Harnessing temporal and spectral dimensionality to map and identify species of individual trees in diverse tropical forests

James G C Ball[1,2], Sadiq Jaffer[3], Anthony Laybros[2], Colin Prieur[2], Toby Jackson[1], Anil Madhavapeddy[3], Nicolas Barbier[2], Gregoire Vincent[2], David A Coomes[1]

[1] Conservation Research Institute, Department of Plant Sciences, University of Cambridge, Cambridge, UK

[2] UMR AMAP, INRAE, CIRAD, IRD, Montpellier, France

[3] Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

Keywords: tree crown delineation; multi-date remote sensing; convolutional neural networks; tropical forests; hyperspectral; imaging spectroscopy; tree species identification

Abstract

To understand how tropical rainforests will adapt to climate change and the extent to which their diversity imparts resilience, precise, taxonomically informed monitoring of individual trees is required. However, the density, diversity and complexity of tropical rainforests present considerable challenges to remote mapping and traditional field-based approaches are limited in scale. This study introduces a new approach for mapping tree species linking a multi-temporal implementation of the convolutional neural network method, *detectree2*, to segment tree-crowns from aerial photographs to machine learning classifiers to identify species from hyperspectral data (416 - 2500 nm). We build upon previous work in two ways. Firstly, we aimed to improve the accuracy of crown delineations by surveying the same patch of forest with UAV-RGB ten times over six months and fusing multi-date information on the location and shape of individual trees. Secondly, we extended the scope of species identification to include far more species than has been previously attempted (169 compared to 20 previously). We trained and tested our algorithms on subsets of a database of 3500 ground truth, labelled tree crown polygons representing 239 species in French Guiana that we had delineated by hand and field verified. We assessed how well our segmentation approach could locate and delineate individual tree crowns and how well our classification approach predicted the species of those crowns. We extracted information on waveband importance for distinguishing species from our classification model. Based on an existing phylogeny of the trees in our dataset, we tested for phylogenetic signal across the hyperspectral bands and probed how species were being classified by comparing the phylogenetic signal to the importance of bands for separating species. The accuracy of delineations increased gradually as additional dates of tree crown maps were stacked and combined. Stacking increased the F_1-score from 0.69 (a single date) to 0.78 (all dates). The overall (weighted) F_1-score for species classification was 0.75. A total of 65 species were predicted from the hyperspectral data with F_1-score > 0.7. The performance for classifying a species increased with the number of crowns in the database available for that species: 8 training crowns were needed to achieve an expected F_1-score = 0.7 for crown level classification. With this new approach, we assessed that 70% of tree crown area at landscape-scale was accurately mapped. The most important wavebands for discriminating species were narrowly clumped on the NIR side of the red edge region (748 - 775 nm). While most wavebands showed some phylogenetic signal, waveband importance for species classification was negatively correlated with phylogenetic signal. Our integrated approach makes a significant contribution to the ongoing development of efficient and accurate methodologies for mapping canopy tree species in tropical forests, providing a framework for mapping trees in diverse tropical forests that is far more comprehensive than its predecessors.

1. Introduction

Tropical moist forests are renowned for their species richness. The ability to map tree species using remote sensing data is of value to ecologists and practitioners of a broad range of other disciplines (Fassnacht et al. 2016). It enables real-time assessment and monitoring of biodiversity and species compositions (Shang and Chisholm 2014) and can provide insights into habitats (Pausas, Austin, and Noble 1997; Jansson and Angelstam 1999; Kennedy and Southwood 1984), tree community dynamics and ecosystem function (Chambers et al. 2013; Ewijk et al. 2014). Such data can support conservation

efforts, by locating species of interest (Baldeck et al. 2015), improving estimates of stored carbon (Bredin, Peres, and Haugaasen 2020), tracking invasive species (Chance et al. 2016; Sabat-Tomala, Raczko, and Zagajewski 2020), helping to manage water stress (Asner et al. 2004; Watt et al. 2021), monitoring the spread of pests and disease (Y. Liu et al. 2021; Chan et al. 2021), and improving our understanding of migration patterns (McGrath, Riper, and Fontaine 2009). Maps of canopy trees are also valuable for assessing the extent to which the extraordinary diversity of tropical forests influences ecosystem processes (Reichstein et al. 2013), including resilience to climate change (Malhi et al. 2008; Corlett 2011) through the presence of species with varied climatic tolerances (Lewis et al. 2009). However, although it is now established that diversity can promote ecosystem stability (Loreau and Mazancourt 2013), it remains unclear whether such diversity is necessary, because of functional redundancy among species (Biggs et al. 2020). Hence, an integrative understanding of the distribution of tree species diversity is vital in understanding the impacts of anthropogenic change and crafting effective strategies for the sustainable management of these globally significant ecosystems (Goetz et al. 2009). From a commercial standpoint, remote sensing of tree species can improve the speed, scope and precision of forestry inventories (Aardt and Wynne 2007; Laybros et al. 2020) which aid in calculating available resources and planning sustainable harvests (Vauhkonen et al. 2014). Together, these insights can support evidence based policy for sustainable forest management, balancing human activity and ecological preservation (FAO 2020).

Mapping individual tree crowns and identifying their species at large scale in diverse and dense tropical forests presents a significant challenge (Asner and Martin 2011), requiring accurate delineation of tree crowns and classification of pixels within those crowns. The first step of locating and delineating individual tree crowns in densely packed tropical forests from above is difficult as crowns can interweave and overlap in complex ways. Remote sensing of individual trees has mostly focused on airborne lidar data, but this is most successful for temperate and boreal forests (Dalponte and Coomes 2016; Hastings et al. 2020) while complex tropical canopies have presented a greater challenge (Aubry-Kientz et al. 2019). RGB photographs offer colour and texture information to distinguish

trees, even if they are structurally similar, but agreement on tree location and crown shape between human analysts is variable (see Section S.1). Mask R-CNN (He et al. 2017) based approaches have shown promise in their ability to harness the colour and texture information of RGB images to precisely separate irregular edges of neighbouring tree crowns (Ball et al. 2023; Gan, Wang, and Iio 2023). However, differences in illumination (resulting in variably shadowed trees), the sway of trees and branches, phenological discrepancies and irregularities in the orthomosaicking process often reveal inconsistent arrangement/delineation of crowns for aerial imagery across dates. Therefore, methods that can draw on consensus between datasets and are tested against a robust ground truth are required.

Assigning species labels to the tree crowns is also challenging due to high species diversity, the rarity of most species, and commonalities between species. Tropical forests have approximately log-normal species abundance distributions: a few species are common, while most are represented by only a few individuals per hectare. A pan-Amazonian study based on 640,000 trees (DBH $\geq$ 10 cm) in 1100 ha of forest plots, 36 % of species were modelled to have a population size of fewer than 1000 individuals across the whole of the Amazon, while 1.4 % of species were estimated to account for half of all trees (Steege et al. 2013). This skewed distribution presents challenges for biodiversity conservation efforts, as it means that many species are always at risk of extinction due to their small populations. It also complicates efforts to identify and study individual tree species using remote sensing techniques, as the large number of rare species (with few training examples) can be difficult to reliably detect and classify. While there is considerable species diversity, the presence of chlorophyll and other photosynthetic pigments is ubiquitous across trees. This further complicates remote species identification, as distinguishing between individual species based on their pigmentation alone is not feasible. Closely related species often share certain chemical, physiological and morphological characteristics that can influence their spectral properties (Meireles et al. 2020; Cavender-Bares et al. 2016). For instance, the leaf structure and chemical properties (e.g. concentrations of chlorophyll and cellulose) can vary between families which in turn can

affect their spectral properties (Serbin et al. 2016). However, there can still be considerable variability in the spectral properties of plants within the same family due to differences at the genus and species level, as well as due to environmental factors (e.g. soil type, water availability, and light exposure) and plant health status or phenological stage (Ollinger 2011). So, while it is possible that members of the same plant family might exhibit some similarities in their spectral properties, these are not definitive, and a comprehensive classification of plant species based on spectral data would likely need to take into account more specific characteristics and use sophisticated machine learning algorithms to capture these complex patterns (Fassnacht et al. 2016).

Hyperspectral remote sensing (imaging spectroscopy), a technique that measures the intensity of incident radiation with many narrow spectral bands across a wide spectrum, is often applied to mapping tree species in forests (Ghiyamat and Shafri 2010; Fassnacht et al. 2016). The spectral properties of plants are influenced by numerous factors, including their evolved biochemical and structural characteristics (Ustin et al. 2004; Ustin 2013; Ollinger 2011; Meireles et al. 2020). Many studies have shown that subtle differences in spectral reflectance arising from these biophysical differences can be detected in hyperspectral data, allowing species to be mapped and their health monitored in low diversity (often temperate) systems (Fassnacht et al. 2016). However, distinguishing species in diverse tropical forests has proven more challenging; Clark, Roberts, and Clark (2005) used hyperspectral data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) to differentiate seven species in a tropical rainforest in Costa Rica, paving the way for further research in this domain. Féret and Asner (2013) were able to distinguish 17 species in the lowland humid tropical forest of Hawaii. Laybros et al. (2019, 2020) showed that a classification rate of around 80% was achievable when classifying within a pool of 20 well represented species in an Amazonian forest in French Guiana. Greater scope has been shown with proximate, leaf level hyperspectral species identification (e.g. 46 species at a tropical wetland in Jamaica (Prospere, McLaren, and Wilson 2014) highlighting the potential for a broader scope with aerial hyperspectral data at the crown level. However, leaf traits are incorporated into spectra in complex ways (Jacquemoud and Baret 1990;

Ollinger 2011; Féret and Asner 2011) and multiple traits can overlap in a given spectral region (Curran 1989). This is further complicated when spectra are measured from a distance as signals of relevant traits are confounded with structural characteristics. Radiative transfer models have shown that signals from canopy structure can dominate over the leaf optical properties and biochemical properties of the vegetation (Knyazikhin et al. 2013; Béland and Kobayashi 2024). However, elsewhere they have been used to show that incorporating the fraction of non-photosynthetic vegetation as leaf brown pigments rather than as woody material gives the better match between simulated and observed canopy spectra (Ebengo et al. 2021). The degree to which these uncertainties place limits on the predictability of tree species from hyperspectral data is as yet unclear. Structural properties such as leaf density, clumping and angle distribution, as well as leaf biochemistry, may, to some extent, be considered functions of species. Rather than attempt to reconstruct a physical representation of the canopy (as in radiative transfer approaches), we can test the separability of species through a data driven analyses. Unpicking the spectral signals in relation to the evolved traits and how species relate to each other is a key challenge in progressing towards robust, transferable remote identification of species in diverse tropical forests (Schweiger et al. 2021).

Applying computer vision approaches in a way that combines information across time may lead to improved detection and segmentation of canopy trees (Shamaoma et al. 2023; Martin et al. 2018). While RGB data lack the spectral resolution of hyperspectral data, its spatial resolution is typically superior, especially when acquired from a drone that can fly close above the forest canopy. This means that textural information can be observed and outlines of individual tree crowns can be distinguished precisely. The relatively low cost of the sensors and UAV systems also means that regular surveys of the same area become feasible. Differences in the canopy (including phenological states), atmosphere, and illumination can mean that predicted crown maps vary considerably across dates. To address this temporal variability, we implemented a consensus-fusion approach to combine tree crown polygons detected on different dates. The tree crown polygons from individual dates were spatially matched and the vertices of their parameters averaged ,

generating output crowns maps that represented the inter-date consensus on the location and shape of tree crowns. Our hope is that this approach would help to mitigate the effects of temporal fluctuations, enhancing the consistency and accuracy of crown delineation.

In this study, we propose a novel approach to rainforest tree mapping that (1) generates precise delineation of individual tree crowns and (2) classifies the species of each tree. The automatic delineation of individual tree crowns was performed with *detectree2*, a tool based on the Mask R-CNN architecture (He et al. 2017; Ball et al. 2023). For the first time, we combine delineated tree maps generated from repeated airborne surveys and analyse whether identifying consensus across stacked tree-crown maps improves segmentation accuracy. Our species identification model was trained and tested using a database of 3256 manually delineated tree crowns, each meticulously verified through several field missions. This dataset, in combination with our novel approach, sets the stage for a more accurate and inclusive tree mapping and species classification system. This two-step approach leverages the strengths of both technologies - the spatial resolution of UAV RGB imagery and the spectral resolution of hyperspectral data - achieving a high level of accuracy in tropical forest mapping.

We address the following research questions:

1. Can the accuracy of tree crown maps from aerial imagery be improved by combining information from segmentations at different dates to build a consensus-based map?
2. Can we accurately classify a representative range of tropical tree species from hyperspectral data?
    a. Which type of machine learning classifier most accurately predicts the species of tree crowns from hyperspectral data?
    b. How many mapped individuals of a given species are needed to achieve a 'good' classification accuracy?
    c. Which wavebands are the most important for determining species?

3. How do species' crown reflectance spectra relate to their phylogenetic relationships?
    a. Do spectra observed within tree crowns exhibit phylogenetic signal?
    b. Are the most important wavebands for species classification those that have the strongest phylogenetic signal?
    c. Are closely related species more often confused in their classification than distantly related species?

Through a novel integration of ground validated manual tree crown generation and machine learning algorithms, we developed a robust and accurate methodology to delineate tree crowns and predict their species. Our methodology combines traditional fieldwork, advanced machine learning techniques, high-resolution remote sensing data, and phylogenetic insights offering an innovative approach to species-level forest mapping.

2. Materials and Methods

2.1 Study site and overview of methods

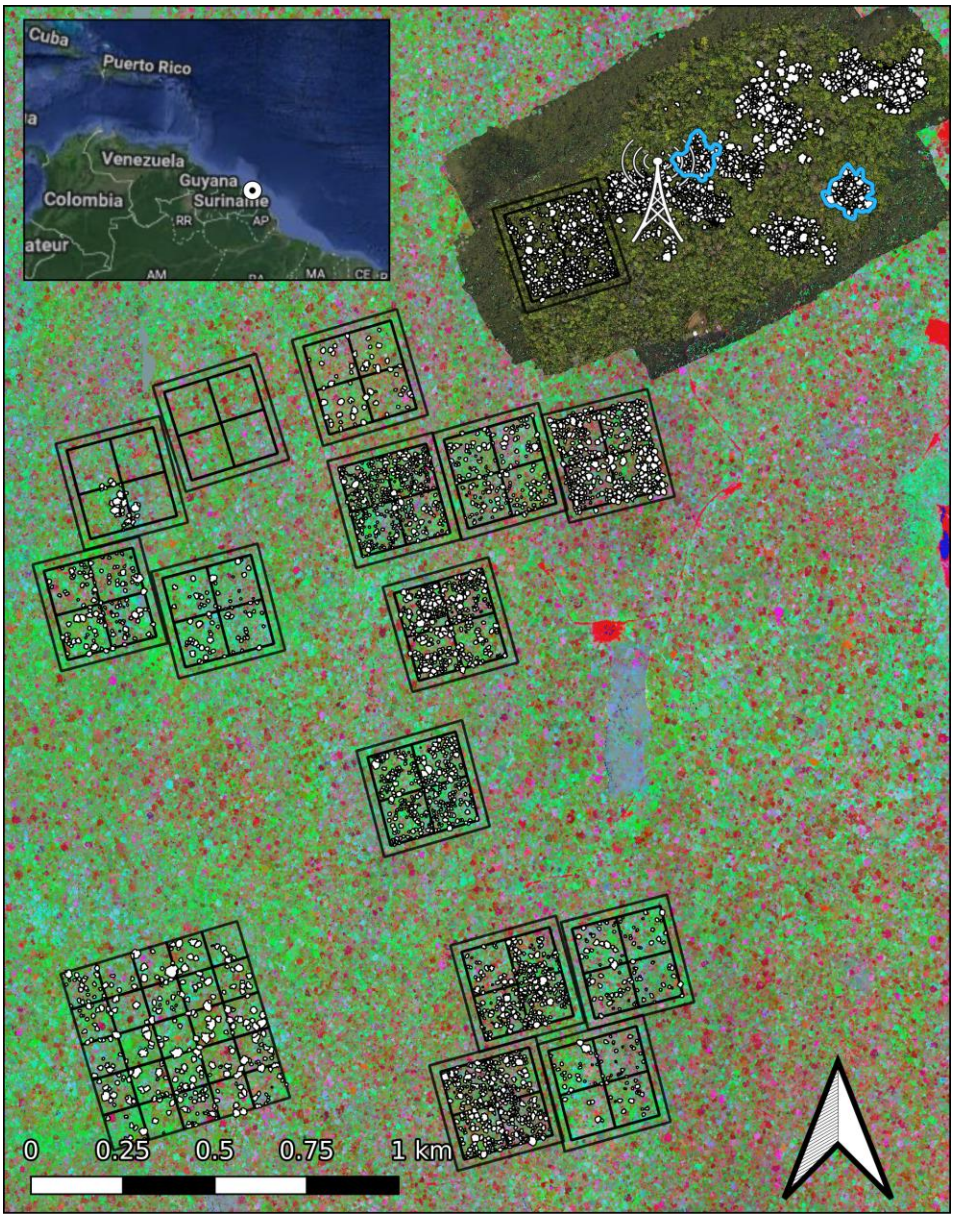

Figure 1*: Site map of Paracou, French Guiana, with crowns and hyperspectral imagery. The manually delineated, labelled crowns are in white. The colourful background scan that covers the entire site is a representation of the hyperspectral data (selected projected PCA bands for illustration). The repeat survey UAV-RGB region is shown in the northwest around the site's flux tower. Within this region the segmentation test data*

areas are delineated in blue - the crowns within these areas were excluded from all training of the segmentation delineation. The black boxes show the plots in which inventories are conducted.

