

Automated detection of an insect-induced keystone vegetation phenotype using airborne LiDAR

Zhengyang Wang^{1,2,3}  | Robert Huben³ | Peter B. Boucher¹  | Chase Van Amburg¹  | Jimmy Zeng³  | Nina Chung¹  | Jocelyn Wang¹  | Jeffrey King³ | Richard J. Knecht^{1,4}  | Ivy Ng'iru^{5,6,7}  | Augustine Baraza⁵ | Christopher C. M. Baker¹  | Dino J. Martins^{5,8}  | Naomi E. Pierce^{1,4}  | Andrew B. Davies¹ 

¹Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA; ²College of Life Sciences, Sichuan University, Chengdu, China; ³Insect Renaissance Group, Boston, Massachusetts, USA; ⁴Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, USA; ⁵Mpala Research Centre, Nanyuki, Laikipia, Kenya; ⁶School of Biosciences, Cardiff University, Cardiff, UK; ⁷UK Centre for Ecology and Hydrology, Wallingford, UK and ⁸Turkana Basin Institute, Stony Brook University, Stony Brook, New York, USA

Correspondence

Zhengyang Wang

Email: zhengyangw@hotmail.com

Naomi E. Pierce

Email: npierce@oeb.harvard.edu

Andrew B. Davies

Email: andrew_davies@g.harvard.edu

Handling Editor: Paul Galpern

Abstract

1. Ecologists, foresters and conservation practitioners need 'biodiversity scanners' to effectively inventory biodiversity, audit conservation progress and track changes in ecosystem function. Quantifying biological diversity using remote sensing methods remains challenging, especially for small invertebrates. However, insect aggregations can drastically alter landscapes and vegetation, and these 'extended phenotypes' could serve as environmental landmarks of insect presence in remotely sensed data.
2. To test the feasibility of this approach, we studied symbiotic ants that alter the canopy shape of whistling thorn acacias (*Acacia* [syn. *Vachellia*] *drepanolobium*), a keystone tree species of the black cotton soils of east African savannas. We demonstrate a protocol for using light detection and ranging (LiDAR) data to collect, prepare (including a customizable tree-segmentation algorithm) and apply a convolutional neural network-based classification for the detection of ant-inhabited acacia tree phenotypic variations. Applying this protocol enabled us to effectively detect intra-specific tree phenotypic variation induced by insects.
3. Surveying ant occupancy across 16 ha and 9680 acacia trees took 1000 work hours, whereas surveyed patterns of ant distribution were replicated by our trained classifier using only an hour-long airborne LiDAR collection time.
4. We suggest that large-scale surveys of insect occupancy (including insect-vector disease) can be automated through a combination of airborne LiDAR and machine learning.

KEYWORDS

ant, extended phenotype, LiDAR, remote sensing, savanna

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

1 | INTRODUCTION

Ecologists, foresters and conservation practitioners need 'biodiversity scanners'—timely, cost-effective landscape observation technology to inventory biodiversity, audit conservation progress and track changes in ecosystem function (Bush et al., 2017; Jetz et al., 2019; Ji et al., 2022). One promising candidate is the use of light detection and ranging (LiDAR) scanners to provide high-resolution scans of terrain and vegetation structure across vast, difficult-to-access landscapes (Boucher et al., 2023; Kellner et al., 2019). LiDAR scanners emit pulses of light into the survey environment and use temporal differences in received light reflection to reconstruct surrounding objects as a collection of point clouds in three-dimensional space (Gatzolis & Andersen, 2008). Such LiDAR scans generate landscape-level measurement of terrain and vegetation structure, such as tree canopy height, cover and density (Atkins et al., 2018; Terryn et al., 2023). When combined with ground data from animals (e.g. GPS tracking, camera traps, field observations), new insights on patterns of animal behaviour and distribution can be deduced (Vierling et al., 2008), including movement strategies (Evans et al., 2020), nest site selection (Davies et al., 2019) and predator–prey interactions (e.g. landscape of fear, Davies et al., 2021). From a community ecology perspective, indices of structural complexity have been shown to correlate with occupancy and species richness of bats, birds and invertebrate species (see Goetz et al., 2007 as one of the earliest examples; see review by Davies & Asner, 2014) in similar ways to how remotely sensed spectral diversity can be used as a proxy for plant species diversity (see review by Wang & Gamon, 2019). However, establishing correlations between structural or spectral diversity and biological diversity is not the same as direct species detection. Directly quantifying biological diversity using LiDAR, therefore, remains a challenging task.

Direct detection of invertebrates with remote sensing is especially difficult. Most insects are small and cannot be distinguished from even high-resolution (i.e. <10cm) remote sensing data, for example using unoccupied aerial vehicles (UAVs) or terrestrial LiDAR (but see Rhodes et al., 2022 for a review of indirect remote sensing methods for insects, and Fisher et al., 2021 or Menz et al., 2022 for direct telemetry-based methods). However, understanding the presence and ecosystem functioning of these 'little things that run the world' (Wilson, 1987) is an integral part of understanding landscape and agricultural health (Dangles & Casas, 2019; Losey & Vaughan, 2006). Large (social or eusocial) insect aggregations can drastically alter landscapes and vegetation (Bentz & Endreson, 2004; Propastin, 2013; Schowalter, 2017), and these 'extended phenotypes' (Dawkins, 1983) could serve as environmental landmarks of insect presence in LiDAR data.

Perhaps the only prominent example of using LiDAR (or other remote sensing techniques) to detect insects' extended phenotype to date is the identification of termite mounds in African savannas (e.g. Davies et al., 2014; Hockridge et al., 2022; Levick et al., 2010), with the mound structure being consistent enough for machine learning

algorithms to detect them in LiDAR-derived digital terrain models (Brodrick et al., 2019; Davies et al., 2020). Just like termites leaving their mark on the land surface, many insect species (and often insect-vectored parasites) induce characteristic phenotypic changes in their vegetative host or habitat, which could possibly be observed with LiDAR: common examples include large hives constructed by bees (Cook et al., 2018), silk webs of spiders, ants and caterpillars that span across canopies (Avilés & Tufiño, 1998; Devarajan, 2016; Fitzgerald & Willer, 1983) and *Taphrina* fungi-induced 'witches' broom' (Christita et al., 2022). Other examples are wasps inducing large galls on tree branches (e.g. oak apples, whose tannin-rich extracts are used to make ink, Stone & Schönrogge, 2003), and *Myrmelachista schumanni* ants using formic acid to poison all plants except their host trees, creating monoculture 'Devil's gardens' amid the Amazonian rainforest (Frederickson et al., 2005). Moreover, insect or insect-vectored pestilence (e.g. *Dendrolimus* caterpillars, fall webworms, pine wood nematodes) can completely defoliate plant canopies (Day & Leather, 1997; Ji et al., 2011; Speight & Wylie, 2001), which will likely be visible in LiDAR point clouds. If such vegetation alterations are (1) morphologically indicative of unique insect symbiont species and (2) detectable in LiDAR scans, machine learning methods could be utilized to identify them (Brodrick et al., 2019).

To test the feasibility of such an approach, we chose the symbiotic ants of the whistling thorn acacia (*Vachellia* [syn. *Acacia*] *drepanolobium*, hereby acacia) as our study system. Acacia is a keystone tree species of the black cotton soils of east African savannas, forming monoculture stands that fix nitrogen and carbon and provide food for large browsers such as elephants (*Loxodonta africana*) and giraffes (*Giraffa* spp.) (Boyle et al., 2019; Palmer & Young, 2017). One such monoculture stand at Mpala Research Centre (MRC), Laikipia, Kenya, has been intensely studied for decades (Figure 1a,b). In this landscape, each acacia tree typically recruits one of four species of symbiotic ants: *Crematogaster sjostedti*, *C. mimosae*, *C. nigriceps* or *Tetraponera penzigi*. Three of the four ant species (*C. mimosae*, *C. nigriceps* and *T. penzigi*) are obligate mutualists, nesting only in acacias and defending their host trees against enemies such as pathogens and herbivores in return for housing and nectar (Palmer & Brody, 2007; Stanton et al., 1999); the fourth species, *C. sjostedti*, is free-living, typically nesting in the soil or rotten logs but regularly patrolling nearby acacias (Palmer & Brody, 2013). Each species of ant engages in territorial warfare against the three other species, with a competitive hierarchy among the four species (Palmer et al., 2000). An equilibrium is thought to be maintained because of a trade-off whereby less competitive species have adaptive advantages in dispersal, colonization, defensive behaviour and energy requirements (Palmer, 2003; Palmer et al., 2013). Since a single acacia tree lives much longer than a colony of any of the ant species, trees experience considerable turnover in their ant occupants over their lifetimes. Simulations have shown that a tree that has been occupied by all four ant species has a higher fitness than one occupied by a fewer number of species (Palmer et al., 2010). Changes in community composition could be indicative of drought-induced vegetation die-off, fire-induced succession events or invasive species incursion

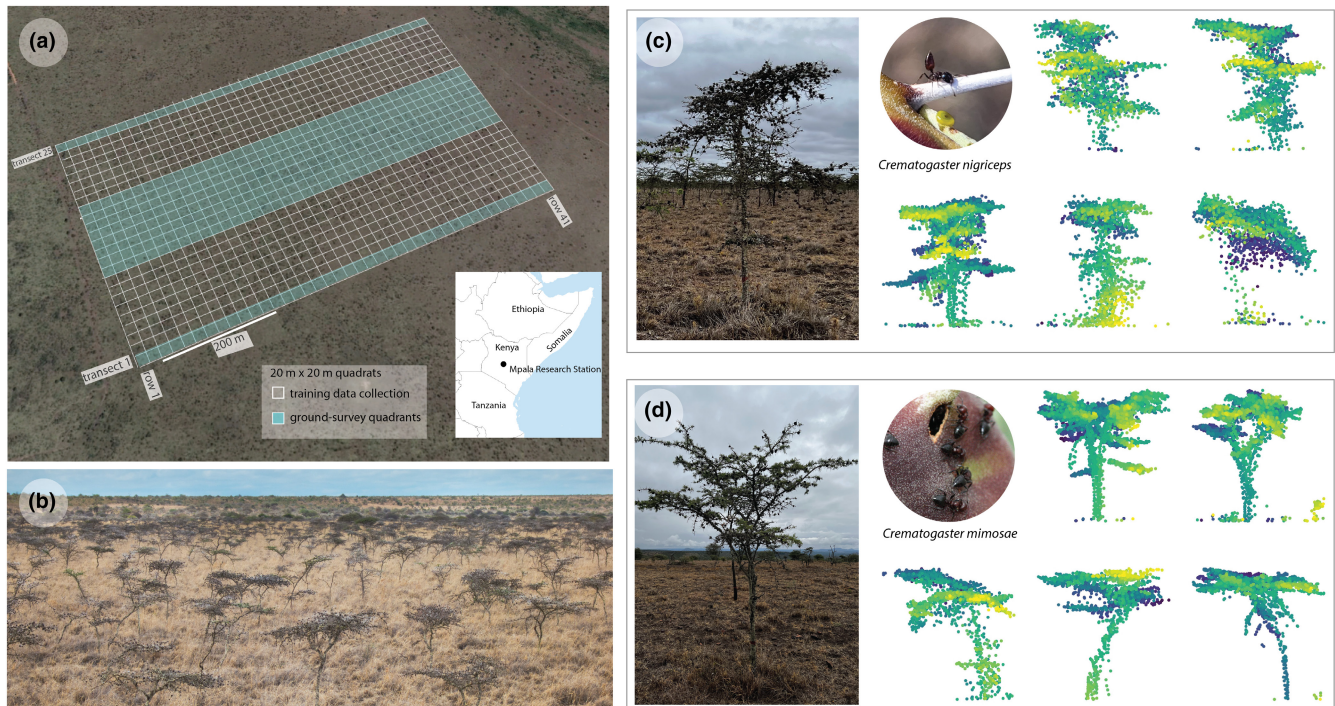


FIGURE 1 Ant-induced canopy shape variations in acacia (*Acacia drepanolobium*) monoculture in Laikipia, Kenya. (a) The study plot (25 transects \times 41 rows, 41 ha) on black cotton soil, above which UAV LiDAR data were collected. Blue-shaded quadrats represent ground-survey areas where the ant occupancy of each tree was recorded 16 ha (410 quadrats). Non-shaded quadrats are areas where training data for the CNN classifier were collected. (b) Aerial view of the acacia monoculture. Each tree contains one species of ant occupant (photo credit: CCMB). (c) Acacia trees occupied by *Crematogaster nigriceps* (our detection target) show stumped and clustered canopies. LiDAR point clouds on the right are exemplars of training input labels for *C. nigriceps*-occupied trees (photo credit: NC). (d) Acacia trees occupied by *Crematogaster mimosae*, the most abundant of the four ant species in the plot (photo credit: DJM). LiDAR point clouds on the right are exemplars of training input labelled as *C. mimosae*-occupied trees. In both (c) and (d), each training image is normalized for tree height, with points coloured to indicate proximity to the viewer.

