Towards comprehensive individual tree species mapping in diverse tropical forests by harnessing temporal and spectral dimensions

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

Corresponding author(s): David Coomes and James Ball

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

Abstract

Monitoring tropical forest diversity and resilience requires taxonomically resolved, tree-scale data, but the structural complexity and species richness of tropical canopies make remote classification challenging. We present a scalable two-step approach to tree crown species mapping, combining CNN-based crown segmentation from multi-date UAV RGB imagery with species identification from high-resolution airborne hyperspectral data (416–2500 nm). Working in moist tropical forests in French Guiana, we improved crown delineation accuracy by evaluating consensus among ten segmented RGB surveys, increasing mean F1 from 0.69 (single date) to 0.78 (ten dates). Using 3,500+ hand-delineated, field-verified crowns, we classified 169 species with a frequency weighted F1 of 0.75, achieving F1 > 0.7 for 65 species. Together ~70% of crown area at landscape-scale was accurately delineated and labelled. Rare species remained difficult to identify, underscoring the need for larger, targeted training datasets. Band-importance analyses isolated a narrow far-red-edge window (748–775 nm) as the most informative region for species discrimination, consistent with spectral convergence in biochemical, anatomical, and hydric leaf and structural traits. Species were easiest to classify when crowns produced a tight, well-separated spectral cluster at acquisition — promoted by pronounced, synchronised leaf-flush cycles, minimal liana load, and few close relatives — indicating signal degradation, rather than specific trait identity, governs separability. Spatial dispersion of conspecifics improved accuracy, suggesting learned features generalised, whereas clustering amplified local artefacts and reduced separability. Our results demonstrate that combining multi-temporal crown segmentation with hyperspectral imaging can map species-level canopy composition in complex tropical forests with unprecedented scope, and highlight spectral–phenological features critical for scaling biodiversity monitoring from airborne to satellite platforms.

1. Introduction

Landscape-scale maps of locations and species identities of individual trees have many applications in ecology, forestry and conservation (Chave et al., 2003; Davies et al., 2021; Fassnacht et al., 2016; Weinstein et al., 2024). They enable tracking of species abundance and health (Baldeck et al., 2015; Park et al., 2019), guide conservation of threatened species (McGrath et al., 2009; Baldeck et al., 2015; Tuomisto et al., 2003), and support the monitoring of invasive species and disease spread (Chance et al., 2016; Liu et al., 2021; Sabat-Tomala et al., 2020; Chan et al., 2021). Precisely mapped inventories underpin sustainable forestry (Laybros et al., 2020; White et al., 2016), improve carbon stock estimation by resolving spatial heterogeneity (Bredin et al., 2020; Slik et al., 2013), and support forest type classification by aggregating tree locations and species identities into broader compositional units (Bohlman, 2015; Jucker et al., 2018; Vaglio Laurin et al., 2014). Species maps help us understand the links between biodiversity and ecosystem functioning (van Ewijk et al., 2014; Kamoske et al., 2022; Reichstein et al., 2013), especially with emerging capacity to track individuals and resolve cryptic tropical phenology and productivity patterns (Park et al., 2019; Schmitt et al., 2024). Tree-level mapping reveals how individual responses to water stress scale to influence broader forest resilience, including the buffering role of species with varied climatic tolerances (Anderegg et al., 2015; Gómez-Candón et al., 2016; McDowell et al., 2008; Nepstad et al., 2007; Rowland et al., 2015; Watt et al., 2021). Spatial data are also key to understanding biodiversity maintenance via mechanisms like density-dependent mortality and dispersal limitation (Kenkel, 1988; Dimov et al., 2013; LaManna et al., 2017; Barry and Schnitzer, 2021), as well as seed dispersal, regeneration (Seidler and Plotkin, 2006), and successional dynamics (Hogan et al., 2016). Large permanent forest inventory plots (e.g., 50 ha ForestGEO) have underpinned spatial forest ecology (Condit et al., 2013; LaManna et al., 2017), but their limited coverage and logistical demands are increasingly complemented by high-resolution remote sensing, which provide rich individual-level data across landscapes and insight at scales relevant to conservation and management (Fassnacht et al., 2016; Park et al., 2019; Goetz et al., 2009; Marconi et al., 2022; FAO, 2020). However, accurate remote mapping of species remains a significant challenge in diverse and dense tropical moist forests (Asner and Martin, 2011; Pereira Martins-Neto et al., 2023).

The first step in tree species mapping is the segmentation of individual tree crowns. Airborne lidar datasets can be highly effective in temperate and boreal forests (e.g. Dalponte and Coomes, 2016; Hastings et al., 2020) but is difficult in complex tropical canopies because crowns can interweave in

complex ways in the upper canopy, and subcanopy trees are all but impossible to detect (Cao et al., 2023; Wielgosz et al., 2024). High spatial resolution RGB photographs offer colour and texture information to distinguish between trees (Aubry-Kientz et al., 2021; Williams et al., 2022), and neural network approaches, such as Mask R-CNN (He et al., 2017), can harness this information to separate irregular edges of neighbouring tree crowns (Ball et al., 2023; Gan et al., 2023). However, differences in illumination (e.g. variably shadowed trees), sway of trees and branches, phenological disparity and irregularities in the orthomosaicking reduce the accuracy of segmentation (Park et al., 2019; Araujo et al., 2021; Gan et al., 2023). A promising strategy for improving accuracy involves segmenting trees in imagery collected from multiple surveys conducted over time, then evaluating which crowns are consistently identified across these datasets (Nuijten et al., 2019). This approach helps distinguish genuine tree crowns from artefacts caused by environmental changes or processing variability (Takahashi Miyoshi et al., 2020), but has yet to be critically evaluated.

The second step is identifying the species of each tree crown, usually working with hyperspectral remote sensing (also known as imaging spectroscopy). Hyperspectral sensors measure the intensity of radiation within many narrow spectral bands reflected off canopies. The reflectance spectra of plants are influenced by multiple factors, including biochemical and structural characteristics (Ustin et al., 2004; Ustin, 2013; Ollinger, 2011; Meireles et al., 2020), allowing species to be mapped and their health monitored in low diversity systems (Fassnacht et al., 2016). Distinguishing species in diverse tropical forests is challenging. Nonetheless, early studies using AVIRIS data distinguished 7 species in Costa Rica (Clark et al., 2005) and 17 in Hawaii (Féret and Asner, 2013) and more recent work in French Guiana classified 20 common species with ~80% accuracy (Laybros et al., 2020, 2019) which is only a fraction of the species present in wet tropical forests (upwards of 100 species per hectare; Valencia et al., 1994; ter Steege et al., 2013). One hope is that the application of advanced machine learning approaches, including neural networks, will lead to improved classification accuracy given the flexibility of these methods (Fricker et al., 2019; Mäyrä et al., 2021) but the effectiveness of these approaches compared with more traditional methods have yet to be demonstrated in the tropics.

One major challenge for hyperspectral species mapping is identifying which spectral regions are most informative, as canopy reflectance integrates signals from multiple interacting factors. Multiple traits can interact to determine reflectance in any given spectral region (Curran, 1989; Jacquemoud and Baret, 1990; Féret and Asner, 2011; Ollinger, 2011), making it difficult to attribute spectral differences to single drivers. Unpicking these relationships is critical for robust,

transferable species identification in diverse tropical forests (Schweiger et al., 2021). Species differ in their leaf-level reflectance spectra (e.g., distinguishing 46 species in a tropical wetland; Prospere et al., 2014) due to variation in physical and biochemical traits (Cavender-Bares et al., 2016; McManus et al., 2016; Meireles et al., 2020), such as chlorophyll, carotenoids, anthocyanins, cellulose, lignin, and water content (Chen et al., 2022; Serbin et al., 2016). These traits can vary substantially even among closely related species, although phylogenetic conservatism often results in shared spectral "syndromes" within lineages (Cavender-Bares et al., 2016; Meireles et al., 2020). Spectral expression is also influenced by environmental conditions (e.g., soil type, water availability, light exposure; Nunes et al., 2017), plant health, and phenology (Ollinger, 2011). At canopy scale, structural effects such as leaf area distribution and crown shape can dominate over biochemical signals in determining reflectance (Knyazikhin et al., 2013; Béland and Kobayashi, 2024). Nonetheless, several studies indicate that the red edge and adjacent NIR region are particularly important for separating species in high-diversity forests (Durgante et al., 2013; Hennessy et al., 2020; Pereira Martins-Neto et al., 2023), although the generality of these findings remains uncertain.

The accuracy with which tree species can be classified from hyperspectral imagery depends on methodological, ecological, and biological determinants of spectral separability, yet we know of no study that has critically evaluated the relative importances of these factors. It is reasonable to assume, for instance, that species represented by a large number of training examples will be identified more accurately. We hypothesize that the inclusion of rare species with limited numbers of crowns in the training dataset would increase confusion and lower overall accuracy (Baldeck and Asner, 2014); this is particularly pertinent when classifying tropical forests, which usually contain a few dominant species and many rare species represented by few individuals (ter Steege et al., 2013). Characteristics of the species themselves may influence their separability. Most directly, one would assume that species which are "spectrally distinct" in multi-dimensional spectral space will be more reliably classified. Species that occur in spatial clusters may appear more separable if classifiers exploit local spectral artefacts or environmental features (i.e. spatial autocorrelation), potentially inflating accuracy, whereas robust performance across widely dispersed individuals would indicate reliance on generalisable, species-specific spectral features; however, the extent to which spatial arrangement influences classification success remains poorly quantified (McManus et al., 2016; Marconi et al., 2022). We hypothesise that species with more prominent crowns are more reliably classified, as their crowns are more likely to be well illuminated, clearly delineated, and less

affected by spectral mixing with neighbouring trees (Clark et al., 2005; Marconi et al., 2022), while the converse is true of smaller trees (Clark and Roberts, 2012; Kalacska et al., 2007). Further, we hypothesise that distinctive leaf and bark traits can aid spectral discrimination (e.g. Asner and Martin, 2009). In addition, phenology—the timing of leaf production, senescence, shedding, and flowering—can strongly influence crown reflectance spectra (Badourdine et al., n.d.; Blackburn and Milton, 1995). We expect that species with synchronous events, such as coordinated leaf flushing across individuals, will produce more homogeneous spectral signals and be more easily separable if imagery coincides with these events (Takahashi Miyoshi et al., 2020; Clark and Roberts, 2012). Conversely, greater within-species phenological variability (e.g., staggered flushing or partial deciduousness) is likely to increase intra-specific spectral variation and reduce classification accuracy (Clark and Roberts, 2012; Hesketh and Sánchez-Azofeifa, 2012). Liana presence also affects crown reflectance spectra (van der Heijden et al., 2022); we expect species with high rates of infestation to have greater within crown spectral variance and therefore to be more difficult to distinguish than those with low rates (Ghiyamat and Shafri, 2010). Finally, we expect that closely related species will be less accurately classified because they often have more similar physico-chemical traits than more distantly related species, which may reduce their spectral separability (Asner and Martin, 2009; McManus et al., 2016; Marconi et al., 2022). Previous studies have explored some of the hypotheses listed above, but we are unaware of any previous work that has comprehensively evaluated their relative impacts on species separability.

In this study, we propose an approach to tropical forest tree mapping that precisely delineates individual tree crowns in multi-temporal RGB images and classifies the species of each tree from hyperspectral data based on training on a broad pool of 169 species. Automatic crown delineation was performed with *detectree2*, a Mask R-CNN-based tool for remote sensing segmentation (He et al., 2017; Ball et al., 2023). For the first time, we combined delineated maps from repeat airborne surveys and evaluated whether consensus across multi-date crown maps improved segmentation accuracy (Martin et al., 2018; Shamaoma et al., 2023). While RGB lacks the spectral richness of hyperspectral sensors, its superior spatial resolution – especially from low-altitude UAV surveys – enables the detection of crown texture and precise boundary delineation. The relatively low cost of the sensors and UAV systems means that regular surveys of the same area become feasible. To address temporal variability in canopy, lighting, and atmosphere, we developed a consensus-fusion method that spatially matched crown polygons across dates and averaged their vertices, producing crown maps that reflect an inter-date consensus on crown shape and position. This reduced

temporal noise and improved delineation consistency. Species classification was trained and tested on 3,256 manually delineated crowns, each field-verified during a decade-long campaign. This dataset, combined with our technical framework, supports scalable, accurate tree mapping and species identification in diverse tropical forests. By leveraging the spatial precision of UAV RGB imagery and the spectral depth of hyperspectral data, our two-step approach achieves high accuracy in complex forest environments. The research aims to:

1. Assess whether the accuracy of tree crown maps can be improved by making repeated aerial surveys, segmenting tree crown in each survey, and combining segmentation results to construct a consensus-based map;

2. Evaluate which machine learning classifier yields the most accurate overall (landscape) predictions of the species identity of individual tree crowns and whether the inclusion of rarer species in training is detrimental to overall accuracy.

3. Determine which wavebands are most important for species determination and understand why;

4. Evaluate the importance of the following factors in influencing the separability of species (i.e. how well we can distinguish one species from the other species present): the number of field-labelled training crowns available; the genetic relatedness of the species; the spatial distribution of individuals of a species across the landscape; the prominence of the species crowns; the distinctiveness of the leaf and bark traits of species; the within and across-species variation in phenology; and the prevalence of liana infestations.

Our paper combines traditional fieldwork, advanced machine learning techniques, high-resolution remote sensing data, and critically evaluates the next steps towards successful species-level forest mapping. We close with a discussion of how our findings inform the potential to scale species classification beyond airborne platforms to satellite-based systems. While satellite sensors (e.g. Sentinel-2, PRISMA, EnMAP) offer broader spatial coverage and more frequent revisits, they are typically limited in either spectral or spatial resolution compared to UAV and airborne systems. By quantifying the spectral regions most critical for species discrimination and the minimum number of training samples required per species, we identify key parameters that could guide the design or selection of future satellite sensors for species mapping.

2. Materials and Methods

2.1 Study site and overview of methods

The research was conducted in forests at Paracou Field Station, French Guiana (5°16'N 52°55'W; see Fig. 1). The lowland tropical rainforests grow mostly on shallow ferralitic soils underlain by a variably transformed loamy saprolite (Gourlet-Fleury et al., 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 which contain approximately 76,000 trees of DBH ≥ 10 cm consisting of over 800 different species (Gourlet-Fleury et al., 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, tuning of model selection, providing a robust assessment of model accuracies.