The research was conducted in forests at Paracou Field Station, French Guiana (5N 52 W) (see Fig. 1). The lowland tropical rainforests grow mostly on shallow ferralitic soils underlain by a variably transformed loamy saprolite (Gourlet-Fleury, Guehl, and Laroussinie 2004). The mean annual rainfall is approximately 3200 mm with a three-month dry season from mid-August to mid-November where rainfall is typically less than 50 mm per month (Bonal et al. 2008; Wagner et al. 2011). The field station has 27 permanent plots ranging in size from 0.5 ha to 25 ha (see Fig. 1) which contain approximately 76,000 trees of DBH $\geq$ 10 cm consisting of over 800 different species (Gourlet-Fleury, Guehl, and Laroussinie 2004). In these plots, inventories are taken every 1-5 years with the species, precise geographic location and DBH of each trunk recorded. The ten most common species account for just over 30% of the inventory's individuals. 90% of the species present have been placed within a time calibrated phylogeny by Baraloto et al. (2012).

Our study uses a combination of remote sensing data from a UAV-mounted camera and a plane-mounted imaging spectrometer both co-registered to a LiDAR-derived Canopy Height Model (CHM; see Table 1). A CNN approach applied to UAV-RGB data recorded with 10 surveys over 6 months was used to locate and delineate the individual tree crowns. Hyperspectral imagery was used to classify the species of the crowns. Each tree species absorbs and reflects light in a specific way across wavelengths in the 416-2500 nm range, generating a spectral 'signature' that can be used to distinguish between species (see Fig. 6). The predictions were evaluated against strict, unseen test sets of manual tree crowns that were not exposed to the algorithms during training, providing a robust assessment of model accuracies.

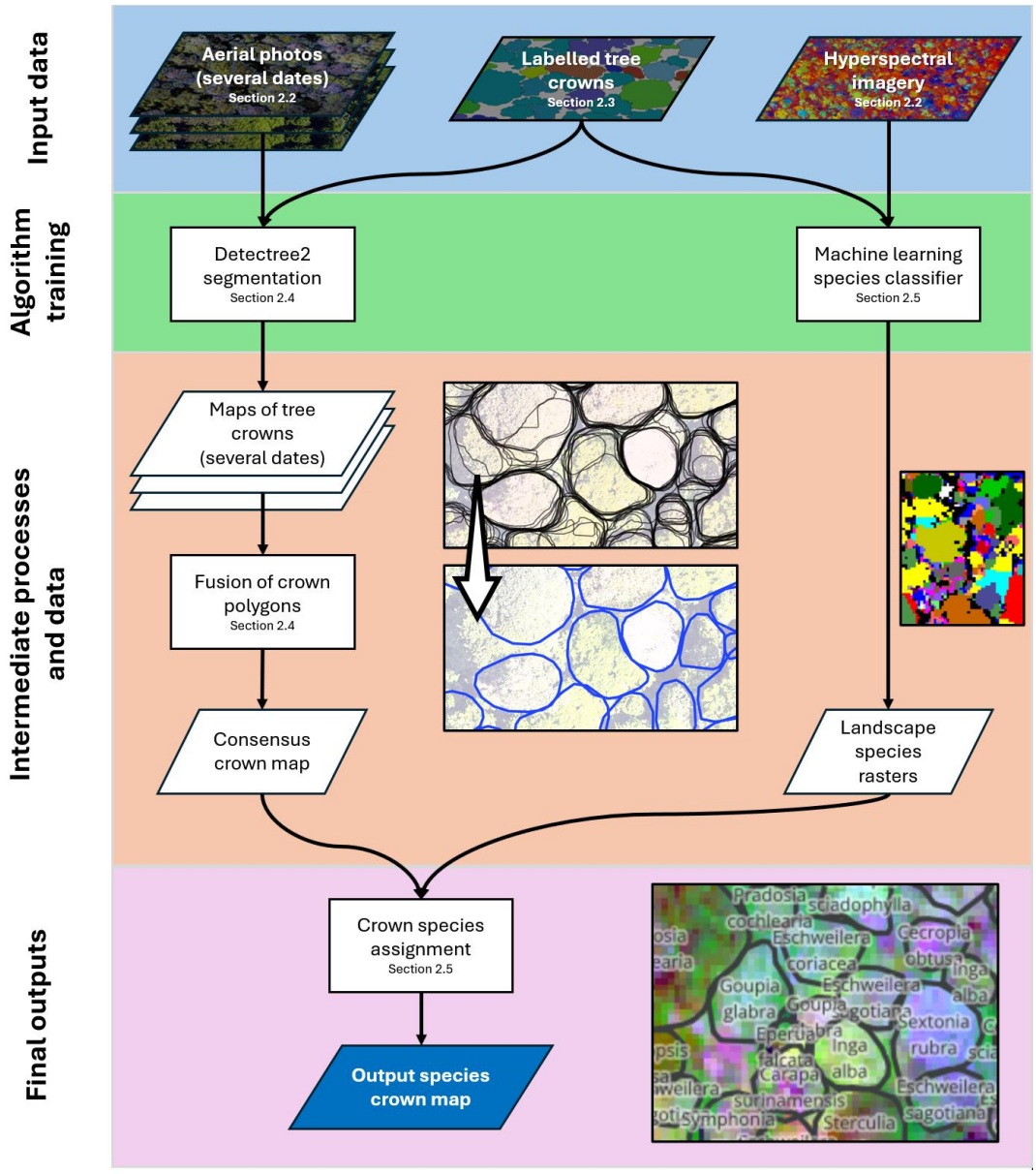

Figure 2*: Simplified schematic of the crown mapping approach showing input data and the intermediate steps to producing a labelled tree crown map. In the centre is an illustration of the process of temporal polygon fusion. The top image shows the overlapping tree crown polygons predicted over multiple dates. Where polygons have a high degree of overlap, each will still have a slightly different shape due to differences in the RGB orthomosaics through time. The bottom image depicts the polygons after the fusion process, effectively averaging the positions of the vertices of the original polygons and discarding those without good consensus through time.*

2.2 Remote sensing data acquisition and co-registration

***Table 1**. Remote sensing data sources used in the study (location Section 2.1). Resolution is given as ground resolution for the RGB orthomosaic and as the processed CHM resolution for the lidar scans. Altitude is given as height above the forest canopy.*

Scan date(s)	Modality	Resolution	Altitude	Spectral range	Sensor
23-Oct-2020—06-Apr-2021	RGB	5 cm	70 m	421-617 nm (3 bands)	1" CMOS (Phantom4 Pro)
19-Sep-2016	Hyperspectral	1 m	900 m	416-992 nm (160 channels)	Hyspex VNIR-1600
19-Sep-2016	Hyperspectral	2 m	900 m	930-2500 nm (288 channels)	Hyspex SWIR-384
15-Nov-2019	LiDAR	0.5 m (CHM)	800 m	1550 nm (active)	RIEGL LMS-Q780

UAVs (DJI Phantom 4 Advanced and DJI P4 Multispectral) were employed to collect high-resolution RGB imagery, with a scan approximately every three weeks over a 6-month period (10 surveys in total) of the region shown in Fig. 1. The RGB orthomosaics were compiled from the raw geotagged UAV photographs using structure from motion (SfM) photogrammetry in AgiSoft Metashape. To improve spatio-temporal coherency, instead of processing each date separately, five date blocks were supplied for the alignment and initial sparse point cloud formation steps establishing a common geometry between dates (Feurer and Vinatier 2018). Details of this processing are given in Section S.2.

Hyperspectral preprocessing is described in detail by Laybros et al. (2019, 2020) and summarised here. Two sensors mounted to an aircraft side-by-side were used to cover the full 416-2500 nm wavelength (see Table 1). To merge the data from the two hyperspectral sensors without degrading the spatial resolution of the VNIR imagery, we resampled the

SWIR imagery to 1 m using nearest-neighbour interpolation. Images were orthorectified and georeferenced at 1 m spatial resolution with the PARGE software using the canopy Digital Surface Model (DSM) produced from the LiDAR point cloud. Bands in the SWIR with a low signal to noise ratio due to water absorption peaks were removed leaving 378 of the 448 total bands (see Fig. 6). Per pixel illumination was calculated using the shadow detection method of Schläpfer, Hueni, and Richter (2018). Spectral information used to train and make predictions with the species classifiers was extracted from the overlapping flight lines rather than from a mosaic. This allows for valuable information to be retained as multiple views of individual crowns within the overlapping flight lines which has been shown to improve the classification performance (Laybros et al. 2019). Reflectance spectrum normalization was applied to each pixel. The normalization consisted of dividing the reflectance value of each band by the spectrum of a pixel, by the sum of all reflectance values, which has been shown to improve tree species classification (Dalponte et al. 2014). Some machine learning classifiers are sensitive to the scale in which each feature (band in this case) is supplied with features that have a higher absolute variability tending to dominate. To address this, we applied the 'standard' scaling approach which standardizes features by removing the mean (centring on zero) and scaling to unit variance (see Fig. 6). Additional details are given in Section S.2.

Accurate co-registration of data from RGB and hyperspectral imagery was important to ensure spatial alignment. We used the LiDAR-derived Canopy Height Model (CHM) as the baseline layer, with all other data registered against it. This choice was due to the CHM's stability and precision in representing the physical landscape, providing a solid reference for co-registration. Eight control points were manually assigned across the different datasets, using identifiable features within the LiDAR CHM, such as the flux tower, roads and dominant trees, affine transformations were applied based on these. This co-registration process ensured that the crowns represented across the datasets corresponded to the same geographical location, serving as the foundation for subsequent analysis steps, including tree crown delineation and species classification.

2.3 Field-derived tree crown database

To train and validate our models we generated a set of hand delineated, 'ground truth' crowns with species labels; this database was built and curated between 2015 and 2023 and validated over eight field missions in this period. An initial delineation of tree crowns was performed in QGIS using a combination of RGB, multispectral, hyperspectral, and LiDAR remote sensing data (see Table 1). The LiDAR CHM was used as the foundational base layer on which the crowns were drawn as it provides the greatest stability and spatial precision for the outlines. We overlaid the RGB, multispectral and hyperspectral data layers, and examined and compared between them to use as much of the spectral, textural and shape information as possible. Where the crowns fell within the inventory plots, an initial guess as to which individual the crown belonged to, based on the location and size of the trunk, was assigned to the polygon. Two provisional confidence scores were assigned to the polygons: (1) a 'crown integrity' score describing the certainty with which the outline defines the complete crown of a single individual (rather than a partial crown or multiple crowns), (2) a 'trunk match' score describing how confident we were that a crown had been correctly assigned to an individual in the inventory. Where there were changes to the crowns through mortality or branch fall events, the date of change was encoded so that the crowns could be filtered to match the remote sensing data source that they are paired with. Subsequent fieldwork further refined and updated the tree crown delineations. By comparing *in situ* observations with the remote sensing data, we either matched the identified crowns to individual trees present in the site inventory (Gourlet-Fleury, Guehl, and Laroussinie 2004), or in cases where crowns were located outside of the known plots, we engaged botanists to assign the appropriate species. We updated the crown outlines and confidence scores based on the field observations and noted where there were liana infestations in the crowns. Unlike in previous studies (Laybros et al. 2019, 2020), we decided to retain infested crowns throughout the analysis despite the risk that it would introduce noise into the crown spectra. Crown infestation is an unavoidable obstacle in practical tree species mapping and this should be reflected in the landscape-scale performance evaluation of such methods. Additional details can be found in Section S.3.

2.4 Automated delineation and fusion of results from repeat surveys

The field-delineated tree crowns were partitioned into a training and testing set based on their geographic location (see Fig. 1). The regional partitioning ensured clear spatial separation between training and testing datasets, thus negating potential inflation of reported accuracy induced by spatial autocorrelative effects providing more reliable and independent assessment of delineation performance (see Kattenborn et al. 2022).

We used *detectree2*, a tool based on the Mask R-CNN deep learning architecture for automated tree crown delineation (Ball et al. 2023), which has been shown to outperform another leading CNN method for tree crown detection (Gan, Wang, and Iio 2023). *Detectree2* was trained on the manual crown delineations and corresponding RGB images from the training dataset (see Fig. 1). Researchers seeking to map trees in their landscape of interest may have some ground truth crowns and RGB surveys to train a model on or they may not have training data available. To reflect this, we tested models trained under different regimes:

1. The first was a 'base' model. This was not trained on the UAV RGB imagery and was just exposed to the plane mounted data and crowns from the range of sites described in Ball et al. (2023). This meant it had been exposed to the Paracou forest but with a different sensor, different resolution imagery and four years separation. This pre-existing model is openly available for anyone to use[1].

2. The '1 date' model took the 'base' model and further trained the model on just the first date of the UAV RGB imagery and manual crowns.

3. The '5 date' model took the 'base' model and further trained the model on the first five dates of the UAV RGB imagery and manual crowns.

[1] https://zenodo.org/records/10522461

Comparing the performance of these models provides an idea of what level of accuracy would be expected for researchers aiming to map their landscape with differing amounts of training data.

Does the fusing of multi-date crown information improve segmentation accuracy?

The trained models were then used to detect and delineate the tree crowns across the entire region of the UAV RGB scans for all 10 acquisitions in the range 23-Oct-2020 to 06-Apr-2021. The RGB images were first tiled, predicted upon and recombined to generate a set of polygons representing tree crowns from each date (see Ball et al. 2023 for details of the detection/delineation method). Each predicted crown polygon was associated with a confidence score (0-1) indicating the reliability of each tree crown prediction. Where spatial overlap between predictions existed ($IoU \geq 0.2$), the most confident prediction was retained, and the less confident predictions removed.

The predictions at the individual dates were then combined into 'fused' (or *consensus*) delineations. This aimed to find a temporal consensus on crown locations and shapes. The fused sets of crowns went from combining just two dates up to combining the full ten dates. This was done to determine the marginal benefit of adding additional dates of data, each of which comes with an associated cost of the survey.

To combine the polygon sets across dates, each individual date polygon was compared to every other predicted polygon to identify significant matches ($IoU \geq 0.75$). Accordingly, in the largest combined predictions set (10 dates), a single polygon could have a maximum of nine matches (in the case of strong agreement across dates) and a minimum of zero matches (where there is no confirmation of the polygon across any of the other dates).

