Benchmarking airborne laser scanning tree segmentation algorithms in broadleaf forests shows high accuracy only for canopy trees

Yujie Cao [a], James G.C. Ball [b,c], David A. Coomes [b], Leon Steinmeier [d], Nikolai Knapp [e,f], Phil Wilkes [g,h], Mathias Disney [g,h], Kim Calders [i], Andrew Burt [g], Yi Lin [a,**], Toby D. Jackson [b,*]

[a] College of Surveying and Geo-informatics, Tongji University, No.1239 Siping Road, Shanghai, 200092, China
[b] Plant Sciences and Conservation Research Institute, University of Cambridge, Cambridge, CB2 3QZ, United Kingdom
[c] UMR AMAP, CIRAD, Montpellier, France
[d] Department of Modelling and Valuation, Helmholtz Institute Freiberg for Resource Technology (HIF), Chemnitzer Str. 40, 09599 Freiberg, Germany
[e] Thünen Institute of Forest Ecosystems, Alfred-Möller-Str. 1, 16225 Eberswalde, Germany
[f] Department of Ecological Modeling, Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany
[g] Department of Geography, University College London, London WC1E 6BT, United Kingdom
[h] NERC National Centre for Earth Observation (NCEO), UCL, Gower Street, London WC1E 6BT, United Kingdom
[i] CAVElab – Computational \& Applied Vegetation Ecology, Department of Environment, Ghent University, Coupure Links 653, 9000 Gent, Belgium

ARTICLE INFO

ABSTRACT

Individual tree segmentation from airborne laser scanning data is a longstanding and important challenge in forest remote sensing. Tree segmentation algorithms are widely available, but robust intercomparison studies are rare due to the difficulty of obtaining reliable reference data. Here we provide a benchmark data set for temperate and tropical broadleaf forests generated from labelled terrestrial laser scanning data. We compared the performance of four widely used tree segmentation algorithms against this benchmark data set. All algorithms performed reasonably well on the canopy trees. The point cloud-based algorithm AMS3D (Adaptive Mean Shift 3D) had the highest overall accuracy, closely followed by the 2D raster based region growing algorithm Dalponte2016 +. However, all algorithms failed to accurately segment the understory trees. This result was consistent across both forest types. This study emphasises the need to assess tree segmentation algorithms directly using benchmark data, rather than comparing with forest indices such as biomass or the number and size distribution of trees. We provide the first openly available benchmark data set for tropical forests and we hope future studies will extend this work to other regions.

1. Introduction

Aerial laser scanning (ALS) is widely used in forest ecology, but automatic individual tree segmentation (ITS) in dense broadleaf forests remains a key outstanding challenge (Qin et al., 2022). Accurate ITS algorithms would enable researchers to study tree growth, mortality, leaf phenology and carbon dynamics remotely, providing opportunities to track changes at landscape scales. These remote sensing methods complement existing field-based methods, which are essential but limited in scale (Ke et al., 2011). Many ITS algorithms have been developed, but robust comparisons of their accuracy are lacking, particularly for broadleaf forests. In this study we address this knowledge gap by generating an ALS tree segmentation benchmark data set for temperate and tropical broadleaf forests and using it to assess the accuracy of leading ITS algorithms.

There are two broad categories of ITS algorithms: 2D raster algorithms and 3D point cloud algorithms. Raster ITS algorithms are based on a 2D top of canopy height matrix (or raster). This means that information about understory trees is excluded and subcanopy trees cannot be segmented. 2D-raster ITS algorithms use clustering or image object detection methods like K-means (Modrsdorf et al., 2003), Pouring algorithm (Weinacker et al., 2004), Centroidal Voronoi Tessellation (Silva et al., 2016) and other variants (Hui et al. 2022; Liu et al., 2015; Chen et al., 2006). A widely used 2D-raster ITS algorithms is the region growing algorithm we refer to as Dalponte2016(Dalponte and Coomes., 2016) which has been applied across many forest types (Coomes et al.,

2017a; Minarik et al., 2020; Junttila et al., 2022).