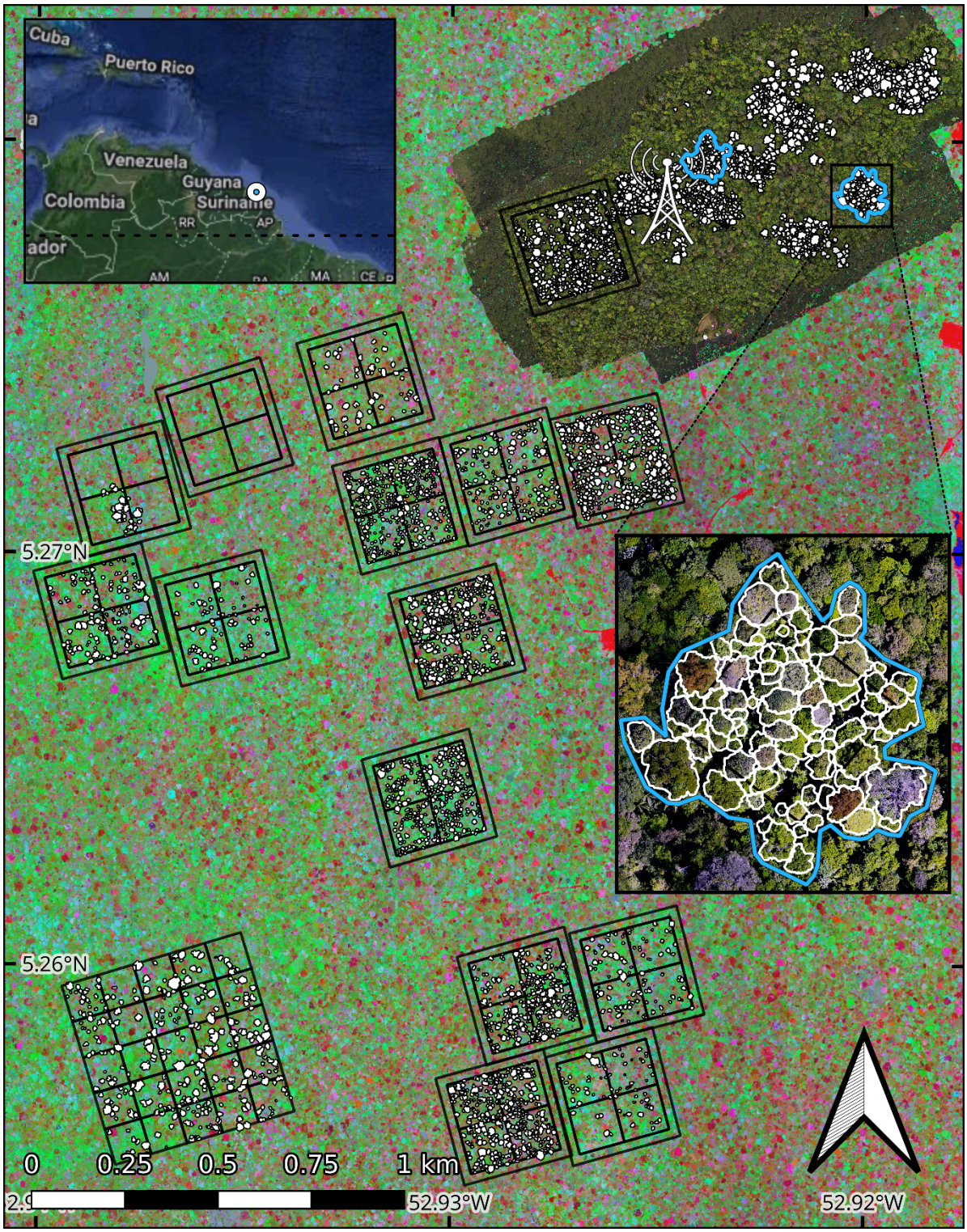

Figure 1*: Site map of Paracou, French Guiana, with crowns and hyperspectral imagery. The manually delineated, labelled crowns are in white. The colourful background scan that covers the entire site is a representation of the hyperspectral data (selected projected PCA bands for illustration). The repeat 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. One of these regions is shown up close in the inset. The black boxes show the plots in which inventories are conducted.*

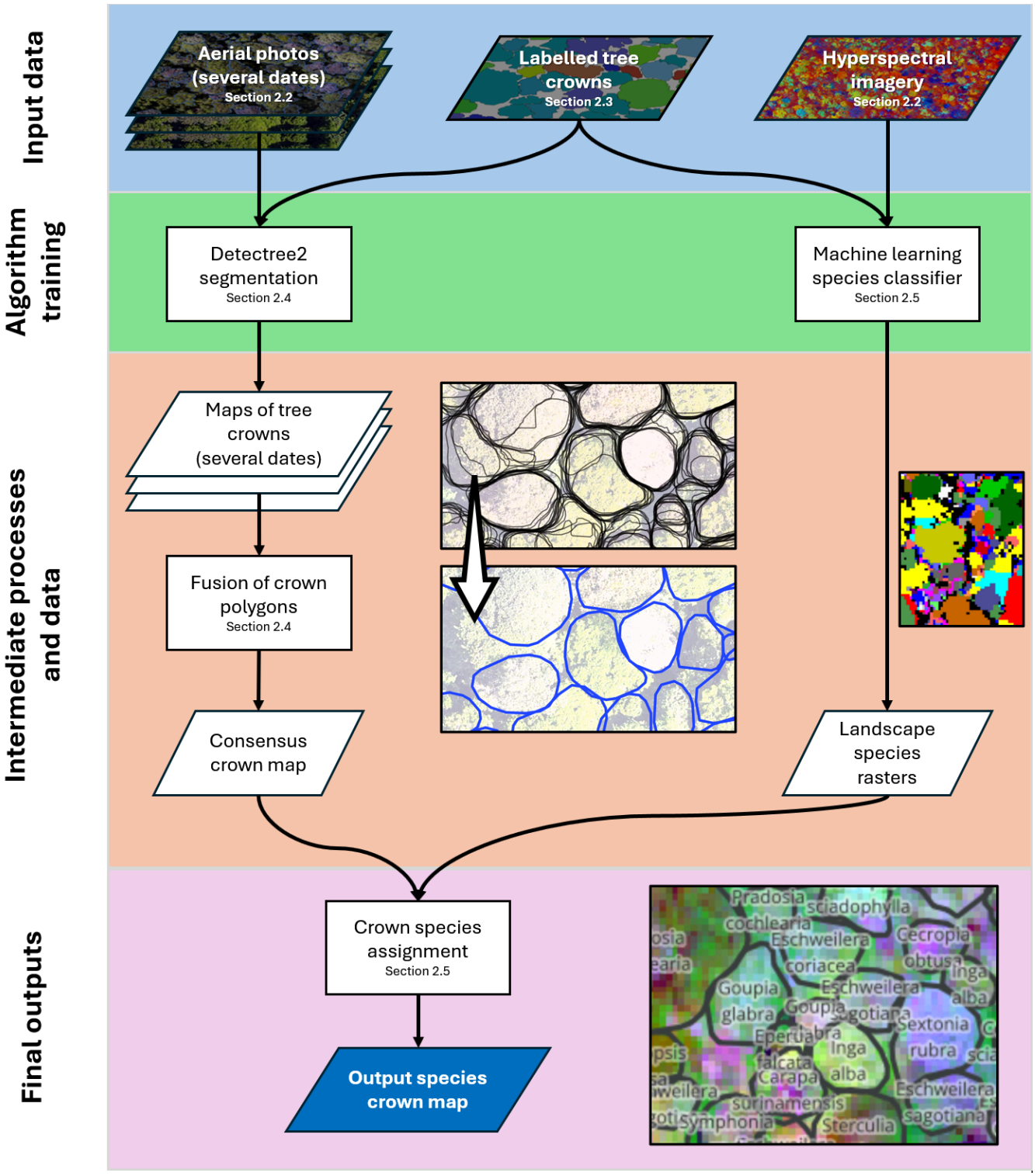

Figure 2*: Simplified schematic of the crown mapping approach showing input data and the intermediate steps to producing a labelled tree crown map. In the centre is an illustration of the process of temporal polygon fusion. The top image shows the overlapping tree crown polygons predicted over multiple 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	1064 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, Fig. S.3). 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 et al. (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, including co-registration, are given in Section S.2.

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 et al., 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, and test the impact of lianas on the accuracy of species recognition under Aim 4. Additional details can be found in Section S.3.

2.4 Consensus mapping of the locations of tree crown using repeated surveys (Aim 1)

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

To test our consensus mapping approach, the trained models were 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 merged to form *consensus* delineations. This aimed to find a temporal agreement on crown locations and shapes. The consensus 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.

Across dates, we matched polygons representing the same crown using an IoU-based criterion and formed consensus crowns by confidence-weighted boundary averaging. We then applied a space-filling selection that placed consensus polygons in descending agreement score while preventing conflicts with already placed crowns, yielding a temporally stable crown map. Full thresholds and implementation details (vertex normalisation, match/placement IoUs, confidence aggregation) are 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 Representative tree species classification (Aim 2)

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 (with adequate confidence) 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. Species with a single recorded individual were excluded.

To test which machine learning classifier yielded the best overall accuracy, models were trained using the hyperspectral data extracted from the delineated tree crowns. In line with commonly practised methodologies, we evaluated Multi-Layer Perceptrons (MLPs), Linear Discriminant Analysis (LDA), Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbours and Logistic Regression algorithms due to their widespread implementations and adaptability. We considered a range of other modelling approaches, including dimensionality reduction, in exploratory analyses that were not taken forward (details in Section S.5.2). We aimed for the classification approach to be broadly reproducible, so we avoided heavy feature engineering. 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.6 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 we observed that it was at a high expense for overall classification accuracy (see Section S.5.2 for details). 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, a situation likely to be true when establishing forest monitoring experiments, so we did not 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 challenges: adjacent pixels may exhibit local effects that risk leakage between training and validation sets when naïve (random) 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 5-fold 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).

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 (by population crown frequency) 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.

To assess whether the inclusion of rarer species in training is detrimental to overall accuracy, we progressively removed the rarest species from the training data and reassessed model performance (Section S.5.3). This allowed us to explore potential gains in less diverse landscapes or in scenarios where species with insufficient representation are excluded. To quantify the relationship between overall accuracy and the number of species represented in training, we reran the best-performing model on a series of artificially restricted datasets. Starting with the full dataset, individuals of the least well-represented species were iteratively removed until only two species remained. We examined two scenarios: (i) the test pool of species was kept complete (169 species) while the training pool was progressively restricted; and (ii) the test pool was reduced in parallel with the training pool.

2.6 Evaluate which wavebands are most important for distinguishing species (Aim 3)

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 validate and 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. See Section S.6.2 for details.

2.7 Factors influencing the separability of species (Aim 4)

In the following section, we quantify the influence of the following factors on the separability of species, as measured by their F1-scores: (a) number of field-labelled tree crowns, (b) spectral distinctiveness, (c) spatial distribution of species, (d) crown size and tree height, (e) leaf and bark traits, (f) strength and synchrony of leafing phenology, and (g) liana presence; (h) genetic relatedness to other species. Factor (a) was tested with the original test set. Factors (b)-(h) were evaluated using a *standardised dataset* in which all species were equally represented by ten individuals to ensure comparability. For this, we selected the 49 species with ≥10 available crowns, randomly sampling 10 crowns per species. We performed 5-fold cross-validation with 8 crowns for training and 2 for validation, repeating this procedure 100 times to capture variability in sample composition. We used random forest modelling to compare the importance of factors (d)-(h) (described below).

(a) Number of field-labelled tree crowns of a species: To estimate how many field-labelled crowns per species are needed for reliable classification, we modelled per-species F1 as a function of training sample size. Following Tötsch and Hoffmann's (2021) Dirichlet-based treatments of classification metric uncertainty, we modelled the joint distribution of true positives, false positives and false negatives with a weakly informative Dirichlet prior (Jeffreys, 1946; Gelman et al., 2013). This approach propagates dependence between precision and recall and yields appropriately wide intervals for rare species. We then related median F1 to the number of training crowns using a beta-regression GAM on log-transformed sample size, as in learning-curve analyses for wildlife image classification (Shahinfar et al., 2020). Species were weighted by the inverse of their posterior F1 variance to reflect differing information content (details in Section S.7.1). We defined "reliable performance" as F1 ≥ 0.7 and reported the smallest training size at which the fitted mean reached this threshold — balancing the need to identify most individuals while keeping mislabelling low, which is critical if mapped labels are used in downstream ecological analyses.

(b) Spectral distinctiveness: To test the assumption that species occupying compact and well-separated regions in spectral space are classified more accurately, we calculated the spectral hypervolume of each species in principal component space over the first five principal components which explained 96.6 % of variance. Hypervolumes were computed from the pixels of each species' crowns (using the *hypervolume* package in R; Blonder et al., 2014), capturing both intra- and inter-crown variation, and estimated as Gaussian hypervolumes, with $\mu \pm \sigma$ encompassing 68% of data points. Hypervolume size indexed intra-species spectral variability, while summed Jaccard indexed inter-species distinctiveness. These two complementary metrics — spread and proximity to others — were tested as predictors of species separability (F1) using beta regression.

Finally, we modelled the influence of factors (c)-(g) on species-specific F1 using Random Forests, because this approach provides flexibility with respect to collinearity and differences in variable distributions (Breiman, 2001). We compared variable importance using percent increase in mean squared error (%IncMSE). Of the 49 species in the standardised dataset ($n_{\text{crowns}} \geq 10$), 47 had the full set of variables and could be included in the analysis. Explanatory variables were calculated as follows. **(c) Spatial distribution**: We examined whether species with tightly clustered individuals yield higher (inflated) separability, exploiting localised spectral artefacts or environmental conditions (spatial autocorrelation). Spatial separation was measured by the mean pairwise distance from a 10-sample subset of individuals within a species averaged over 1000 random draws. Other clustering metrics were considered but not employed as they were distorted by the non-uniform arrangement of the field plots. **(d) Crown size and tree height**: We assessed whether classification success increased with tree stature. We tested whether species with larger crowns (by crown polygon area) or a tendency to be taller (by the median value of the CHM within the crown) were more accurately classified. **(e) Leaf and bark traits**: Using trait data from Vleminckx et al. (2021), we assessed whether specific functional traits were associated with classification performance. The leaf traits included nitrogen, carbon, $\delta^{13}C$, phosphorus, calcium and chlorophyll content, toughness, area, thickness, and specific leaf area (SLA). We also included bark thickness. **(f) Leaf phenology**: we worked with estimates of the signal strength (similarity of leaf-cycling patterns among individuals of a species) of leaf phenology (average time shift required to align cycles for maximum similarity) and signal asynchrony that were calculated by Ball (2023). These were calculated using a time series of crown-level greenness (i.e. Green Chromatic Coordinate) measured every three weeks over two years. Some species show strong seasonality in greenness, related to the time of leaf flushing and senescence, while other species have a weaker signal (as illustrated in

the inset of Fig. 9). Species with consistent evergreen canopies or highly synchronous leaf flushing/shedding were expected to yield more homogeneous spectra and higher classification accuracy, whereas staggered or variable phenology (e.g. partial deciduousness) was expected to increase within-species spectral variance and reduce separability. **(g) Liana presence**: We assessed whether species separability decreased with liana infestation of crowns. For each species, we calculated the proportion of crowns recorded as having lianas, based on the presence–absence field in the tree crown database. A value of 1 indicated lianas were observed, whereas 0 indicated none were recorded, but did not guarantee absence. **(h) Phylogenetic relatedness:** We asked whether phylogenetically distinctive species are more spectrally separable, using three complementary metrics computed on the time-calibrated phylogeny of Baraloto et al. (2012). Nearest-neighbour phylogenetic (or patristic) distance (NNPD) captures *local* proximity — the tree distance from a species to its closest relative in the study set (i.e. the most likely confuser). Mean phylogenetic distance (MPD) summarises *global isolation* by averaging distances to all other species, reflecting how crowded or sparse its broader phylogenetic neighbourhood is. Evolutionary distinctiveness (ED) integrates along the entire root-tip path, allocating branch length in proportion to how many tips share each branch, quantifying how much *unique evolutionary history* the species represents (details in Section S.7.1).