Where polygons had been assigned to a matched, inter-date group, the group of polygons was then 'averaged' to produce a single, representative polygon. This was done by normalising the polygons to have 300 boundary vertices, matching vertices across polygons based on their position and taking a weighted average of the *x*- and *y*-coordinates (weighted by the confidence score of the individual polygons; details in Section S.4). An outputted polygon would be assigned a *summed confidence* (i.e. the sum of each individual

polygon output confidence scores) to provide an estimate of the reliability of each prediction, taking into account both the confidence of the original predictions and the degree of agreement between them.

A space filling algorithm iteratively selected the fused polygons of highest *summed confidence* to fill regions of the landscape. A polygon could not be assigned to the landscape if it overlapped ($IoU < 0.2$) with a polygon of higher *summed confidence*. The resulting set of averaged polygons provided a spatial-temporal integration of the tree crown predictions, with each polygon representing the average location and outline of a tree crown across multiple time steps. The algorithm parameters, including confidence thresholds, IoU threshold for matching, and number of vertices on a normalised polygon, were tuned on the training crowns prior to testing on the unseen test regions - the optimised values are those given above. Details of the fusion algorithm and landscape filling approach are given in Section S.4.

To evaluate the performance of the segmentation algorithm, we measured the overlap between predictions and reference crowns. An IoU of an overlapping pair of more than 0.5 was considered a match. This is a standard threshold used in the comparison of tree crown segmentation algorithms (Aubry-Kientz et al. 2021) that allows for small discrepancies in alignment and outline. These true positives, as well as the unmatched predictions (false positives) and unmatched manual crowns (false negatives), were used to calculate the precision, recall and F_1 score of the predictions. To determine whether combining tree crown segmentation predictions across dates improved the segmentation accuracy through consensus building, we assessed the F_1-score of each combination of dates, from each single date prediction to the combination of all ten dates. By taking all possible date combinations we estimated a mean and standard deviation for the F_1-score for each level of multi-date combination (single date through to ten dates).

2.5 Tree species classification

Once we had mapped tree crowns as objects in the landscape it was necessary to assign them a species label. The aim was to train a classifier that could, based on the spectral

bands of the hyperspectral data, classify tree crowns by species. Of the 3500 manual crowns in the database, 3256 were labelled down to species level (239 total species) and delineated, with sufficient confidence to use as a reference for species classification. Within this set, 169 species had at least two crowns which meant these species could have at least one crown for training and one crown reserved for independent testing. Crowns that were the sole representative of their species (70 in total) were dropped from training and testing of the models but their proportional representation in the overall population was accounted for in the final evaluation of the performance metrics, correcting any potential inflation in reported performance from dropping them. Specifically, the weightings of the weighted F_1-scores were adjusted (based on the original, unfiltered dataset) to account for the fact these species were missing from the test set allowing for a representative estimate of the landscape level classification accuracy (as opposed to performance metrics evaluated within a subset of species in the landscape). Due to the scarcity of data, particularly for the less well represented species, it was necessary to take a pixel-based approach for training and prediction that would then be aggregated for crown-level predictions (as opposed to taking each crown as an individual unit). Taking a pixel level approach was also likely to improve the spatial transferability across a landscape which may have different atmospheric and illumination conditions. Each pixel was labelled with a class (species) and contained a reflectance value for each hyperspectral band. Pixels with an illumination of less than 60% were discarded. The problem was approached as a supervised learning problem so pixels without a class label were not included in model training.

Whereas the crown delineation required a regional partitioning of crowns between train and test sets to give a robust estimate of performance, for tree species identification, a species-stratified crown level partitioning was more suitable. Due to the diversity and mixing of the forest the average distance between crowns of the same species was large enough to not require additional spatial constraints that would control for spatial autocorrelation. Therefore, a test set containing a (stratified) random sample of 20% of crowns for each species (except where a species had just four individuals or less, in which

case a single test crown was randomly selected) was separated and withheld until the final performance evaluations.

Which machine learning classifier has the greatest predictive power? Classifier models were trained using the hyperspectral data extracted from the delineated tree crowns. In line with commonly practised methodologies, we evaluated several algorithms such as Multi-Layer Perceptrons (MLPs), Linear Discriminant Analysis (LDA), Random Forest (RF), Linear SVM, k-Nearest Neighbours and Logistic Regression due to their widespread implementations and adaptability. The hyperspectral pixel-data proved challenging to train on, having 378 spectral bands and spanning 169 species, 99 genera, and 41 families. The dataset exhibited significant imbalance; for instance, the most populous tree species boasted 55,448 pixels while the least populated class possessed just 223. Fig. S.5 plots the pixel count by species, illustrating the imbalance. Techniques exist for boosting the classification accuracy of less well represented classes (e.g. data resampling, class weight adjustments) but typically at the expense of overall classification accuracy. Rather than target the identification of rare species, we aimed to generate a map of the canopy that was comprehensive and representative of observed population species prevalence so we chose not to employ these methods. While classifiers were ultimately evaluated at crown level via pixel-wise majority voting, the scarcity of crown labels required training at the pixel level. Such an approach, however, encounters pitfalls: adjacent pixels may exhibit local effects that risk leakage between training and validation sets when naive cross validation is applied. An uninformed species-stratified split at the pixel level yielded classifiers with high pixel-level accuracy but low crown-level accuracy on unseen data. To counter this, a stratified group k-fold (k=5) cross-validation strategy was adopted. Here, pixels were grouped by crown (each pixel was associated with a crown ID) and stratified by training target, ensuring that (for any fold) pixels within any singular crown only appeared in either the validation or training set, avoiding the impression of high performance achieved through fitting to local effects (i.e. not transferrable between crowns).

Further complicating matters, the data exhibits significant inter-crown and intra-species noise. Fig. S.6 depicts the spectral value distribution for the two largest tree crowns of species *Pradosia cochlearia*, illustrating noteworthy variance. Meanwhile, Fig. S.7 shows the standard deviation across each spectral band of the top ten species (by pixel frequency) in the training dataset, alongside the standard deviation spanning all pixels from those species. Note that these spectra underwent standardized scaling, ensuring unit variance across all pixels.

After cross-validation, models were trained on the full training set and used to predict species classes across the entire landscape. Crown-level classifications were determined by majority vote of pixels comprising the crown. A more complex approach that averaged pixel wise class predictions was tested but this gave almost identical results to the simpler majority-vote approach and so was not taken further. Final performance evaluations were conducted using the held-out test set (as described above), using the (weighted average) F_1-score for model performance comparisons. Additionally, weighted average precision, recall, and the macro-average F_1-score were recorded, the latter assisting in gauging performance for less prevalent classes.

How does classification accuracy depend on the number of training crowns? To understand how many individual training crowns from a species were required to achieve a good classification accuracy we modelled the relationship between F_1-score and training crown number with a beta regression. As scores could take 0 or 1 values, this was done as a Bayesian zero-and-one-inflated beta regression with the *zoib-1.6* R package (F. Liu and Kong 2015; F. Liu and Li 2016). We assigned a threshold of F_1-score ≥ 0.7 for a 'good' performance. At this value, we could be confident that we correctly identify the majority of the individuals of a species present without assigning many trees of a different species a label of that species. This would be important in cases where tree species mapping is used to expand the sample size in a way that improves the signal-to-noise ratio to answer species specific ecological questions (e.g. monitoring rates of mortality and phenological patterns). The expected minimum number of training crowns required to achieve this performance was recorded.

Which wavebands are the most important for determining species? To determine which wavebands were most important for distinguishing species, we selected the best predictive model (based on the test set F_1-score) and extracted feature importance in a method suitable for that model type. In the case of LDA, this was done by extracting the scalings (or *loadings*) for each waveband across all linear discriminant functions. The scalings determine how each feature contributes to the axes that are constructed to maximise the separation between classes in the transformed space. They are useful for understanding which features contribute most to the separation between classes along each discriminant axis. A larger absolute value suggests a feature is important for separating the classes and a value near zero suggests that the feature does little to separate classes. The average absolute scaling value at each band across discriminants gave an overall importance value for classification, aggregating how important each feature is for separating the classes in the dataset.

To understand how the resultant waveband importances transferred to prediction accuracy we performed ablation experiments. Ablation involves systematically removing (ablating) parts of the model and observing how this affects its performance. We progressively removed the number of bands (in increments of 10 bands) available to the classifier by setting their standardised values to 0 prior to model training with four removal procedures. The first removed bands in a ranked order of importance, starting with the most important through to the least important. The second reversed the order (i.e. removing from least to most important). The third shuffled randomly 15 (k-means) clusters of band importance (retaining band order within the clusters and therefore a realistic data structure as feature importance tended to cluster in specific regions). The fourth, removed bands after a random shuffle, which meant all spectral regions were available to the classifier for longer than in the other methods.

Do spectra observed within tree crowns exhibit phylogenetic signal? To test for phylogenetic signal in the hyperspectral data, we calculated the mean and standard error of the standardised reflectance for the pixels belonging to each species, and, using the time calibrated phylogeny of Baraloto et al. (2012) and the *phytools* R package (Revell 2012),

calculated Pagel's λ (Pagel 1999), and Blomberg's K (and test) (Blomberg, Garland, and Ives 2003) on each of the bands of the hyperspectral data included in the species classification. Pagel's λ and Blomberg's K are statistical measures that quantify the degree of phylogenetic signal in continuous trait data. Pagel's λ measures the extent to which trait variation among species is correlated with their phylogenetic relationships, with λ scaling from no phylogenetic signal (0) to a signal that matches the expected pattern under a Brownian motion model of evolution (1). Blomberg's K quantifies how the observed trait variance among species deviates from what would be expected under a Brownian motion model of trait evolution. A K value greater than 1 suggests that closely related species are more similar than would be expected under a Brownian motion model, while a K value less than 1 suggests less similarity among related species than expected under a Brownian motion model. Of the 169 species tested in the classifier task, 153 were represented in the phylogeny and could be used to test for signal.

Are the most important wavebands for species classification those that have the strongest phylogenetic signal? We tested the correlation (Spearman's rank ρ) between the strength of the phylogenetic signal (λ and K) of each spectral band to the feature importance values of each band. This was to test whether those bands that were helpful in discriminating species also contained a stronger phylogenetic signal.

Are closely related species more often confused in their classification than distantly related species? To test whether more closely related species were more likely to be confused with one another than distantly related species we tested the correlation between pairwise phylogenetic distance (as Myr of independent evolution between pairs of plant taxa as provided in the time calibrated phylogeny of Baraloto et al. 2012) and pairwise mis-classification rate. This was done with the rank coefficient as the distributions were highly non-normal.

3. Results

3.1 Diversity of Paracou's canopy

We labelled with confidence crowns of 3,256 individuals to species level. The dataset comprised 239 unique species, 124 genera and 43 families. We removed 70 of these species from subsequent analyses because they were only encountered once in our survey, so could not be included in the classifier training and testing (see Fig. 3). The 70 species represented just 2% of the crowns we recorded. The filtering process left 169 species for further analysis.

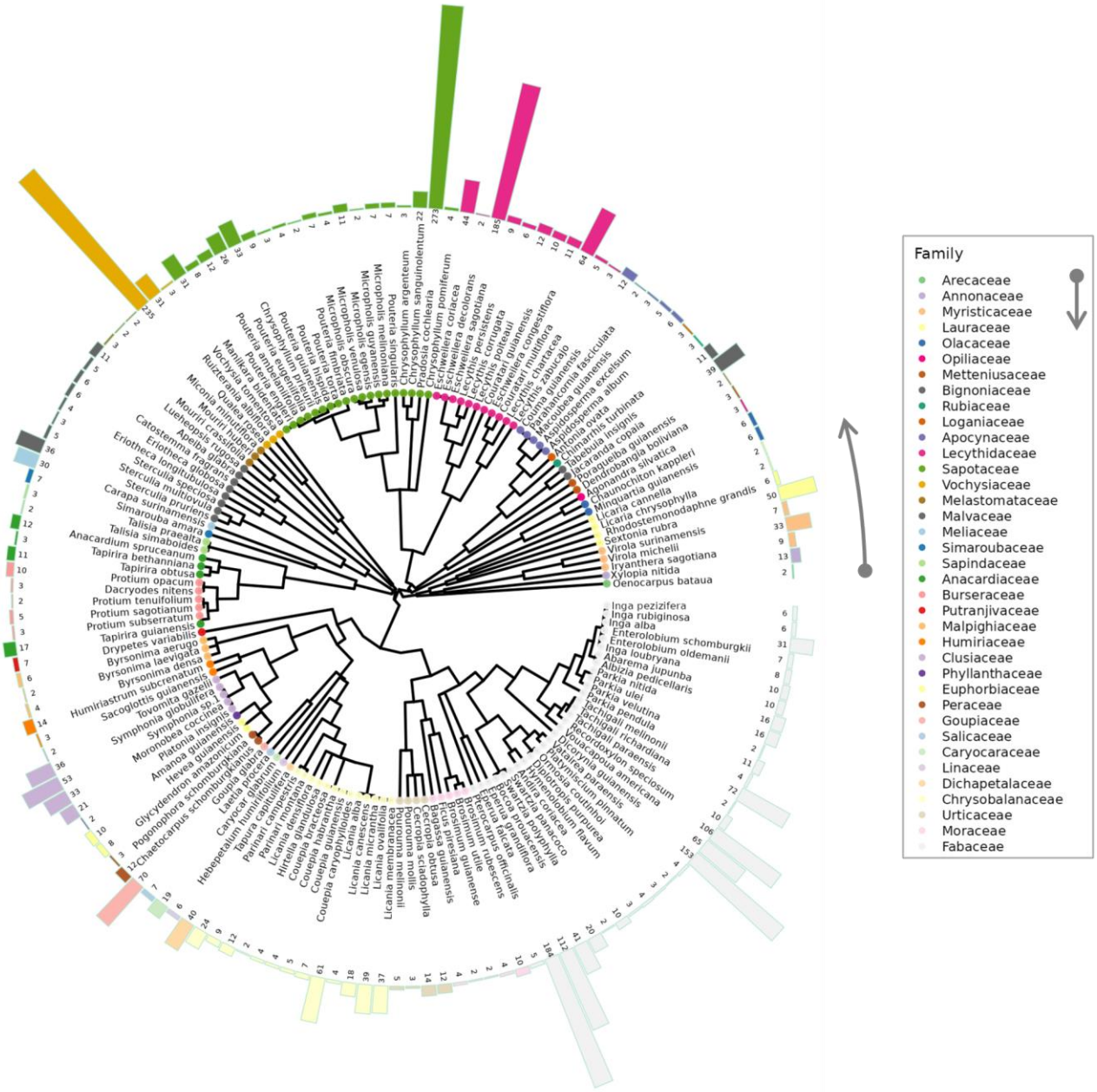

Figure 3*: Abundance of 153 species in the ground validated tree crown dataset mapped onto the phylogeny of Baraloto et al. (2012). Bars represent the total number of individuals sampled in the field. The legend is arranged in the order families appear on the tree (anti-clockwise from 3 o'clock). Species with just a sole representative crown record (70 in total) were not included in the classification task and are not presented here.*

3.2 Tree crown segmentation

Does the fusing of multi-date crown information improve segmentation accuracy?

The accuracy of the tree crown segmentation was improved by combining multiple dates of tree crown segmentation predictions and retaining crowns that had good confidence and agreement between dates (see Fig. 4). We compared the accuracy of segmentation from using a single time step and multiple time steps with an unseen test set of 169 test crowns across two spatially separate test zones. The accuracy of delineations increased as more dates were combined. The best performing model overall was the one trained on five dates' worth of imagery. The accuracy of its delineation was boosted significantly by combining information from different time steps from mean F_1-score of 0.68 for a single date prediction, to a peak at the combination of nine time steps with a mean F_1-score of 0.78. In terms of total crown area, approximately 86% of the test region had well located and delineated crowns (see Fig. 5). Accuracy tended to increase with tree crown area.

The 'base' model and the model trained on additional single date of UAV data had a comparable performance at a single date prediction and combinations of less than seven dates (see Fig. 4). However, after combining seven dates, the combined delineations of the 'base' model became substantially better, surpassing the single time step prediction accuracy of the model trained on five dates. The model trained on a single time step failed to improve in accuracy by combining dates to the same degree as the other two models, suggesting it became overfitted to the limited available data. The 'base' model was trained on a range of non-UAV RGB imagery (see Ball et al. 2023) without the additional focused training on UAV data; this seemingly led to better temporal transferability than was achieved with the additional focused training on just a single date.