(e.g. invasive *Pheidole megacephala* ants out-competes symbiotic ants, leading to increased herbivore and pathogen damage and loss in above-ground carbon sequestration in the region; see Milligan et al., 2021, 2022). Therefore, monitoring the ant occupancy of each tree can be used as a means of assessing ecosystem stability (Câmara et al., 2019; Houadria & Menzel, 2017; Vandermeer et al., 2008). However, censusing the ant occupancy of each of the acacia trees in MRC is time-consuming, as each tree needs to be observed in the field to determine its ant occupant.

Field observations have suggested that certain acacia canopy shapes are indicative of their associated ant species. In particular, workers of *C. nigriceps* actively trim their host trees to reduce canopy contact with neighbouring trees occupied by other (more aggressive) ant species, resulting in a stumped and clustered canopy architecture (Stanton et al., 1999). The trimming behaviour reduces acacia flower production to the extent of sterilizing the host trees (Young et al., 1996)—blurring the line between a mutualist and a parasite. We hypothesized that the distinctive and relatively consistent phenotypic difference between trees occupied by *C. nigriceps* and those occupied by the other three ant species could be captured in LiDAR point clouds (Figure 1c,d), which would allow us to algorithmically identify *C. nigriceps*-occupied acacias using machine learning models.

Convolutional neural networks (CNN), a style of deep learning model for image and video machine learning, has increasingly been used to identify biophysical features of ecological interest, such as termite mounds (Davies et al., 2020), coral reefs (Arsad et al., 2023) and oil palm plantations (Li et al., 2019), from large remotely sensed datasets. Moreover, CNN-based classifiers have been used to identify different species of trees from LiDAR point clouds of forests in the Mediterranean (Allen et al., 2022), temperate Europe and the United States (Seidel et al., 2021) and subtropical Asia (Zou et al., 2017). The overall classification accuracy of these studies ranges from 80% to 96%, but they have all used segmented, pre-labelled point clouds as training data. To our knowledge, (1) there has not been a demonstrably feasible protocol covering the complete process from the collection of LiDAR data, to the labelling of training data, to the construction of a CNN classifier and (2) the effectiveness of bespoke in silico high-accuracy tree classification models has not been verified in a field setting. Furthermore, (3) while existing CNN classifiers focus on tree species delineation, an arguably more challenging application would be to classify intra-species phenotypic variation, such as within trees of *A. drepanolobium* occupied by different ant symbionts.

Here, we demonstrate a protocol for using LiDAR data to collect, prepare (including a customizable tree-segmentation algorithm), train

and apply a CNN-based classifier for the detection of *C. nigriceps*-occupied, within-species acacia phenotypic variation. Crucially, to validate the ecological relevance of our methods, we surveyed 16 ha of acacia savanna for patterns of *C. nigriceps* occupancy and compared our ground-truth data with that generated from LiDAR-based automated detection. After describing our pipeline and evaluating its effectiveness for recapitulating ground-based ecosystem inferences, we suggest how large-scale surveys of insect occupancy (or insect-vector-disease) can be automated through a combination of airborne LiDAR and machine learning.

2 | MATERIALS AND METHODS

2.1 | Overview of methodology

Field work was conducted under Government of Kenya NACOSTI permits (NACOSTI/P/22/18988, NACOSTI/P/21/12633). All UAV flights were conducted with permission from the Kenya Civil Aviation Authority (KCAA/OPS/2117/4).

We studied the distribution of *C. nigriceps* with field surveys across 16 ha of savanna acacia monoculture (Figure 1a, areas in blue

shade) as described in Section 2.2. We then examined whether the distribution patterns of *C. nigriceps* obtained from these ground surveys could be recapitulated by our airborne LiDAR analysis pipeline, which included LiDAR data collection and processing (Section 2.3, Figure 2a,b), tree segmentation (Sections 2.4 and 2.5, Figure 2c) and CNN classifier training (Sections 2.6 and 2.7, Figure 2d). Finally, analysis comparing community dynamics between ground surveys and classifier predictions are described in Section 2.8.

Ground-survey results (Supporting Information S1), with code and LiDAR point clouds for CNN training (Supporting Information S2), are deposited at Zenodo (Wang, 2023, <https://doi.org/10.5281/zenodo.8175736>). Our script for tree segmentation, with user instructions, is deposited on Github (Huben, 2024, <https://doi.org/10.5281/zenodo.10570318>).

2.2 | Ground survey of ant occupancy

Our study area, the western portion of the Forest Global Earth Observatory (ForestGEO) plot at Mpala Research Centre, on the Laikipia Plateau in central Kenya (0°17'22"N; 36°51'56"E; mean annual temperature=17.7°C; mean annual precipitation=842mm), is a

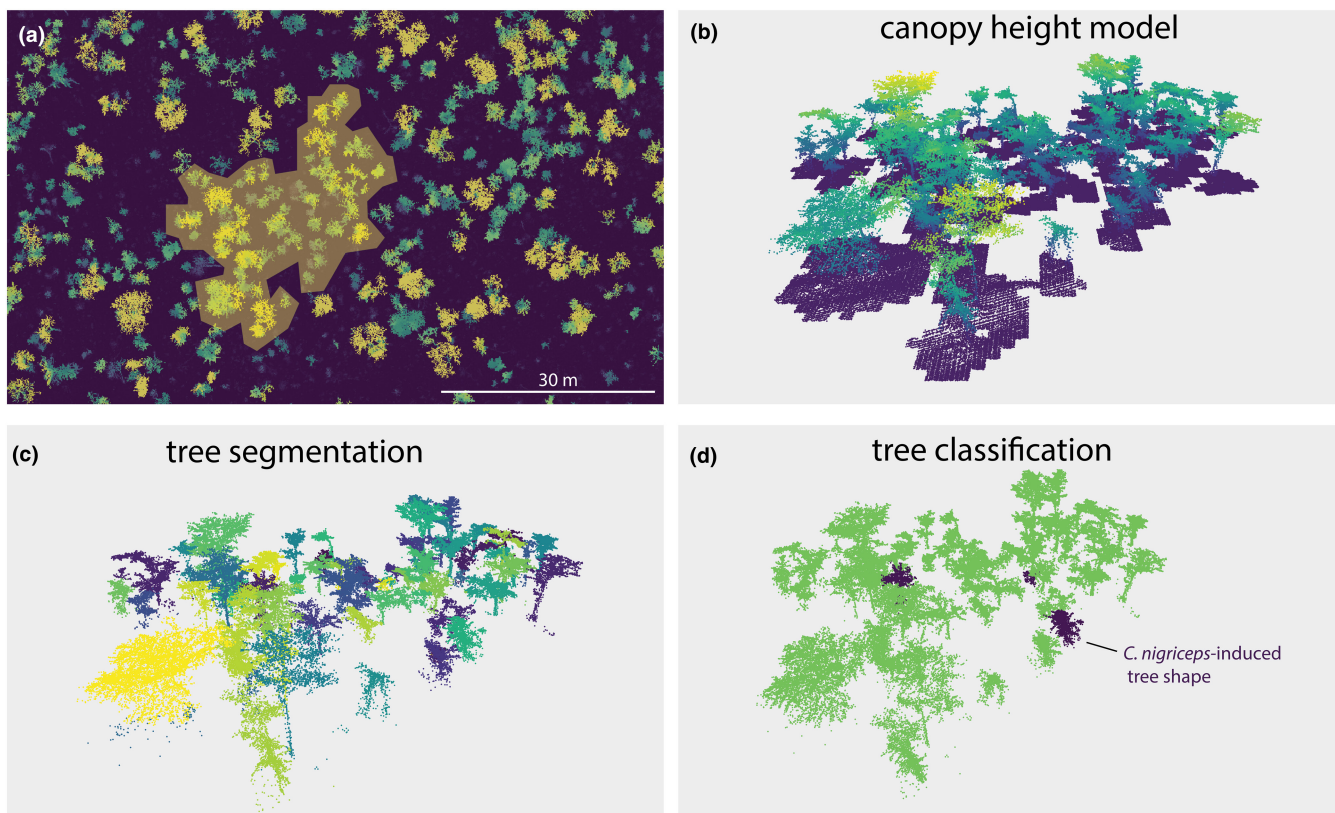


FIGURE 2 Pipeline for automated detection of insect-induced vegetation phenotype from airborne LiDAR data. (a) A top-down view of a part of the canopy height model generated from LiDAR data. Colours indicate tree height, with yellow representing taller trees. An exemplar ‘tree patch’ of closely located trees with overlapping canopies (yellow shade) is selected for tree segmentation. (b) Point cloud of the selected tree patch. (c) After implementing the customized tree segmentation, point clouds representing each individual tree in the patch are identified and shown in distinct colours. (d) Each individual tree point cloud was input into the CNN classifier. Dark colours indicate trees predicted as occupied by *Crematogaster nigriceps*.

A. drepanolobium savanna monoculture, located at 1800m above sea level on black cotton soil west of an escarpment (the eastern side of the escarpment, at 1600m, is red soil with more diverse vegetation composition). The site is divided into 25 transects and 41 rows, forming a total of 1025 quadrats, each 20m×20m (Figure 1a, quadrats delineated by lines). Acacia density increases from the escarpment edge (row 41) to the western quadrats (row 1), with as many as 90 trees per quadrat on the western edge. Each acacia is typically occupied by only one of the four ant symbionts at any one time (see Section 1).

From June to July 2022, we surveyed the ant occupancy of each acacia in transects 10–17, as well as in transects 1 and 25 (a total of 410 quadrats, Figure 1a, areas in blue shade; hereby referred to as 'ground survey area' or 'survey area'). It was logistically infeasible to survey all 1025 quadrats. Our survey scheme sampled a gradient of acacia density from east to west (rows 1–41), and covered both the northernmost and southernmost transects. We disturbed each acacia by gently tapping its branches and recorded the species of the alarmed ants defending their host tree. Overlap between tree canopies prevented us from geotagging every surveyed tree (Figure S1A,B).

2.3 | UAV LiDAR data collection and processing

We collected high-resolution LiDAR data over the study site using the Harvard Animal Landscape Observatory (HALO) in January 2022. HALO is equipped with a Riegl VUX-1LR LiDAR sensor (Phoenix LiDAR Systems, Austin, Texas, USA), which for this study was flown using a Freefly Alta-X rotary-wing unoccupied aerial vehicle (UAV; Freefly Systems, Woodinville, Washington, USA). Multiple flights of ~15–20min each were performed to cover the site. The UAV was flown 50m above-ground level at 8m/s in a serpentine flight pattern with a line speed of 114 lines/s and an 820kHz pulse rate. Flight trajectories were corrected during post-processing using GPS data from a nearby mobile base station. LiDAR data were then denoised, classified (following Axelsson, 2000) and aligned using the Terrasolid software suite (Terrasolid Ltd, Espoo, Finland). The average point density was ~300 points/m² after denoising. We created a digital terrain model (DTM) of the site using a triangulated model of ground points. The height above-ground was then computed for each point based on its vertical distance to the DTM, and a canopy height model (CHM) was constructed using the maximum height of each point at a 0.25 m resolution.