To examine the broader links between phylogeny and spectra, we tested for phylogenetic signal in the spectral data by calculating species mean and standard error of standardised reflectance across hyperspectral bands. We calculated Pagel's λ and Blomberg's K for each band using the *phytools* R package (Revell, 2012), including all 153 species present in the phylogeny; these metrics have different theoretical underpinnings but both give a score of zero if there is no phylogenetic signal and a score of 1 if there is a strong phylogenetic signal (details in Section S.7.2). We also assessed whether classification errors (of the full dataset) correlated with phylogenetic relatedness by comparing pairwise misclassification rates (how often individuals from species A are misclassified as species B and vice versa) and phylogenetic distances (in Myr), using Spearman's ρ to accommodate non-normality. Together, these analyses evaluated if spectral reflectance contains an evolutionary signal relevant to species classification.

3. Results

3.1 Tree crown segmentation using multiple surveys (Aim 1)

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. 3). 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 the broadest range of imagery (five individual dates' worth). The accuracy of its delineation was boosted significantly by combining the delineations from different time steps (consensus mapping) from a mean F_1-score of 0.68 for a single date prediction, to a peak at the combination across 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. 4). Accuracy tended to increase with tree crown area.

The 'base' model and the model trained on additional single date (1 date) of UAV data had a comparable performance at a single date prediction and combinations of less than seven dates (see Fig. 3). 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 did not improve in accuracy by combining dates to the same degree as the other two models, suggesting overfitting. 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 led to better temporal transferability than was achieved with the additional focused training on just a single date.

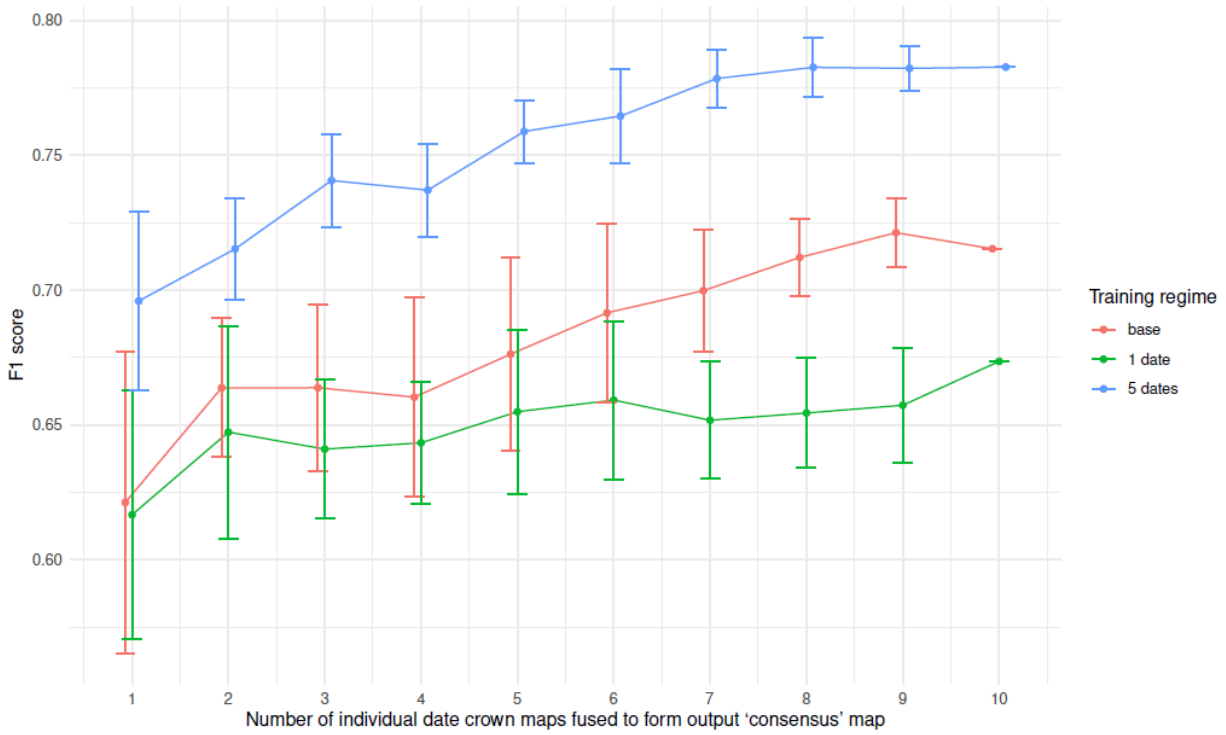

Figure 3*: The performance of the tree crown delineations with the number of individual date crown maps that were combined to form the output 'consensus' map. Researchers may or may not have training crowns available and a varying number of RGB surveys. To reflect this we tested three models trained 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. Only a single combination of dates was possible for the 10-date combination.*

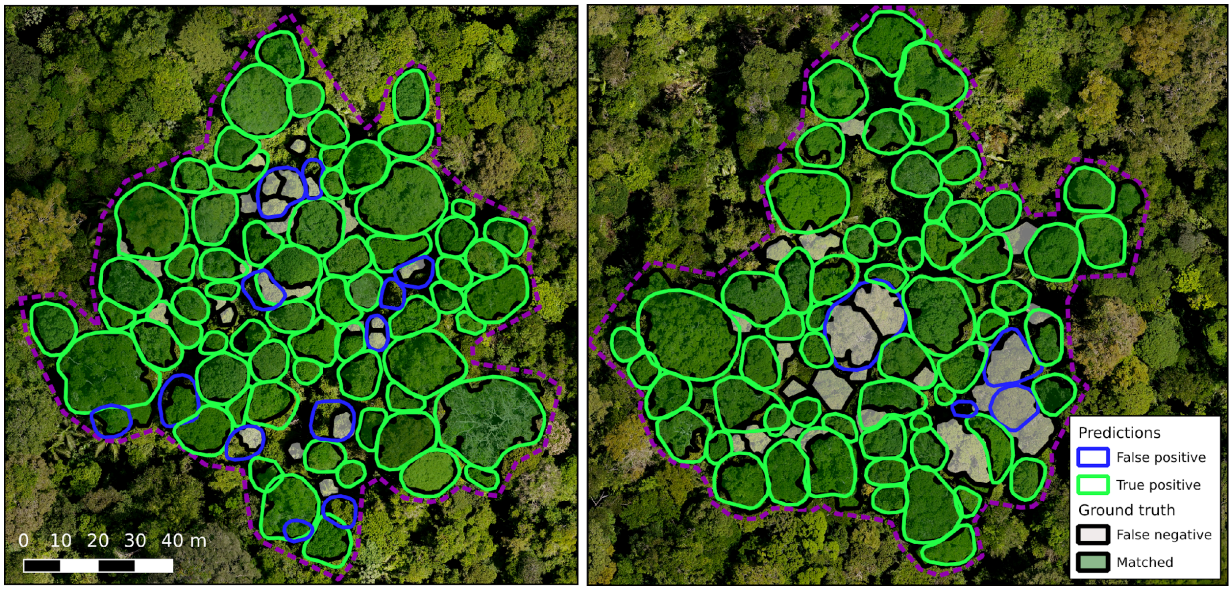

Figure 4*: Predictions and ground truth crowns in the unseen test regions. The reference set of crowns are shown with black border and their fill colour depends on whether a crown was matched (green) or unmatched (grey) with a prediction crown. The predictions that match with a reference crown have a 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* $IoU \geq 0.5$.

3.2 Tree species classification using hyperspectral imagery (Aim 2)

The field teams in Paracou labelled 3,256 individual tree crowns to species level with sufficient confidence. 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 independently in the classifier training and testing. The filtering process left 169 species for further analysis (see Fig. 5).

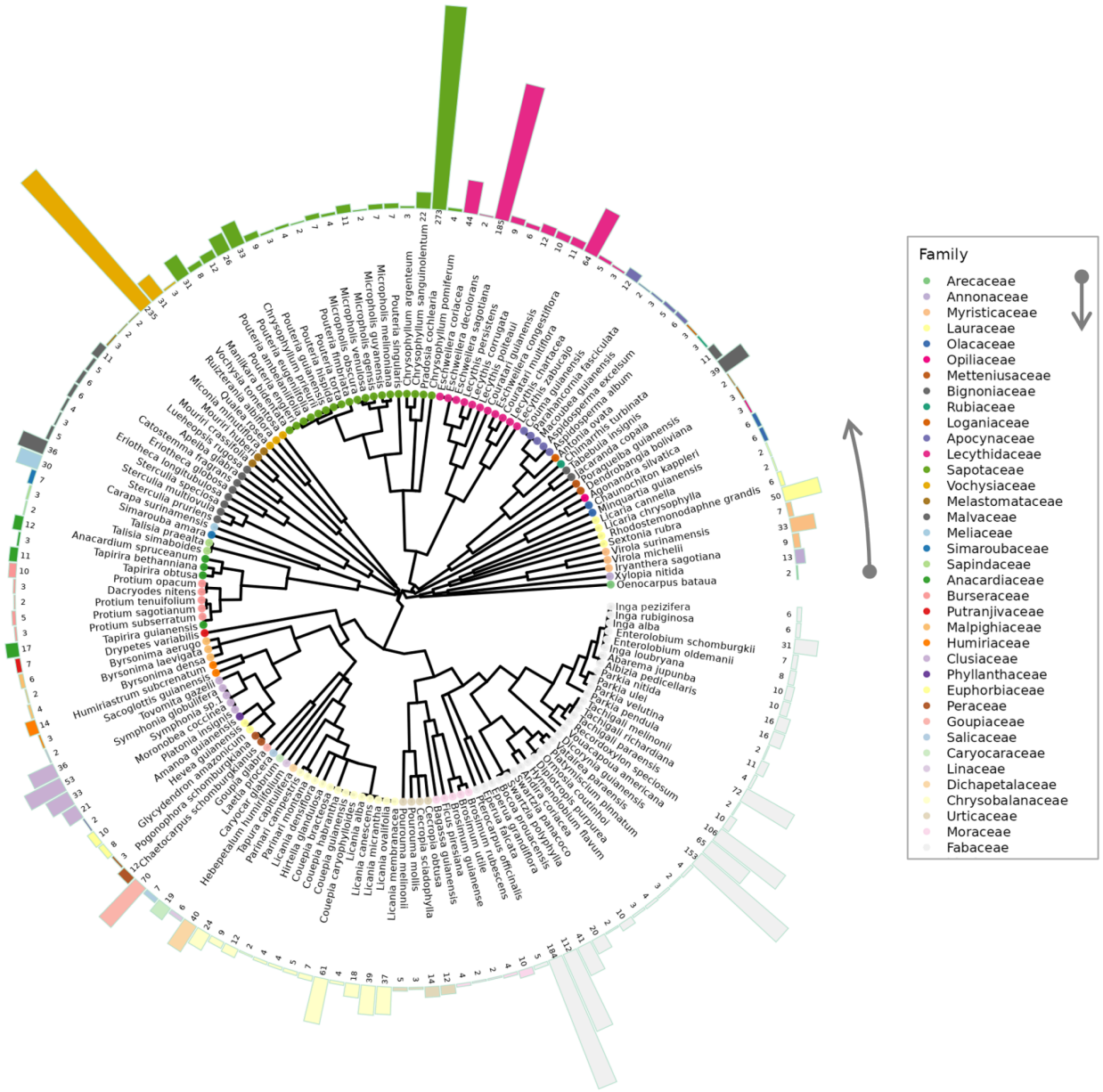

Figure 5*: Abundance of 169 species in the ground validated tree crown dataset mapped onto the phylogeny of Baraloto et al. (2012). Bars represent the total number of individuals sampled in the field. The legend is arranged in the order families appear on the tree (anti-clockwise from 3 o'clock). Species with just a sole representative crown record (70 in total) were not included in the classification task*

Comparison of machine learning classifiers showed that the LDA performed best at classifying species in the test set of tree crowns from their hyperspectral signal (weighted average F_1- score = 0.75; see Table 2). 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 less flexible approach to separating the classes led to more robust transferability between crowns. Furthermore, the LDA classifier was far quicker taking 20 seconds to train whereas the SVM took more than 7 hours. SVM was the slowest method due to its quadratic to cubic complexity with respect to the number of training samples and the large number of training pixels[2].