Recent developments have focused on analysing the 3D point cloud ITS algorithms searching for clusters of points which may represent tree crowns including those of understory trees. These methods include rule-based and data-driven methods. Rule-based methods extract individual trees with a series of user-defined spatial constraints such as relative distances between trees, shape indices calculated from horizontal

projections and point density changes (Xu et al., 2018; Li et al., 2012; Harmraz et al., 2016). However, these methods often oversimplify forest structure and are thus not applicable to diverse forest types. Data-driven approaches rely on unsupervised clustering algorithms to extract tree boundaries. They include prior knowledge of forest structure before (Ferraz et al., 2016; Ayrey et al., 2017; Dai et al., 2018; Strimbu et al., 2015) or after (Amiri et al., 2018) the clustering stage for the purpose of
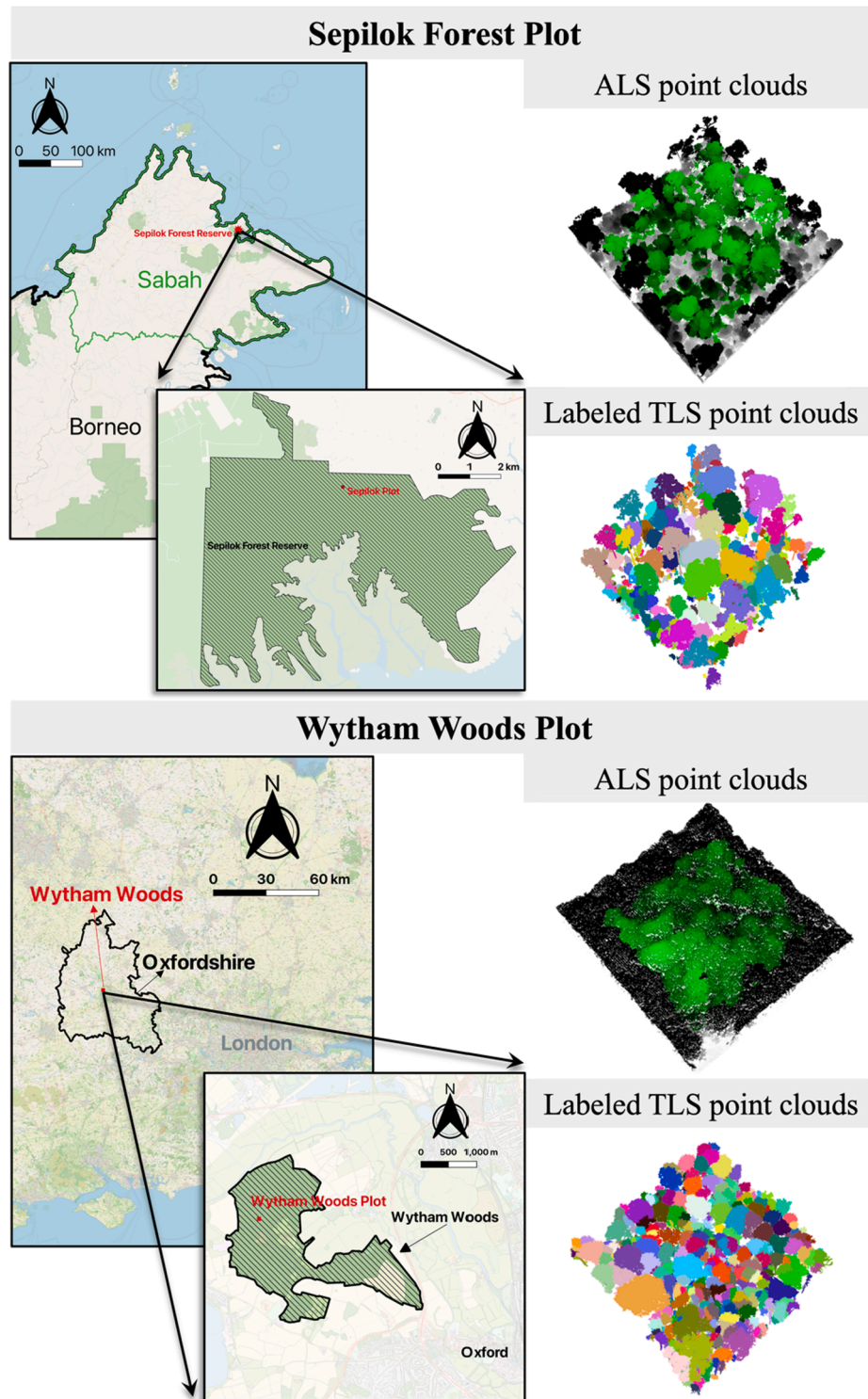
Fig. 1. Plot Location, raw ALS data and segmented TLS data of Sepilok Forest (upper panel) and Wytham Woods (bottom panel): Green areas are core areas of raw ALS datasets; Black areas are buffered to obtain complete ALS individual tree point clouds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

initialization or crown refinement.