2.4 | CHM-based segmentation

To provide an initial segmentation of trees, we used the watershed transform propagation algorithm of Meyer and Beucher (1990), implemented as the SegmentCrowns function in the R package ForestTools (Plowright & Roussel, 2021), with a linear variable radius window function of 0.3 times tree height (Popescu & Wynne, 2004) and no restriction on minimum canopy height in tree location

detection. The segmented crowns were then converted to polygons to delineate an individual .las file for point clouds of each tree.

2.5 | Tree segmentation

The CHM-based segmentation in the previous step isolated point clouds representing acacia trees, but introduced two sources of error in delineating the boundaries of individual trees: a single tree with branching canopies could be split into multiple 'trees'; multiple closely located trees, especially those with overlapping canopies, could be interpreted as a single 'tree' (Figure S2A,B). Since both error types resulted from incorrect assignment of canopy boundaries, we merged all neighbouring 'trees' in our CHM-based results into 'tree patches' (merge command in PDAL; PDAL Contributors, 2020) and devised a new algorithm for tree segmentation within each tree patch.

Instead of segmenting point clouds into trees using canopy height models (e.g. Dalponte & Coomes, 2016; Meyer & Beucher, 1990) or from a top-down clustering of point clouds (e.g. Kaartinen et al., 2012; Li et al., 2012), we devised an algorithm, conceptually similar to Burt et al. (2019), that first identifies the available tree stems. Our algorithm started with the highest point in the patch and identified a path (i.e. stem) within a 20cm diameter of points that extended to the ground (-dr argument for DESCENDENTS_RADIUS). This descending path was extended by 30cm during each iteration (-dh argument for DESCENDENTS_HEIGHT). If a path (i.e. a stem) was found, point clouds within 70cm of the stem were excluded (-fsdr argument for FOUND_STEM_DISABLE_RADIUS). Once all the stems in a patch had been found, the non-stem point clouds were iteratively 'grown back' in a 20cm radius, 40cm high cylinder from points that belonged to the tree (-gr and -gh arguments for GROW_RADIUS and GROW_HEIGHT). This process was iterated until no more points belonging to the tree could be found. All other unclassified points were then assigned to a tree by proximity. Our tree-segmentation code is deposited on Github (Huben, 2024, <https://doi.org/10.5281/zenodo.10570318>).

2.6 | Training data collection

In July 2022, we surveyed the ant occupancy of trees outside of the transects where we collected ground-survey data previously (Figure 1a, areas without blue shade), and recorded their coordinates at sub-1 m accuracy using a Trimble GeoXT Differential GPS (dGPS). In areas where trees were sparsely distributed, this level of accuracy allowed for matching geotagged tree occupancy information with the LiDAR data (Figure S1C). During the survey, we loaded CHM polygon shapefiles (see Section 2.4) onto the dGPS screen to ensure that real-time geotags landed on (or were in close proximity to) their corresponding tree (Figure S1D,E). These LiDAR point clouds, represented as polygons, are matched with occupancy information and were used as labelled training data.

2.7 | Ant occupancy classification

Classification of 3D point clouds is often achieved through projection-based classification of 2D images (reviewed in Goyal et al., 2021). We therefore transformed 3D point clouds of segmented trees with associated ant occupancy labels (see Sections 2.5 and 2.6) into 2D projections following a protocol similar to that employed by Goyal et al. (2021) and Allen et al., 2022. For each point cloud representation of an acacia tree, we selected three horizontal projections at 0, 60 and 120 degree angles. Each image was presented in 256×256 pixels that precisely cropped the highest point of the tree canopy (transforms.Resize function in PyTorch; Paszke et al., 2019). In other words, tree height was not represented in the classifier input. The pixel value of the image scaled linearly with a point's proximity to the viewer along the y-axis (Figure 1c,d; see Table 1 for other representation schemes tested). We used a 70%–15%–15% training-validation-testing split (while keeping images projected from the same point cloud in the same split). Since point clouds with labels of *C. nigriceps* occupancy were rare in our dataset, we augmented *C. nigriceps* data in our training and validation pool by generating projections at –10, 10, 50, 70, 110 and 130 degrees to account for training imbalance (Beery et al., 2018; Huang et al., 2016).

We started classification training with a pretrained ResNet18 CNN network (models.resnet18() option in torchvision, Marcel & Rodriguez, 2010). The pretrained network contained 11.6 million parameters entrained on 1000 categories of images from ImageNet 1K (Deng et al., 2009), with accuracy ranging from 69% to 89% per category (He et al., 2015). We reset the final layer (the last of the fully connected classification layers) of the ResNet18 network down to two output features: *C. nigriceps* and non-*C. nigriceps* (nn.

Linear function in PyTorch). Within the non-*C. nigriceps* class, *C. mimosae*, *C. sjostedti* and *T. penzigi* were not further distinguished. For each image input into ResNet18, after feature extraction and classification, we computed the cross-entropy loss (negative log-likelihood of model output probabilities) as our minimization target (Good, 1952, CrossEntropyLoss function in PyTorch). The gradient (i.e. change in network weights) computed based on the loss function was back-propagated across the network (backward function in PyTorch). We implemented a stochastic gradient descent algorithm as our optimizer to fine-tune the weights of the network (Ruder, 2017, optim.SGD function in PyTorch). For hyperparameters, we decayed our learning rate (set at 0.001) every seven epochs by a factor of 0.1 (StepLR (optimizer_ft, step_size = 7, gamma = 0.1) in PyTorch) to help converge on a local minimum (You et al., 2019). We obtained the iteration of model weights that provided the highest accuracy in the validation split. To generate a prediction, the trained network input an image (from the 15% test split), and its output was decided by choosing the category in the final layer of the network (*C. nigriceps* or non-*C. nigriceps*) with the highest 'probability' (i.e. the element with the highest value in the first row of the output tensor; torch.max (outputs, 1) in PyTorch). Final training accuracy was reported based on the result of the 15% testing split.

The entrained classifier was then applied to predict every segmented tree point cloud that occurred in areas where we had ascertained ant occupancy during the ground surveys. We removed point clouds less than 1.2 m in maximum height (such short saplings were likewise not recorded during ground surveys) and with less than 100 points. For each tree point cloud representation, we generated eight 2D projections at equidistant angles, and analysed the classification output of each projection. We designated a tree point cloud as *C.*

TABLE 1 Comparison of 3D point cloud representation schemes used as convolutional neural network (CNN) model inputs.

| | Height representation | Proximity representation | Rare class augmentation | Overall accuracy | Rare class accuracy | Common class accuracy |
|----------------|-----------------------|--------------------------|-------------------------------|------------------|---------------------|-----------------------|
| Present study | | | | | | |
| Selected model | | | | 0.74 | 0.31 | 0.91 |
| | Colour gradient | Transparency | | 0.8 | 0.4 | 0.9 |
| | Colour gradient | Transparency | Training split | 0.79 | 0.48 | 0.87 |
| | | Colour gradient | Training split | 0.82 | 0.42 | 0.91 |
| | Image at bottom | Colour gradient | Training split | 0.77 | 0.15 | 0.9 |
| | Image at centre | Colour gradient | Training split | 0.75 | 0.47 | 0.82 |
| | | Colour gradient | Training and validation split | 0.82 | 0.67 | 0.87 |
| Other studies | | | | | | |
| | Allen et al. (2022) | Colour gradient | | 0.81 | 0.62 | 0.84 |
| | Seidel et al. (2021) | | Training and validation split | 0.86 | 0.63 | 0.94 |

Note: When representing 3D point clouds as multiple horizontal view 2D images, considerations include whether to represent object height, proximity of each point to the viewer and to augment input images for rare classes. Rare and common classes in this study refer to classification results of *C. nigriceps* and non-*C. nigriceps*-occupied trees, while they refer to the classification results of the rarest and most common tree species in other studies. Classification accuracy refers to the percentage of predictions that are correct.

nigriceps if it had no less than 50% of its 2D projections classified as such (i.e. majority ensemble voting; Dietterich, 2000).

2.8 | Ecological validation

To evaluate the consistency between predicted patterns of *C. nigriceps* occupancy (as derived in Section 2.7) and that of the ground survey (as derived in Section 2.2), we first conducted dimensionality reduction analyses of quadrat-level species composition. Specifically, we built a site \times species matrix (410 quadrats with the count of *C. nigriceps* and non-*C. nigriceps* of each quadrat) for ground-survey and classifier-predicted data. We then performed a principal component analysis (PCA; princomp function in R; R Core Team, 2021) and principal coordinate analysis (PCoA; pcoa function in R package ape; Paradis & Schliep, 2018) on the Bray–Curtis dissimilarity matrices of ground-survey and classifier-predicted site \times species matrices (Bray & Curtis, 1957). To investigate whether changes in quadrat-level species composition are reflected visually in such dimensionality reduction analyses, we simulated a scenario in which each *Crematogaster nigriceps*-occupied tree in the 200 quadrats on the western side of the plot (which is more densely vegetated) has an 80% chance of turnover to other species that do not produce distinct canopy shapes. We generated PCoA visualizations before and after the simulation using both ground-survey and classifier-predicted data.

Next, we tallied the total number of trees, the number of *C. nigriceps*-occupied trees and the proportion of trees occupied by *C. nigriceps* in each quadrat and calculated the correlation of each of these counts between the ground survey and the classifier-predicted data. We calculated both the Pearson correlation coefficients between the numerical count per quadrat (which does not account for the spatial positioning of the quadrats, cor.test function in R; R Core Team, 2021) and a Mantel correlation between the Bray–Curtis dissimilarity matrices among quadrats (Mantel, 1967; mantel.test function in R package ade4, Dray & Dufour, 2007). To account for spatial autocorrelation, we generated spatial weighting matrices (SWMs) based on the grid design of our plot (implemented using the listw.candidates function in the R package adespatial, Dray et al., 2023), and from these SWMs further selected a subset of Moran's Eigenvector Maps (MEMs, Dray et al., 2006, implemented using the listw.select function in the R package adespatial), following the optimization protocol of Bauman et al. (2018). We conducted variance partitioning to identify the proportion of marginal variance in the ground-survey distribution patterns uniquely explained by classifier-generated matrices versus MEMs (implemented using the varpart function in the R package vegan, Oksanen et al., 2022, set up as described in Peres-Neto et al., 2006). Finally, we simulated sets of null hypotheses, in which the counts of *C. nigriceps*-occupied trees per quadrat were randomly assigned as a proportion of total trees per quadrat (we followed the proportion of *C. nigriceps* recorded in the ground survey). We then calculated the correlation between the null hypothesis simulation and the ground surveys. We repeated the

variance partitioning analysis using simulated distribution patterns as response variables.

3 | RESULTS

3.1 | Ground survey of ant occupancy

We surveyed 410 quadrats for ant occupancy, eight of which had labelling mismatches and were dropped from subsequent analysis (Figure 3; blank squares in 'total tree count' represent quadrats dropped from analysis). Among a total of 9680 acacias assessed, 1044 (10.8%) were occupied by *C. nigriceps* (the target of our classification). There were 4847 (50.1%) trees occupied by *C. mimosae*, 2734 (28.2%) by *C. sjostedti* and 1054 (10.9%) by *T. penzigi*. Quadrats on the western side of the plot (lower row numbers) had higher tree densities, while there was no change in tree density along the north–south (transect) gradient (Figure 4; Figure S3A). Surveying 410 quadrats took our eight-person team approximately 1000 work hours (125 work hours per person).