Table 2*. 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
SVM	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

[2] https://scikit-learn.org/stable/modules/svm.html#complexity

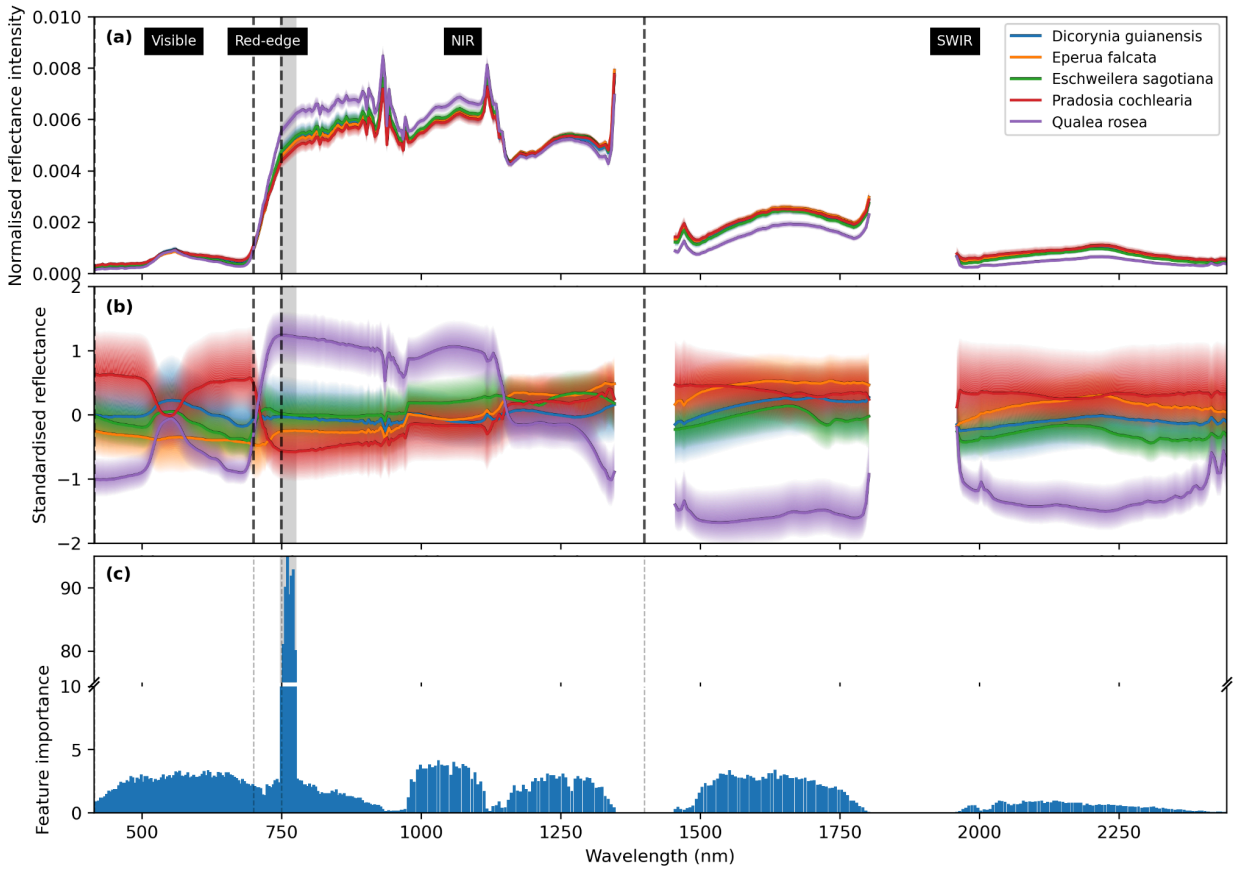

Figure 6*: Spectral signatures of five common species and per-band importance for classification. Lines show the median reflectance per band across crown pixels; shading shows the interquartile range. (a) Normalised reflectance (each pixel spectrum divided by its total reflectance). (b) Standardised reflectance (per-band mean subtracted and divided by the per-band standard deviation across valid crown pixels), allowing comparison on a common scale. (c) LDA band importance (best model, trained on the standardised data): taller bars indicate greater contribution to interspecific separation. The grey vertical band highlights the far-red edge (748–775 nm), identified as most important for species classification. SWIR gaps correspond to bands removed because of atmospheric humidity effects.*

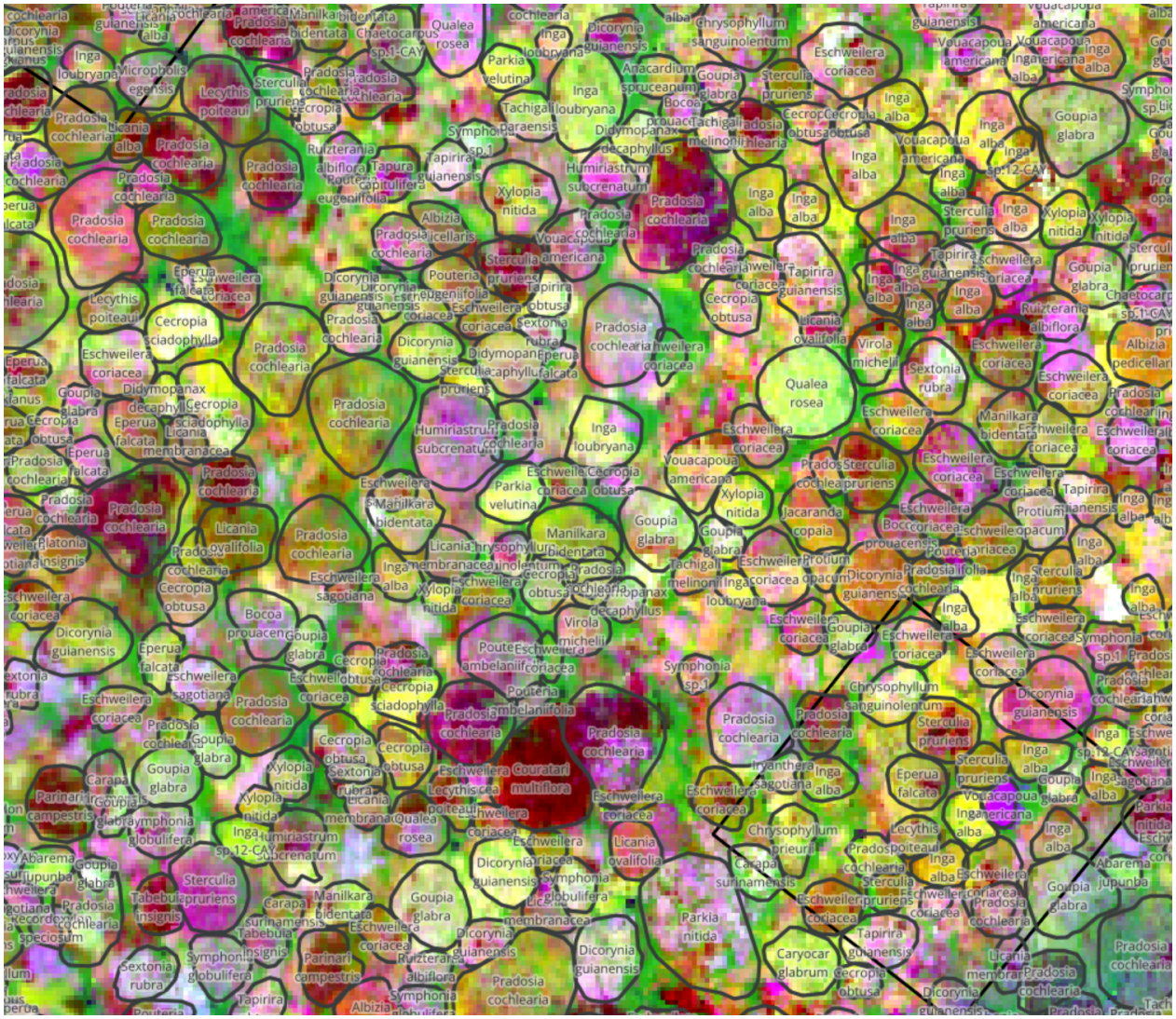

Figure 7*: Final crown map over a hyperspectral PCA backdrop. Background shows an RGB composite of three principal components (PC1–PC3) from the HSI for visualisation only (PCA outputs were not used in the analysis). Predicted crowns are outlined in black and labelled with species; black squares denote forest inventory plots.*

Using the best-performing classifier, we found that 81% of the total tree crown area in the test set was correctly identified to species level. Combining this with the percentage of the total crown area that was well located and delineated (86%), we conservatively estimated that 70% of the landscape's crown area was mapped correctly. For reference, crowns of the twenty most abundant species make < 60% of the total crown area of the reference dataset (Fig. 7).

Inclusion of rare species: Contrary to our expectations, overall (weighted) F1-score decreased as the less-well represented species were removed from the training pool, whilst being retained in the test set (see Fig. S.8). This decline in performance with the removal of rare species indicated there was no penalty overall accuracy for including as many species as possible in training.

3.3 Which wavebands are important for determining species? (Aim 3)

Eight bands between 748 and 775 nm, on the "far red-edge" transition between the red and near-infrared ranges, dominated in terms of relative feature importance for separating species (see Fig. 6). 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.5). 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 (as there was shared information between adjacent bands).

3.4 What factors determine whether a species can be accurately identified from its spectra? (Aim 4)

We distinguished 65 species with an F_1-score of at least 0.7. However, species with a low number of training (and correspondingly test crowns) had large confidence intervals (Fig. 8). Within the subsample of species that had at least 10 training crowns, we cross validated the LDA model with a 5-fold split of 8 training crowns to 2 validation crowns (with resampling) and the resulting species-level F1-scores had a range of 0.6 – 1.0 (see Fig. S.13), consistent with the F1-score for the full set of 169 species being centred around ~0.7.

Number of labels in training dataset: The accuracy of classification increased with the number of training crowns in the class, meaning more common species were more accurately identified (see Fig. 8) in agreement with similar patterns observed by Féret and Asner (2013) and Baldeck and Asner (2014). There was a sharp increase between 1 and 15 training crowns, after which the

performance improved more gradually. Within this regime, species with at least 13 training crowns would be expected to achieve an F1-score of 0.7.

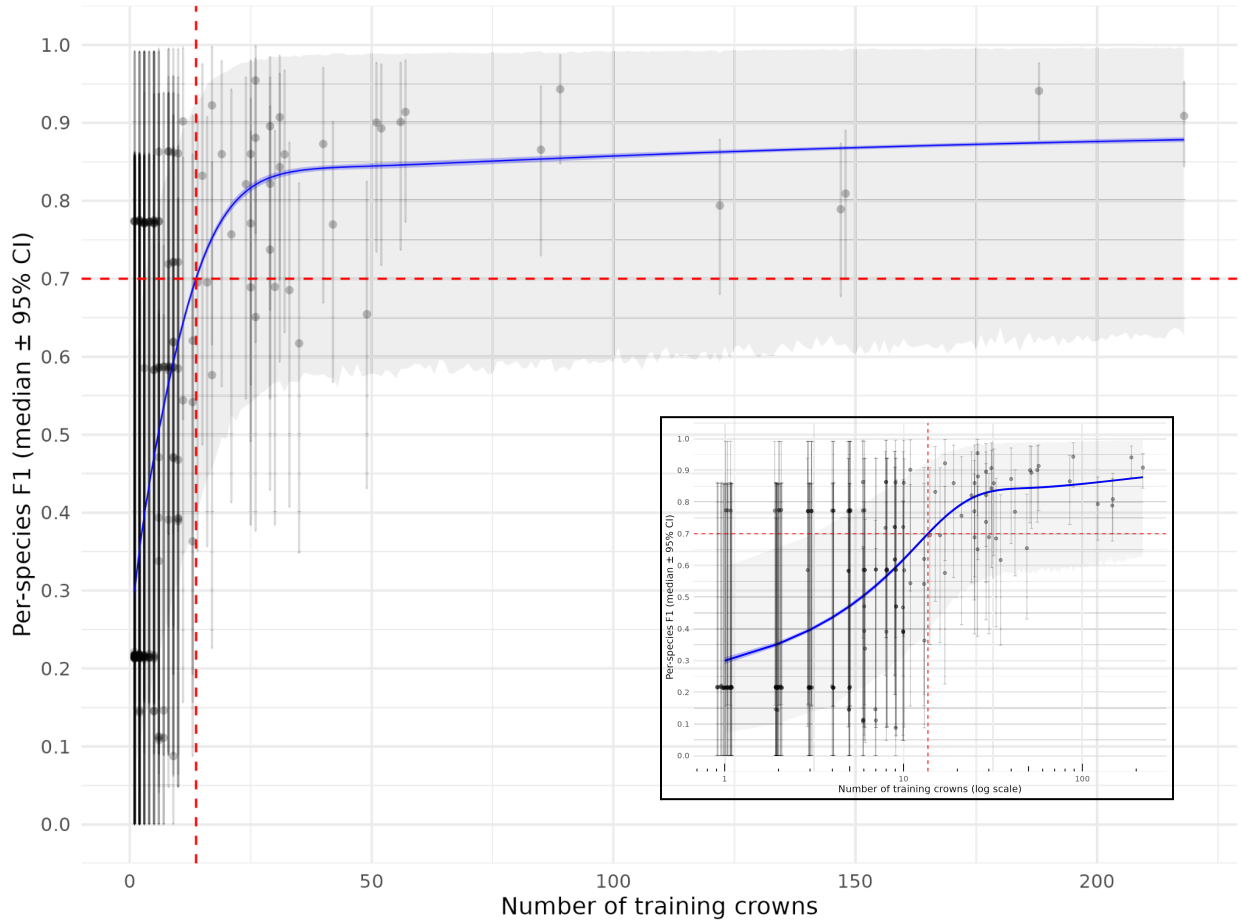

Figure 8*: Relationship between the number of training crowns per species and classification performance. Points are per-species posterior median F1 (harmonic mean of precision and recall) with 95% Bayesian credible intervals. The solid line is a beta-regression GAM fit; the inner shaded (thin blue) band is the 95% confidence interval for the mean trend, and the outer faint (grey) band is the 95% predictive interval (mean-trend uncertainty plus residual variation from the Beta likelihood). Point transparency reveals overplotting, which is common at small* n_{train} *where only a few discrete F1 values are possible. The vertical dashed red line at x=13 marks the smallest number of training crowns at which the fitted mean F1 reaches 0.7. Inset shows the same plot on a logarithmic x-axis to give more resolution to low numbers of training crowns.*

Spectral distinctiveness: In line with expectations, Beta regression indicated that both spectral hypervolume size ($z = -3.67$, $p = 0.00025$) and spectral overlap (summed Jaccard overlap; $z = -2.70$,

p = 0.0069) were significantly and negatively associated with per-species F1-score. A one-SD increase in either predictor reduced expected F1-score from ~0.81 to ~0.77–0.78 ($p \leq 0.007$).

Spatial distribution: Contrary to expectations, species whose individuals were more spatially dispersed (mean distance) across the landscape tended to exhibit higher classification accuracy (see Fig. 9). This finding indicates that the model did not overfit to localised spectral artefacts or environmental conditions (i.e. spatial autocorrelation), which was predicted to inflate performance, but instead benefited from training on individuals sampled across a wider range of environmental and scanning conditions. Broad spatial coverage likely exposed the classifier to more stable, species-specific spectral features, enhancing generalisability.

Crown size and tree height: Separability was higher for species with taller individuals and larger crowns, likely due to better illumination and a greater number of high-quality pixels. While consistent with expectations, the effect size was modest relative to other explanatory variables.

Leaf and bark traits: Contrary to expectations, leaf traits had relatively little influence on species separability. Leaf nitrogen was the only influential leaf trait but with negligible effect. Unexpectedly, species with thicker bark were more separable than those with thin bark.

Leaf phenology emerged as a key determinant of classification accuracy. The strongest predictor of species separability was the signal strength of leaf phenology (see Ball, 2023 for more details). Additionally, species with greatest among-tree alignment (minimal signal asynchrony) tended to be most accurately classified. This implies that evergreen species with weak leaf phenology signal and leaves of a wide range of ages were less separable than those with obvious cyclical leaf renewal, including deciduous and near-deciduous species.

Liana presence: As anticipated, species that incurred higher rates of liana infestation were more likely to be poorly classified. Lianas induce vegetative spectral mixing to a tree crown. The effect on species separability was moderate compared to other covariates.