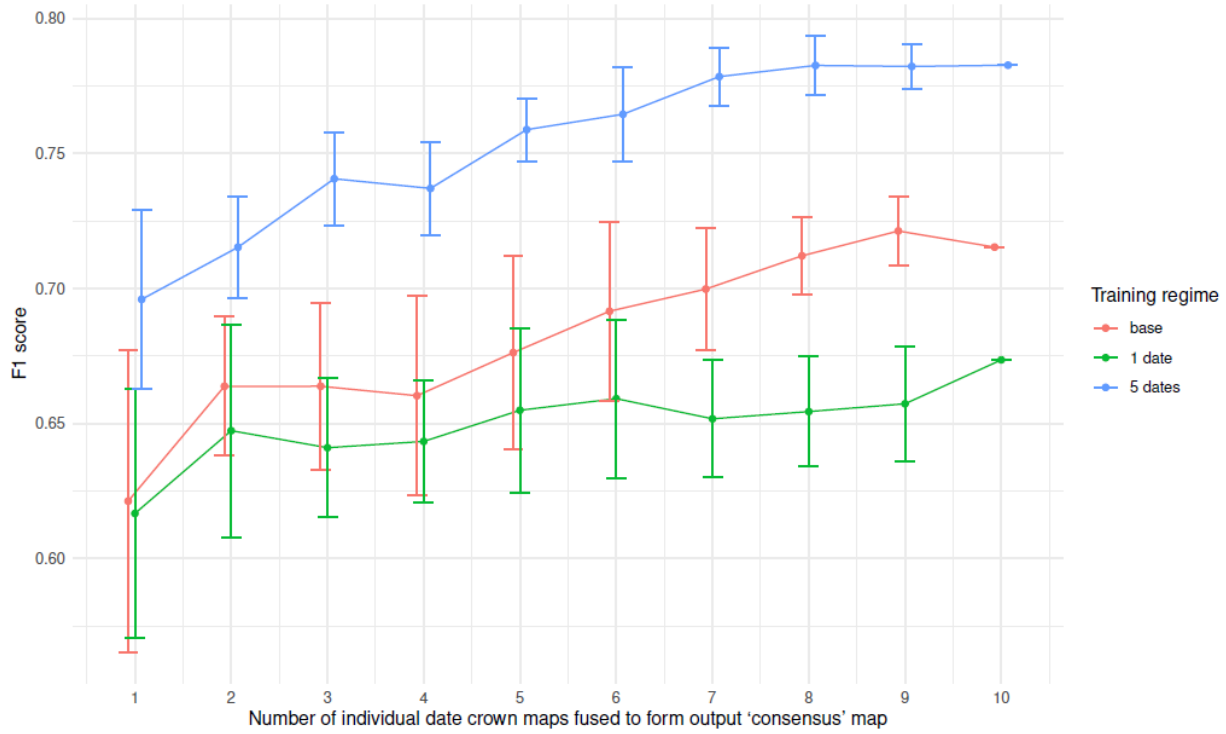

Figure 4*: The performance of the tree crown delineations with the number of individual date crown maps that were combined to form the output 'consensus' map. Researchers may or may not have training crowns available and a varying number of RGB surveys. To reflect this we tested three models trained under different data regimes: (1) the 'base' which was freely available online (trained on different crowns and imagery) (2) the '1 date model' - the base model then trained on manual crowns with a single date of RGB imagery; (3) the '5 date model' - the base model then trained on manual crowns with five UAV-RGB surveys of the same location. The mean and standard deviation of the F_1-score was calculated by taking all possible date combinations. Note that only a single combination of dates was possible for the 10-date combination.*

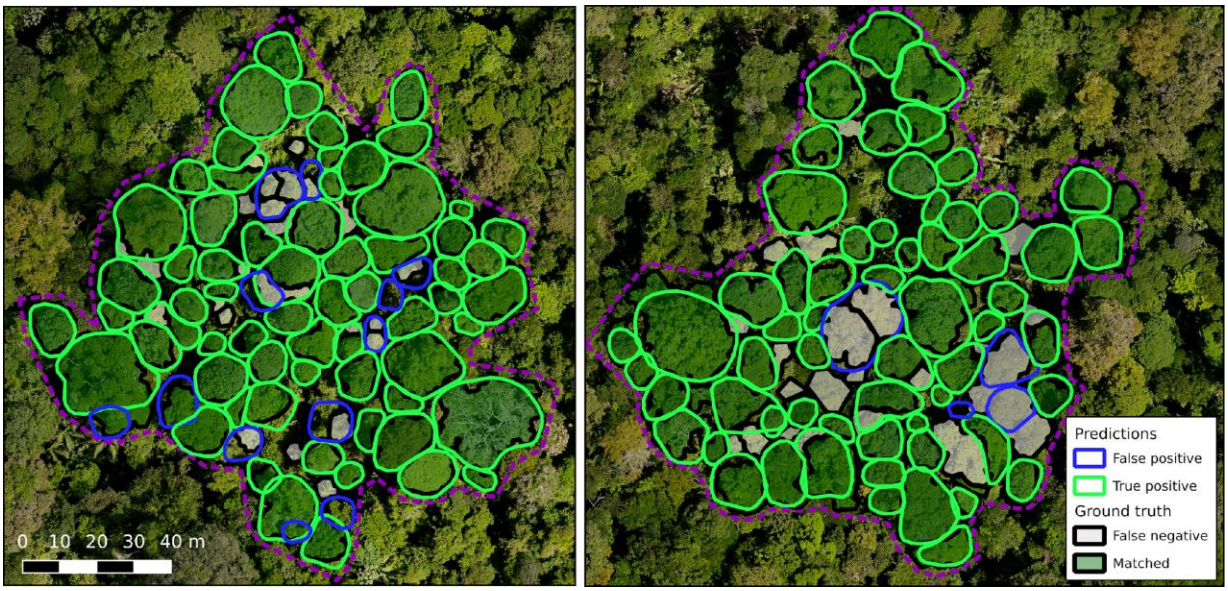

Figure 5*: Predictions and ground truth crowns in the unseen test regions. The reference set of crowns are shown with black border and their fill colour depends on whether a crown was matched (green) or unmatched (grey) with a prediction crown. The predictions that match with a reference crown have a green border and those that did not match a reference crown have a blue border. A match was granted in the case that a reference crown and a prediction crown had* $IoU \geq 0.5$*.*

3.3 Tree species classification

Which type of machine learning classifier can most accurately predict the species of tree crowns from hyperspectral data? The LDA classifier performed best at classifying the species of the test set of tree crowns from their hyperspectral signal (weighted average F_1- score = 0.75; see Table 2). Of the 169 species included, 65 of the species (38%) were classified with an F_1-score of over 0.7. The logistic regression model had a slightly lower overall performance than the LDA but with a substantially lower macro-average F_1-score suggesting it struggled with less well represented classes. The more flexible (and expensive to train) MLP and SVM classifiers failed to match the performance of the LDA classifier, highlighting that the LDA's approach to separating the classes led to more robust transferability between crowns. Furthermore, the LDA classifier was also far quicker taking 20 seconds to train whereas the SVM took more than 7 hours.

As a percentage of total crown area of the test set, about 81% was assigned with the correct species. By combining the percentage of the total crown area that was well located and delineated (86%) and the percentage of the total crown area that was assigned the correct species (81%), we estimate conservatively (given that incorrect areas of each step are more likely to coincide than not) that 70% of the landscape's crown area was mapped correctly. For reference, the crowns of the top 20 most abundant species make up less than 60% of the total crown area of the reference dataset.

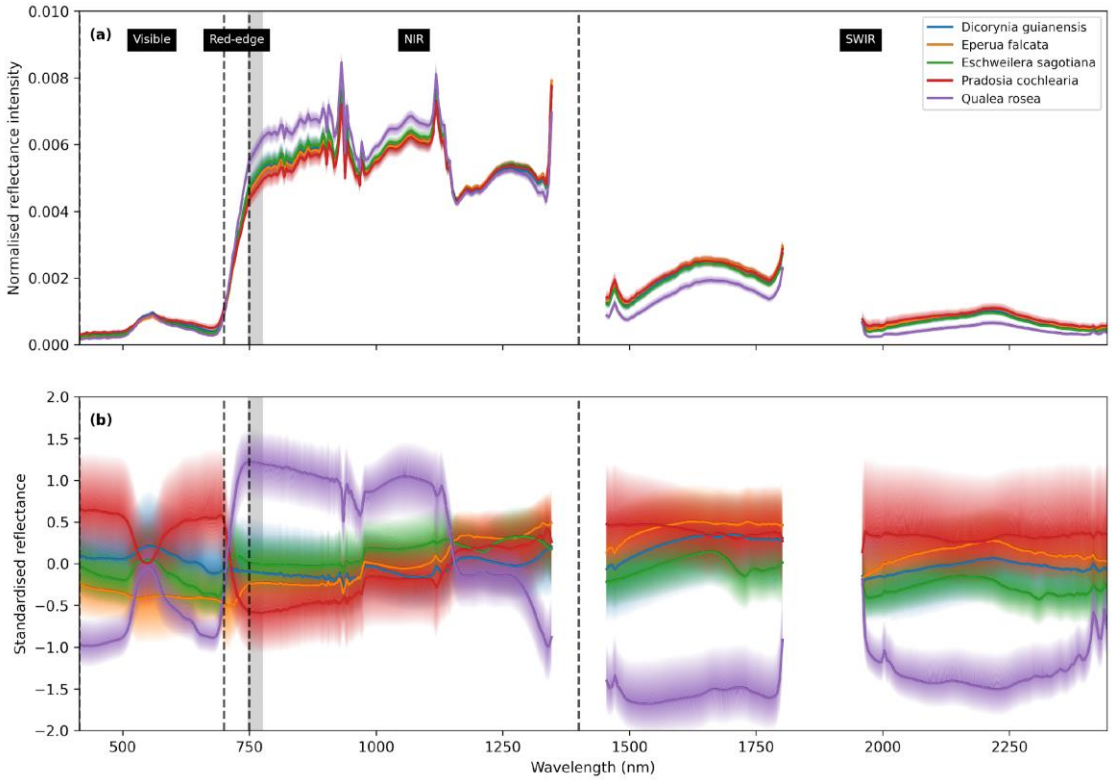

Figure 6*: The observed spectrum of five common species which together comprise 31% of the total crowns in this study. The median line of pixel values is plotted and the IQR is shaded to show spread. (a) Shows the reflectance spectrum where pixels have been normalised by dividing the reflectance intensity by the summed reflectance over all bands. (b) The mean value is subtracted from the standardised reflectance, and the resulting value is divided by the standard deviation for each band across valid tree crown pixels, so that all bands are shown on the same scale. The plots illustrate how species might be identified from spectral information (spectral signatures). The two gaps in the SWIR region are the result of removing bands influenced by air humidity. A grey band at 748 nm to 775 nm shows the spectral region in which bands were most important for classification.*

How many mapped individuals of a given species are needed to achieve a 'good' classification accuracy? We distinguished 65 species with an F_1-score of at least 0.7, which is a far more species than previous studies which are typically limited to no more than 20 species (Féret and Asner 2013; Laybros et al. 2019, 2020). We were particularly successful at distinguishing the commoner species. The accuracy of classification increased with the number of training crowns in the class (see Fig. 7) in agreement with similar patterns observed by Baldeck and Asner (2014) and Féret and Asner (2013). There was a sharp increase between 1 and 10 training crowns, after which the performance improved more gradually. Species with at least 8 training crowns could reasonably be expected to be classified with an F_1-score of 0.7.

Table 2*. The accuracy statistics for the classification models based on the unseen test set of crowns (169 species). The weightings for the (weighted average) precision, recall and F1-score were adjusted to reflect species occurrence in the crown dataset prior to the removal of species with a single representative and the train-test split thereby correcting any potential inflation in performance from dropping these species. Train time is the amount of time it took to train the final tuned model of the model class on a 128 core 2 x AMD EPYC 9534 with 1.5TB of ram and NVIDIA A30 GPU.*

Classifier	Precision	Recall	Weighted F_1-score	Macro-av. F_1-score	Classification accuracy	Train time
LDA	0.74	0.79	0.75	0.49	0.78	20s
MLP	0.65	0.74	0.68	0.37	0.72	2m26s
Logistic	0.66	0.74	0.68	0.35	0.72	45m20s
LinearSVM	0.64	0.71	0.65	0.33	0.70	7h11m6s
QDA	0.52	0.58	0.53	0.21	0.57	14s
RdmForest	0.41	0.43	0.35	0.11	0.42	41s
kNN	0.35	0.39	0.31	0.08	0.38	3s

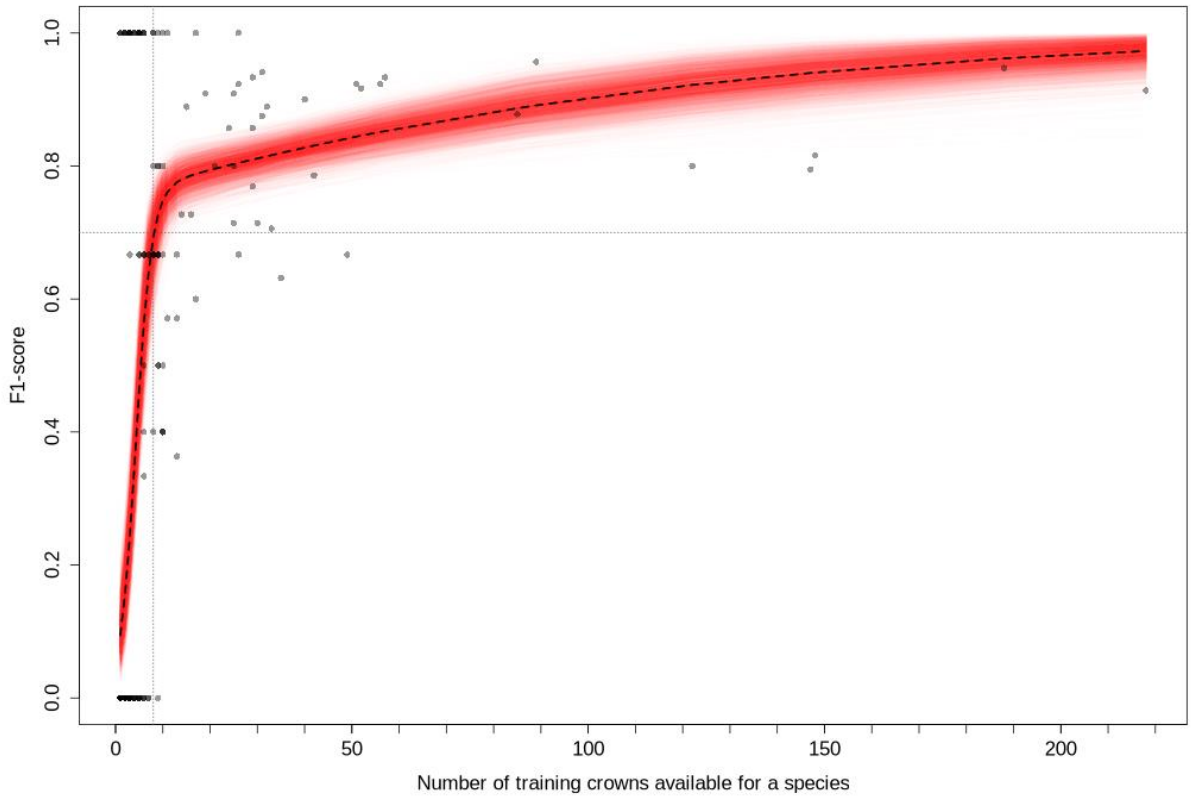

Figure 7*: Classifier performance for individual classes (species) in relation to the number of training crowns in the class. Each dot represents the F_1-score for the classification of an individual species. A zero-and-one inflated beta regression was performed to find the expected F_1-score by number of classes (the black dashed line). Dots are slightly transparent so where they appear darker there are several overlapping species with that score. This is common with low numbers of training crowns as there are fewer discrete scores that can be attained. The dotted line at x=8 shows the minimum number of training crowns required before an expected classification performance of F_1-score = 0.7 is acquired for a species.*