3.2 | UAV LiDAR data collection and processing

HALO flights obtained LiDAR data across 84.46 ha, covering the entire western extent of (and beyond) the ForestGEO plot. At a realized density of 276.85 pulses per m², we obtained an average of 401.30 points per m² at an average of 0.04 m spacing among points.

3.3 | Tree segmentation

The CHM-based segmentation resulted in 43,709 'trees' (coverage extended beyond the ForestGEO plot), from which 18,674 were within or bordering our ForestGEO quadrats and were selected for further processing (Figure 1a, delineated by white lines). We obtained 9885 'tree patches' after merging neighbouring 'trees', 2750 of which were in the area that had been ground surveyed (Figure 1a, areas in blue shade). Our largest 'tree patch' contained 990,273 points spanning six quadrats (merged from 107 neighbouring 'trees'), which outputted 258 trees after customized segmentation. On average, after tree segmentation, each patch was split into 3.09 trees (SD = 7.95). After maximum height and point density filtering, we recovered a total of 8364 trees (compared with 9680 trees in the ground survey).

3.4 | Training data

Outside the survey area, we collected a total of 812 dGPS geotagged tree locations associated with their ant occupant (Figure 1a, areas without blue shade). After LiDAR data processing and tree segmentation, 674 (83%) of these points could be unequivocally associated

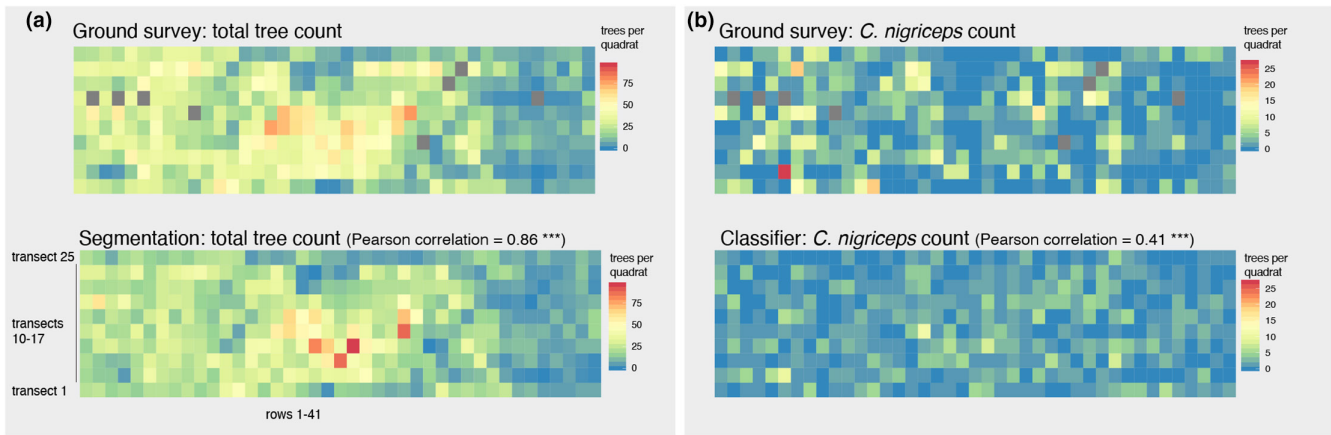


FIGURE 3 Distributional correlation between ground survey and classifier predictions across 16 ha (410 quadrats, only surveyed transects shown). (a) Significant correlation between the per quadrat number of trees counted in the ground survey and the number of trees counted with the tree segmentation (Pearson coefficients=0.86, Mantel correlation=0.73, $p < 0.001$ for both). (b) Significant correlation between the per quadrat number of *C. nigriceps*-occupied trees recorded in the ground survey and that predicted by the CNN classifier (Pearson coefficients=0.41, Mantel correlation=0.31, $p < 0.001$ for both).

with a point cloud representation of a tree. In total, we obtained 127 point clouds representing *C. nigriceps* (target species) and 547 point clouds representing non-*C. nigriceps* (447 *C. mimosae*, 82 *C. sjostedti* and 18 *T. penzigi*).

3.5 | Ant occupancy classification

Each *C. nigriceps*-labelled point cloud generated nine horizontal projections, and each non-*C. nigriceps*-labelled point cloud generated three horizontal projections. After a random 70%–15%–15% training-validation-testing split, our input training split consisted of 756 images labelled *C. nigriceps* and 1107 images labelled non-*C. nigriceps*; the validation split consisted of 162 images labelled *C. nigriceps* and 234 images labelled non-*C. nigriceps*; the testing split consisted of 57 images labelled *C. nigriceps* and 225 images labelled non-*C. nigriceps* (see Supporting Information S2).

The input representation scheme with the highest overall accuracy was one similar to that described in Goyal et al. (2021) and Allen et al. (2022) (Table 1). On the 15% testing split, this model achieved 82% overall accuracy and 66.7% accuracy at detecting images of *C. nigriceps* trees. We applied the model to post-segmentation point clouds representing trees in the surveyed area (Section 3.3, a total of 8364 trees input as 66,912 horizontal view images): 665 (8.0%) segmented point clouds were identified as *C. nigriceps*. In comparison, 10.8% of a total of 9680 trees in the ground survey were recorded as *C. nigriceps*-occupied.

3.6 | Ecological validation

Dimensionality reduction analyses (both PCA and PCoA) of species composition using ground-survey and classifier-predicted data showed near-identical patterns: quadrat similarity was driven

by total tree count (Figure 4a) and the proportion of *C. nigriceps* in each quadrat (Figure S3B). Quadrats with lower row counts on the western side of the plot were more similar to each other, whereas quadrats with higher row numbers were compositionally more distinct (Figure 4b). There was no north–south distribution pattern (Figure S3A; see Figure S4 for PCA results). In the simulated ant species turnover scenario, a decrease in plot diversity (due to the replacement of rare *C. nigriceps*) is reflected as a thinning of PCoA spread on the second principal coordinate axis in the dimensionality reduction analysis using ground-survey data (Figure S5A). This change was mirrored when we conducted the same simulation on classifier-generated data (Figure S5B).

The total number of trees across the 410 quadrats generated as a result of the tree segmentation showed a significant correlation with counts obtained from the ground survey (Pearson coefficients=0.86, Mantel correlation=0.73, $p < 0.001$ for both comparisons). The number of *C. nigriceps*-occupied trees predicted by the classification showed a lower, but significant, correlation with counts from the ground survey (Pearson coefficients=0.41, Mantel correlation=0.31, $p < 0.001$ for both comparisons), as did the proportion of *C. nigriceps* trees per quadrat (Pearson coefficients=0.34, Mantel correlation=0.30, $p < 0.001$ for both comparisons) (Figure 3). In contrast, the randomly simulated dataset of *C. nigriceps* as 8% (the proportion of *C. nigriceps* predicted by the classifier) and 10% (the proportion of *C. nigriceps* recorded in the ground survey) of the total tree composition did not achieve a high level of correlation with ground-survey counts (Table 2). In the variance partitioning analysis, 11% of the variance in ground-survey species composition is uniquely explained by the classifier-generated predictions; 7% of the variance is uniquely explained by spatial autocorrelation; and a total of 60% of the variance is explained jointly by the classifier prediction and spatial autocorrelation (Table 2). The fraction of variance explained by partitioning the simulated distribution data is lower than that from the classifier, at 8%–9%.

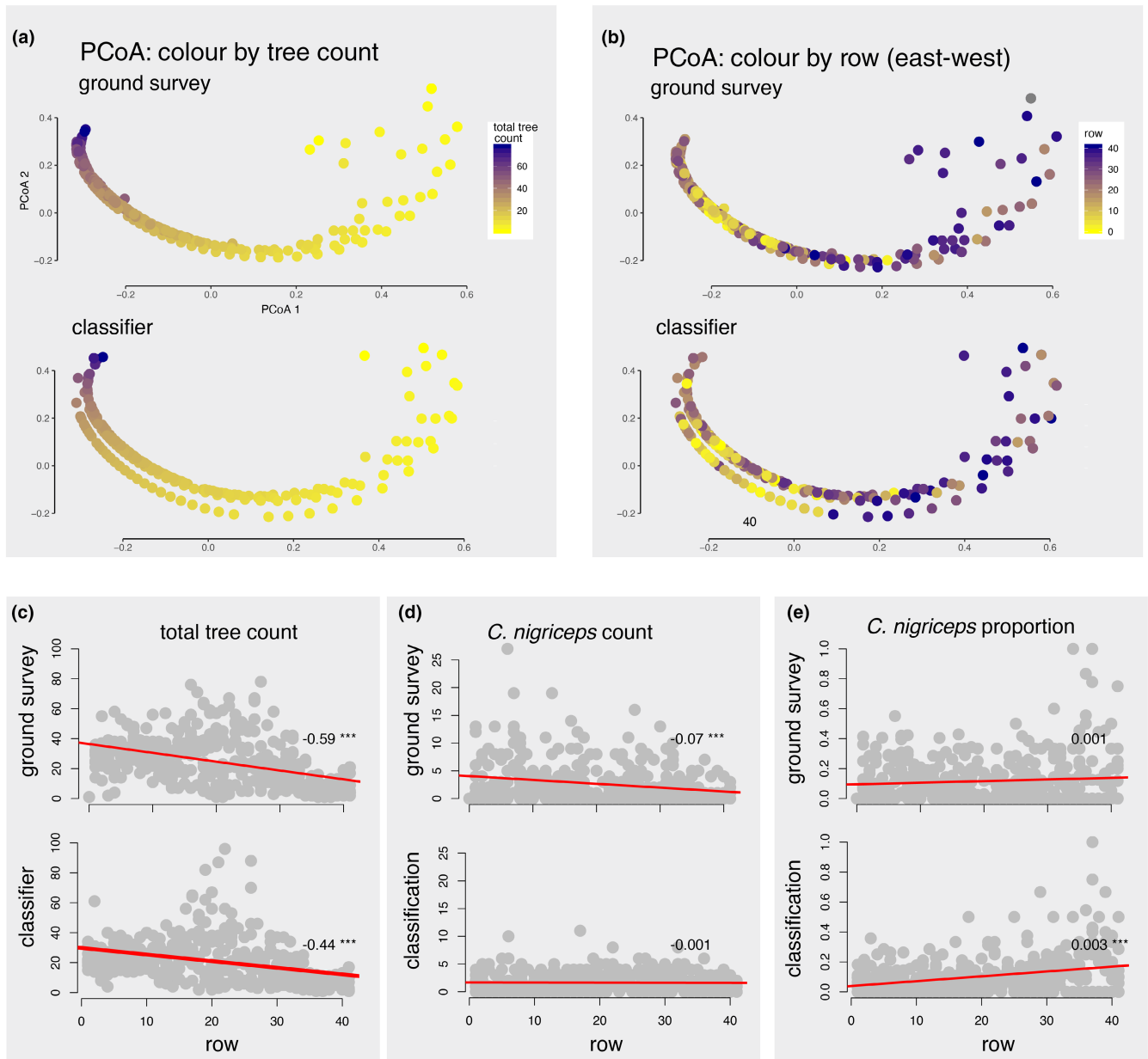


FIGURE 4 Consistency of principal coordinate analysis (PCoA) between ground survey and classifier predictions across 410 quadrats. (a) Each dot represents the ant community composition of a single quadrat, shown on the first and second principal coordinate axes. Colours indicate the total number of trees in each quadrat. (b) Same as (a), but coloured by each quadrat's row number. See Figure S3 for the same figure coloured by *C. nigriceps* proportion and transect numbers. See Figure S4 for similar results from principal component analysis (PCA). (c–e) Trends of tree count, *C. nigriceps* count and *C. nigriceps* proportion from the western to the eastern side of the plot (low to high row numbers) are consistent between ground survey and classifier predictions (same signs for the estimates), although the significance of linear coefficients differ. Each dot represents a quadrat, and the red line indicates a linear regression fit.