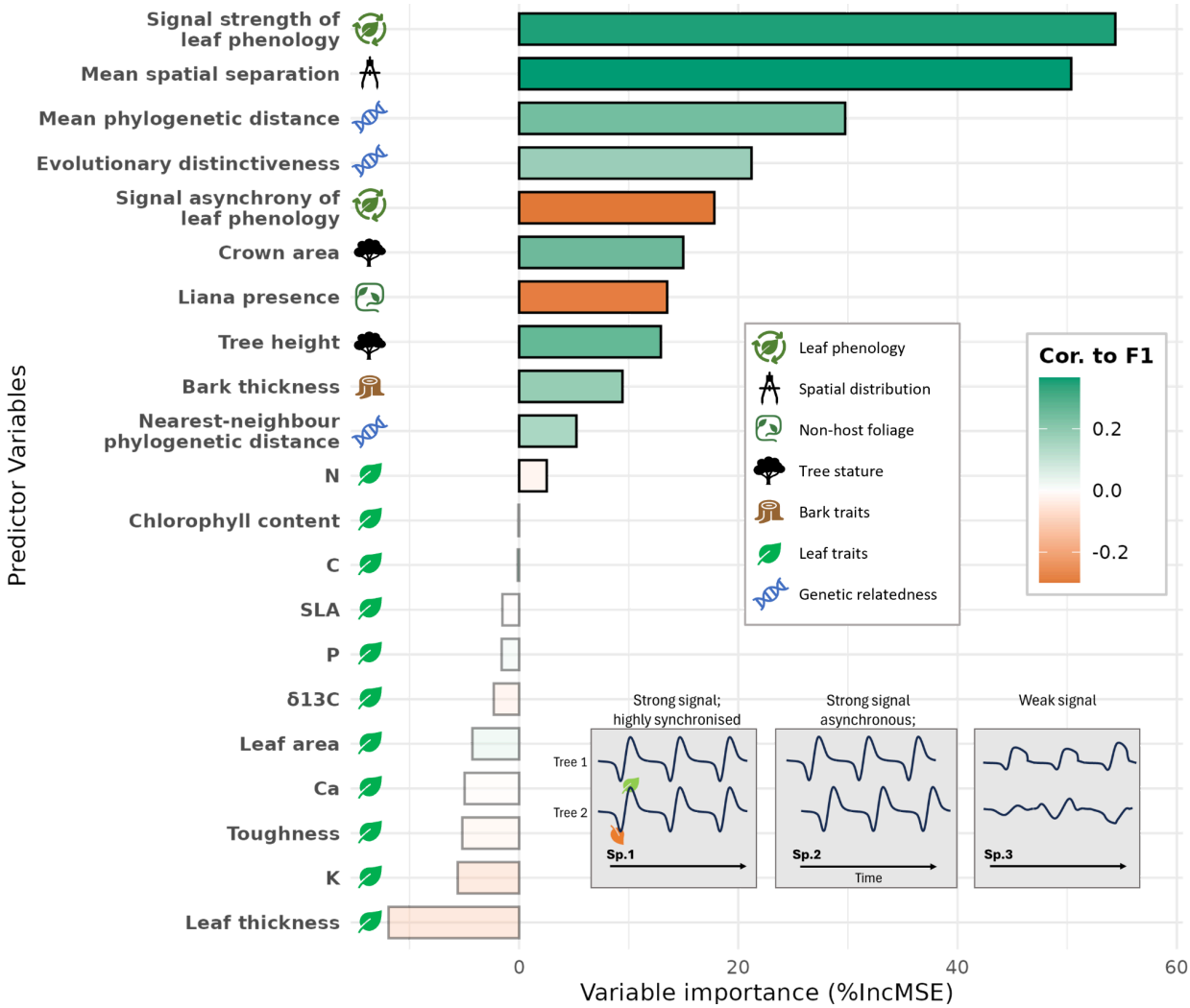

Figure 9. *Determinants of species classification separability (one-vs-rest F1-score) based on random forest modelling of species with at least 10 sampled crowns. The more positive the value of %IncMSE, the more helpful the variable was in predicting species classification accuracy, and those with near zero or negative values were uninformative. The direction of the correlation is indicated by the colour of the bars (green = positive, red = negative). The inset illustrates how the leaf phenology metrics were derived from the Green Chromatic Coordinate UAV time series of individual tree crowns.*

Relatedness to other species: As expected, species with high degrees of genetic relatedness to other species were more difficult to distinguish compared to those that were more evolutionarily distinct. Global isolation of a species (represented by MPD) had a strong effect whereas local proximity (NNPD) had little to no influence (i.e. deeper divergence was more influential than shallow).

The majority of wavebands showed a significant phylogenetic structure, including the FRE region (Fig. S.7): the Blomberg's K values (range 0.015-0.04) were very weak, whereas Pagel's λ values (0.28-0.75) suggested moderately strong phylogenetic signals (see discussion Section 4.5). As expected, comparison of phylogenetic distances with pairwise misclassification rates, we found that species with smaller phylogenetic distances had significantly higher confusion rates in classification ($\rho = -0.0437$, $p = 0.0008$; see Fig. S.18). Accordingly, misclassifications within the same genus or family occurred significantly more often than expected under random misassignment (see Fig. 10; full normalised confusion matrix as attached supplementary file).

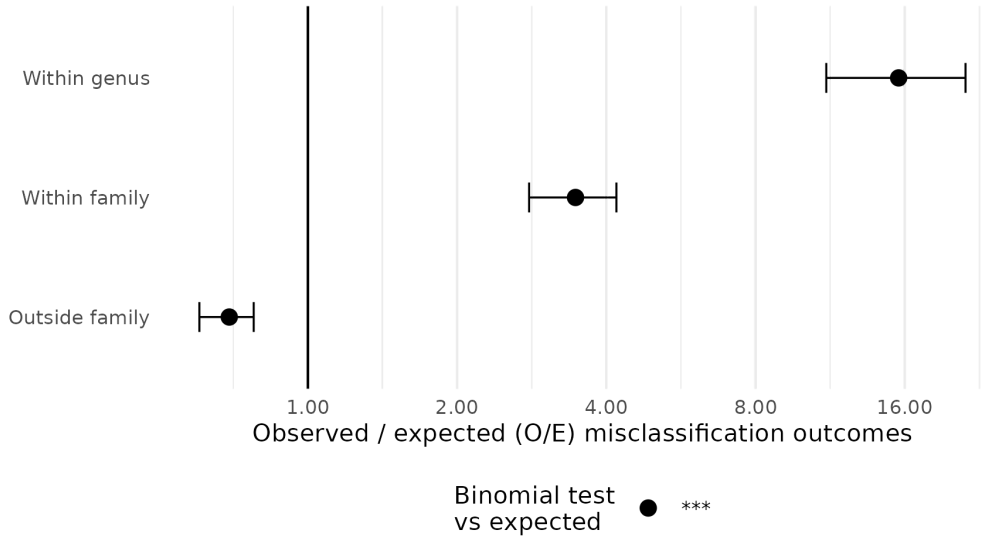

***Figure 10**: Overall observed vs expected composition of misclassification outcomes. Points show observed/expected (O/E) ratios for errors that fall within genus, within family (inclusive of genus), and outside family across the entire test set; horizontal bars are 95% Wilson confidence intervals, and the vertical line marks the null (O/E = 1). Expected values were obtained by distributing each species' misclassification count across candidate labels in proportion to label prevalence (excluding the true species), preserving the overall misclassification rate. Point shapes indicate exact binomial test significance versus this prevalence-weighted expectation. Note overlap: "within family" includes "within genus". Full confusion matrix provided as Supplementary File.*

4. Discussion

4.1 Overview

We demonstrate that species-level mapping of tropical forest canopies at landscape scale is now achievable at substantially greater scope and reliability than previously reported. While earlier airborne hyperspectral studies in the wet tropics typically resolved 20 or fewer common canopy species (Féret and Asner, 2013; Garzon-Lopez and Lasso, 2020; Laybros et al., 2019), our approach achieved robust accuracy (F1 > 0.7) for 65 species — representing the majority of visible canopy crown area in a hyperdiverse forest where hundreds of species may occur per hectare (Lee et al., 2002; Bohlman, 2015; Duque et al., 2017) . This advance rests on four interacting factors: (1) a dense, well-curated reference dataset of geo-located, confidently (ground) identified crowns; (2) precise field-to-image co-registration; (3) high-resolution, co-registered hyperspectral imagery; and (4) improved crown segmentation via a CNN with temporal consensus mapping. Together, these elements allowed us to move beyond the long-standing bottleneck of limited species scope and to quantify, in unprecedented detail, the spectral and ecological determinants of classification success. In the sections that follow, we examine the methodological gains and the physiological, morphological, spectral and phenological drivers of separability revealed by these analyses.

4.2 Advances in Crown Segmentation (Aim 1)

Temporal consensus mapping improved crown segmentation even when using a model trained elsewhere, showing that ensembling predictions across dates can substitute for limited local training data. We used detectree2 (Ball et al., 2023), a Mask R-CNN–based model (He et al., 2017) that integrates spectral, textural, and geometric cues to delineate individual tree crowns from UAV RGB imagery. Trained on a carefully validated set of field-mapped crowns, the model achieved state-of-the-art performance (Gan et al., 2023), even surpassing human annotators in complex canopies (Section S.1). The key innovation was temporal fusion: combining multiple UAV acquisitions to generate consensus crown maps. Remarkably, a model trained entirely on other sites matched or exceeded the performance of locally trained models once predictions were aggregated across dates (Fig. 4), underscoring the value of both temporal diversity in imagery and breadth in training data. This is, to our knowledge, the first demonstration in tropical forests that multi-date aggregation improves segmentation accuracy, offering a scalable route to monitoring forest structure, growth, mortality, and phenology across large landscapes.

4.3 Improved mapping of species (Aim 2)

High classification accuracy was achieved without resorting to highly complex algorithms: the relatively simple Linear Discriminant Analysis (LDA) consistently outperformed more flexible models such as SVMs and MLPs, echoing (Féret and Asner, 2013). While higher-capacity models captured fine-grained, within-crown variation, they tended to overfit to crown-specific artefacts, reducing transferability across individuals. This reinforces the need to constrain complex models — through approaches such as regularisation, representation learning, or multi-instance methods — so they capture robust, crown-level signals rather than noise or artefacts. Performance gains were instead driven by the quality of the input data and preprocessing. The reference dataset — over 3,500 geo-located crowns — was developed, curated, and field-validated over multiple years of intensive mapping campaigns. Labour-intensive image–field alignment yielded a rare, high-confidence training and evaluation resource. Such spatially explicit field datasets are essential for species-level ecological inference from remote sensing (Chave et al., 2019; Davies et al., 2021), yet remain scarce for testing mapping approaches at this taxonomic and spatial resolution (Laliberté et al., 2020).

High-quality hyperspectral imagery further underpinned performance: acquisitions were near-nadir and precisely co-registered to crowns, enabling fine-scale analyses of trait-linked separability (see Section 4.4; Fig. 6). Hyperdiverse forests present the additional challenge of many rare species (Jackson and Adam, 2021; ter Steege et al., 2013). Our inclusive training approach — retaining these taxa — did not reduce overall landscape-level accuracy, suggesting training pools should remain as broad as possible (Fig. S.4). Nevertheless, accurate identification of rare species remains difficult and will require targeted strategies (see Section 4.5).

4.4 Critical Spectral Regions for Species Classification (Aim 3)

Although retaining information across the broader spectrum was beneficial (Fig. S.9), the far-red edge (FRE; 748-775 nm) emerged as by far the most important region for species discrimination (Fig. 6) aligning with Badourdine et al. (2023). While prior work broadly emphasised red-edge and NIR reflectance for vegetation classification (e.g., Cochrane, 2000; Clark et al., 2005; Clark and Roberts, 2012; Dalponte et al., 2012; Fassnacht et al., 2016; Marconi et al., 2022; Hennessy et al., 2020), our fine-grained analysis pinpoints a specific zone just beyond the conventional red-edge inflection of 700–750 nm (Guo et al., 2023).

The FRE spans the transition between dominant chlorophyll absorption and internal leaf scattering characteristics of the NIR (Horler et al., 1983; Filella and Peñuelas, 1994; Gitelson et al., 2003; Hennessy et al., 2020). Reflectance here is shaped by pigment concentrations and species-specific microstructural attributes — mesophyll architecture, refractive index gradients, and intercellular airspace — that control scattering efficiency (Boochs et al., 1990; Ustin et al., 2009). Leaf water content, via effects on cellular turgor and optical path length, can further modulate absorption–scattering balance (Ustin et al., 2009). The lower importance of adjacent wavelengths indicates that the FRE functions as a spectral trait "confluence" where biochemical, structural, and physiological axes intersect to amplify interspecific contrast (Imran et al., 2020). This helps explain why species with broadly similar chlorophyll concentrations nevertheless diverge spectrally in this range (Curran et al., 1991; Blackburn, 1998).

Secondary regions also contributed. In the visible, 640-660 nm (red) likely reflects variation in chlorophyll b while 560-575nm (green) may capture differences in leaf structure, carotenoids, anthocyanins, and/or stress responses (Gitelson et al., 2003; Sims and Gamon, 2002; Carter and Knapp, 2001). In the SWIR, 1630–1700 nm corresponds to water absorption features and variation in cellulose/lignin composition (Gao, 1996; Ceccato et al., 2001; Serrano et al., 2002; Kokaly et al., 2009). These signals, while weaker than the FRE, underline the value of retaining information across the full spectrum to complement narrowband trait cues (Fig. S.5 ablation analysis).

Phenological and physiological differences among individuals can shift the red-edge position — and thus the FRE (Thomas and Gausman, 1977; Curran, 1989; Gitelson et al., 2003; Clark and Roberts, 2012). We treat timing and transferability in detail in Section 4.6.

4.5 Drivers of Species Separability (Aim 4)

Our analyses partitioned the factors underlying variation in species separability.

Number of crowns: Within this site, 13 training crowns per species were typically sufficient to achieve reliable classification (Fig. 7; F1-score > 0.7) (Ramezan et al., 2021; Shahinfar et al., 2020). Paracou's high species richness, with substantial spectral overlap among congeners, and complex phenological dynamics (Sabatier, 1983; Loubry, 1994) likely make this threshold conservative (however see influence of phylogeny below). Similar performance was sustained when an artificial dataset capped training crowns at 8 for species with ≥10 individuals (8:2 split with 5-fold CV). Thresholds will be sensor-, species-, and site-specific; analogous analyses elsewhere should guide

data collection. Many taxa fell below this threshold, often offering too few examples for the model to learn distinctive spectral patterns (Graves et al., 2023; Mouret et al., 2024). Improving predictions for such taxa will require not only advances in model architecture but also larger and more geographically representative labelled crown datasets, targeted field sampling, data augmentation, emerging one-/few-shot learning (Gevaert et al., 2025; Mu et al., 2025) and foundation-model (Huo et al., 2025) approaches. Without such strategies, mapping will remain biased toward dominant canopy species, limiting biodiversity assessments and understanding of the full species pool.

Spectral distinctiveness: Broad spectral hypervolumes and high overlap with other taxa were consistently harder to distinguish, consistent with the expectation that high intra-species variability dilutes a consistent signal, while close proximity to others increases confusion (Clark et al., 2005; Féret and Asner, 2013). The ecological and physical factors discussed below help explain why some species occupy broader or more crowded regions of spectral space.

Spatial dispersion of individuals across varied environmental conditions positively influenced accuracy, indicating the model benefited from exposure to species-specific spectral traits rather than localised artefacts (cf. Carlson et al., 2007; Tuia et al., 2011). This finding reinforces confidence in the generalisability of the spectral features learned by the classifier.