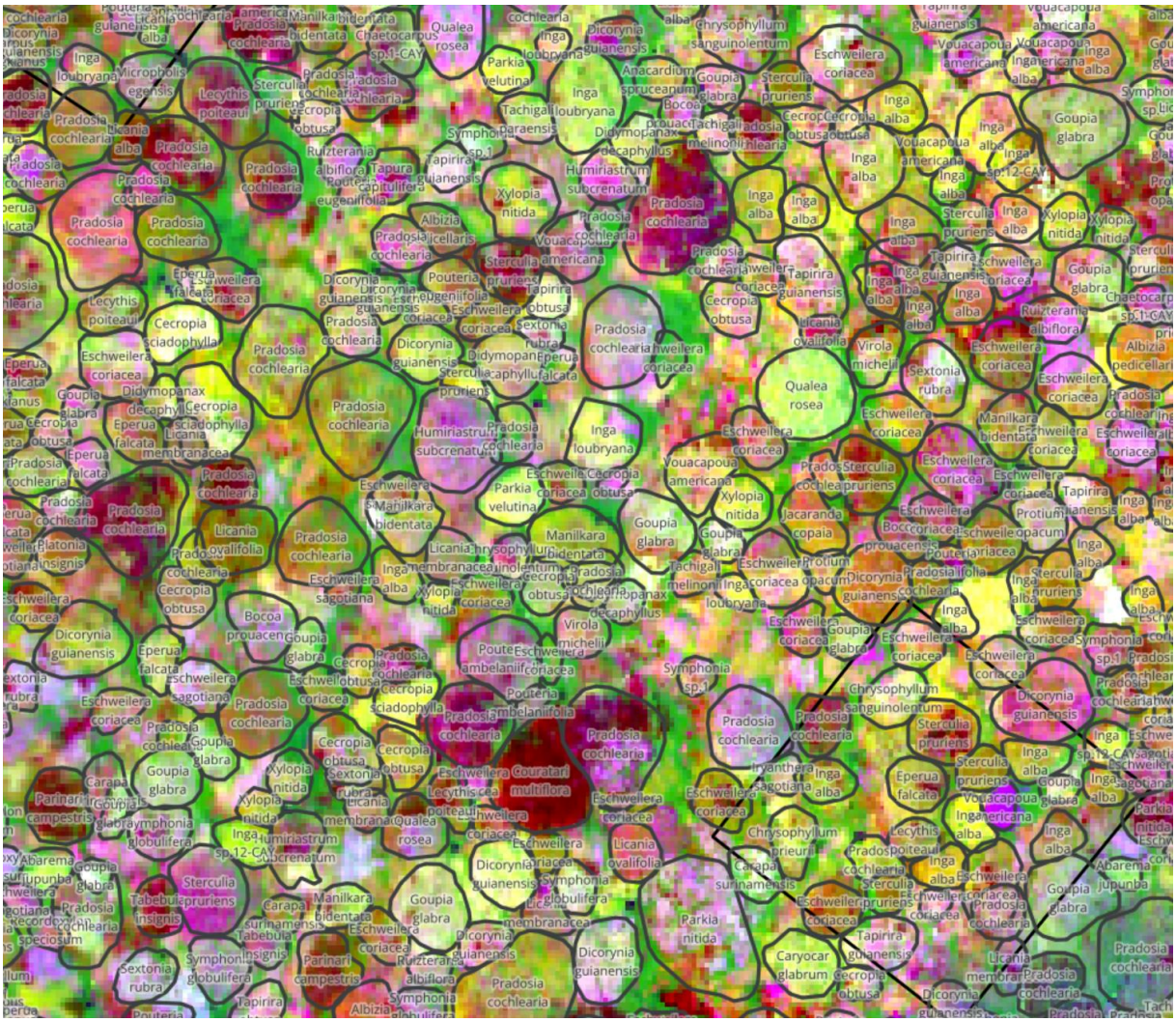

Figure 8*: A portion of the finalised crown map overlaid on three selected bands of the PCA projection of the HSI. The PCA projection of the HSI is purely a means to visualise the hyperspectral data and was not used at any stage of the analysis. Crowns identified are outlined in black and labelled with predicted species. The black squares are forest plots.*

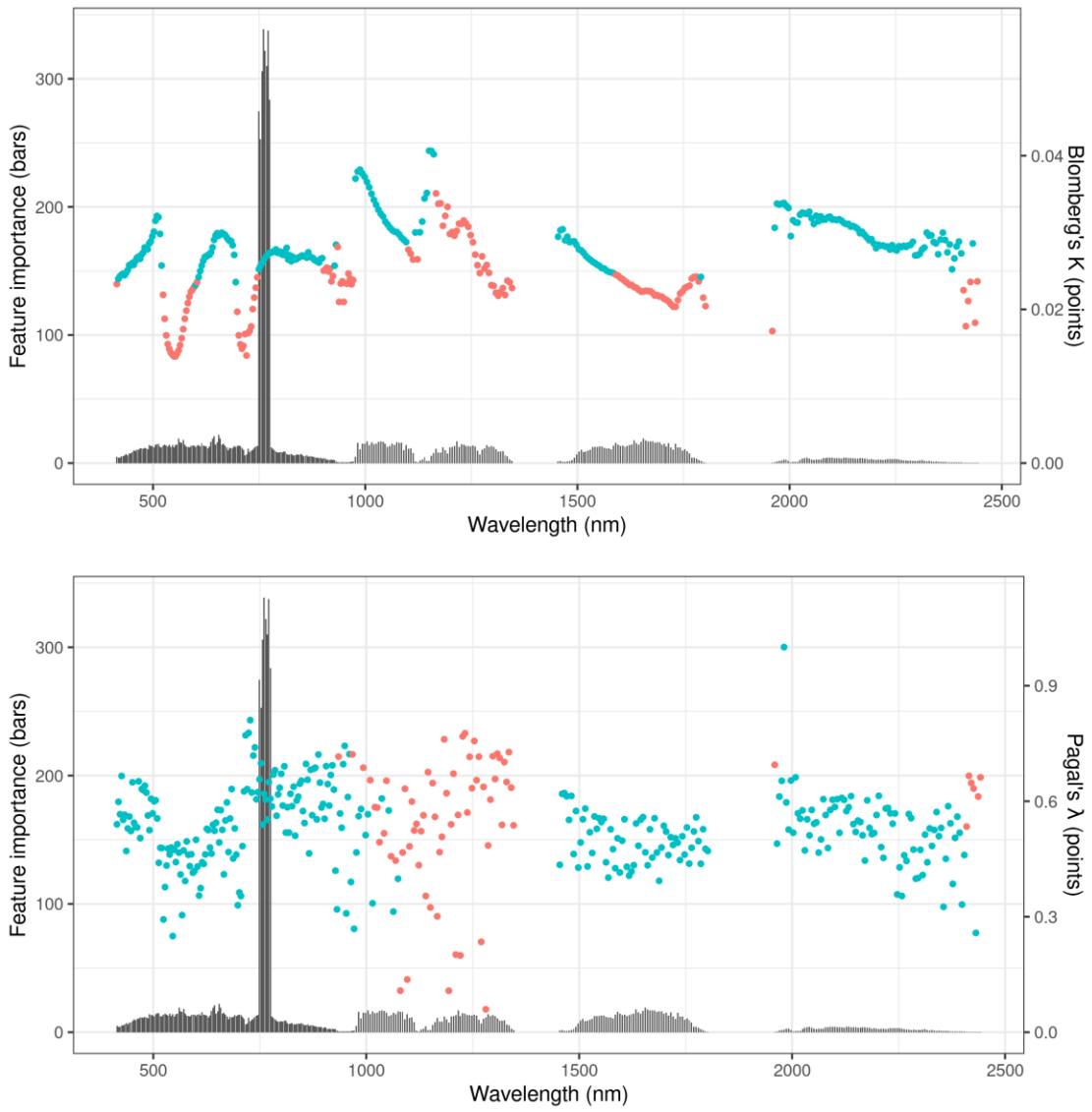

Figure 9*: The relative feature importance (bars) and phylogenetic signal (points) of each band used in the classifications. The blue circles indicate where the phylogenetic signal is statistically significant (at* $p \leq 0.05$ *) while orange circles show values that are not statistically significant.*

Which wavebands are important for determining species? Eight bands between 748 and 775 nm, on the edge of the "red edge" transition between the red and near-infrared ranges, dominated in terms of relative feature importance for separating species (see Fig. 9). The next most important region was 640-660 nm which fell in the red region of the visible spectrum. The 560-575 nm (green), 1630-1680 nm (SWIR) and 1000-1100 nm

(NIR) regions were also relatively important for discriminating between the species. While the importance peak at 748-775 nm appeared stark, ablation tests demonstrated that removal of the ten most important bands did not lead to an immediate, dramatic drop in classifier performance with the removal compensated for elsewhere in the spectrum. However, progressive band ablation showed that removal of the most important bands did lead to a more substantial drop in performance than by removing bands after shuffling the clusters of band importance and removing bands in reverse feature importance order (see Fig. S.4). Removing bands in a randomised order maintained classifier performance for the longest as it left all spectral regions available until later in the ablation process (and there was shared information between adjacent bands).

Do spectra observed within tree crowns exhibit phylogenetic signal? Most standardised and normalised bands showed a statistically significant phylogenetic signal ($p \leq 0.05$; see Fig. 9). Out of a total of 378, Pagel's λ test gave 310 significant bands while Blomberg's K test gave 239 (whereas one would expect 19 significant bands by random chance). This indicated that closely related species were more likely to have similar reflectance values at spectral bands than distantly related species along most of the spectrum included in this analysis.

Are the most important wavebands for species classification those that have the strongest phylogenetic signal? From inspecting Fig. 9, there was no obvious relationship between feature importance and phylogenetic signal of the bands. However, both λ ($\rho = -0.14, p = 0.005$) and K ($\rho = -0.16, p = 0.001$) had a statistically significant negative correlation with feature importance suggesting that the more phylogenetic signal presented by a band the less helpful it was to discriminate between species.

Are closely related species more often confused in their classification than distantly related species? By comparing phylogenetic distances between pairs of species with their pairwise mis-classification rates, we determined that the more closely related the species the more likely they were to be confused with one another in the classification ($\rho =$

$-0.0437, p = 0.0008$). The confusion matrix of the species is given in Fig. 10. Note where confusion between species occurs within genera.

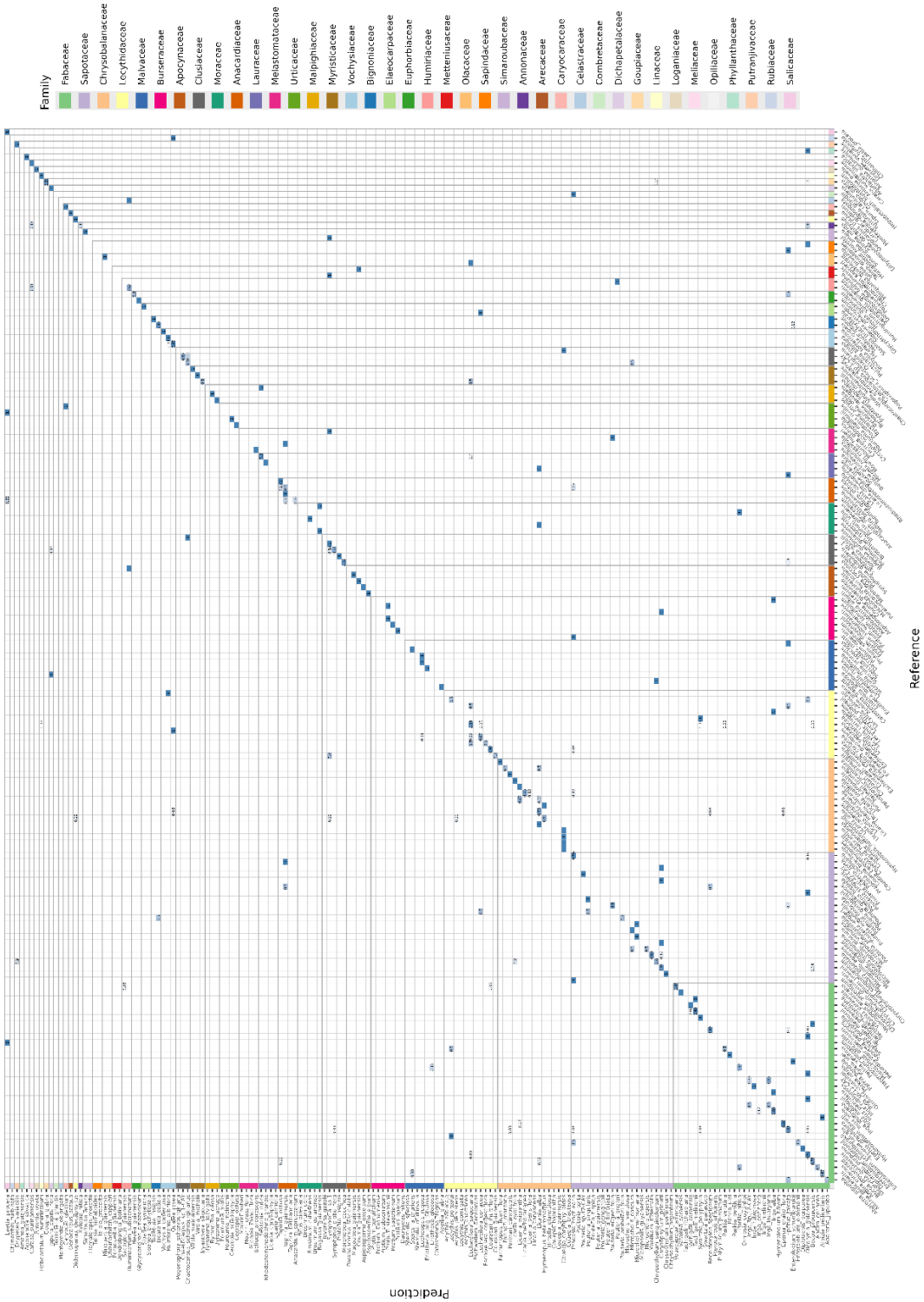

Figure 10*: Normalised confusion matrix for predictions on the test set crowns. The families are ordered by the number of species included in the study that they contain. Deeper blue values show higher proportions (closer to 1) and the faint colours show proportions closer to 0.*

4. Discussion

Mapping tropical forest canopy species is essential for gaining a granular understanding of large-scale ecological processes. We have extended the scope and reliability of species identification from aerial hyperspectral data in diverse tropical forests, surpassing prior levels of identification accuracy for a far greater variety of species. Previous studies have managed to map around 20 species with accuracy (Féret and Asner 2013; Laybros et al. 2019; Garzon-Lopez and Lasso 2020) with greater scope in diverse tropical forest species classification only coming from direct leaf spectroscopy (Prospere, McLaren, and Wilson 2014; Harrison, Rivard, and Sánchez-Azofeifa 2018). Since tropical forests typically contain several hundred species per hectare (Lee et al. 2002; Duque et al. 2017), of which around 30%-60% make it to the canopy (Bohlman 2015), many of the crowns have been left unidentified. By extending this number to aim to map comprehensively (178 species with 64 species $F_1 > 0.7$) we get much closer to complete landscape mapping of upper canopy trees. In our study, the top 20 species covered less than 60% of the total crown area. Assuming perfect segmentation and classification accuracy, this would be the upper limit of accuracy (by area) that could be achieved at a landscape scale if only those 20 species were included in the mapping process. Assuming realistic classification and segmentation accuracies of 80% (Laybros et al. 2020; Aubry-Kientz et al. 2021) this drops to less than 40% correct crown area mapping at the landscape-scale. In comparison, by expanding the pool of species (to 169), we were able to achieve an F1-score > 0.7 in 65 species and accurately map over 70% of the total crown area of the landscape. This more complete coverage is not the result of more sophisticated species classification algorithms, as we found that the well-established LDA approach transferred better between crowns than more flexible methods (SVM, MLP; in agreement with Féret and Asner, 2013). Instead, improvements were achieved by (1) creating a large reference database of labelled geo-located crowns; (2) improving training/testing datasets by careful mapping and identification of trees in the field; (3) high-quality hyperspectral imagery that accurately co-aligned with tree crown maps, allowing species to be distinguished from upper red edge bands; (4) improved methods for segmenting tree crowns, which is vital for mapping

across landscapes. We discuss these improvements, the fundamental basis for species identification based on knowledge of the evolution of plant physiology and morphology, then consider the obstacles to achieving affordable, easy-to-use and transferable approaches to tree species identification.