4 | DISCUSSION

We present a protocol for detecting insect-induced intra-specific vegetation phenotypic variation using airborne LiDAR. We tested the validity of our ecological inferences on 16 ha (410 quadrats) of savanna containing 9680 acacia trees, and showed that the distribution pattern of a relatively rare ant symbiont can be predicted by a custom-trained convolutional neural network (CNN) classifier based on tree morphology. Our claim to the effectiveness of

the CNN classifier at detecting ecological dynamics rests principally not on the absolute accuracy of individual tree phenotype classification, but rather on the demonstrated high congruence of beta-diversity patterns between ground survey and classifier-generated predictions, even after accounting for spatial autocorrelation (Table 2). For practical management purposes, such congruence can be visualized as similarity between dimensionality reduction plots, and we show that under simulated environmental disturbance scenarios, visible changes in intra-specific phenotypic

TABLE 2 Correlation and variance partitioning of the classifier and simulation-generated predictions from the ground survey.

| | Segmentation and classifier | | Simulation: 8% <i>C. Nigriceps</i> | | Simulation: 10% <i>C. Nigriceps</i> | |
|--------------------------------|-----------------------------|---------|------------------------------------|---------|-------------------------------------|---------|
| | Pearson | Mantel | Pearson | Mantel | Pearson | Mantel |
| Total tree count | 0.86*** | 0.73*** | | | | |
| <i>C. nigriceps</i> count | 0.41*** | 0.31*** | 0.22*** | 0.12** | 0.15*** | 0.03 |
| <i>C. nigriceps</i> proportion | 0.34*** | 0.30*** | -0.04 | 0.01 | 0.01 | 0.04 |
| | Environmental | Spatial | Environmental | Spatial | Environmental | Spatial |
| Marginal variance fraction | 0.11 | 0.07 | 0.08 | 0.07 | 0.09 | 0.08 |

Note: Simulation values of 8% and 10% were derived from the percentage of total *C. nigriceps* proportion predicted by the classifier and the percentage of *C. nigriceps* recorded in the ground survey. Top panel: Pearson correlation coefficients were calculated from numerical counts per quadrat, while Mantel correlations were calculated from Bray–Curtis dissimilarity matrices among quadrats (p -values: * < 0.05; ** < 0.01; *** < 0.001). Bottom panel: proportion of variance in ground-survey distribution patterns uniquely explained by classifier-generated (or simulated) environmental matrices versus those generated by subsets of Moran's Eigenvector Maps corresponding to plot spatial designs.

variation can be tracked through our automated monitoring pipeline (Figure 4; Figures S3–S5).

We suggest that similar procedures can be adapted to enable the identification of other biologically relevant 'extended phenotypes' using LiDAR or other remotely sensed data, increasing the efficiency of monitoring ecological dynamics. We first discuss the effectiveness of our pipeline as applied to the *A. drepanolobium*-ant symbiosis system (Section 4.1). We then discuss some of the limitations of our procedure and plans for future improvements (Section 4.2). Finally, we provide examples of how our protocol can be applied to other large-scale ecosystem monitoring endeavours (Section 4.3).

4.1 | Evaluating a classifier to detect intra-specific phenotypic variation

Consistent with previous results (Palmer et al., 2000) at the same site, *Crematogaster nigriceps* occupied only 10% of all trees in our survey area. When collecting training data outside our ground-survey plot, we biased our search towards finding *C. nigriceps*-occupied trees, with the resulting phenotype accounting for 18% of our training data. At this proportion, *C. nigriceps* detection still constitutes a 'rare class classification' problem: without sufficient training data, accuracy for classifying rare classes is low (Beery et al., 2018; Huang et al., 2016). In the context of classifying tree species from LiDAR data, small amounts of training data usually result in the rarest tree species having the lowest classification accuracy. For example, in classifying five tree species in a Mediterranean forest (Allen et al., 2022), *Pinus pinaster*, the rarest species, comprised 5% of the total tree data and achieved 62% accuracy, compared with an 81% total accuracy for the dataset. Similarly, Seidel et al. (2021) classified seven species of trees from labelled LiDAR data across temperate Europe and the United States: *Quercus rubra* (red oak) was the rarest class, constituting 14% of the data and achieving 63% accuracy compared with an average dataset accuracy of 86%. Our 66.7% accuracy for *C. nigriceps*, among an overall accuracy of 82%, is therefore comparable to these previous studies (Table 1), suggesting a generalizable limitation in rare class detection accuracy. However, our study

represents the first attempt to distinguish intra-specific phenotypic variation, rather than tree species, using a custom-trained classifier.

The above-mentioned model accuracies were estimated based on the 15% 'testing split' of all training data collected. In practical ecological and conservation applications, ecologists and managers are typically less concerned with a model's reported accuracy than its ability to recapitulate ecosystem dynamics. Ground-truthing 16 ha (410 quadrats) of *C. nigriceps* occupancy took us 1000 work hours, while a UAV survey took less than an hour. A model that would allow us to detect patterns of species composition change based on future UAV flights would be immensely convenient for tracking the ecosystem health of this savanna, as changes in ant occupancy are closely linked to savanna productivity, herbivore activity and carbon cycling (Milligan et al., 2021; Palmer et al., 2008, 2010).

Ant occupancy in the 16-ha (410 quadrats) ground-survey area was tallied at the quadrat level (rather than at the level of individual trees, due to limited dGPS accuracy; see Figure S1); correspondingly, we tallied our model prediction of *C. nigriceps*-occupied trees by quadrats. In effect, this approach treats each quadrat as an ecological community, and we therefore compared measured beta-diversity patterns (an informative index of ecological dynamics and conservation planning; see Socolar et al., 2016) between the ground surveys and those predicted by our CNN classifier, finding a high level of congruence in the total tree count, *C. nigriceps* spatial distribution, and *C. nigriceps* proportion. Our direct approaches to calculating the 'correlation' between ground-survey and classifier-predicted beta-diversity patterns have limitations: (1) Pearson correlation treats each quadrat as independent and does not account for spatial autocorrelation; (2) while the Mantel test is a valid method for demonstrating the degree of difference between multivariate dissimilarity matrices (in our case, Bray–Curtis dissimilarity matrices), it does not reveal correlation between geographical distributional patterns per se (see Legendre et al., 2015 for caution against the use of the Mantel test in spatial analysis). Indirectly, a variance partitioning approach with spatial MEMs as a proxy for autocorrelation (Peres-Neto et al., 2006) explains the unique contribution of the classifier-generated distribution pattern (conditioned on spatial pattern) towards the

ground-surveyed distribution pattern; in this context, we interpret a high partitioned contribution as high congruence. In our study, a combination of all three approaches show congruence between the ground survey and modelled predictions, even after accounting for spatial autocorrelation; this finding attests to the validity of our classification approach (Figure 4b; Table 2). Although the predicted proportion of *C. nigriceps*-occupied trees per quadrat was significantly correlated with the ground-survey results, these correlations had smaller coefficients compared with the *C. nigriceps* count correlations, likely because of error introduced from both the total tree count (acquired through tree segmentation) and the *C. nigriceps* counts (acquired through tree classification). Importantly, however, these spatial correlations could not be derived from simply randomly simulating a proportion of trees occupied by *C. nigriceps* (Table 2, 'null hypothesis')—strictly speaking, even our choices of simulating *C. nigriceps* at 8% or 10% occurrence are biologically informed, which might account for the high (and sometimes significant) correlation between our simulated patterns and those from the ground survey. To achieve an even higher correlation with ground-survey beta-diversity patterns than these simulations did, our classifier must have correctly identified a high proportion of the targeted phenotype.

Although we were able to replicate ecologically meaningful patterns with our model classification, we acknowledge that a map with a 0.41 correlation to ground-survey counts (e.g. Figure 3b) is not very helpful for pinpointing *C. nigriceps* in the field (but see our suggestions for improving model accuracy in Section 4.2). Rather than static 'treasure maps', our approach is best suited to detect dynamic ecological community changes along spatial or temporal axes. Using dimensionality reduction analysis, ground-surveyed communities were arrayed along the axis of total tree count and proportion of *C. nigriceps* per community (Figure 4a; Figure S3B). Quadrats on the western side of the plot were more similar to each other than those on the eastern side (Figure 4b, lower row numbers cluster more closely), while there were no distinguishable patterns along the north–south axis (Figure S3A, no pattern by transect number). These patterns likely reflect a soil nutritional gradient, which is present along the east–west but not the north–south axis of the plot (Childers, 2021; Palmer & Young, 2017). Soil nutrition either directly drives ant species composition and tree use (Farji-Brener & Werenkraut, 2017; Kenfack et al., 2021; Wagner & Fleur Nicklen, 2010), or affects herbivore use of the landscape, which drives the patterning of ant occupancy (Kaspari, 2020; Palmer et al., 2008). Regardless of the reasons behind this distribution pattern, it is almost identically recapitulated by our classifier-predicted datasets (Figure 4a,b). This successful recapitulation of patterns suggests that our model was sensitive enough to replicate patterns of species distribution at a community level, even though individual-level prediction accuracy left considerable room for improvement. In our simulated turnover scenario (reflecting a loss of focus species due to invasive ant encroachment; see next paragraph), the dimensional reduction results of all quadrats change visibly; the same pattern of change was

mirrored when the simulation was applied to classifier-generated data (Figure S5). These results suggest that if a different plot were surveyed or if the pattern of *C. nigriceps* distribution of the current plot changed (e.g. due to drought-induced vegetation die-off, fire-induced succession events or invasive species incursion), we can be confident that observable community changes would be reflected in our LiDAR-derived classifier.

Such an application of our protocol is timely and relevant: an invasive ant species, the big-headed ant (*Pheidole megacephala*), established an invasion front less than 2 km from our plot. *Pheidole megacephala* out-competes all four mutualist ant species and does not defend acacias from browsers, leading to increased herbivore and pathogen damage and loss in above-ground carbon sequestration (Milligan et al., 2021, 2022). Such shifting invasion fronts are difficult to detect on the ground but could be delineated from monitoring the retreat of *C. nigriceps*-occupied trees with future LiDAR surveys.

4.2 | Limitations and future improvements

Due to logistic constraints, our collection of on the ground ant occupancy data (Sections 2.2 and 2.6) lagged 6 months behind UAV data collection. During this interval, it is possible (but unlikely) that ant occupancy on acacias could have changed (no data on ant turnover rate on host trees have been published, but Palmer et al., 2010 estimated that a 20-year-old acacia tree would have experienced 3–7 turnover events). Such potential turnover would decrease the correlation between the ground-surveyed *C. nigriceps* distribution and that predicted by the classifier (although it is also unclear how long *C. nigriceps*-induced structural change would persist after a turnover event).

Collecting more training data of *C. nigriceps*-occupied trees will likely increase the accuracy of our model (Huang et al., 2016), which would involve either geotagging more *C. nigriceps*-occupied trees or finding creative ways of bootstrapping inputs from graphic engines (e.g. Beery et al., 2018; Das et al., 2021). Better representation schemes for input point clouds is another option to explore: in reducing 3D point clouds to multi-aspect 2D representations, structural information is lost. However, classifiers that directly input 3D point clouds can outperform image-based representation schemes (Budei et al., 2018; Lin et al., 2023). Lastly, given sufficient computational resources, hyperparameters of machine learning, such as those determining the network learning rate and its stepwise decay, could be optimized in a Bayesian framework (e.g. Frazier, 2018). While these enhancements would quantitatively improve the accuracy of our model, ant occupancy is one of many factors in predicting acacia canopy topology. Tree age, soil type, climate, herbivory and competition also influence tree shapes; it remains challenging to predict all intra-species phenotypic variation with a single variable (Auger & Shipley, 2013; Dube et al., 2014; Jiménez & Díaz-Delgado, 2015).