Taller trees with larger crowns were modestly easier to classify — likely due to illumination and pixel availability — but this effect was secondary. While this bias favours the upper canopy, its small magnitude is encouraging for species lower in the visible canopy.

Leaf traits, analysed singly or in combination, had surprisingly little influence on separability suggesting that this approach is suitable for a broad array of tree species. The association of thicker bark with improved classification merits further investigation, as it could directly or indirectly influence detectable crown-level spectra (Hadlich et al., 2018).

Phenology emerged as a key predictor of species separability. Species exhibiting consistent seasonal leafing patterns ("signal strength of leaf phenology") and synchronous phenological timing among individuals (low "signal asynchrony") showed markedly higher separability. Phenological synchrony likely constrains within-species spectral variability associated with leaf developmental stages, thus enhancing spectral comparability across individuals (Blackburn and Milton, 1995; Clark and Roberts, 2012; Takahashi Miyoshi et al., 2020). In contrast, species lacking clear phenological cues — such as evergreen taxa without distinct seasonal renewal — tended to produce more

confusion. Unique events like mass flowering or fruiting can also create strong spectral contrasts, potentially explaining the exceptionally high classification reliability observed in certain species (e.g., Couma guianensis, Xylopia nitida, and Manilkara bidentata; see Fig. S.13). The acute lack of detailed phenological records in the wet tropics calls for expanded collection of leaf and reproductive phenology data (Ball, 2023). See Section 4.6 for mechanisms and acquisition timing.

Lianas: A tendency for crowns to be infested by lianas reduced classification accuracy, likely due to spectral mixing between the vegetation of the host tree and that of the liana (van der Heijden et al., 2022). While this could reflect a species-level trait (e.g., liana susceptibility), it may also be a sampling artefact, with some individuals more affected than others.

Relatedness to other species: As expected, species embedded within clades of close relatives were harder to distinguish (Fig. 9). Misclassification (error rate) increased as mean phylogenetic distance (MPD) to the candidate community decreased, whereas nearest-neighbour phylogenetic distance (NNPD) explained little to no variation — indicating that deeper, clade-wide divergence matters more than the presence of a single close congener. Practically, phylogenetically isolated species (high MPD) were more separable even when a close relative was present, whereas species nested in dense clades remained challenging despite moderate NNPD. Accordingly, canopy mapping should be more tractable where dominant taxa are phylogenetically dispersed (e.g. Amazonian terra firme forests where "hyperdominants" span multiple families; ter Steege et al., 2013), and comparatively error-prone in forests with tight within-family radiations (e.g. Bornean mixed dipterocarp forests; Slik et al., 2003), implying larger training sets and more conservative accuracy expectations in the latter. This aligns with our finding that closely related species show more similar reflectance than distant relatives (Cavender-Bares et al., 2016; Madritch et al., 2014; McManus et al., 2016; Meireles et al., 2020; Schweiger et al., 2021). High Pagel's λ but low Blomberg's K in the spectral bands (Section 3.4) indicates a deep rather than shallow signal: covariance follows phylogeny overall, yet close relatives are not more similar than expected under Brownian motion, consistent with MPD — but not NNPD — predicting errors. A likely explanation is that lineage-conserved structural traits (e.g. lignin, cellulose, leaf anatomy) are stable across clades (Ackerly and Donoghue, 1998; Weng and Chapple, 2010), whereas pigments, water content, and stress responses are more labile (Donovan et al., 2011; Gitelson et al., 2003). Integrating phylogenetic, trait-based, and spectral frameworks (Baldeck and Asner, 2014; Nunes et al., 2017; Schmitt et al., 2022) will be key to disentangling lineage-conserved from context-dependent sources of spectral variation.

Taken together, these results show that spectral distinctiveness, allowing for species separability, reflects an interplay between species attributes and assemblage and survey factors. Phenology — especially clear, synchronous leaf cycles — was the strongest influence, while leaf traits had little predictive value. On the survey side, sufficient training sample size was critical, and broader spatial dispersion improved performance by exposing the model to varied conditions. Liana infestation and close phylogenetic relatedness tended to reduce separability. Recognising how biological and sampling influences interact helps target effort toward phenologically informative acquisition windows, greater representation of poorly sampled taxa, and designs that capture environmental breadth.

4.6 Limitations and Future Directions

Biological variability and timing: Within-species spectral variability — arising from leaf ontogeny (age-dependent shifts in pigment concentrations and water content), physiological stress (e.g. from drought, pests, or disturbance), liana infestation, crown position and microenvironment, and phenological stage — can obscure consistent species-level signals (Ollinger, 2011; Yang et al., 2014; Chavana-Bryant et al., 2019; Chen et al., 2022; Hesketh and Sánchez-Azofeifa, 2012). Even at a single time point, individuals of the same species may spectrally diverge (Reich, 1995; see Fig. S.8). Biological drivers interact with external sources of variation — including atmosphere, sun–sensor geometry, and illumination — which further shift or distort recorded spectra (Schläpfer et al., 2018; Theiler et al., 2019). Fully modelling the joint biochemical, structural and phenological controls with radiative-transfer frameworks remains intractable in hyperdiverse tropical canopies because trait covariation and crown architecture are difficult to parameterise at scale (André et al., 2021; Liu et al., 2022; Qi et al., 2022; Viskari et al., 2019; Yanagi and Costa, 2011). This limitation highlights the need for empirical, multi-temporal datasets and representation learning approaches that can absorb and disentangle such variability.

Phenological and physiological asynchrony among individuals can shift the red-edge position — and thus the FRE — through changes in pigments and stress state (Thomas and Gausman, 1977; Curran, 1989; Gitelson et al., 2003; Clark and Roberts, 2012), reducing separability, especially when imagery is acquired during periods of low leaf area or poor synchrony (Clark and Roberts, 2012; Hesketh and Sánchez-Azofeifa, 2012). Conversely, well-defined cycles constrain within-species variance by aligning leaf developmental stages, and distinct events (e.g. synchronous leaf flush or flowering) can accentuate between-species contrasts (Clark and Roberts, 2012; Takahashi Miyoshi et al., 2020). In

our case, the dry-season acquisition coincided with substantial leaf flushing. Future surveys aimed at species mapping should be timed to phenophases that maximise within-species alignment, thereby improving spectral comparability and classification reliability.

Multimodal classification: Our approach focused on spectral separation; additional data streams can further aid species with overlapping spectra. Structural LiDAR metrics (Quan et al., 2023; Wang et al., 2023; Zhong et al., 2022), high-resolution texture (Williams et al., 2022), and multi-season imagery (Takahashi Miyoshi et al., 2020; Modzelewska et al., 2021; Grybas and Congalton, 2021) provide complementary information on crown architecture, leaf angle distribution and phenological cycling. However, acquiring these modalities at matched spatial/temporal fidelity across large or remote sites is logistically demanding, and data fusion increases alignment and model-complexity burdens — practical constraints for operational tropical monitoring.

Addressing transferability: The goal is robust classifiers across time (seasonal/interannual variation) and space (different forests). Performance can degrade with modest shifts in acquisition conditions or community composition (Laybros et al., 2019). Some errors — image/label misalignment, occasional annotation mistakes – are tractable; others — atmospheric effects, geometry, illumination — require stronger radiometric and geometric corrections grounded in physical modelling. Beyond correction, phenology-aware methods and self-supervised representation learning (e.g. Barlow Twins; (Zbontar et al., 2021) applied to multi-temporal hyperspectral data can learn encodings that are invariant to nuisance variation and domain shift yet sensitive to phenological/physiological identity. With modest fine-tuning on local data, these approaches can support scalable biodiversity pipelines. Because phenological signal strength and alignment strongly predicted success here, encoding phenological fingerprints via multi-season time-series features should improve both within-site accuracy and cross-site transferability. Ultimately, progress requires integrating: (1) trait-informed spectral interpretation, (2) sensor harmonisation/noise modelling, and (3) temporal-aware machine learning.

Data priorities: Progress depends on well-labelled hyperspectral datasets spanning multiple seasons, years and sites, ideally with overlapping taxa to test generalisation (cf. NEON in the North American context; Marconi et al., 2022). Paracou has recently undergone a second hyperspectral scan which will be further complemented by equivalent acquisitions at Nouragues (French Guiana), providing a rare opportunity to quantify spatial and temporal transferability under realistic

conditions and to identify which spectral features or representations remain stable across environmental and technical variability.

Implications for hyperspectral sensing from space: Our findings have direct consequences for using multispectral (e.g., Sentinel-2) and hyperspectral (e.g., EnMAP; forthcoming CHIME, HYPXIM) satellites to map forest composition, traits, and phenological diversity at scale. Section 4.4 shows that the FRE (748–775 nm) is most informative. This window falls between Sentinel-2's red-edge bands (B6 = 740 nm; B7 = 783 nm) and, even with ideal placement, broad bandwidths (~15–20 nm) smear the subtle differences critical for species discrimination. Thus, while Sentinel-2's ~10 m pixels can capture large crowns, its spectral granularity is insufficient. Current and planned hyperspectral missions offer the narrow, contiguous bands needed for FRE-scale cues, but their coarser pixels (≥30 m) under-resolve individual tropical crowns. Absent robust spectral unmixing, this spectral–spatial trade-off is a first-order bottleneck. Additional variability from atmosphere, view geometry, phenology, illumination and stress will often be amplified in satellite data. Consequently, while trait or community-level patterns can be detected (Fassnacht et al., 2016; Féret and Asner, 2013; Aguirre-Gutiérrez et al., 2025), fine-grained classification of species remains elusive under current capabilities.

Nevertheless, fusion of data from multiple sensors would provide richer information. Ablation analyses showed that some spectral information is partially redundant across adjacent bands, suggesting potential for leveraging broader-band satellite data with appropriate feature engineering. Field-validated airborne models, co-registered with temporally matched satellite observations, could support the development of hybrid classifiers that bridge scales. In the longer term, fusing UAV, airborne, and satellite data may offer a pathway to pan-tropical species mapping. The mismatch in spatial resolution, spectral precision, and atmospheric distortion is profound. High-resolution airborne platforms will continue to serve as essential tools for advancing species-level forest mapping and establishing the benchmarks that will guide the development of next-generation satellite systems.

Conclusions

This study demonstrates that accurate, species-level mapping of tropical forest canopy trees at landscape scales is achievable by combining temporally fused CNN-based crown segmentation with high-resolution airborne hyperspectral imagery. At the Paracou site in French Guiana, we classified 169 species, achieving robust accuracy (F1-score > 0.7) for 65 species, accounting for over 70% of

the canopy area—significantly exceeding the scope of previous studies typically limited to fewer than 20 species (Féret and Asner, 2013; Laybros et al., 2019; Garzon-Lopez and Lasso, 2020). Importantly, this performance relied on straightforward, interpretable modelling supported by high-quality spatially explicit datasets and precise image-to-field alignment. A key insight was the identification of the narrow FRE spectral region (748–775 nm), which integrates multiple biochemical, structural, and physiological traits essential for species discrimination (Curran, 1989; Gitelson et al., 2003; Ustin et al., 2009). Despite these advances, scaling species-level mapping to satellite platforms remains challenging due to spatial and spectral limitations. Current and upcoming satellite sensors either lack sufficient spectral resolution or have inadequate spatial resolution, creating a fundamental barrier for direct airborne-to-satellite model transferability (Fassnacht et al., 2016; Saberioon et al., 2023). Hybrid approaches integrating UAV, airborne, and satellite data, anchored by temporally consistent spectral features, offer a promising strategy forward. Continued investment in extensive annotated datasets and advances in self-supervised machine learning techniques will enable robust, scalable biodiversity monitoring. Achieving this vision requires coordinated efforts in ground-truth data collection across diverse sites and conditions, transitioning from mapping known species distributions to discovering novel ecological patterns and monitoring global forest biodiversity—tree by tree, crown by crown. As forests face intensifying pressures from climate change and land-use conversion, the ability to track individual species at scale will be vital for informing conservation priorities, restoration planning, and biodiversity policy.

ACKNOWLEDGEMENTS

J.G.C.B. was supported by the NERC C-CLEAR doctoral training programme (PDAG/501) and the Franklinia Foundation. 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 Verley (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 CP 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 [INSERT ZENODO LINK].

All analysis code and scripts 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

Ackerly, D.D., and Donoghue, M.J. (1998) Leaf size, sapling allometry, and Corner's rules: phylogeny and correlated evolution in maples (Acer). *Am Nat*. 152: 767–791.

Aguirre-Gutiérrez, J., Rifai, S.W., Deng, X., ter Steege, H., Thomson, E., Corral-Rivas, J.J., et al. (2025) Canopy functional trait variation across Earth's tropical forests. *Nature*. 641: 129–136.

Anderegg, W.R.L., Hicke, J.A., Fisher, R.A., Allen, C.D., Aukema, J., Bentz, B., et al. (2015) Tree mortality from drought, insects, and their interactions in a changing climate. *New Phytol*. 208: 674–683.

André, F., de Wergifosse, L., de Coligny, F., Beudez, N., Ligot, G., Gauthray-Guyénet, V., et al. (2021) Radiative transfer modeling in structurally complex stands: towards a better understanding of parametrization. *Ann For Sci*. 78: 1–21.

Araujo, R.F., Grubinger, S., Celes, C.H.S., Negrón-Juárez, R.I., Garcia, M., Dandois, J.P., et al. (2021) Strong temporal variation in treefall and branchfall rates in a tropical forest is related to extreme rainfall: results from 5 years of monthly drone data for a 50 ha plot. *Biogeosciences*. 18: 6517–6531.

Asner, G.P., and Martin, R.E. (2009) Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. *Front Ecol Environ*. 7: 269–276.

Asner, G.P., and Martin, R.E. (2011) Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian forest. *New Phytol*. 189: 999–1012.

Aubry-Kientz, M., Laybros, A., Weinstein, B., Ball, J.G.C., Jackson, T., Coomes, D., et al. (2021) Multisensor Data Fusion for Improved Segmentation of Individual Tree Crowns in Dense Tropical Forests. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 14: 3927–3936.

Badourdine, C., Féret, J.-B., Pélissier, R., and Vincent, G. (2023) Exploring the link between spectral variance and upper canopy taxonomic diversity in a tropical forest: influence of spectral processing and feature selection. *Remote Sens Ecol Conserv*. 9: 235–250.

Badourdine, C., Féret, J.-B., Pélissier, R., and Vincent, G. (n.d.) Key challenges in estimating taxonomic diversity using hyperspectral data: findings from spectral simulations in a hyper-diverse tropical forest. *J Veg Sci*.

Baldeck, C.A., and Asner, G.P. (2014) Improving Remote Species Identification through Efficient Training Data Collection. *Remote Sens*. 6: 2682–2698.

Baldeck, C.A., Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., Kellner, J.R., et al. (2015) Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLoS One*. 10: e0118403.

Ball, J.G.C. (2023).

Ball, J.G.C., Hickman, S.H.M., Jackson, T.D., Koay, X.J., Hirst, J., Jay, W., et al. (2023) Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. *Remote Sens Ecol Conserv*. 9: 641–655.