Abundant high quality field data. The field dataset was developed, curated and ground validated over a number of years. The careful mapping of what was observed from above to the reality on the ground was labour intensive but the resulting database of 3500 crowns provided a robust basis upon which this study could be built. Without substantial, high quality ground datasets, remote sensing is limited in the inferences it is able to make (Chave et al. 2019; Davies et al. 2021). Despite this, there are few studies that provide benchmark data upon which tropical tree species mapping approaches can be tested (Laliberté, Schweiger, and Legendre 2020).

The importance of high-quality hyperspectral imagery. Aerial hyperspectral data captures spectral intensities across hundreds of contiguous, narrow wavelengths, allowing signals of biochemical and morphological properties in foliage, such as chlorophyll, water content and leaf structure to be observed (Clark and Roberts 2012). This depth of information is indispensable for the remote identification of species, particularly in diverse tropical forests where conventional, broader band multispectral imagery falls short (Zhang et al. 2006). The hyperspectral images used in this study were collected close to nadir, ensuring minimal distortion and noise. Additionally, these images were co-aligned with ground-truthed tree crown maps, thereby significantly enhancing the reliability and precision of species identification. Through this approach we identified the narrow upper red edge band range of 748 to 775 nm as the most important region for discriminating species (Fig. 9) but that it is valuable to retain information across a broad spectrum (Fig. S.4).

Improved segmentation of tree crowns. The emergence of CNN methods has allowed for considerable progress in automatic processing of images across a range of fields. By integrating geospatial features with the sophisticated Mask R-CNN architecture (He et al.

2017), the *detectree2* Python package (Ball et al. 2023) can harness subtle spectral and textural clues to delineate trees precisely achieving state-of-the-art performance on tree detection in aerial RGB data (Gan, Wang, and Iio 2023). Even humans struggle to agree where trees are located when looking at the same data layers (Section S.1). By allowing a machine to learn on a carefully validated manual (field verified) dataset, it has synthesised past human attempts to produce quicker and more consistent predictions. Accuracy of segmentation can be significantly increased by combining maps of segmented tree crowns over time. However, forest canopies can appear very different across dates due to atmospheric perturbations, differences in illumination (resulting in variably shadowed trees), the sway of trees and branches, phenological changes, death of individual trees and branches and irregularities in the orthomosaicking process. UAV RGB sensors are cheap and provide a source of repeated, high-resolution scans over the course of a few months (during which there are few mortality events and little growth). We found that, by combining predictions across dates, a model that had never been exposed to the specification of imagery that it was predicting on (in this case UAV RGB) in training could reach a comparable accuracy to a model trained on a high volume of the specific imagery (Fig. 4). There is no substitute for gathering high quality training data but, if this is not available, repeat predictions with a pre-trained, freely available model can give excellent tree crown delineations on a new site. This can support the establishment of new studies to track tree growth, mortality and phenology over large areas. This was the first time that information across dates has been combined to improve the accuracy of tree crown delineations.

Towards a spectral-taxonomic understanding of species differentiation

We found upper red edge bands in the 748 to 775 nm range were the most important bands for discriminating species (Fig. 6), a finding not picked up in Laybros et al. (2019) as predictors were not standardised prior to model training which is an important step if model scalings are to translate into feature importance. However, it aligned with Badourdine et al. (2023)'s assessment of band importance for assessing canopy taxonomic diversity at the same site. This pattern has not been picked up explicitly in other

773 previously published studies but is not precluded by them as importance has generally
774 been assessed over broader regions of the spectrum (with red edge to NIR generally being
775 considered important) (Clark, Roberts, and Clark 2005; Dalponte, Bruzzone, and Gianelle
776 2012; Marconi et al. 2022; Fassnacht et al. 2016; Hennessy, Clarke, and Lewis 2020). Leaf
777 traits are incorporated into spectra in complex ways (Jacquemoud and Baret 1990; Féret
778 and Asner 2011) and multiple traits can superimpose in a given spectral region (Curran
779 1989). The *red edge*, defined loosely as the 700-750 nm region, is widely recognised as an
780 important region for classifying vegetation and is linked to chlorophyll content, leaf area
781 density, water content and overall plant health (Thomas and Gausman 1977; Horler,
782 Dockray, and Barber 1983; Filella and Penuelas 1994; Boochs et al. 1990; Gitelson, Gritz,
783 and Merzlyak 2003; Hennessy, Clarke, and Lewis 2020). However, the red edge is usually
784 defined as the 700-750 nm range. Here we found that the region immediately beyond the
785 red edge (748-775 nm) was particularly sensitive to differences among species in a tropical
786 forest. The "upper red edge" is at the transition zone between chlorophyll absorption (in
787 the red) and cellular structure scattering (in the NIR), potentially capturing information
788 from both the biochemical and structural aspects of the vegetation. The wavelengths
789 adjacent to this range are not showing strong feature importance, which means the unique
790 reflectance in the 748-775 nm range could be capturing some species-specific anatomical
791 features, possibly related to internal leaf structure affecting scattering of near-infrared
792 light (Ustin et al. 2009).

793 Other wavelengths were also of secondary importance. In the visible range, red
794 wavelengths of 640-660 nm were relatively important: Chlorophyll a and b have peak
795 absorbance at different wavelengths of red light (660-680 nm vs 640-660 nm respectively),
796 suggesting that variation in Chlorophyll-b may be important for discriminating
797 species (Gitelson, Gritz, and Merzlyak 2003). Bands within the green region of 560-575 nm
798 showed some importance which could be due to chlorophyll reflectance
799 differences (Gitelson, Gritz, and Merzlyak 2003) but could equally be due to leaf structure,
800 carotenoids and anthocyanins (Sims and Gamon 2002), or even stress (Carter and Knapp
801 2001). Water has strong absorption features in the SWIR region. The relative importance of

bands in the 1630-1700 nm region suggests detection of variation in leaf water content (Ceccato et al. 2001; Gao 1996) and/or differences in cellulose and lignin composition (Kokaly et al. 2009; Serrano, Peñuelas, and Ustin 2002).

We found "crown reflectance spectra" (i.e. spectra influenced by leaf reflectance spectra plus influences of absorption, reflectance and transmission by leaves in a multi-layered tree crown) showed phylogenetic structure. Madritch et al. (2014) showed the capacity for aerial hyperspectral imagery to characterise genotypic identity while Schweiger et al. (2021) showed a correlation between leaf level spectral dissimilarity with phylogenetic distance (Schweiger et al. 2021). Other studies have shown phylogenetic structure of foliar spectral traits (Cavender-Bares et al. 2016; Meireles et al. 2020) including for leaves in tropical forest canopies (McManus et al. 2016). For the first time, we have linked crown reflectance spectra to the phylogenetic signal to help explain species classification from aerial hyperspectral data. Most bands exhibited some phylogenetic signal but the importance for classification was negatively correlated to the signal. This could be for a number of reasons including spectral overlap between closely related species, the importance of ecological, environmental and stress factors, and convergent adaptive traits that occur broadly across the phylogeny. Looking within lineages instead of across the whole phylogeny may be a way to probe this relationship further (Meireles et al. 2020).

Traits may vary in their degree of phylogenetic conservation depending on a variety of factors including environmental pressures, mutation rates, and the evolutionary history of the species in question. Of those traits discussed above, some might be expected to be more conserved than others. Polymers including lignin and cellulose are critical for plant structure, and their relative concentrations are generally highly conserved within lineages (Weng and Chapple 2010). Additionally, leaf structure, including traits such as leaf thickness or specific leaf area, is conserved to some extent within phylogenetic lineages (Ackerly and Donoghue 1998). Other traits may be more labile and lack phylogenetic conservatism. For instance, while the ability to retain water might be conserved within specific lineages adapted to particular environments, there can be significant variability in this trait both within and between species based on immediate

environmental conditions (Donovan et al. 2011). Other traits are highly plastic and have weak phylogenetic signals. Although chlorophyll is essential for photosynthesis in all plants, the specific amount and ratio between Chla and Chlb concentration can vary greatly even within a single species based on a variety of factors, including age, health, and immediate environmental conditions like light and nutrient availability (Gitelson, Gritz, and Merzlyak 2003). Stress responses are highly variable traits that can differ significantly even within a species based on environmental pressures and are likely to be among the least conserved traits phylogenetically. Furthermore, the classifier may be getting clues from soil and other external factors that the plant interacts with, rather than intrinsic traits of the plant. With this in mind, it may be that the classifier is basing its decisions more on how species are responding to environmental conditions rather than on their intrinsic biophysical properties. Analyses of functional traits variation (e.g. Asner et al. 2014; Asner 2014; Schmitt et al. 2022) may hold the key to understanding the extent to which hyperspectral sensing is detecting interspecific vs intraspecific variation in biophysical traits (e.g. see Nunes, Davey, and Coomes 2017).

Our species-focused approach differs from biodiversity metric approach adopted by many studies, which focuses on mapping taxonomic diversity using spatial variance in the hyperspectral signal (Laliberté, Schweiger, and Legendre 2020; Vaglio Laurin et al. 2014; Jucker et al. 2018; Kamoske et al. 2022). Mapping diversity in this way is challenging because of uneven spectral distances among species (e.g. because related species have more similar spectra), and the large variance in spectral properties of single species when compared across landscapes (Badourdine et al. 2023).

Future work

Improvements in classification. Accurate classification of species from hyperspectral data required learning the spectral features exhibited across crowns of the same species. To close the gap between realised and intrinsic predictability of species classification we need to understand the sources of error. Simple mis-classifications due to the misalignment of data sources or mistakes in the field labelling, are trivial to address. External sources of

signal variation (e.g. atmospheric disturbance, sun-sensor geometry) may be addressed with improved data processing based on physical models. Variation of crown spectra within species driven by biological factors such as water stress, or phenology may be addressed with more informed feature selection/engineering and more sophisticated classification approaches.

The more flexible models, including MLP and SVM were better able to fit the local features of the pixels within crowns, but this failed to transfer to performance across crowns where the simpler LDA did best. This highlights the challenge in applying cutting-edge machine learning methods to this task. More work needs to be done on understanding how best to constrain the more advanced methods so they can learn to encode the key crown-level features from limited training data.

Addressing transferability. The ultimate aim is to train classifiers that can accurately predict the species of tree crowns when transferred in time (e.g. different seasons) and space (e.g. new forest locations). While we were able to identify a wide range of species at a single date and location from hyperspectral data, we have not demonstrated that it is possible to transfer this to other locations. It is known that classification degrades with time between training and prediction scans (Laybros et al. 2019) and if applied to new regions.

Trees exhibit temporal variability in their spectral signatures due to seasonal phenological changes (Hesketh and Sánchez-Azofeifa 2012; Chen et al. 2022) and external stressors like pests or drought, complicating year-to-year or season-to-season species identification. Even at a fixed point in time, individuals of the same species can be at different phenological stages (Reich 1995) and can have different spectral signatures based on geographical factors like soil type, local climate and topographic position (see e.g. Fig. S.6). Atmospheric conditions, from clouds, gases constitution, to airborne particulates, further modify the spectral data acquired by airborne sensors (Theiler et al. 2019; Arroyo-Mora et al. 2021; Schläpfer, Hueni, and Richter 2018). This problem is exacerbated by variations in atmospheric conditions between airborne sensor and the Earth's surface, which can modify

the spectral signatures, making hyperspectral data from different times and locations difficult to compare directly (Theiler et al. 2019; Prieur et al. 2024, manuscript submitted for publication). Additionally, spectral responses can differ between sensors, and even the same sensor can vary over time due to calibrations or degradations (Baumgartner et al. 2012), complicating data comparison. Finally, the spectral data acquired can be influenced by the illumination (Theiler et al. 2019; Arroyo-Mora et al. 2021; Schläpfer, Hueni, and Richter 2018) and viewing geometry (Theiler et al. 2019; Schläpfer and Richter 2014; Montes and Ureña 2012; Duthoit et al. 2008; Schläpfer, Richter, and Feingersh 2015; Lyapustin et al. 2012), including the angle of sunlight and sensor viewing angle, introducing additional variability across space and time and necessitating more sophisticated data handling and analysis approaches.

More work needs to be done on collecting hyperspectral data (and labelled tree crowns) at different dates with a broader range of locations but also across a wider range of acquisition conditions. With this it may be possible to identify features/encodings that remain stable through time and space so that models may be flexible enough to be applied broadly across an ecosystem. Most features could have some innate biochemical explanation, or have an explanation more indicative of different levels of stress among different species. Indeed, there are likely many interacting effects between bands that we have analysed here. However, it is also possible that machine learning models applied to real data may find importance in specific spectral ranges due to noise or collinearity in the data that may not have an easy biological explanation. Without further analyses, explanations are speculative. Confirming the reason for the peaks in feature importance would likely require controlled studies involving leaf-level spectroscopy, coupled with biochemical assays to identify the specific compounds or structure responsible for these spectral features.

Conclusions

Three key ingredients were required to achieve the accuracy and comprehensiveness of the mapping presented here: images collected across several dates providing alternate views of

trees within a dense, complex canopy; high-quality field data that matches botanical and structural information against remotely sensed spectral information; a precise, automated segmentation approach capable of imitating human perceptiveness while drawing on rich information across different dates. With this detailed mapping, we began to probe the relationship between evolutionary histories and spectral separability of species. Mapping species in this detail, across large forest extents, opens opportunities for addressing persistent ecological questions around the current and future states of tropical forests. With the advent of freely available, high spectral (e.g. EnMAP, CHIME, HYPXIM) and spatial resolution satellite imagery imagery (e.g. PlanetLabs), the local approaches presented in this study can give a foundation from which methods for pan-tropical analyses can be developed. We move towards being able to monitor entire biomes down to the granularity of individual trees (see e.g. Reiner et al. 2023), potentially ushering in a new era for studying forest dynamics, including tree growth and mortality, biodiversity, and phenology.

ACKNOWLEDGEMENTS

J.G.C.B. was supported by the NERC C-CLEAR doctoral training programme (PDAG/501). The work (including access to high performance computing clusters) was supported by the Cambridge Centre for Earth Observation. Data collection in French Guiana was supported by CNES who funded the 2016 hyperspectral data over Paracou and Labex CEBA (ANR-10-LABX-25) who funded the UAV RGB collections and the field validation of manual crown segmentations as part of the Phenobs project. Thanks to Jean-Louis Smock (IRD), Ilona Clocher (CNRS), Isabelle Maréchaux (INRA), Chantal Geniez (IRD), Julien Engel (IRD), Tom Hattermann (CNRS), Géraldine Derroire (CIRAD), Patrick Heuret (INRA), and all staff at Paracou Research Station for help in the field. Thanks to Philippe Verly (IRD) for support in data processing.

AUTHOR CONTRIBUTIONS

JGCB wrote the manuscript and all authors contributed to the final version. JGCB, SJ, GV and DAC conceived of the study design. JGCB and SJ ran the model training and supporting experiments. AL and PR developed the HS data from raw to analysable states. GV supervised the fieldwork and HS data collection. NB supervised the UAV data collection.

CODE AND DATA AVAILABILITY

Hyperspectral, RGB and tree crown data will be made available upon acceptance for publication [ZENODO LINK].

All scripts for analysis will be made available upon acceptance for publication [GITHUB] https://github.com/sadiqj/hyperspectral-nns; https://github.com/sadiqj/hyper-analysis;

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

References

Aardt, J A N van, and R H Wynne. 2007. "Examining Pine Spectral Separability Using Hyperspectral Data from an Airborne Sensor: An Extension of Field-based Results." *Int. J. Remote Sens.* 28 (2): 431–36.