At our field site, *C. nigriceps* is one of four ant species found on *A. drepanolobium*. Although *C. nigriceps*-occupied trees exhibit the

most distinguishable phenotypic changes, the other three ant species might also be identifiable with high-resolution LiDAR data. For example: (1) *C. sjostedti*, the most competitive of the four ant species, tends to occupy older, taller trees (Palmer & Young, 2017). While our current input representation scheme is height-independent (see Section 2.7), tree height could be used to identify *C. sjostedti*. (2) Ant species that aggressively trim and/or consume acacia leaves (i.e. *C. nigriceps* and *T. penzigi*) make their host trees less verdant, which would likely be detected with infrared or RGB colour sensors (Madawy et al., 2019). (3) *Tetraponera penzigi* is the best disperser among the four ant species (Stanton et al., 2002). Spatial variables such as distance to the nearest neighbouring tree could be useful for predicting *T. penzigi* occupancy. We emphasize that a nuanced appreciation of ant life history and behavioural ecology is the foundation of such data-driven ecological inferences.

4.3 | Future applications

Vegetation, when interacting with insects, exhibits phenotypic plasticity (reviewed in Ananthkrishnan, 2000; Ashra & Nair, 2022). Our study demonstrates that these intra-specific phenotypic variations can be identified in ultra-high-resolution (10cm pixels) LiDAR data and used to reliably infer the occupancy of small invertebrates. Modern-day agriculture and forestry typically consist of large-scale monoculture plantations that are highly concerned with monitoring insect-vectored infections. Our approach could be extended to these settings to provide rapid assessment of agricultural and forestry health at reduced costs compared with ground-based surveys. For example, the hemlock woolly adelgid (*Adelges tsugae*) is a major forestry pest in the northeastern United States, defoliating the mid-storey and understorey of eastern hemlock trees (*Tsuga canadensis*) before signs of infection occur in the forest canopy. This damage makes quantification of infection difficult from satellite images, costly from the ground, but convenient using 3D-LiDAR point clouds (Boucher et al., 2020). Following the protocol outlined in our study, a 'woolly adelgid infestation classifier' could be trained on topologies of infested trees and then applied to surveil pestilence using UAV LiDAR flights. More broadly, the most economically important tree crop diseases, such as wilt, rust, powdery mildew and anthracnose, exhibit consistent, identifiable phenotypic changes in canopy shape or colour in their early stages on pine, rubber, chocolate, cacao and palms before defoliation. A combination of high-resolution LiDAR data, CNN training and ground-survey data collection hold promise for identifying these diseases at large scales (Abdulridha et al., 2020; Liu et al., 2021; Moriya et al., 2021; Yu et al., 2021; Zeng et al., 2022).

AUTHOR CONTRIBUTIONS

Zhengyang Wang, Christopher C. M. Baker, Dino J. Martins, Naomi E. Pierce and Andrew B. Davies conceived the ideas for this project. Zhengyang Wang, Robert Huben, Peter B. Boucher, Chase Van Amburg, Jimmy Zeng and Andrew B. Davies designed and developed the methodology. Zhengyang Wang, Peter B. Boucher, Chase

Van Amburg, Nina Chung, Jocelyn Wang, Jeffrey King, Richard J. Knecht, Ivy Ng'iru, Augustine Baraza, Naomi E. Pierce and Andrew B. Davies collected the data. Zhengyang Wang, Robert Huben, Peter B. Boucher, Chase Van Amburg and Jimmy Zeng analysed the data. Zhengyang Wang led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

NEP and ABD's research groups provided valuable feedback on the manuscript. This work was supported by the William F. Milton Fund of Harvard University. CVA was supported by the Harvard Museum of Comparative Zoology Grant-In-Aid of Undergraduate Research fund. NC was supported by the Harvard College Research Program and a Harvard Center for African Studies summer travel grant. JW was supported by a Herchel Smith Undergraduate Science Research Program grant from Harvard University. NEP, DJM and RJK were supported by a grant from the Putnam Expeditionary Fund of the MCZ (to NEP). We thank Jeff Blossom from the Harvard Center for Geographic Analysis for assistance with dGPS data collection and processing, Pat Milligan for discussions on invasive ants, and Weilin Meng and Allison King for discussions on machine learning methods. We thank the Government of Kenya (NACOSTI/P/22/18988, NACOSTI/P/21/12633) and Mpala Research Centre for permission to perform the study, and the Kenyan Civil Aviation Authority for authorization to perform the UAV surveys (KCAA/OPS/2117/4). We thank the staff at Mpala Research Centre, especially Jackson Ekwam Etele, Wilson Mureyian, Steven Gitau Githuthu, Godfrey Amoni, Godfrey Gitimu, Patrick Kamukunji, Aimee Gaitho, Cosmas Nzomo and Peter Lochomin, for their assistance.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interests.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14298>.

DATA AVAILABILITY STATEMENT

Ground-survey results (Supporting Information S1), with code and LiDAR point clouds for CNN training (Supporting Information S2), are deposited at Zenodo (Wang, 2023, <https://doi.org/10.5281/zenodo.8175736>). Our script for tree segmentation, with user instructions, is deposited on Github (Huben, 2024, <https://doi.org/10.5281/zenodo.10570318>).

STATEMENT ON INCLUSION

Our study brings together authors from four continents, including scientists based in Kenya, where the field work was carried out. All authors were engaged early on with the research and study design to ensure that the diverse sets of perspectives they represent were considered from the onset.

ORCID

Zhengyang Wang  <https://orcid.org/0000-0003-3244-1954>
 Peter B. Boucher  <https://orcid.org/0000-0003-2585-6705>
 Chase Van Amburg  <https://orcid.org/0000-0001-9970-0207>
 Jimmy Zeng  <https://orcid.org/0000-0001-8773-4856>
 Nina Chung  <https://orcid.org/0000-0002-3112-3236>
 Jocelyn Wang  <https://orcid.org/0000-0002-4923-1518>
 Richard J. Knecht  <https://orcid.org/0000-0002-2324-701X>
 Ivy Ng'iru  <https://orcid.org/0000-0002-4766-6251>
 Christopher C. M. Baker  <https://orcid.org/0000-0002-2675-1078>
 Dino J. Martins  <https://orcid.org/0000-0003-1581-1642>
 Naomi E. Pierce  <https://orcid.org/0000-0003-3366-1625>
 Andrew B. Davies  <https://orcid.org/0000-0002-0003-1435>

REFERENCES

- Abdulridha, J., Ampatzidis, Y., Roberts, P., & Kakarla, S. C. (2020). Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. *Biosystems Engineering*, 197, 135–148. <https://doi.org/10.1016/j.biosystemseng.2020.07.001>
- Allen, M. J., Grieve, S. W. D., Owen, H. J. F., & Lines, E. R. (2022). Tree species classification from complex laser scanning data in Mediterranean forests using deep learning. *Methods in Ecology and Evolution*, 14, 1657–1667. <https://doi.org/10.1111/2041-210X.13981>
- Ananthakrishnan, T. N. (2000). Phenotypic flexibility of plants and adaptive dynamics of specialist-generalist insects. *Current Science*, 78(11), 1303–1306.
- Arsad, T. N. T., Awalludin, E. A., Bachok, Z., Yusoff, W. N. J. H. W., & Hitam, M. S. (2023). A review of coral reef classification study using deep learning approach. *AIP Conference Proceedings*, 2484(1), 050005. <https://doi.org/10.1063/5.0110245>
- Ashra, H., & Nair, S. (2022). Review: Trait plasticity during plant-insect interactions: From molecular mechanisms to impact on community dynamics. *Plant Science*, 317, 111188. <https://doi.org/10.1016/j.plantsci.2022.111188>
- Atkins, J. W., Bohrer, G., Fahey, R. T., Hardiman, B. S., Morin, T. H., Stovall, A. E. L., Zimmerman, N., & Gough, C. M. (2018). Quantifying vegetation and canopy structural complexity from terrestrial LiDAR data using the *forestr* package. *Methods in Ecology and Evolution*, 9(10), 2057–2066. <https://doi.org/10.1111/2041-210X.13061>
- Auger, S., & Shipley, B. (2013). Inter-specific and intra-specific trait variation along short environmental gradients in an old-growth temperate forest. *Journal of Vegetation Science*, 24(3), 419–428. <https://doi.org/10.1111/j.1654-1103.2012.01473.x>
- Avilés, L., & Tufiño, P. (1998). Colony size and individual fitness in the social spider *Anelosimus eximius*. *The American Naturalist*, 152(3), 403–418. <https://doi.org/10.1086/286178>
- Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. <https://www.semanticscholar.org/paper/DEM-Generation-from-Laser-Scanner-Data-Using-TIN-Axelsson/98bbd72676af97cfd351506c650fc175c7317adb>
- Bauman, D., Drouet, T., Fortin, M.-J., & Dray, S. (2018). Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. *Ecology*, 99(10), 2159–2166. <https://doi.org/10.1002/ecy.2469>
- Beery, S., van Horn, G., & Perona, P. (2018). Recognition in terra incognita. *arXiv*, arXiv:1807.04975. <https://doi.org/10.48550/arXiv.1807.04975>
- Bentz, B. J., & Endreson, D. (2004). Evaluating satellite imagery for estimating mountain pine beetle-caused lodgepole pine mortality: Current status. In T. L. Shore, J. E. Brooks, & J. E. Stone (Eds.), *Mountain pine beetle symposium: Challenges and solutions*; October 30–31, 2003; Kelowna, British Columbia. Information Report BC-X-399 (pp. 154–163). Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre.
- Boucher, P. B., Hancock, S., Orwig, D. A., Duncanson, L., Armston, J., Tang, H., Krause, K., Cook, B., Paynter, I., Li, Z., Elmes, A., & Schaaf, C. (2020). Detecting change in forest structure with simulated GEDI lidar waveforms: A case study of the hemlock woolly adelgid (HWA; *Adelges tsugae*) Infestation. *Remote Sensing*, 12(8), 1304. <https://doi.org/10.3390/rs12081304>
- Boucher, P. B., Hockridge, E. G., Singh, J., & Davies, A. B. (2023). Flying high: Sampling savanna vegetation with UAV-lidar. *Methods in Ecology and Evolution*, 14, 1668–1686. <https://doi.org/10.1111/2041-210X.14081>
- Boyle, J. H., Martins, D., Musili, P. M., & Pierce, N. E. (2019). Population genomics and demographic sampling of the ant-plant *Vachellia drepanolobium* and its symbiotic ants from sites across its range in east Africa. *Frontiers in Ecology and Evolution*, 7, 206. <https://doi.org/10.3389/fevo.2019.00206>
- Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27(4), 325–349. <https://doi.org/10.2307/1942268>
- Brodrick, P. G., Davies, A. B., & Asner, G. P. (2019). Uncovering ecological patterns with convolutional neural networks. *Trends in Ecology & Evolution*, 34(8), 734–745. <https://doi.org/10.1016/j.tree.2019.03.006>
- Budei, B. C., St-Onge, B., Hopkinson, C., & Audet, F.-A. (2018). Identifying the genus or species of individual trees using a three-wavelength airborne lidar system. *Remote Sensing of Environment*, 204, 632–647. <https://doi.org/10.1016/j.rse.2017.09.037>
- Burt, A., Disney, M., & Calders, K. (2019). Extracting individual trees from lidar point clouds using treeseg. *Methods in Ecology and Evolution*, 10(3), 438–445. <https://doi.org/10.1111/2041-210X.13121>
- Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C. A., Dawson, T. P., Emerson, B. C., Ferrier, S., Gilbert, M. T. P., Herold, M., Jones, L., Leendertz, F. H., Matthews, L., Millington, J. D. A., ... Yu, D. W. (2017). Connecting earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution*, 1(7), 0176. <https://doi.org/10.1038/s41559-017-0176>
- Câmara, T., Leal, I. R., Blüthgen, N., Oliveira, F. M. P., & Arnan, X. (2019). Anthropogenic disturbance and rainfall variation threaten the stability of plant-ant interactions in the Brazilian Caatinga. *Ecography*, 42(11), 1960–1972. <https://doi.org/10.1111/ecog.04531>
- Childers, R. A. (2021). Ecology and evolution of the African ant acacia, *Vachellia drepanolobium*, and its multiple symbionts. <https://dash.harvard.edu/handle/1/37368410>
- Christita, M., Sipilä, T. P., Auzane, A., & Overmyer, K. (2022). Distinct Taphrina strains from the phyllosphere of birch exhibiting a range of witches' broom disease symptoms. *Environmental Microbiology*, 24(8), 3549–3564. <https://doi.org/10.1111/1462-2920.16037>
- Cook, R. M., Parrini, F., King, L. E., Witkowski, E. T. F., & Henley, M. D. (2018). African honeybees as a mitigation method for elephant impact on trees. *Biological Conservation*, 217, 329–336. <https://doi.org/10.1016/j.biocon.2017.11.024>
- Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution*, 7(10), 1236–1245. <https://doi.org/10.1111/2041-210X.12575>
- Dangles, O., & Casas, J. (2019). Ecosystem services provided by insects for achieving sustainable development goals. *Ecosystem Services*, 35, 109–115. <https://doi.org/10.1016/j.ecoser.2018.12.002>

