F D and Schaepman M E 2020 Intraspecific genetic variation of a Fagus sylvatica population in a temperate forest derived from airborne imaging spectroscopy time series *Ecol. Evol.* **10** 7419–30

[26] Seeley M M, Martin R E, Vaughn N R, Thompson D R, Dai J and Asner G P 2023 Quantifying the variation in reflectance spectra of metrosideros polymorpha canopies across environmental gradients *Remote Sens.* **15** 1614

[27] Araújo I, Marimon B S, Scalon M C, Cruz W J A, Fauset S, Vieira T C S, Galbraith D R and Gloor M U 2021 Intraspecific variation in leaf traits facilitates the occurrence of trees at the Amazonia–Cerrado transition *Flora* **279** 151829

[28] Gotsch S G, Powers J S and Lerdau M T 2010 Leaf traits and water relations of 12 evergreen species in Costa Rican wet and dry forests: patterns of intra-specific variation across forests and seasons *Plant Ecol.* **211** 133–46

[29] Albert C H, Thuiller W, Yoccoz N G, Soudant A, Boucher F, Saccone P and Lavorel S 2010 Intraspecific functional variability: extent, structure and sources of variation *J. Ecol.* **98** 604–13

[30] Kichenin E, Wardle D A, Peltzer D A, Morse C W, Freschet G T and Kitajima K 2013 Contrasting effects of plant inter- and intraspecific variation on community-level trait measures along an environmental gradient *Funct. Ecol.* **27** 1254–61

[31] Martin R E and Asner G P 2009 Leaf chemical and optical properties of metrosideros polymorpha across environmental gradients in Hawaii *Biotropica* **41** 292–301

[32] Cordell S, Goldstein G, Mueller-Dombois D, Webb D and Vitousek P M 1998 Physiological and morphological variation in Metrosideros polymorpha, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity *Oecologia* **113** 188–96

[33] Martin R E, Asner G P and Sack L 2007 Genetic variation in leaf pigment, optical and photosynthetic function among diverse phenotypes of Metrosideros polymorpha grown in a common garden *Oecologia* **151** 387–400

[34] Stacy E A, Johansen J B, Sakishima T, Price D K and Pillon Y 2014 Incipient radiation within the dominant Hawaiian tree Metrosideros polymorpha *Heredity* **113** 334–42

[35] Stacy E A, Johansen J B, Sakishima T and Price D K 2016 Genetic analysis of an ephemeral intraspecific hybrid zone in the hypervariable tree, Metrosideros polymorpha, on Hawaiʻi Island *Heredity* **117** 173–83

[36] Des Roches S, Post D M, Turley N E, Bailey J K, Hendry A P, Kinnison M T, Schweitzer J A and Palkovacs E P 2018 The ecological importance of intraspecific variation *Nat. Ecol. Evol.* **2** 57–64

[37] Clark J S 2010 Individuals and the variation needed for high species diversity in forest trees *Science* **327** 1129–32

[38] Jung V, Violle C, Mondy C, Hoffmann L and Muller S 2010 Intraspecific variability and trait-based community assembly *J. Ecol.* **98** 1134–40

[39] Reich P B, Ellsworth D S, Walters M B, Vose J M, Gresham C, Volin J C and Bowman W D 1999 Generality of leaf trait relationships: a test across six biomes *Ecology* **80** 1955–69

[40] Joel G, Aplet G and Vitousek P M 1994 Leaf morphology along environmental gradients in hawaiian metrosideros polymorpha *Biotropica* **26** 17–22

[41] Seeley M M, Vaughn N R, Shanks B L, Martin R E, König M and Asner G P 2023 Classifying a highly polymorphic tree species across landscapes using airborne imaging spectroscopy *Remote Sens.* **15** 4365

[42] Asner G P, Knapp D E, Boardman J, Green R O, Kennedy-Bowdoin T, Eastwood M, Martin R E, Anderson C and Field C B 2012 Carnegie airborne observatory-2: increasing science data dimensionality via high-fidelity multi-sensor fusion *Remote Sens. Environ.* **124** 454–65

[43] Miller C J 2002 Performance assessment of ACORN atmospheric correction algorithm *Proc. SPIE* **4725** 438–49

[44] Schaepman-Strub G, Schaepman M, Martonchik J and Schaaf C 2006 Whats in a satellite albedo product? *2006 IEEE Int. Symp. on Geoscience and Remote Sensing* pp 2848–52

[45] Asner G P, Knapp D E, Kennedy-Bowdoin T, Jones M O, Martin R E, Boardman J W and Field C B 2007 Carnegie airborne observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems *Proc. SPIE* **1** 013536