Ackerly, D D, and M J Donoghue. 1998. "Leaf Size, Sapling Allometry, and Corner's Rules: Phylogeny and Correlated Evolution in Maples (Acer)." *Am. Nat.* 152 (6): 767–91.

Arroyo-Mora, J Pablo, Margaret Kalacska, Trond Løke, Daniel Schläpfer, Nicholas C Coops, Oliver Lucanus, and George Leblanc. 2021. "Assessing the Impact of Illumination on UAV Pushbroom Hyperspectral Imagery Collected Under Various Cloud Cover Conditions." *Remote Sens. Environ.* 258: 112396.

Asner, Gregory P. 2014. "A Chemical-Evolutionary Basis for Remote Sensing of Tropical Forest Diversity." In *Forests and Global Change*, edited by David A. Coomes, David F. R. P. Burslem, William D. Simonson, 343–58. Ecological Reviews. Cambridge University Press.

Asner, Gregory P, and Roberta E Martin. 2011. "Canopy Phylogenetic, Chemical and Spectral Assembly in a Lowland Amazonian Forest." *New Phytol.* 189 (4): 999–1012.

Asner, Gregory P, Roberta E Martin, Raul Tupayachi, Christopher B Anderson, Felipe Sinca, Loreli Carranza-Jiménez, and Paola Martinez. 2014. "Amazonian Functional Diversity from Forest Canopy Chemical Assembly." *Proc. Natl. Acad. Sci. U. S. A.* 111 (15): 5604–9.

Asner, Gregory P, Daniel Nepstad, Gina Cardinot, and David Ray. 2004. "Drought Stress and Carbon Uptake in an Amazon Forest Measured with Spaceborne Imaging Spectroscopy." *Proceedings of the National Academy of Sciences* 101 (16): 6039–44.

Aubry-Kientz, Mélaine, Raphaël Dutrieux, Antonio Ferraz, Sassan Saatchi, Hamid Hamraz, Jonathan Williams, David Coomes, Alexandre Piboule, and Grégoire Vincent. 2019. "A Comparative Assessment of the Performance of Individual Tree Crowns Delineation Algorithms from ALS Data in Tropical Forests." *Remote Sensing* 11 (9): 1086.

Aubry-Kientz, Mélaine, Anthony Laybros, Ben Weinstein, James G C Ball, Tobias Jackson, David Coomes, and Grégoire Vincent. 2021. "Multisensor Data Fusion for Improved Segmentation of Individual Tree Crowns in Dense Tropical Forests." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 3927–36.

Badourdine, Colette, Jean-Baptiste Féret, Raphaël Pélissier, and Grégoire Vincent. 2023. "Exploring the Link Between Spectral Variance and Upper Canopy Taxonomic Diversity in a Tropical Forest: Influence of Spectral Processing and Feature Selection." *Remote Sensing in Ecology and Conservation* 9 (2): 235–50.

Baldeck, Claire A, and Gregory P Asner. 2014. "Improving Remote Species Identification Through Efficient Training Data Collection." *Remote Sensing* 6 (4): 2682–98.

Baldeck, Claire A, Gregory P Asner, Robin E Martin, Christopher B Anderson, David E Knapp, James R Kellner, and S Joseph Wright. 2015. "Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy." *PLoS One* 10 (7): e0118403.

Ball, James G C, Sebastian H M Hickman, Tobias D Jackson, Xian Jing Koay, James Hirst, William Jay, Matthew Archer, Mélaine Aubry-Kientz, Grégoire Vincent, and David A Coomes. 2023. "Accurate Delineation of Individual Tree Crowns in Tropical Forests from Aerial RGB Imagery Using Mask R-CNN." *Remote Sensing in Ecology and Conservation* 9 (5): 641–55.

Baraloto, Christopher, Olivier J Hardy, C E Timothy Paine, Kyle G Dexter, Corinne Cruaud, Luke T Dunning, Mailyn-Adriana Gonzalez, et al. 2012. "Using Functional Traits and Phylogenetic Trees to Examine the Assembly of Tropical Tree Communities." *J. Ecol.* 100 (3): 690–701.

Baumgartner, Andreas, Peter Gege, Claas Köhler, Karim Lenhard, and Thomas Schwarzmaier. 2012. "Characterisation Methods for the Hyperspectral Sensor HySpex at DLR's Calibration Home Base." In *Sensors, Systems, and Next-Generation Satellites XVI*, 8533:371–78. SPIE.

Béland, Martin, and Hideki Kobayashi. 2024. "Drivers of Deciduous Forest Near-Infrared Reflectance: A 3D Radiative Transfer Modeling Exercise Based on Ground Lidar." *Remote Sens. Environ.* 302 (March): 113951.

Biggs, Christopher R, Lauren A Yeager, Derek G Bolser, Christina Bonsell, Angelina M Dichiera, Zhenxin Hou, Spencer R Keyser, et al. 2020. "Does Functional Redundancy Affect Ecological Stability and Resilience? A Review and Meta-analysis." *Ecosphere* 11 (7).

Blomberg, Simon P, Theodore Garland Jr, and Anthony R Ives. 2003. "Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile." *Evolution* 57 (4): 717–45.

Bohlman, Stephanie A. 2015. "Species Diversity of Canopy Versus Understory Trees in a Neotropical Forest: Implications for Forest Structure, Function and Monitoring." *Ecosystems* 18 (4): 658–70.

Bonal, Damien, Alexandre Bosc, Stéphane Ponton, Jean-Yves Goret, Benoît Burban, Patrick Gross, Jean-Marc Bonnefond, et al. 2008. "Impact of Severe Dry Season on Net Ecosystem Exchange in the Neotropical Rainforest of French Guiana." *Glob. Chang. Biol.* 14 (8): 1917–33.

Boochs, F, G Kupfer, K Dockter, and W Kühbauch. 1990. "Shape of the Red Edge as Vitality Indicator for Plants." *Int. J. Remote Sens.* 11 (10): 1741–53.

Bredin, Yennie K, Carlos A Peres, and Torbjørn Haugaasen. 2020. "Forest Type Affects the Capacity of Amazonian Tree Species to Store Carbon as Woody Biomass." *For. Ecol. Manage.* 473 (October): 118297.

Carter, G A, and A K Knapp. 2001. "Leaf Optical Properties in Higher Plants: Linking Spectral Characteristics to Stress and Chlorophyll Concentration." *Am. J. Bot.* 88 (4): 677–84.

Cavender-Bares, Jeannine, Jose Eduardo Meireles, John J Couture, Matthew A Kaproth, Clayton C Kingdon, Aditya Singh, Shawn P Serbin, et al. 2016. "Associations of Leaf Spectra

with Genetic and Phylogenetic Variation in Oaks: Prospects for Remote Detection of Biodiversity." *Remote Sensing* 8 (3): 221.

Ceccato, Pietro, Stéphane Flasse, Stefano Tarantola, Stéphane Jacquemoud, and Jean-Marie Grégoire. 2001. "Detecting Vegetation Leaf Water Content Using Reflectance in the Optical Domain." *Remote Sens. Environ.* 77 (1): 22–33.

Chambers, Dominic, Catherine Périé, Nicolas Casajus, and Sylvie de Blois. 2013. "Challenges in Modelling the Abundance of 105 Tree Species in Eastern North America Using Climate, Edaphic, and Topographic Variables." *For. Ecol. Manage.* 291 (March): 20–29.

Chan, Aland H Y, Chloe Barnes, Tom Swinfield, and David A Coomes. 2021. "Monitoring Ash Dieback (Hymenoscyphus fraxineus) in British Forests Using Hyperspectral Remote Sensing." *Remote Sens. Ecol. Conserv.* 7 (2): 306–20.

Chance, Curtis M, Nicholas C Coops, Ken Crosby, and Neal Aven. 2016. "Spectral Wavelength Selection and Detection of Two Invasive Plant Species in an Urban Area." *Can. J. Remote Sens.* 42 (1): 27–40.

Chave, Jérôme, Stuart J Davies, Oliver L Phillips, Simon L Lewis, Plinio Sist, Dmitry Schepaschenko, John Armston, et al. 2019. "Ground Data Are Essential for Biomass Remote Sensing Missions." *Surv. Geophys.* 40 (4): 863–80.

Chen, Litong, Yi Zhang, Matheus Henrique Nunes, Jaz Stoddart, Sacha Khoury, Aland H Y Chan, and David A Coomes. 2022. "Predicting Leaf Traits of Temperate Broadleaf Deciduous Trees from Hyperspectral Reflectance: Can a General Model Be Applied Across a Growing Season?" *Remote Sens. Environ.* 269 (February): 112767.

Clark, Matthew L, and Dar A Roberts. 2012. "Species-Level Differences in Hyperspectral Metrics Among Tropical Rainforest Trees as Determined by a Tree-Based Classifier." *Remote Sensing* 4 (6): 1820–55.

Clark, Matthew L, Dar A Roberts, and David B Clark. 2005. "Hyperspectral Discrimination of Tropical Rain Forest Tree Species at Leaf to Crown Scales." *Remote Sens. Environ.* 96 (3): 375–98.

Corlett, Richard T. 2011. "Impacts of Warming on Tropical Lowland Rainforests." *Trends Ecol. Evol.* 26 (11): 606–13.

Curran, Paul J. 1989. "Remote Sensing of Foliar Chemistry." *Remote Sens. Environ.* 30 (3): 271–78.

Dalponte, Michele, Lorenzo Bruzzone, and Damiano Gianelle. 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 (August): 258–70.

Dalponte, Michele, and David A Coomes. 2016. "Tree-Centric Mapping of Forest Carbon Density from Airborne Laser Scanning and Hyperspectral Data." *Methods Ecol. Evol.* 7 (10): 1236–45.

Dalponte, Michele, Hans Ole Ørka, Liviu Theodor Ene, Terje Gobakken, and Erik Næsset. 2014. "Tree Crown Delineation and Tree Species Classification in Boreal Forests Using Hyperspectral and ALS Data." *Remote Sens. Environ.* 140 (January): 306–17.

Davies, Stuart J, Iveren Abiem, Kamariah Abu Salim, Salomón Aguilar, David Allen, Alfonso Alonso, Kristina Anderson-Teixeira, et al. 2021. "ForestGEO: Understanding Forest Diversity and Dynamics Through a Global Observatory Network." *Biol. Conserv.* 253 (January): 108907.

Donovan, Lisa A, Hafiz Maherali, Christina M Caruso, Heidrun Huber, and Hans de Kroon. 2011. "The Evolution of the Worldwide Leaf Economics Spectrum." *Trends Ecol. Evol.* 26 (2): 88–95.

Duque, Alvaro, Helene C Muller-Landau, Renato Valencia, Dairon Cardenas, Stuart Davies, Alexandre de Oliveira, Álvaro J Pérez, Hugo Romero-Saltos, and Alberto Vicentini. 2017.

"Insights into Regional Patterns of Amazonian Forest Structure, Diversity, and Dominance from Three Large Terra-Firme Forest Dynamics Plots." *Biodivers. Conserv.* 26 (3): 669–86.

Duthoit, Sylvie, Valérie Demarez, Jean-Philippe Gastellu-Etchegorry, Emmanuel Martin, and Jean-Louis Roujean. 2008. "Assessing the Effects of the Clumping Phenomenon on BRDF of a Maize Crop Based on 3D Numerical Scenes Using DART Model." *Agric. For. Meteorol.* 148 (8-9): 1341–52.

Ebengo, Dav M, Florian de Boissieu, Grégoire Vincent, Christiane Weber, and Jean-Baptiste Féret. 2021. "Simulating Imaging Spectroscopy in Tropical Forest with 3D Radiative Transfer Modeling." *Remote Sensing* 13 (11): 2120.

Ewijk, Karin Y van, Christophe F Randin, Paul M Treitz, and Neal A Scott. 2014. "Predicting Fine-Scale Tree Species Abundance Patterns Using Biotic Variables Derived from LiDAR and High Spatial Resolution Imagery." *Remote Sens. Environ.* 150 (July): 120–31.

FAO. 2020. *Global Forest Resources Assessment 2020: Main Report*. Food & Agriculture Organization of the United Nations.

Fassnacht, Fabian Ewald, Hooman Latifi, Krzysztof Stereńczak, Aneta Modzelewska, Michael Lefsky, Lars T Waser, Christoph Straub, and Aniruddha Ghosh. 2016. "Review of Studies on Tree Species Classification from Remotely Sensed Data." *Remote Sens. Environ.* 186: 64–87.

Féret, Jean-Baptiste, and Gregory P Asner. 2011. "Spectroscopic Classification of Tropical Forest Species Using Radiative Transfer Modeling." *Remote Sens. Environ.* 115 (9): 2415–22.

Féret, Jean-Baptiste, and Gregory P. Asner. 2012 "Tree species discrimination in tropical forests using airborne imaging spectroscopy." IEEE Transactions on Geoscience and Remote Sensing 51 (1): 73-84.

Feurer, Denis, and Fabrice Vinatier. 2018. "The Time-SIFT Method : Detecting 3-D Changes from Archival Photogrammetric Analysis with Almost Exclusively Image Information," July. https://arxiv.org/abs/1807.09700.

Filella, I, and J Penuelas. 1994. "The Red Edge Position and Shape as Indicators of Plant Chlorophyll Content, Biomass and Hydric Status." *Int. J. Remote Sens.* 15 (7): 1459–70.

Gan, Yi, Quan Wang, and Atsuhiro Iio. 2023. "Tree Crown Detection and Delineation in a Temperate Deciduous Forest from UAV RGB Imagery Using Deep Learning Approaches: Effects of Spatial Resolution and Species Characteristics." *Remote Sensing* 15 (3).

Gao, Bo-Cai. 1996. "NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space." *Remote Sens. Environ.* 58 (3): 257–66.

Garzon-Lopez, Carol X, and Eloisa Lasso. 2020. "Species Classification in a Tropical Alpine Ecosystem Using UAV-Borne RGB and Hyperspectral Imagery." *Drones* 4 (4): 69.

Ghiyamat, Azadeh, and Helmi Z M Shafri. 2010. "A Review on Hyperspectral Remote Sensing for Homogeneous and Heterogeneous Forest Biodiversity Assessment." *Int. J. Remote Sens.* 31 (7): 1837–56.

Gitelson, Anatoly A, Yuri Gritz, and Mark N Merzlyak. 2003. "Relationships Between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves." *J. Plant Physiol.* 160 (3): 271–82.

Goetz, Scott J, Alessandro Baccini, Nadine T Laporte, Tracy Johns, Wayne Walker, Josef Kellndorfer, Richard A Houghton, and Mindy Sun. 2009. "Mapping and Monitoring Carbon Stocks with Satellite Observations: A Comparison of Methods." *Carbon Balance Manag.* 4 (March): 2.

Gourlet-Fleury, Sylvie, Jean-Marc Guehl, and Olivier Laroussinie. 2004. *Ecology and Management of a Neotropical Rainforest : Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana*. Elsevier.

Harrison, D, B Rivard, and G Arturo Sánchez-Azofeifa. 2018. "Classification of Tree Species Based on Longwave Hyperspectral Data from Leaves, a Case Study for a Tropical Dry Forest." *Int. J. Appl. Earth Obs. Geoinf.* 66 (April): 93–105.

Hastings, Jack H, Scott V Ollinger, Andrew P Ouimette, Rebecca Sanders-DeMott, Michael W Palace, Mark J Ducey, Franklin B Sullivan, David Basler, and David A Orwig. 2020. "Tree Species Traits Determine the Success of LiDAR-Based Crown Mapping in a Mixed Temperate Forest." *Remote Sensing* 12 (2): 309.

He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. "Mask R-CNN," March. https://arxiv.org/abs/1703.06870.

Hennessy, Andrew, Kenneth Clarke, and Megan Lewis. 2020. "Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability." *Remote Sensing* 12 (1): 113.

Hesketh, Michael, and G Arturo Sánchez-Azofeifa. 2012. "The Effect of Seasonal Spectral Variation on Species Classification in the Panamanian Tropical Forest." *Remote Sens. Environ.* 118 (March): 73–82.