Previous efforts to assess ITS algorithms have been limited by the available ground truth data. Many ITS algorithms were assessed based on their ability to predict some property of the forest such as: the number of trees (Ma et al., 2022; Vega et al., 2014), tree trunk position (Mongus and Zalik et al., 2015), tree height (Jakubowski et al., 2013), crown size (Duncanson et al., 2014; Lee et al., 2010), diameter at breast (Dalponte and Coomes., 2016; Williams et al., 2019), stem volume (Hyyppa et al., 2020) or above ground biomass (Ferraz et al., 2020). However, ITS algorithms tuned to predict these forest properties may segment individual trees with low accuracy. In addition, ITS algorithms tuned in this indirect way are unlikely to generalize well and may produce unexpected results when applied to a new data set.

Direct ITS algorithm assessment studies have used manually interpreted point clouds (Dai et al., 2018) and segmented tree crowns (Ke and Quackenbush., 2018), which allows the algorithms to be assessed directly on their ability to segment trees. However, manually segmented tree crowns are usually biased towards the large visible, canopy trees only (Aubry-Keintz et al., 2019). To the best of our knowledge all available benchmark data sets focus on conifers or temperate deciduous forests with relatively open canopy (Liang et al., 2018; Wang et al., 2016). Virtual laser scanning provides highly detailed simulated ALS data and has been used to assess ITS algorithms (Xiao et al., 2019; Wang, 2020). However, assessment on real data (i.e., not simulated data) is still required to evaluate the performance of ITS algorithms (Winiwarter et al., 2022) in practical terms.

In this study, our key objectives are to:

1. Generate an ALS benchmark dataset with individual trees labelled in both the canopy and understory and make this openly available online.
2. Use this benchmark data to robustly compare the accuracy of ITS algorithms, which is a key limitation of previous work.
3. Explore the sensitivity of each ITS algorithm to its key input parameters.

2. Materials and methods

2.1. Field data

Temperate broadleaf forests (Wytham Woods, UK) and Tropical rainforest (Sepilok Reserve, Malaysia) were included in this study (see Fig. 1.).

Sepilok Forest Reserve (117 56′ E, 5 10′ N) is a tropical rain forest located close to the north-east coast of Sabah, Malaysia (Jucker et al., 2018). It is one of the oldest protected tropical forests in Southeast Asia, covering around 4500 ha with elevation ranging from 50 m to 250 m. The 1 ha plot used in our study is in the tall and highly diverse alluvial dipterocarp forest. The maximum diameter at breast height is 159.2 cm, while the tallest tree is around 66 m.

Wytham Woods is a semi-natural woodland, located on a gentle hill with the maximum elevation rising to 165 m in Oxfordshire, England (1 20′ W, 51 47′ N) (Butt et al., 2009). Since 2008, an 18-ha permanent inventory plot was established. The woodland is mixed deciduous forest, dominated by Acer pseudoplatanus (Sycamore), Fraxinus excelsior (Ash), Corylus avellana (Hazel) and Quercus robur (Oak). The trees are diverse in terms of size and distribution with a maximum diameter at breast height of 141.2 cm, and the tallest tree is about 40 m.

2.2. LiDAR data

2.2.1. ALS data

In Sepilok, ALS data was acquired in February 2020 using a RIEGL LMSQ560 with the flight height of 200 m. the instrument has a beam divergence of<0.5 mrad and ± 30-degree scanning angle. For the focal area, the ALS data has a point density of about 139 per m-2. In Wytham

Woods, ALS data was collected in summer 2014 using a Leica ALS50-II with the altitude of 2000 m. It has a beam divergence of<0.22 mrad and ± 14-degree scanning angle. Over the area of interest, it has a point density of around 6.26 per m-2 (see Table 1).