[46] Asner G P, Martin R E, Anderson C B and Knapp D E 2015 Quantifying forest canopy traits: imaging spectroscopy versus field survey *Remote Sens. Environ.* **158** 15–27

[47] Weingarten E, Martin R, Hughes F, Vaughn N, Schafron E and Asner G P 2021 Early detection of a tree pathogen using airborne remote sensing *Ecol. Appl.* **32** e2519

[48] U.S. Geological Survey Gap Analysis Program 2011 GAP/LANDFIRE national terrestrial ecosystems 2011: U.S. Geological survey (https://doi.org/10.5066/F7ZS2TM0) (Accessed 13 May 2016)

[49] Giambelluca T W, Chen Q, Frazier A G, Price J P, Chen Y-L, Chu P-S, Eischeid J K and Delparte D M 2013 Online rainfall atlas of Hawaiʻi *Bull. Am. Meteorol. Soc.* **94** 313–6

[50] Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72

[51] Balzotti C S and Asner G P 2018 Biotic and abiotic controls over canopy function and structure in humid Hawaiian forests *Ecosystems* **21** 331–48

[52] Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Martin-Guerrero J D, Soria-Olivas E, Alonso-Chorda L and Moreno J 2004 Robust support vector method for hyperspectral data classification and knowledge discovery *IEEE Trans. Geosci. Remote Sens.* **42** 1530–42

[53] Dalponte M, Ørka H O, Gobakken T, Gianelle D and Næsset E 2013 Tree species classification in boreal forests with hyperspectral data *IEEE Trans. Geosci. Remote Sens.* **51** 2632–45

[54] Melgani F and Bruzzone L 2004 Classification of hyperspectral remote sensing images with support vector machines *IEEE Trans. Geosci. Remote Sens.* **42** 1778–90

[55] Piiroinen R, Heiskanen J, Maeda E, Viinikka A and Pellikka P 2017 Classification of tree species in a diverse African agroforestry landscape using imaging spectroscopy and laser scanning *Remote Sens.* **9** 875

[56] Pedregosa F *et al* 2011 Scikit-learn: machine Learning in Python *J. Mach. Learn. Res.* **12** 2825–30

[57] Belgiu M and Drăguţ L 2016 Random forest in remote sensing: a review of applications and future directions *ISPRS J. Photogramm. Remote Sens.* **114** 24–31

[58] Zhang X M, He G J, Zhang Z M, Peng Y and Long T F 2017 Spectral-spatial multi-feature classification of remote sensing big data based on a random forest classifier for land cover mapping *Clust. Comput.* **20** 2311–21

[59] Czyż E A, Schmid B, Hueni A, Eppinga M B, Schuman M C, Schneider F D, Guillén-Escribà C and Schaepman M E 2023 Genetic constraints on temporal variation of airborne reflectance spectra and their uncertainties over a temperate forest *Remote Sens. Environ.* **284** 113338

[60] Madritch M D, Kingdon C C, Singh A, Mock K E, Lindroth R L and Townsend P A 2014 Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales *Phil. Trans. R. Soc.* B **369** 20130194

[61] Omumbo J A, Hay S I, Goetz S J, Snow R W and Rogers D J 2002 The effect of training strategies on supervised classification at different spatial resolutions *Photogramm. Eng. Remote Sens.* **68** 161–6

[62] Hixson M 1980 Evaluation of several schemes for classification of remotely sensed data *Photogramm. Eng. Remote Sens.* **46** 1547–53

[63] Foody G M and Mathur A 2004 Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification *Remote Sens. Environ.* **93** 107–17

[64] Mathur A and Foody G M 2008 Crop classification by support vector machine with intelligently selected training data for an operational application *Int. J. Remote Sens.* **29** 2227–40

[65] Puillandre N, Dupas S, Dangles O, Zeddam J-L, Capdevielle-Dulac C, Barbin K, Torres-Leguizamon M and Silvain J-F 2008 Genetic bottleneck in invasive species: the potato tuber moth adds to the list *Biol. Invasions* **10** 319–33

[66] Culver T *et al* 2020 SBG user needs and valuation study (RTI Innovation Advisors)

[67] Iwasaki A, Tanii J, Kashimura O and Ito Y 2019 Prelaunch status of hyperspectral imager suite (Hisui) *IGARSS 2019–2019 IEEE Int. Geoscience and Remote Sensing Symp.* pp 5887–90

[68] Lopinto E and Ananasso C 2020 The prisma hyperspectral mission p 12

[69] Müller R *et al* 2018 Overview and status of the DESIS mission (available at: https://elib.dlr.de/123992/) (Accessed 5 January 2021)