Baraloto, C., Hardy, O.J., Paine, C.E.T., Dexter, K.G., Cruaud, C., Dunning, L.T., et al. (2012) Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *J Ecol*. 100: 690–701.

Barry, K.E., and Schnitzer, S.A. (2021) Are we missing the forest for the trees? Conspecific negative density dependence in a temperate deciduous forest. *PLoS One*. 16: e0245639.

Béland, M., and Kobayashi, H. (2024) Drivers of deciduous forest near-infrared reflectance: A 3D radiative transfer modeling exercise based on ground lidar. *Remote Sens Env*. 302: 113951.

Blackburn, G.A. (1998) Quantifying Chlorophylls and Caroteniods at Leaf and Canopy Scales: An Evaluation of Some Hyperspectral Approaches. *Remote Sens Environ*. 66: 273–285.

Blackburn, G.A., and Milton, E.J. (1995) Seasonal variations in the spectral reflectance of deciduous tree canopies. *Int J Remote Sens*. 16: 709–720.

Blonder, B., Lamanna, C., Violle, C., and Enquist, B.J. (2014) The n-dimensional hypervolume. *Glob Ecol Biogeogr*. 23: 595–609.
Bohlman, S.A. (2015) Species Diversity of Canopy Versus Understory Trees in a Neotropical Forest: Implications for Forest Structure, Function and Monitoring. *Ecosystems*. 18: 658–670.
Bonal, D., Bosc, A., Ponton, S., Goret, J.-Y., Burban, B., Gross, P., et al. (2008) Impact of severe dry season on net ecosystem exchange in the Neotropical rainforest of French Guiana. *Glob Chang Biol*. 14: 1917–1933.
Boochs, F., Kupfer, G., Dockter, K., and Kühbauch, W. (1990) Shape of the red edge as vitality indicator for plants. *Int J Remote Sens*. 11: 1741–1753.
Bredin, Y.K., Peres, C.A., and Haugaasen, T. (2020) Forest type affects the capacity of Amazonian tree species to store carbon as woody biomass. *Ecol Manage*. 473: 118297.
Breiman, L. (2001) Random Forests. *Mach Learn*. 45: 5–32.
Cao, Y., Ball, J.G.C., Coomes, D.A., Steinmeier, L., Knapp, N., Wilkes, P., et al. (2023) Benchmarking airborne laser scanning tree segmentation algorithms in broadleaf forests shows high accuracy only for canopy trees. *Int J Appl Earth Obs Geoinf*. 123: 103490.
Carlson, K.M., Asner, G.P., Hughes, R.F., Ostertag, R., and Martin, R.E. (2007) Hyperspectral Remote Sensing of Canopy Biodiversity in Hawaiian Lowland Rainforests. *Ecosystems*. 10: 536–549.
Carter, G.A., and Knapp, A.K. (2001) Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am J Bot*. 88: 677–684.
Cavender-Bares, J., Meireles, J.E., Couture, J.J., Kaproth, M.A., Kingdon, C.C., Singh, A., et al. (2016) Associations of Leaf Spectra with Genetic and Phylogenetic Variation in Oaks: Prospects for Remote Detection of Biodiversity. *Remote Sens*. 8: 221.
Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., and Grégoire, J.-M. (2001) Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sens Env*. 77: 22–33.
Chan, A.H.Y., Barnes, C., Swinfield, T., and Coomes, D.A. (2021) Monitoring ash dieback (Hymenoscyphus fraxineus) in British forests using hyperspectral remote sensing. *Remote Sens Ecol Conserv*. 7: 306–320.
Chance, C.M., Coops, N.C., Crosby, K., and Aven, N. (2016) Spectral Wavelength Selection and Detection of Two Invasive Plant Species in an Urban Area. *Can J Remote Sens*. 42: 27–40.
Chavana-Bryant, C., Malhi, Y., Anastasiou, A., Enquist, B.J., Cosio, E.G., Keenan, T.F., et al. (2019) Leaf age effects on the spectral predictability of leaf traits in Amazonian canopy trees. *Sci Total Env*. 666: 1301–1315.
Chave, J., Condit, R., Lao, S., Caspersen, J.P., Foster, R.B., and Hubbell, S.P. (2003) Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *J Ecol*. 91: 240–252.
Chave, J., Davies, S.J., Phillips, O.L., Lewis, S.L., Sist, P., Schepaschenko, D., et al. (2019) Ground Data are Essential for Biomass Remote Sensing Missions. *Surv Geophys*. 40: 863–880.
Chen, L., Zhang, Y., Nunes, M.H., Stoddart, J., Khoury, S., Chan, A.H.Y., et al. (2022) Predicting leaf traits of temperate broadleaf deciduous trees from hyperspectral reflectance: can a general model be applied across a growing season? *Remote Sens Env*. 269: 112767.
Clark, M.L., and Roberts, D.A. (2012) Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sens*. 4: 1820–1855.
Clark, M.L., Roberts, D.A., and Clark, D.B. (2005) Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sens Env*. 96: 375–398.
Cochrane, M.A. (2000) Using vegetation reflectance variability for species level classification of hyperspectral data. *Int J Remote Sens*. 21: 2075–2087.
Condit, R., Engelbrecht, B.M.J., Pino, D., Pérez, R., and Turner, B.L. (2013) Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical

trees. *Proc Natl Acad Sci U A*. 110: 5064–5068.
Curran, P.J. (1989) Remote sensing of foliar chemistry. *Remote Sens Env*. 30: 271–278.
Curran, P.J., Dungan, J.L., Macler, B.A., and Plummer, S.E. (1991) The effect of a red leaf pigment on the relationship between red edge and chlorophyll concentration. *Remote Sens Environ*. 35: 69–76.
Dalponte, M., Bruzzone, L., and Gianelle, D. (2012) Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sens Env*. 123: 258–270.
Dalponte, M., and Coomes, D.A. (2016) Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods Ecol Evol*. 7: 1236–1245.
Dalponte, M., Ørka, H.O., Ene, L.T., Gobakken, T., and Næsset, E. (2014) Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sens Env*. 140: 306–317.
Davies, S.J., Abiem, I., Abu Salim, K., Aguilar, S., Allen, D., Alonso, A., et al. (2021) ForestGEO: Understanding forest diversity and dynamics through a global observatory network. *Biol Conserv*. 253: 108907.
Dimov, L.D., Chambers, J.L., and Lockhart, B.R. (2013) Tree species exhibit complex patterns of distribution in bottomland hardwood forests. *Ann Sci*. 70: 813–823.
Donovan, L.A., Maherali, H., Caruso, C.M., Huber, H., and de Kroon, H. (2011) The evolution of the worldwide leaf economics spectrum. *Trends Ecol Evol*. 26: 88–95.
Duque, A., Muller-Landau, H.C., Valencia, R., Cardenas, D., Davies, S., de Oliveira, A., et al. (2017) Insights into regional patterns of Amazonian forest structure, diversity, and dominance from three large terra-firme forest dynamics plots. *Biodivers Conserv*. 26: 669–686.
Durgante, F.M., Higuchi, N., Almeida, A., and Vicentini, A. (2013) Species Spectral Signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy. *For Ecol Manag*. 291: 240–248.
van Ewijk, K.Y., Randin, C.F., Treitz, P.M., and Scott, N.A. (2014) Predicting fine-scale tree species abundance patterns using biotic variables derived from LiDAR and high spatial resolution imagery. *Remote Sens Env*. 150: 120–131.
FAO (2020) Global forest resources assessment 2020: Main report. Food & Agriculture Organization of the United Nations.
Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., et al. (2016) Review of studies on tree species classification from remotely sensed data. *Remote Sens Env*. 186: 64–87.
Féret, J.-B., and Asner, G.P. (2011) Spectroscopic classification of tropical forest species using radiative transfer modeling. *Remote Sens Env*. 115: 2415–2422.
Féret, J.-B., and Asner, G.P. (2013) Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Trans Geosci Remote Sens*. 51: 73–84.
Feurer, D., and Vinatier, F. (2018) Joining multi-epoch archival aerial images in a single SfM block allows 3-D change detection with almost exclusively image information. *ISPRS J Photogramm Remote Sens*. 146: 495–506.
Filella, I., and Peñuelas, J. (1994) The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *Int J Remote Sens*. 15: 1459–1470.
Fricker, G.A., Ventura, J.D., Wolf, J.A., North, M.P., Davis, F.W., and Franklin, J. (2019) A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. *Remote Sens Basel*. 11: 2326.
Gan, Y., Wang, Q., and Iio, A. (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 Sens*. 15.

Gao, B.-C. (1996) NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Env*. 58: 257–266.
Garzon-Lopez, C.X., and Lasso, E. (2020) Species Classification in a Tropical Alpine Ecosystem Using UAV-Borne RGB and Hyperspectral Imagery. *Drones*. 4: 69.
Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013) Bayesian Data Analysis, 3rd edn. Chapman and Hall/CRC, New York.
Gevaert, C.M., Aguiar Pedro, A., Ku, O., Cheng, H., Chandramouli, P., Dadrass Javan, F., et al. (2025) Explainable few-shot learning workflow for detecting invasive and exotic tree species. *Sci Rep*. 15: 23238.
Ghiyamat, A., and Shafri, H.Z.M. (2010) A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *Int J Remote Sens*. 31: 1837–1856.
Gitelson, A.A., Gritz, Y., and Merzlyak, M.N. (2003) Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J Plant Physiol*. 160: 271–282.
Goetz, S.J., Baccini, A., Laporte, N.T., Johns, T., Walker, W., Kellndorfer, J., et al. (2009) Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance Manag*. 4: 2.
Gómez-Candón, D., Virlet, N., Labbé, S., Jolivot, A., and Regnard, J.-L. (2016) Field phenotyping of water stress at tree scale by UAV-sensed imagery: new insights for thermal acquisition and calibration. *Precis Agric*. 17: 786–800.
Gourlet-Fleury, S., Guehl, J.-M., and Laroussinie, O. (2004) Ecology and management of a neotropical rainforest : lessons drawn from Paracou, a long-term experimental research site in French Guiana. Elsevier.
Graves, S.J., Marconi, S., Stewart, D., Harmon, I., Weinstein, B., Kanazawa, Y., et al. (2023) Data science competition for cross-site individual tree species identification from airborne remote sensing data. *PeerJ*. 11: e16578.
Grybas, H., and Congalton, R.G. (2021) A Comparison of Multi-Temporal RGB and Multispectral UAS Imagery for Tree Species Classification in Heterogeneous New Hampshire Forests. *Remote Sens*. 13: 2631.
Guo, Y., Mokany, K., Ong, C., Moghadam, P., Ferrier, S., and Levick, S.R. (2023) Plant species richness prediction from DESIS hyperspectral data: A comparison study on feature extraction procedures and regression models. *ISPRS J Photogramm Remote Sens*. 196: 120–133.
Hadlich, H.L., Durgante, F.M., dos Santos, J., Higuchi, N., Chambers, J.Q., and Vicentini, A. (2018) Recognizing Amazonian tree species in the field using bark tissues spectra. *For Ecol Manag*. 427: 296–304.
Hastings, J.H., Ollinger, S.V., Ouimette, A.P., Sanders-DeMott, R., Palace, M.W., Ducey, M.J., et al. (2020) Tree Species Traits Determine the Success of LiDAR-Based Crown Mapping in a Mixed Temperate Forest. *Remote Sens*. 12: 309.
He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017) . In *Proceedings of the IEEE International Conference on Computer Vision*. pp. 2961–2969.
van der Heijden, G.M.F., Proctor, A.D.C., Calders, K., Chandler, C.J., Field, R., Foody, G.M., et al. (2022) Making (remote) sense of lianas. *J Ecol*. 110: 498–513.
Hennessy, A., Clarke, K., and Lewis, M. (2020) Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. *Remote Sens*. 12: 113.
Hesketh, M., and Sánchez-Azofeifa, G.A. (2012) The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. *Remote Sens Env*. 118: 73–82.
Hogan, J.A., Zimmerman, J.K., Uriarte, M., Turner, B.L., and Thompson, J. (2016) Land-use history augments environment–plant community relationship strength in a Puerto Rican wet forest. *J Ecol*. 104: 1466–1477.

Horler, D.N.H., Dockray, M., and Barber, J. (1983) The red edge of plant leaf reflectance. *Int J Remote Sens*. 4: 273–288.
Huo, C., Chen, K., Zhang, S., Wang, Zeyu, Yan, H., Shen, J., et al. (2025) When Remote Sensing Meets Foundation Model: A Survey and Beyond. *Remote Sens*. 17: 179.
Imran, H.A., Gianelle, D., Rocchini, D., Dalponte, M., Martín, M.P., Sakowska, K., et al. (2020) VIS-NIR, Red-Edge and NIR-Shoulder Based Normalized Vegetation Indices Response to Co-Varying Leaf and Canopy Structural Traits in Heterogeneous Grasslands. *Remote Sens*. 12: 2254.
Jackson, C.M., and Adam, E. (2021) Machine Learning Classification of Endangered Tree Species in a Tropical Submontane Forest Using WorldView-2 Multispectral Satellite Imagery and Imbalanced Dataset. *Remote Sens*. 13: 4970.
Jacquemoud, S., and Baret, F. (1990) PROSPECT: A model of leaf optical properties spectra. *Remote Sens Env*. 34: 75–91.
Jeffreys, H. (1946) An invariant form for the prior probability in estimation problems. *Proc R Soc Lond Ser Math Phys Sci*. 186: 453–461.
Jucker, T., Bongalov, B., Burslem, D.F.R.P., Nilus, R., Dalponte, M., Lewis, S.L., et al. (2018) Topography shapes the structure, composition and function of tropical forest landscapes. *Ecol Lett*. 21: 989–1000.
Kalacska, M., Bohlman, S., Sanchez-Azofeifa, G.A., Castro-Esau, K., and Caelli, T. (2007) Hyperspectral discrimination of tropical dry forest lianas and trees: Comparative data reduction approaches at the leaf and canopy levels. *Remote Sens Environ*. 109: 406–415.
Kamoske, A.G., Dahlin, K.M., Read, Q.D., Record, S., Stark, S.C., Serbin, S.P., et al. (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: 1440–1460.
Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M.D., and Dormann, C.F. (2022) Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. *ISPRS Open J Photogramm Remote Sens*. 5: 100018.
Kenkel, N.C. (1988) Pattern of self-thinning in Jack Pine: Testing the random mortality hypothesis. *Ecology*. 69: 1017–1024.
Knyazikhin, Y., Schull, M.A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., et al. (2013) Hyperspectral remote sensing of foliar nitrogen content. *Proc Natl Acad Sci U A*. 110: E185-92.
Kokaly, R.F., Asner, G.P., Ollinger, S.V., Martin, M.E., and Wessman, C.A. (2009) Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sens Env*. 113: S78–S91.
Laliberté, E., Schweiger, A.K., and Legendre, P. (2020) Partitioning plant spectral diversity into alpha and beta components. *Ecol Lett*. 23: 370–380.
LaManna, J.A., Mangan, S.A., Alonso, A., Bourg, N.A., Brockelman, W.Y., Bunyavejchewin, S., et al. (2017) Plant diversity increases with the strength of negative density dependence at the global scale. *Science*. 356: 1389–1392.
Laybros, A., Aubry-Kientz, M., Féret, J.-B., Bedeau, C., Brunaux, O., Derroire, G., et al. (2020) Quantitative Airborne Inventories in Dense Tropical Forest Using Imaging Spectroscopy. *Remote Sens*. 12.
Laybros, A., Schläpfer, D., Féret, J.-B., Descroix, L., Bedeau, C., Lefevre, M.-J., et al. (2019) Across Date Species Detection Using Airborne Imaging Spectroscopy. *Remote Sens*. 11.
Lee, H.S., Davies, S.J., LaFrankie, J.V., Tan, S., Yamakura, T., Itoh, A., et al. (2002) Floristic and structural diversity of mixed Dipterocarp forest in Lambir Hills National Park, Sarawak, Malaysia. *J Trop Sci*. 14: 379–400.
Liu, C., Calders, K., Meunier, F., Gastellu-Etchegorry, J.P., Nightingale, J., Disney, M., et al. (2022)