Horler, D N H, M Dockray, and J Barber. 1983. "The Red Edge of Plant Leaf Reflectance." *Int. J. Remote Sens.* 4 (2): 273–88.

Jacquemoud, S, and F Baret. 1990. "PROSPECT: A Model of Leaf Optical Properties Spectra." *Remote Sens. Environ.* 34 (2): 75–91.

Jansson, G, and P Angelstam. 1999. "Threshold Levels of Habitat Composition for the Presence of the Long-Tailed Tit (Aegithalos Caudatus) in a Boreal Landscape." *Landsc. Ecol.* 14 (3): 283–90.

Jucker, Tommaso, Boris Bongalov, David F R P Burslem, Reuben Nilus, Michele Dalponte, Simon L Lewis, Oliver L Phillips, Lan Qie, and David A Coomes. 2018. "Topography Shapes the Structure, Composition and Function of Tropical Forest Landscapes." *Ecol. Lett.* 21 (7): 989–1000.

Kamoske, Aaron G, Kyla M Dahlin, Quentin D Read, Sydne Record, Scott C Stark, Shawn P Serbin, Phoebe L Zarnetske, and Maria Dornelas. 2022. "Towards Mapping Biodiversity from Above: Can Fusing Lidar and Hyperspectral Remote Sensing Predict Taxonomic, Functional, and Phylogenetic Tree Diversity in Temperate Forests?" *Glob. Ecol. Biogeogr.* 31 (7): 1440–60.

Kattenborn, Teja, Felix Schiefer, Julian Frey, Hannes Feilhauer, Miguel D Mahecha, and Carsten F Dormann. 2022. "Spatially Autocorrelated Training and Validation Samples Inflate Performance Assessment of Convolutional Neural Networks." *ISPRS Open Journal of Photogrammetry and Remote Sensing* 5 (August): 100018.

Kennedy, C E J, and T R E Southwood. 1984. "The Number of Species of Insects Associated with British Trees: A Re-Analysis." *J. Anim. Ecol.* 53 (2): 455–78.

Knyazikhin, Yuri, Mitchell A Schull, Pauline Stenberg, Matti Mõttus, Miina Rautiainen, Yan Yang, Alexander Marshak, et al. 2013. "Hyperspectral Remote Sensing of Foliar Nitrogen Content." *Proc. Natl. Acad. Sci. U. S. A.* 110 (3): E185–92.

Kokaly, Raymond F, Gregory P Asner, Scott V Ollinger, Mary E Martin, and Carol A Wessman. 2009. "Characterizing Canopy Biochemistry from Imaging Spectroscopy and Its Application to Ecosystem Studies." *Remote Sens. Environ.* 113 (September): S78–91.

Laliberté, Etienne, Anna K Schweiger, and Pierre Legendre. 2020. "Partitioning Plant Spectral Diversity into Alpha and Beta Components." *Ecol. Lett.* 23 (2): 370–80.

Laybros, Anthony, Mélaine Aubry-Kientz, Jean-Baptiste Féret, Caroline Bedeau, Olivier Brunaux, Géraldine Derroire, and Grégoire Vincent. 2020. "Quantitative Airborne Inventories in Dense Tropical Forest Using Imaging Spectroscopy." *Remote Sensing* 12 (10).

Laybros, Anthony, Daniel Schläpfer, Jean-Baptiste Féret, Laurent Descroix, Caroline Bedeau, Marie-Jose Lefevre, and Grégoire Vincent. 2019. "Across Date Species Detection Using Airborne Imaging Spectroscopy." *Remote Sensing* 11 (7).

Lee, H S, S J Davies, J V LaFrankie, S Tan, T Yamakura, A Itoh, T Ohkubo, and P S Ashton. 2002. "Floristic and Structural Diversity of Mixed Dipterocarp Forest in Lambir Hills National Park, Sarawak, Malaysia." *J. Trop. For. Sci.* 14 (3): 379–400.

Lewis, Simon L, Jon Lloyd, Stephen Sitch, Edward T A Mitchard, and William F Laurance. 2009. "Changing Ecology of Tropical Forests: Evidence and Drivers." *Annu. Rev. Ecol. Evol. Syst.* 40 (1): 529–49.

Liu, Fang, and Yunchuan Kong. 2015. "Zoib: An R Package for Bayesian Inference for Beta Regression and Zero/One Inflated Beta Regression." *R J.* 7 (2): 34.

Liu, Fang, and Qing Li. 2016. "A Bayesian Model for Joint Analysis of Multivariate Repeated Measures and Time to Event Data in Crossover Trials." *Stat. Methods Med. Res.* 25 (5): 2180–92.

Liu, Yujie, Zhongyi Zhan, Lili Ren, Sangzi Ze, Linfeng Yu, Qi Jiang, and Youqing Luo. 2021. "Hyperspectral Evidence of Early-Stage Pine Shoot Beetle Attack in Yunnan Pine." *For. Ecol. Manage.* 497 (October): 119505.

Loreau, Michel, and Claire de Mazancourt. 2013. "Biodiversity and Ecosystem Stability: A Synthesis of Underlying Mechanisms." *Ecol. Lett.* 16 Suppl 1 (May): 106–15.

Lyapustin, Alexei I, Yujie Wang, Istvan Laszlo, Thomas Hilker, Forrest G Hall, Piers J Sellers, Compton J Tucker, and Sergey V Korkin. 2012. "Multi-Angle Implementation of Atmospheric Correction for MODIS (MAIAC): 3. Atmospheric Correction." *Remote Sens. Environ.* 127: 385–93.

Madritch, Michael D, Clayton C Kingdon, Aditya Singh, Karen E Mock, Richard L Lindroth, and Philip A Townsend. 2014. "Imaging Spectroscopy Links Aspen Genotype with Below-Ground Processes at Landscape Scales." *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369 (1643): 20130194.

Malhi, Yadvinder, J Timmons Roberts, Richard A Betts, Timothy J Killeen, Wenhong Li, and Carlos A Nobre. 2008. "Climate Change, Deforestation, and the Fate of the Amazon." *Science* 319 (5860): 169–72.

Marconi, Sergio, Ben G Weinstein, Sheng Zou, Stephanie A Bohlman, Alina Zare, Aditya Singh, Dylan Stewart, Ira Harmon, Ashley Steinkraus, and Ethan P White. 2022. "Continental-Scale Hyperspectral Tree Species Classification in the United States National Ecological Observatory Network." *Remote Sens. Environ.* 282 (December): 113264.

Martin, François-Marie, Jana Müllerová, Laurent Borgniet, Fanny Dommanget, Vincent Breton, and André Evette. 2018. "Using Single- and Multi-Date UAV and Satellite Imagery to Accurately Monitor Invasive Knotweed Species." *Remote Sensing* 10 (10): 1662.

McGrath, Laura J, Charles van Riper 3rd, and Joseph J Fontaine. 2009. "Flower Power: Tree Flowering Phenology as a Settlement Cue for Migrating Birds." *J. Anim. Ecol.* 78 (1): 22–30.

McManus, Kelly M, Gregory P Asner, Roberta E Martin, Kyle G Dexter, W John Kress, and Christopher B Field. 2016. "Phylogenetic Structure of Foliar Spectral Traits in Tropical Forest Canopies." *Remote Sensing* 8 (3): 196.

Meireles, José Eduardo, Jeannine Cavender-Bares, Philip A Townsend, Susan Ustin, John A Gamon, Anna K Schweiger, Michael E Schaepman, et al. 2020. "Leaf Reflectance Spectra Capture the Evolutionary History of Seed Plants." *New Phytol.* 228 (2): 485–93.

Montes, Rosana, and Carlos Ureña. 2012. "An Overview of BRDF Models." *University of Grenada, Technical Report LSI-2012* 1: 19.

Nunes, Matheus H, Matthew P Davey, and David A Coomes. 2017. "On the Challenges of Using Field Spectroscopy to Measure the Impact of Soil Type on Leaf Traits." *Biogeosciences* 14 (13): 3371–85.

Ollinger, S V. 2011. "Sources of Variability in Canopy Reflectance and the Convergent Properties of Plants." *New Phytol.* 189 (2): 375–94.

Pagel, M. 1999. “Inferring the Historical Patterns of Biological Evolution.” *Nature* 401 (6756): 877–84.

Pausas, Juli G, Mike P Austin, and Ian R Noble. 1997. “A Forest Simulation Model for Predicting Eucalypt Dynamics and Habitat Quality for Arboreal Marsupials.” *Ecol. Appl.* 7 (3): 921–33.

Prieur, Colin, Antony Laybros, Giovanni Frati, Daniel Schlaepfer, Jocelyn Chanussot, and Gregoire Vincent. 2024. “Investigating abiotic sources of spectral variability from multitemporal hyperspectral airborne acquisitions over the French Guyana canopy”. Manuscript submitted for publication, Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Prospere, Kurt, Kurt McLaren, and Byron Wilson. 2014. “Plant Species Discrimination in a Tropical Wetland Using in Situ Hyperspectral Data.” *Remote Sensing* 6 (9): 8494–8523.

Reich, P. B. 1995. “Phenology of tropical forests: Patterns, causes, and consequences.” *Can. J. Bot.*, 73(2), 164–174.

Reichstein, Markus, Michael Bahn, Philippe Ciais, Dorothea Frank, Miguel D Mahecha, Sonia I Seneviratne, Jakob Zscheischler, et al. 2013. “Climate Extremes and the Carbon Cycle.” *Nature* 500 (7462): 287–95.

Reiner, Florian, Martin Brandt, Xiaoye Tong et al. 2023. “More than one quarter of Africa’s tree cover is found outside areas previously classified as forest.” *Nature Communications* 14, 2258.

Revell, Liam J. 2012. “Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things).” *Methods in Ecology and Evolution* 3 (2): 217–23.

Sabat-Tomala, Anita, Edwin Raczko, and Bogdan Zagajewski. 2020. “Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data.” *Remote Sensing* 12 (3): 516.

Schläpfer, Daniel, Andreas Hueni, and Rudolf Richter. 2018. “Cast Shadow Detection to Quantify the Aerosol Optical Thickness for Atmospheric Correction of High Spatial Resolution Optical Imagery.” *Remote Sensing* 10 (2).

Schläpfer, Daniel, and Rudolf Richter. 2014. “Evaluation of Brefcor BRDF Effects Correction for HYSPEX, CASI, and APEX Imaging Spectroscopy Data.” In *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–4. IEEE.

Schläpfer, Daniel, Rudolf Richter, and Tal Feingersh. 2015. “Operational BRDF Effects Correction for Wide-Field-of-View Optical Scanners (BREFCOR).” *IEEE Trans. Geosci. Remote Sens.* 53 (4): 1855–64.

Schmitt, S, S Trueba, S Coste, É Ducouret, N Tysklind, M Heuertz, D Bonal, B Burban, B Hérault, and G Derroire. 2022. “Seasonal Variation of Leaf Thickness: An Overlooked Component of Functional Trait Variability.” *Plant Biol.* 24 (3): 458–63.

Schweiger, Anna K, Jeannine Cavender-Bares, Shan Kothari, Philip A Townsend, Michael D Madritch, Jake J Grossman, Hamed Gholizadeh, Ran Wang, and John A Gamon. 2021. “Coupling Spectral and Resource-Use Complementarity in Experimental Grassland and Forest Communities.” *Proc. Biol. Sci.* 288 (1958): 20211290.

Serbin, Shawn P, Aditya Singh, Brenden E McNeil, Clayton C Kingdon, and Philip A Townsend. 2016. “Spectroscopic Determination of Leaf Morphological and Biochemical Traits for Northern Temperate and Boreal Tree Species.” *Ecol. Appl.* 24 (7): 1651–69.

Serrano, Lydia, Josep Peñuelas, and Susan L Ustin. 2002. “Remote Sensing of Nitrogen and Lignin in Mediterranean Vegetation from AVIRIS Data: Decomposing Biochemical from Structural Signals.” *Remote Sens. Environ.* 81 (2): 355–64.

Shamaoma, Hastings, Paxie W Chirwa, Jules C Zekeng, Abel Ramoelo, Andrew T Hudak, Ferdinand Handavu, and Stephen Syampungani. 2023. “Use of Multi-Date and Multi-Spectral UAS Imagery to Classify Dominant Tree Species in the Wet Miombo Woodlands of Zambia.” *Sensors* 23 (4).

Shang, Xiao, and Laurie A Chisholm. 2014. "Classification of Australian Native Forest Species Using Hyperspectral Remote Sensing and Machine-Learning Classification Algorithms." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (6): 2481–89.

Sims, Daniel A, and John A Gamon. 2002. "Relationships Between Leaf Pigment Content and Spectral Reflectance Across a Wide Range of Species, Leaf Structures and Developmental Stages." *Remote Sens. Environ.* 81 (2): 337–54.

Steege, Hans ter, Nigel C A Pitman, Daniel Sabatier, Christopher Baraloto, Rafael P Salomão, Juan Ernesto Guevara, Oliver L Phillips, et al. 2013. "Hyperdominance in the Amazonian Tree Flora." *Science* 342 (6156): 1243092.

Theiler, James, Amanda Ziemann, Stefania Matteoli, and Marco Diani. 2019. "Spectral Variability of Remotely Sensed Target Materials: Causes, Models, and Strategies for Mitigation and Robust Exploitation." *IEEE Geoscience and Remote Sensing Magazine* 7 (2): 8–30.

Thomas, J R, and H W Gausman. 1977. "Leaf Reflectance Vs. Leaf Chlorophyll and Carotenoid Concentrations for Eight Crops [1]." *Agron. J.* 69 (5): 799–802.

Ustin, Susan L. 2013. "Remote Sensing of Canopy Chemistry." *Proc. Natl. Acad. Sci. U. S. A.* 110 (3): 804–5.

Ustin, Susan L, A A Gitelson, Stéphane Jacquemoud, Michael Schaepman, Gregory P Asner, John A Gamon, and Pablo Zarco-Tejada. 2009. "Retrieval of Foliar Information about Plant Pigment Systems from High Resolution Spectroscopy." *Remote Sens. Environ.* 113 (September): S67–77.

Ustin, Susan L, Dar A Roberts, John A Gamon, Gregory P Asner, and Robert O Green. 2004. "Using Imaging Spectroscopy to Study Ecosystem Processes and Properties." *Bioscience* 54 (6): 523–34.

Vaglio Laurin, Gaia, Jonathan Cheung-Wai Chan, Qi Chen, Jeremy A Lindsell, David A Coomes, Leila Guerriero, Fabio Del Frate, Franco Miglietta, and Riccardo Valentini. 2014. "Biodiversity Mapping in a Tropical West African Forest with Airborne Hyperspectral Data." *PLoS One* 9 (6): e97910.

Vauhkonen, Jari, Hans Ole Ørka, Johan Holmgren, Michele Dalponte, Johannes Heinzel, and Barbara Koch. 2014. "Tree Species Recognition Based on Airborne Laser Scanning and Complementary Data Sources." In *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*, edited by Matti Maltamo, Erik Næsset, and Jari Vauhkonen, 135–56. Dordrecht: Springer Netherlands.

Wagner, Fabien, Bruno Hérault, Clément Stahl, Damien Bonal, and Vivien Rossi. 2011. "Modeling Water Availability for Trees in Tropical Forests." *Agric. For. Meteorol.* 151 (9): 1202–13.

Watt, Michael S, Ellen Mae C Leonardo, Honey Jane C Estarija, Peter Massam, Dilshan de Silva, Renelle O'Neill, David Lane, Rebecca McDougal, Henning Buddenbaum, and Pablo J Zarco-Tejada. 2021. "Long-Term Effects of Water Stress on Hyperspectral Remote Sensing Indicators in Young Radiata Pine." *For. Ecol. Manage.* 502 (December): 119707.

Weng, Jing-Ke, and Clint Chapple. 2010. "The Origin and Evolution of Lignin Biosynthesis." *New Phytol.* 187 (2): 273–85.

Zhang, Jinkai, Benoit Rivard, G Arturo Sánchez-Azofeifa, and Karen Castro-Esau. 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 (2): 129–41.