- Das, T., Brintjes, R.-J., Lengyel, A., van Gemert, J., & Beery, S. (2021). Domain adaptation for rare classes augmented with synthetic samples. *arXiv*, arXiv:2110.12216. <https://doi.org/10.48550/arXiv.2110.12216>
- Davies, A. B., & Asner, G. P. (2014). Advances in animal ecology from 3D-LiDAR ecosystem mapping. *Trends in Ecology & Evolution*, 29(12), 681–691. <https://doi.org/10.1016/j.tree.2014.10.005>
- Davies, A. B., Brodrick, P. G., Parr, C. L., & Asner, G. P. (2020). Resistance of mound-building termites to anthropogenic land-use change. *Environmental Research Letters*, 15(9), 094038. <https://doi.org/10.1088/1748-9326/aba0ff>
- Davies, A. B., Crowsigt, J. P. G. M., Tambling, C. J., Le Roux, E., Vaughn, N., Druce, D. J., Marneweck, D. G., & Asner, G. P. (2021). Environmental controls on African herbivore responses to landscapes of fear. *Oikos*, 130(2), 171–186. <https://doi.org/10.1111/oik.07559>
- Davies, A. B., Levick, S. R., Asner, G. P., Robertson, M. P., van Rensburg, B. J., & Parr, C. L. (2014). Spatial variability and abiotic determinants of termite mounds throughout a savanna catchment. *Ecography*, 37(9), 852–862. <https://doi.org/10.1111/ecog.00532>
- Davies, A. B., Oram, F., Ancrenaz, M., & Asner, G. P. (2019). Combining behavioural and LiDAR data to reveal relationships between canopy structure and orangutan nest site selection in disturbed forests. *Biological Conservation*, 232, 97–107. <https://doi.org/10.1016/j.biocon.2019.01.032>
- Dawkins, R. (1983). *The extended phenotype: The long reach of the gene*. Oxford University Press.
- Day, K. R., & Leather, S. R. (1997). Threats to forestry by insect pests in Europe. In A. D. Watt, N. E. Stork, & M. D. Hunter (Eds.), *Forests and insects* (pp. 177–205). Chapman and Hall.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Devarajan, K. (2016). The antsy social network: Determinants of nest structure and arrangement in Asian weaver ants. *PLoS ONE*, 11(6), e0156681. <https://doi.org/10.1371/journal.pone.0156681>
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Springer. https://doi.org/10.1007/3-540-45014-9_1
- Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guénard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. H. (2023). *adespatial: Multivariate multiscale spatial analysis*. <https://CRAN.R-project.org/package=adespatial>
- Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4). <https://doi.org/10.18637/jss.v022.i04>
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196(3), 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Dube, T., Mutanga, O., Elhadi, A., & Ismail, R. (2014). Intra-and-inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution spaceborne multispectral rapideye sensor and advanced machine learning algorithms. *Sensors*, 14(8), 15348–15370. <https://doi.org/10.3390/s140815348>
- Evans, L. J., Goossens, B., Davies, A. B., Reynolds, G., & Asner, G. P. (2020). Natural and anthropogenic drivers of Bornean elephant movement strategies. *Global Ecology and Conservation*, 22, e00906. <https://doi.org/10.1016/j.gecco.2020.e00906>
- Farji-Brener, A. G., & Werenkraut, V. (2017). The effects of ant nests on soil fertility and plant performance: A meta-analysis. *Journal of Animal Ecology*, 86(4), 866–877. <https://doi.org/10.1111/1365-2656.12672>
- Fisher, K. E., Dixon, P. M., Han, G., Adelman, J. S., & Bradbury, S. P. (2021). Locating large insects using automated VHF radio telemetry with a multi-antennae array. *Methods in Ecology and Evolution*, 12(3), 494–506. <https://doi.org/10.1111/2041-210X.13529>
- Fitzgerald, T. D., & Willer, D. E. (1983). Tent-building behavior of the eastern tent caterpillar *Malacosoma americanum* (Lepidoptera: Lasiocampidae). *Journal of the Kansas Entomological Society*, 56(1), 20–31.
- Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv*, arXiv:1807.02811. <https://doi.org/10.48550/arXiv.1807.02811>
- Frederickson, M. E., Greene, M. J., & Gordon, D. M. (2005). 'Devil's gardens' bedevilled by ants. *Nature*, 437(7058), 495–496. <https://doi.org/10.1038/437495a>
- Gatzliolis, D., & Andersen, H.-E. (2008). A guide to LIDAR data acquisition and processing for the forests of the Pacific Northwest. *Gen. Tech. Rep. PNW-GTR-768*. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 32 p, 768. <https://doi.org/10.2737/PNW-GTR-768>
- Goetz, S., Steinberg, D., Dubayah, R., & Blair, B. (2007). Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sensing of Environment*, 108(3), 254–263. <https://doi.org/10.1016/j.rse.2006.11.016>
- Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society: Series B: Methodological*, 14(1), 107–114.
- Goyal, A., Law, H., Liu, B., Newell, A., & Deng, J. (2021). Revisiting point cloud shape classification with a simple and effective baseline. *arXiv*, arXiv:2106.05304. <http://arxiv.org/abs/2106.05304>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv*, arXiv:1512.03385. <https://doi.org/10.48550/arXiv.1512.03385>
- Hockridge, E. G., Singh, J., Boucher, P. B., & Davies, A. B. (2022). Resistance of termite mounds to variation in long-term fire regimes across semi-arid African savannas. *Journal of Animal Ecology*, 92, 1294–1305. <https://doi.org/10.1111/1365-2656.13830>
- Houadria, M., & Menzel, F. (2017). What determines the importance of a species for ecosystem processes? Insights from tropical ant assemblages. *Oecologia*, 184(4), 885–899. <https://doi.org/10.1007/s00442-017-3900-x>
- Huang, C., Li, Y., Loy, C. C., & Tang, X. (2016). Learning deep representation for imbalanced classification. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5375–5384. <https://doi.org/10.1109/CVPR.2016.580>
- Huben, R. (2024). RobertHuben/tree_point_clouds: Publication Release (v1.0). *Zenodo*, <https://doi.org/10.5281/zenodo.10570318>
- Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, 3(4), 539–551. <https://doi.org/10.1038/s41559-019-0826-1>
- Ji, L., Wang, Z., Wang, X., & An, L. (2011). Forest insect pest management and forest management in China: An overview. *Environmental Management*, 48(6), 1107–1121. <https://doi.org/10.1007/s00267-011-9697-1>
- Ji, Y., Baker, C. C. M., Popescu, V. D., Wang, J., Wu, C., Wang, Z., Li, Y., Wang, L., Hua, C., Yang, Z., Yang, C., Xu, C. C. Y., Diana, A., Wen, Q., Pierce, N. E., & Yu, D. W. (2022). Measuring protected-area effectiveness using vertebrate distributions from leech IDNA. *Nature Communications*, 13(1), 1555. <https://doi.org/10.1038/s41467-022-28778-8>
- Jiménez, M., & Díaz-Delgado, R. (2015). Towards a standard plant species spectral library protocol for vegetation mapping: A case study in the Shrubland of Doñana National Park. *ISPRS International Journal of Geo-Information*, 4(4), 2472–2495. <https://doi.org/10.3390/ijgi4042472>
- Kaartinen, H., Hyypää, J., Yu, X., Vastaranta, M., Hyypää, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset,