Implications of 3D Forest Stand Reconstruction Methods for Radiative Transfer Modeling: A Case Study in the Temperate Deciduous Forest. *J Geophys Res Atmospheres*. 127: e2021JD036175.
Liu, Y., Zhan, Z., Ren, L., Ze, S., Yu, L., Jiang, Q., et al. (2021) Hyperspectral evidence of early-stage pine shoot beetle attack in Yunnan pine. *Ecol Manage*. 497: 119505.
Loubry, D. (1994) La phénologie des arbres caducifoliés en forêt guyanaise (5° de latitude nord): illustration d'un déterminisme à composantes endogène et exogène. *Can J Bot*. 72: 1943–1957.
Madritch, M.D., Kingdon, C.C., Singh, A., Mock, K.E., Lindroth, R.L., and Townsend, P.A. (2014) Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. *Philos Trans R Soc Lond B Biol Sci*. 369: 20130194.
Marconi, S., Weinstein, B.G., Zou, S., Bohlman, S.A., Zare, A., Singh, A., et al. (2022) Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network. *Remote Sens Env*. 282: 113264.
Martin, F.-M., Müllerová, J., Borgniet, L., Dommanget, F., Breton, V., and Evette, A. (2018) Using Single- and Multi-Date UAV and Satellite Imagery to Accurately Monitor Invasive Knotweed Species. *Remote Sens*. 10: 1662.
Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., et al. (2021) Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens Env*. 256: 112322.
McDowell, N., Pockman, W.T., Allen, C.D., Breshears, D.D., Cobb, N., Kolb, T., et al. (2008) Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New Phytol*. 178: 719–739.
McGrath, L.J., van Riper, C., 3rd, and Fontaine, J.J. (2009) Flower power: tree flowering phenology as a settlement cue for migrating birds. *J Anim Ecol*. 78: 22–30.
McManus, K.M., Asner, G.P., Martin, R.E., Dexter, K.G., Kress, W.J., and Field, C.B. (2016) Phylogenetic Structure of Foliar Spectral Traits in Tropical Forest Canopies. *Remote Sens*. 8: 196.
Meireles, J.E., Cavender-Bares, J., Townsend, P.A., Ustin, S., Gamon, J.A., Schweiger, A.K., et al. (2020) Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytol*. 228: 485–493.
Modzelewska, A., Kamińska, A., Fassnacht, F.E., and Stereńczak, K. (2021) Multitemporal hyperspectral tree species classification in the Białowieża Forest World Heritage site. *For Int J For Res*. 94: 464–476.
Mouret, F., Morin, D., Planells, M., and Vincent-Barbaroux, C. (2024).
Mu, Y., Xiong, Z., Wang, Y., Shahzad, M., Essl, F., Kleunen, M. van, et al. (2025).
Nepstad, D.C., Tohver, I.M., Ray, D., Moutinho, P., and Cardinot, G. (2007) Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology*. 88: 2259–2269.
Nuijten, R.J.G., Coops, N.C., Goodbody, T.R.H., and Pelletier, G. (2019) Examining the Multi-Seasonal Consistency of Individual Tree Segmentation on Deciduous Stands Using Digital Aerial Photogrammetry (DAP) and Unmanned Aerial Systems (UAS). *Remote Sens*. 11: 739.
Nunes, M.H., Davey, M.P., and Coomes, D.A. (2017) On the challenges of using field spectroscopy to measure the impact of soil type on leaf traits. *Biogeosciences*. 14: 3371–3385.
Ollinger, S.V. (2011) Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytol*. 189: 375–394.
Park, J.Y., Muller-Landau, H.C., Lichstein, J.W., Rifai, S.W., Dandois, J.P., and Bohlman, S.A. (2019) Quantifying Leaf Phenology of Individual Trees and Species in a Tropical Forest Using Unmanned Aerial Vehicle (UAV) Images. *Remote Sens*. 11: 1534.
Pereira Martins-Neto, R., Garcia Tommaselli, A.M., Imai, N.N., Honkavaara, E., Miltiadou, M., Saito Moriya, E.A., et al. (2023) Tree Species Classification in a Complex Brazilian Tropical Forest

Using Hyperspectral and LiDAR Data. *Forests*. 14: 945.
Prospere, K., McLaren, K., and Wilson, B. (2014) Plant Species Discrimination in a Tropical Wetland Using In Situ Hyperspectral Data. *Remote Sens*. 6: 8494–8523.
Qi, J., Xie, D., Jiang, J., and Huang, H. (2022) 3D radiative transfer modeling of structurally complex forest canopies through a lightweight boundary-based description of leaf clusters. *Remote Sens Environ*. 283: 113301.
Quan, Y., Li, M., Hao, Y., Liu, J., and Wang, B. (2023) Tree species classification in a typical natural secondary forest using UAV-borne LiDAR and hyperspectral data. *GIScience Remote Sens*. 60: 2171706.
Ramezan, C.A., Warner, T.A., Maxwell, A.E., and Price, B.S. (2021) Effects of Training Set Size on Supervised Machine-Learning Land-Cover Classification of Large-Area High-Resolution Remotely Sensed Data. *Remote Sens*. 13: 368.
Reich, P.B. (1995) Phenology of tropical forests: patterns, causes, and consequences. *Can J Bot*. 73: 164–174.
Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., et al. (2013) Climate extremes and the carbon cycle. *Nature*. 500: 287–295.
Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3: 217–223.
Rowland, L., da Costa, A.C.L., Galbraith, D.R., Oliveira, R.S., Binks, O.J., Oliveira, A. a. R., et al. (2015) Death from drought in tropical forests is triggered by hydraulics not carbon starvation. *Nature*. 528: 119–122.
Sabatier, D. (1983) Fructification et dissémination en forêt guyanaise : l'exemple de quelques espèces ligneuses. l'Université des Sciences et Techniques du Languedoc.
Sabat-Tomala, A., Raczko, E., and Zagajewski, B. (2020) Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sens*. 12: 516.
Saberioon, M., Khosravi, V., Brom, J., Gholizadeh, A., and Segl, K. (2023) Examining the sensitivity of simulated EnMAP data for estimating chlorophyll-a and total suspended solids in inland waters. *Ecol Inform*. 75: 102058.
Schläpfer, D., Hueni, A., and Richter, R. (2018) Cast Shadow Detection to Quantify the Aerosol Optical Thickness for Atmospheric Correction of High Spatial Resolution Optical Imagery. *Remote Sens*. 10.
Schmitt, S., Fischer, F., Ball, J., Barbier, N., Boisseaux, M., Bonal, D., et al. (2024) TROLL 4.0: representing water and carbon fluxes, leaf phenology, and intraspecific trait variation in a mixed-species individual-based forest dynamics model – Part 2: Model evaluation for two Amazonian sites. *EGUsphere*. 1–54.
Schmitt, S., Trueba, S., Coste, S., Ducouret, É., Tysklind, N., Heuertz, M., et al. (2022) Seasonal variation of leaf thickness: An overlooked component of functional trait variability. *Plant Biol*. 24: 458–463.
Schweiger, A.K., Cavender-Bares, J., Kothari, S., Townsend, P.A., Madritch, M.D., Grossman, J.J., et al. (2021) Coupling spectral and resource-use complementarity in experimental grassland and forest communities. *Proc Biol Sci*. 288: 20211290.
Seidler, T.G., and Plotkin, J.B. (2006) Seed dispersal and spatial pattern in tropical trees. *PLoS Biol*. 4: e344.
Serbin, S.P., Singh, A., McNeil, B.E., Kingdon, C.C., and Townsend, P.A. (2016) Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecol Appl*. 24: 1651–1669.
Serrano, L., Peñuelas, J., and Ustin, S.L. (2002) Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural

signals. *Remote Sens Env*. 81: 355–364.
Shahinfar, S., Meek, P., and Falzon, G. (2020).
Shamaoma, H., Chirwa, P.W., Zekeng, J.C., Ramoelo, A., Hudak, A.T., Handavu, F., et al. (2023) Use of Multi-Date and Multi-Spectral UAS Imagery to Classify Dominant Tree Species in the Wet Miombo Woodlands of Zambia. *Sensors*. 23.
Sims, D.A., and Gamon, J.A. (2002) Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens Env*. 81: 337–354.
Slik, J.W.F., Paoli, G., McGuire, K., Amaral, I., Barroso, J., Bastian, M., et al. (2013) Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics: Large trees and tropical forest biomass. *Glob Ecol Biogeogr*. 22: 1261–1271.
Slik, J.W.F., Poulsen, A.D., Ashton, P.S., Cannon, C.H., Eichhorn, K. a. O., Kartawinata, K., et al. (2003) A floristic analysis of the lowland dipterocarp forests of Borneo. *J Biogeogr*. 30: 1517–1531.
ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomão, R.P., Guevara, J.E., et al. (2013) Hyperdominance in the Amazonian tree flora. *Science*. 342: 1243092.
Takahashi Miyoshi, G., Imai, N.N., Garcia Tommaselli, A.M., Antunes de Moraes, M.V., and Honkavaara, E. (2020) Evaluation of Hyperspectral Multitemporal Information to Improve Tree Species Identification in the Highly Diverse Atlantic Forest. *Remote Sens*. 12: 244.
Theiler, J., Ziemann, A., Matteoli, S., and Diani, M. (2019) Spectral variability of remotely sensed target materials: Causes, models, and strategies for mitigation and robust exploitation. *IEEE Geosci Remote Sens Mag*. 7: 8–30.
Thomas, J.R., and Gausman, H.W. (1977) Leaf reflectance vs. Leaf chlorophyll and carotenoid concentrations for eight crops [1]. *Agron J*. 69: 799–802.
Tötsch, N., and Hoffmann, D. (2021) Classifier uncertainty: evidence, potential impact, and probabilistic treatment. *PeerJ Comput Sci*. 7: e398.
Tuia, D., Volpi, M., Copa, L., Kanevski, M., and Munoz-Mari, J. (2011) A survey of active learning algorithms for supervised remote sensing image classification. *IEEE J Sel Top Signal Process*. 5: 606–617.
Tuomisto, H., Ruokolainen, K., and Yli-Halla, M. (2003) Dispersal, environment, and floristic variation of western Amazonian forests. *Science*. 299: 241–244.
Ustin, S.L. (2013) Remote sensing of canopy chemistry. *Proc Natl Acad Sci U A*. 110: 804–805.
Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., et al. (2009) Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sens Env*. 113: S67–S77.
Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., and Green, R.O. (2004) Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. *Bioscience*. 54: 523–534.
Vaglio Laurin, G., Cheung-Wai Chan, J., Chen, Q., Lindsell, J.A., Coomes, D.A., Guerriero, L., et al. (2014) Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. *PLoS One*. 9: e97910.
Valencia, R., Balslev, H., and Paz Y Miño C, G. (1994) High tree alpha-diversity in Amazonian Ecuador. *Biodivers Conserv*. 3: 21–28.
Viskari, T., Shiklomanov, A., Dietze, M.C., and Serbin, S.P. (2019) The influence of canopy radiation parameter uncertainty on model projections of terrestrial carbon and energy cycling. *PLoS ONE*. 14: e0216512.
Vleminckx, J., Fortunel, C., Valverde-Barrantes, O., Timothy Paine, C.E., Engel, J., Petronelli, P., et al. (2021) Resolving whole-plant economics from leaf, stem and root traits of 1467 Amazonian tree species. *Oikos*. 130: 1193–1208.
Wagner, F., Hérault, B., Stahl, C., Bonal, D., and Rossi, V. (2011) Modeling water availability for trees in tropical forests. *Agric Meteorol*. 151: 1202–1213.

Wang, B., Liu, J., Li, J., and Li, M. (2023) UAV LiDAR and Hyperspectral Data Synergy for Tree Species Classification in the Maoershan Forest Farm Region. *Remote Sens*. 15: 1000.

Watt, M.S., Leonardo, E.M.C., Estarija, H.J.C., Massam, P., de Silva, D., O'Neill, R., et al. (2021) Long-term effects of water stress on hyperspectral remote sensing indicators in young radiata pine. *Ecol Manage*. 502: 119707.

Weinstein, B.G., Marconi, S., Zare, A., Bohlman, S.A., Singh, A., Graves, S.J., et al. (2024) Individual canopy tree species maps for the National Ecological Observatory Network. *PLOS Biol*. 22: e3002700.

Weng, J.-K., and Chapple, C. (2010) The origin and evolution of lignin biosynthesis. *New Phytol*. 187: 273–285.

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., and Tompalski, P. (2016) Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Can J Remote Sens*. 42: 619–641.

Wielgosz, M., Puliti, S., Xiang, B., Schindler, K., and Astrup, R. (2024) SegmentAnyTree: A sensor and platform agnostic deep learning model for tree segmentation using laser scanning data. *Remote Sens Environ*. 313: 114367.

Williams, J., Jackson, T.D., Schönlieb, C.-B., Swinfield, T., Irawan, B., Achmad, E., et al. (2022) Monitoring early-successional trees for tropical forest restoration using low-cost UAV-based species classification. *Front Glob Chang*. 5.

Yanagi, S.N.M., and Costa, M.H. (2011) Modeling radiative transfer in tropical rainforest canopies: sensitivity of simulated albedo to canopy architectural and optical parameters. *An Acad Bras Cienc*. 83: 1231–1242.

Yang, X., Tang, J., and Mustard, J.F. (2014) Beyond leaf color: Comparing camera-based phenological metrics with leaf biochemical, biophysical, and spectral properties throughout the growing season of a temperate deciduous forest. *J Geophys Res Biogeosciences*. 119: 181–191.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. (2021).

Zhong, H., Lin, W., Liu, H., Ma, N., Liu, K., Cao, R., et al. (2022) Identification of tree species based on the fusion of UAV hyperspectral image and LiDAR data in a coniferous and broad-leaved mixed forest in Northeast China. *Front Plant Sci*. 13.