2.2.2. TLS data

At both locations terrestrial laser scanning (TLS) data was captured using a RIEGL VZ-400 (RIEGL, Horn, Austra) following standard protocols described in (Wilkes et al., 2017). The instrument has a beam divergence of 0.35 mrad and operates in the infrared (wavelength 1550 nm) with a range up to 350 m, a pulse repetition rate of 300 kHz and the angular sampling resolution was 0.04°. At each scan position, two scans were acquired where the scanner rotation axis was perpendicular and parallel to the ground surface respectively. In both sites, reflective targets were located between scan positions to aid with co-registration (Wilkes et al., 2017). Scans were registered in RiSCAN Pro in a two-step process where (1) the reflective targets were used to generate a coarse registration (2) a set of planes generated from the point cloud were used in a multi-station adjustment to improve the registration. The point cloud was then downsampled using a voxel size of 0.026 m and 0.02 m for Wytham and Sepilok respectively.

At Wytham, scans were done within a larger 6 ha area to ensure the best possible data quality within our 1 ha study area (Calders et al. 2022). TLS data were collected throughout December 2015 and January 2016 (in leaf-off conditions) on a 20 m × 20 m grid (Calders et al.,2018). Individual trees were segmented from the larger point cloud using *treeseg* (Burt et al., 2019) and manually checked for quality assurance.

In Sepilok, the study area was in the tall alluvial forest and the data were collected in 2017. The local 1 ha plot reference is RP292/1 and SEP12 in the forestplots.net data base. TLS data was captured across a 1 ha plot from 121 scan positions on a 10 m × 10 m grid. Trees were segmented from the point cloud using *TLS2trees* (Wilkes et al., 2022). Briefly, trees are segmented using a 2-step process where a semantic segmentation is used to classify points into ground, leaf, wood, and coarse woody debris. Using the wood and leaf classes only, a graph-based instance segmentation is used where woody stems are first identified, leaf points were subsequently added to individual stems. Resulting segmented trees were manually checked for quality assurance.

Note that TLS scanning parameters and tree segmentation workflow differed between the two sites. However, in both cases the TLS data contains more than enough detail to accurately label the individual trees in the ALS data.

2.3. Creating ALS benchmark data sets

We aligned the ALS and TLS data using a geo-coordinate transformation with 9 manually selected feature points. These features were selected as the intersection of two branches in the upper canopy, so they could be easily identified in both ALS and TLS. The root mean square error in alignment for Sepilok was 0.192 m and for Wytham was 0.625 m. We created a polygon to outline the core area in which the TLS (1 ha) overlaps with the ALS (>100 ha). A kdTree was used for fast 3D data retrieval.

We labelled the individual trees in the ALS data using the segmented TLS data to produce a benchmark ALS data. Here, we used an iterative nearest neighbour voting strategy to label the ALS data. The labelling

Table 1
TLS and ALS data details for Sepilok Forest and Wytham Woods.

	Sepilok Forest	Wytham Woods
Plot Area (ha)	1	1
Number of trees	430	523
TLS point density(pts/m²)	>1000	>1000
ALS point density(pts/m²)	139	7
ALS-TLS alignment RMSE(m)	0.192	0.625

strategy consisted of the following steps (see Fig. 2 for workflow and Fig. 3 for resulting benchmark data set).

(1) For each ALS point, we selected a maximum of 900 (Sepilok) or 300 (Wytham) surrounding points within a radius of 1.4 m (Sepilok) or 2.0 m (Wytham).

(2) The label of each ALS point was then assigned as the label of majority of these neighbour points. This process was repeated until it converged (i.e. the number of newly labelled ALS points

was<5). The pseudo-code demo can be found in Supplementary L.