- E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B. M., & Wu, J.-C. (2012). An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sensing*, 4(4), 950–974. <https://doi.org/10.3390/rs4040950>
- Kaspari, M. (2020). The seventh macronutrient: How sodium shortfall ramifies through populations, food webs and ecosystems. *Ecology Letters*, 23(7), 1153–1168. <https://doi.org/10.1111/ele.13517>
- Kellner, J. R., Armston, J., Birrer, M., Cushman, K. C., Duncanson, L., Eck, C., Falger, C., Imbach, B., Král, K., Krůček, M., Trochta, J., Vrška, T., & Zraggen, C. (2019). New opportunities for forest remote sensing through ultra-high-density drone lidar. *Surveys in Geophysics*, 40(4), 959–977. <https://doi.org/10.1007/s10712-019-09529-9>
- Kenfack, D., Arellano, G., Kibet, S., Kimuyu, D., & Musili, P. (2021). Understanding the monodominance of *Acacia drepanolobium* in East African savannas: Insights from demographic data. *Trees*, 35(5), 1439–1450. <https://doi.org/10.1007/s00468-021-02127-6>
- Legendre, P., Fortin, M.-J., & Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution*, 6(11), 1239–1247. <https://doi.org/10.1111/2041-210X.12425>
- Levick, S. R., Asner, G. P., Chadwick, O. A., Khomo, L. M., Rogers, K. H., Hartshorn, A. S., Kennedy-Bowdoin, T., & Knapp, D. E. (2010). Regional insight into savanna hydrogeomorphology from termite mounds. *Nature Communications*, 1(1), 65. <https://doi.org/10.1038/ncomms1066>
- Li, W., Dong, R., Fu, H., & Yu, L. (2019). Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. *Remote Sensing*, 11(1), 11. <https://doi.org/10.3390/rs11010011>
- Li, W., Guo, Q., Jakubowski, M. K., & Kelly, M. (2012). A new method for segmenting individual trees from the lidar point cloud. *Photogrammetric Engineering & Remote Sensing*, 78(1), 75–84. <https://doi.org/10.14358/PERS.78.1.75>
- Lin, H., Nazari, M., & Zheng, D. (2023). *Pctrees—3d point cloud tree species classification using airborne lidar images*.
- Liu, Y., Zhan, Z., Ren, L., Ze, S., Yu, L., Jiang, Q., & Luo, Y. (2021). Hyperspectral evidence of early-stage pine shoot beetle attack in Yunnan pine. *Forest Ecology and Management*, 497, 119505. <https://doi.org/10.1016/j.foreco.2021.119505>
- Losey, J. E., & Vaughan, M. (2006). The economic value of ecological services provided by insects. *BioScience*, 56(4), 311–323. [https://doi.org/10.1641/0006-3568\(2006\)56\[311:TEVOES\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)56[311:TEVOES]2.0.CO;2)
- Madawy, K. E., Rashed, H., Sallab, A. E., Nasr, O., Kamel, H., & Yogamani, S. (2019). RGB and lidar fusion based 3d semantic segmentation for autonomous driving. *arXiv*, arXiv:1906.00208. <https://doi.org/10.48550/arXiv.1906.00208>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2_Part_1), 209–220.
- Marcel, S., & Rodriguez, Y. (2010). Torchvision the machine-vision package of torch. *Proceedings of the 18th ACM International Conference on Multimedia*, pp. 1485–1488. <https://doi.org/10.1145/1873951.1874254>
- Menz, M. H. M., Scacco, M., Bürki-Spycher, H.-M., Williams, H. J., Reynolds, D. R., Chapman, J. W., & Wikelski, M. (2022). Individual tracking reveals long-distance flight-path control in a nocturnally migrating moth. *Science (New York, N.Y.)*, 377(6607), 764–768. <https://doi.org/10.1126/science.abn1663>
- Meyer, F., & Beucher, S. (1990). Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1), 21–46. [https://doi.org/10.1016/1047-3203\(90\)90014-M](https://doi.org/10.1016/1047-3203(90)90014-M)
- Milligan, P. D., Martin, T. A., John, G. P., Riginos, C., Goheen, J. R., Carpenter, S. M., & Palmer, T. M. (2021). Mutualism disruption by an invasive ant reduces carbon fixation for a foundational east African ant-plant. *Ecology Letters*, 24(5), 1052–1062. <https://doi.org/10.1111/ele.13725>
- Milligan, P. D., Martin, T. A., Pringle, E. G., Riginos, C., Mizell, G. M., & Palmer, T. M. (2022). A soil-nesting invasive ant disrupts carbon dynamics in saplings of a foundational ant-plant. *Journal of Ecology*, 110(2), 359–373. <https://doi.org/10.1111/1365-2745.13803>
- Moriya, É. A. S., Imai, N. N., Tommaselli, A. M. G., Berveglieri, A., Santos, G. H., Soares, M. A., Marino, M., & Reis, T. T. (2021). Detection and mapping of trees infected with citrus gummosis using UAV hyperspectral data. *Computers and Electronics in Agriculture*, 188, 106298. <https://doi.org/10.1016/j.compag.2021.106298>
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., ... Weedon, J. (2022). *Vegan: Community ecology package*. <https://CRAN.R-project.org/package=vegan>
- Palmer, T. M. (2003). Spatial habitat heterogeneity influences competition and coexistence in an African acacia ant guild. *Ecology*, 84(11), 2843–2855. <https://doi.org/10.1890/02-0528>
- Palmer, T. M., & Brody, A. K. (2007). Mutualism as reciprocal exploitation: African plant-ants defend foliar but not reproductive structures. *Ecology*, 88(12), 3004–3011. <https://doi.org/10.1890/07-0133.1>
- Palmer, T. M., & Brody, A. K. (2013). Enough is enough: The effects of symbiotic ant abundance on herbivory, growth, and reproduction in an African acacia. *Ecology*, 94(3), 683–691. <https://doi.org/10.1890/12-1413.1>
- Palmer, T. M., Doak, D. F., Stanton, M. L., Bronstein, J. L., Kiers, E. T., Young, T. P., Goheen, J. R., & Pringle, R. M. (2010). Synergy of multiple partners, including freeloaders, increases host fitness in a multispecies mutualism. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40), 17234–17239. <https://doi.org/10.1073/pnas.1006872107>
- Palmer, T. M., Stanton, M. L., Young, T. P., Goheen, J. R., Pringle, R. M., & Karban, R. (2008). Breakdown of an ant-plant mutualism follows the loss of large herbivores from an African savanna. *Science*, 319(5860), 192–195. <https://doi.org/10.1126/science.1151579>
- Palmer, T. M., Stanton, M. L., Young, T. P., Lemboi, J. S., Goheen, J. R., & Pringle, R. M. (2013). A role for indirect facilitation in maintaining diversity in a guild of African acacia ants. *Ecology*, 94(7), 1531–1539. <https://doi.org/10.1890/12-1873.1>
- Palmer, T. M., & Young, T. P. (2017). Integrating ecological complexity into our understanding of ant-plant mutualism: Ant-acacia interactions in African savannas. In P. S. Oliveira & S. Koptur (Eds.), *Ant-plant interactions* (1st ed., pp. 200–222). Cambridge University Press. <https://doi.org/10.1017/9781316671825.011>
- Palmer, T. M., Young, T. P., Stanton, M. L., & Wenk, E. (2000). Short-term dynamics of an acacia ant community in Laikipia, Kenya. *Oecologia*, 123(3), 425–435. <https://doi.org/10.1007/s004420051030>
- Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). *Pytorch: An imperative style, high-performance deep learning library*. *arXiv*, arXiv:1912.01703. <https://doi.org/10.48550/arXiv.1912.01703>
- PDAL Contributors. (2020). Pdal point data abstraction library. *Zenodo*. <https://doi.org/10.5281/zenodo.2556737>
- Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology*, 87(10), 2614–2625. [https://doi.org/10.1890/0012-9658\(2006\)87\[2614:VPOSDM\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2)

- Plowright, A., & Roussel, J.-R. (2021). *ForestTools: Analyzing remotely sensed forest data* (0.2.5) [Computer software]. <https://cran.r-project.org/web/packages/ForestTools/index.html>
- Popescu, S. C., & Wynne, R. H. (2004). Seeing the trees in the forest. *Photogrammetric Engineering & Remote Sensing*, 70(5), 589–604. <https://doi.org/10.14358/PERS.70.5.589>
- Propastin, P. A. (2013). Satellite-based monitoring system for assessment of vegetation vulnerability to locust hazard in the river Ili delta (Lake Balkhash, Kazakhstan). *Journal of Applied Remote Sensing*, 7(1), 075094. <https://doi.org/10.1117/1.JRS.7.075094>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rhodes, M. W., Bennie, J. J., Spalding, A., French-Constant, R. H., & Maclean, I. M. D. (2022). Recent advances in the remote sensing of insects. *Biological Reviews*, 97(1), 343–360. <https://doi.org/10.1111/brv.12802>
- Ruder, S. (2017). *An overview of gradient descent optimization algorithms*. arXiv, arXiv:1609.04747. <https://doi.org/10.48550/arXiv.1609.04747>
- Schowalter, T. D. (2017). Biology and management of the forest tent caterpillar (Lepidoptera: Lasiocampidae). *Journal of Integrated Pest Management*, 8(1), 24. <https://doi.org/10.1093/jipm/pmx022>
- Seidel, D., Annighöfer, P., Thielman, A., Seifert, Q. E., Thauer, J.-H., Glatthorn, J., Ehbrecht, M., Kneib, T., & Ammer, C. (2021). Predicting tree species from 3d laser scanning point clouds using deep learning. *Frontiers in Plant Science*, 12, 635440. <https://doi.org/10.3389/fpls.2021.635440>
- Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution*, 31(1), 67–80. <https://doi.org/10.1016/j.tree.2015.11.005>
- Speight, M. R., & Wylie, F. R. (2001). *Insect pests of tropical forestry*. CABI Pub.
- Stanton, M. L., Palmer, T. M., & Young, T. P. (2002). Competition–colonization trade-offs in a guild of African acacia-ants. *Ecological Monographs*, 72(3), 347–363. [https://doi.org/10.1890/0012-9615\(2002\)072\[0347:CCTOIA\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2002)072[0347:CCTOIA]2.0.CO;2)
- Stanton, M. L., Palmer, T. M., Young, T. P., Evans, A., & Turner, M. L. (1999). Sterilization and canopy modification of a swollen thorn acacia tree by a plant-ant. *Nature*, 401(6753), 578–581. <https://doi.org/10.1038/44119>
- Stone, G. N., & Schönrogge, K. (2003). The adaptive significance of insect gall morphology. *Trends in Ecology & Evolution*, 18(10), 512–522. [https://doi.org/10.1016/S0169-5347\(03\)00247-7](https://doi.org/10.1016/S0169-5347(03)00247-7)
- Terryn, L., Calders, K., Åkerblom, M., Bartholomeus, H., Disney, M., Levick, S., Origo, N., Raunonen, P., & Verbeeck, H. (2023). Analysing individual 3D tree structure using the R package ITSM. *Methods in Ecology and Evolution*, 14(1), 231–241. <https://doi.org/10.1111/2041-210X.14026>
- Vandermeer, J., Perfecto, I., & Philpott, S. M. (2008). Clusters of ant colonies and robust criticality in a tropical agroecosystem. *Nature*, 451(7177), 457–459. <https://doi.org/10.1038/nature06477>
- Vierling, K. T., Vierling, L. A., Gould, W. A., Martinuzzi, S., & Clawges, R. M. (2008). Lidar: Shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment*, 6, 11–22. <https://doi.org/10.1890/070001>
- Wagner, D., & Fleur Nicklen, E. (2010). Ant nest location, soil nutrients and nutrient uptake by ant-associated plants: Does extrafloral nectar attract ant nests and thereby enhance plant nutrition? *Journal of Ecology*, 98(3), 614–624. <https://doi.org/10.1111/j.1365-2745.2010.01640.x>
- Wang, R., & Gamon, J. A. (2019). Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, 231, 111218. <https://doi.org/10.1016/j.rse.2019.111218>
- Wang, Z. (2023). Supplementary material for “Automated detection of an insect-infested keystone vegetation phenotype using airborne LiDAR” [Data set]. Zenodo, <https://doi.org/10.5281/zenodo.8175736>
- Wilson, E. O. (1987). The little things that run the world (the importance and conservation of invertebrates). *Conservation Biology*, 1(4), 344–346.
- You, K., Long, M., Wang, J., & Jordan, M. I. (2019). How does learning rate decay help modern neural networks? arXiv, arXiv:1908.01878. <https://doi.org/10.48550/arXiv.1908.01878>
- Young, T. P., Stubblefield, C. H., & Isbell, L. A. (1996). Ants on swollen-thorn acacias: Species coexistence in a simple system. *Oecologia*, 109(1), 98–107. <https://doi.org/10.1007/s004420050063>
- Yu, R., Luo, Y., Zhou, Q., Zhang, X., Wu, D., & Ren, L. (2021). A machine learning algorithm to detect pine wilt disease using UAV-based hyperspectral imagery and LiDAR data at the tree level. *International Journal of Applied Earth Observation and Geoinformation*, 101, 102363. <https://doi.org/10.1016/j.jag.2021.102363>
- Zeng, T., Li, C., Zhang, B., Wang, R., Fu, W., Wang, J., & Zhang, X. (2022). Rubber leaf disease recognition based on improved deep convolutional neural networks with a cross-scale attention mechanism. *Frontiers in Plant Science*, 13, 829479. <https://doi.org/10.3389/fpls.2022.829479>
- Zou, X., Cheng, M., Wang, C., Xia, Y., & Li, J. (2017). Tree classification in complex forest point clouds based on deep learning. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2360–2364. <https://doi.org/10.1109/LGRS.2017.2764938>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1: Geo-tagging acacia trees to match ant occupancy data with LiDAR using sub-1m accuracy dGPS.

Figure S2: Tree segmentation based on canopy height models (CHM) identifies tree presence but introduces two kinds of error that need to be addressed with a customized tree segmentation pipeline.

Figure S3: Consistency of principal coordinate analysis (PCoA) between ground survey and classifier predictions across 410 quadrats.

Figure S4: Consistency of principal component analysis (PCA) between ground survey and classifier predictions across 410 quadrats.

Figure S5: Changes in principal coordinates under simulated turn over scenario are tracked by those under the same simulation on classifier-generated predictions.

How to cite this article: Wang, Z., Huben, R., Boucher, P. B., Van Amburg, C., Zeng, J., Chung, N., Wang, J., King, J., Knecht, R. J., Ng'iru, I., Baraza, A., Baker, C. C. M., Martins, D. J., Pierce, N. E., & Davies, A. B. (2024). Automated detection of an insect-induced keystone vegetation phenotype using airborne LiDAR. *Methods in Ecology and Evolution*, 00, 1–16. <https://doi.org/10.1111/2041-210X.14298>