2.4. Manually rating ALS benchmark trees with confidence score

A confidence score system was introduced to control the data quality of the labelled ALS benchmark tree (see Fig. 2, Supplementary C, D). We categorized the trees as high, medium, and low confidence. The scores were defined in terms of the number of points contributing to a crown
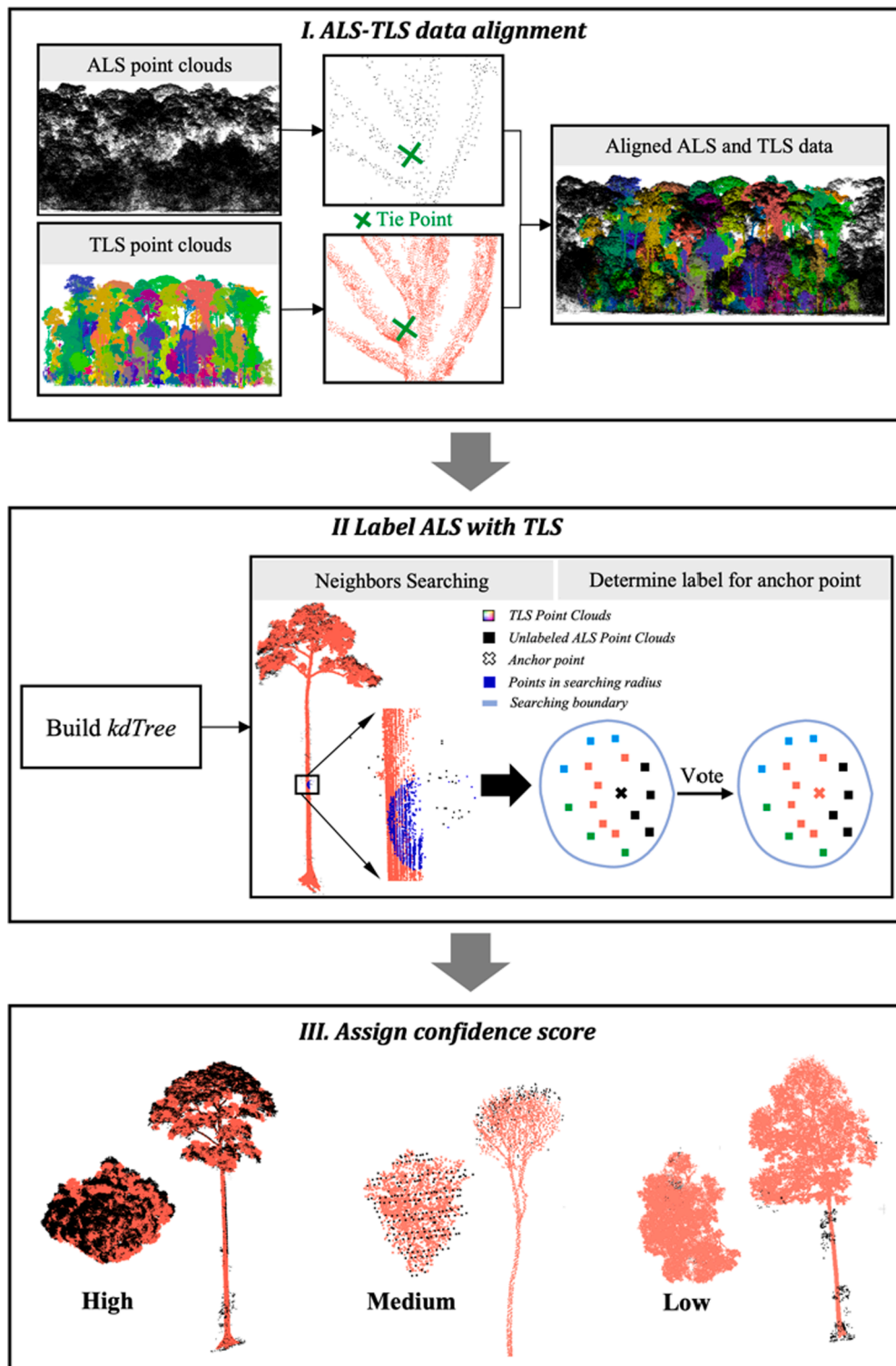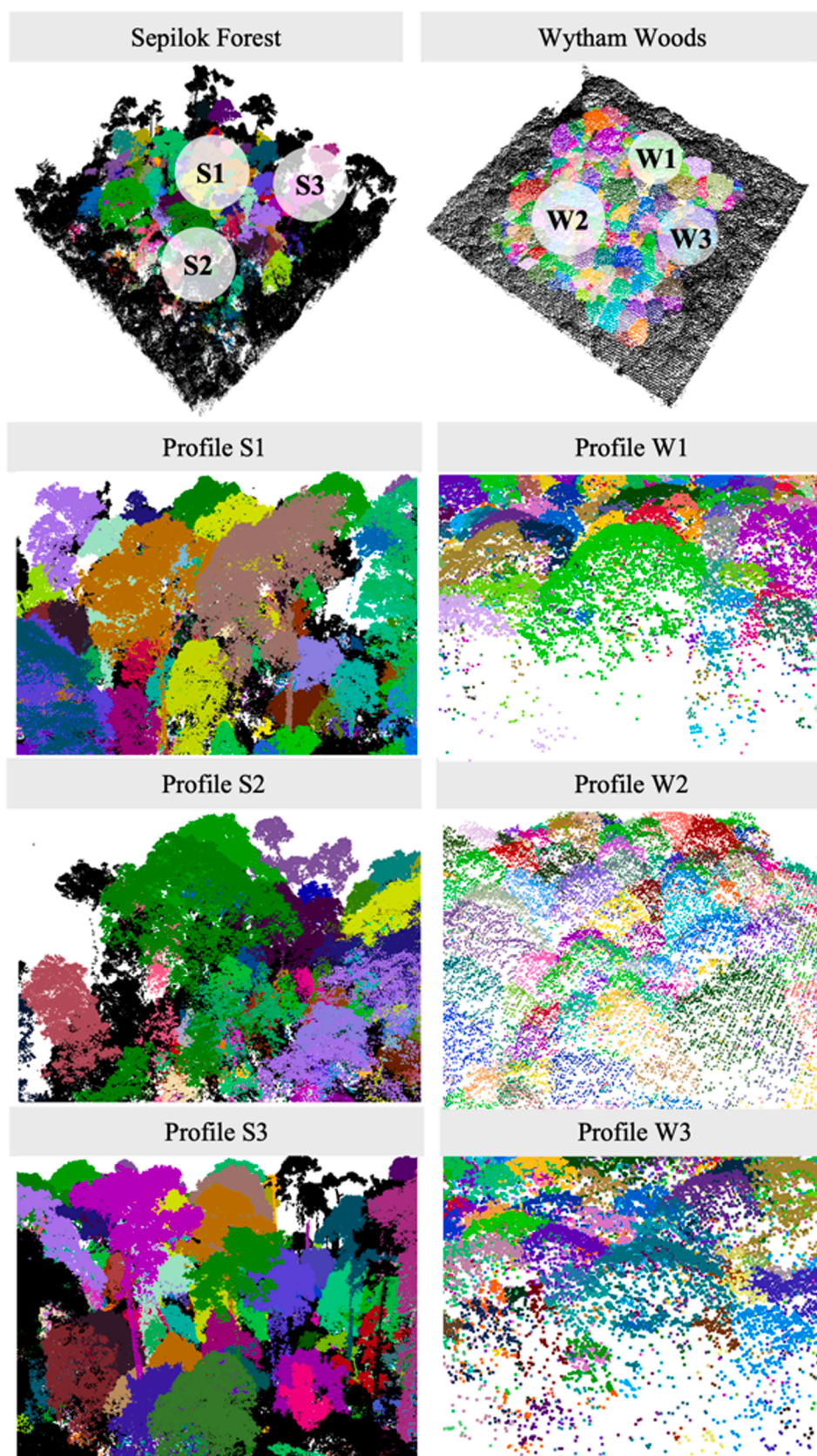

Fig. 2. Workflow to create ALS benchmark with detailed TLS data.

Fig. 3. ALS benchmark data: Left panel displays ALS data for Sepilok Forest, and Profile S1-S3 and W1-W3 show details of the two datasets.

polygon, the distribution of ALS point clouds with comparison to corresponding TLS data and bias of individual tree feature representation. In the benchmark data set, all points within a tree were assigned the tree-level confidence score. Importantly, ITS algorithms segmented trees on the raw ALS data, but only trees belonging to high and medium classes were used in our assessment. The confidence score of the predicted trees was assigned as the most common confidence score of the points within it. We filtered by confidence score during our assessment process to ensure the results are not due to poor quality reference data.

For a detailed description of the confidence score system see Supplementary B. Note that this confidence score system is intended to improve the benchmark data set and therefore the accuracy assessment, not for large scale tree segmentation projects.

2.5. ITS algorithm selection

We reviewed all highly cited ITS algorithms and summarized our findings in Table A1 of Supplementary A. We noted the algorithm type (2D or 3D), key methods, assessment method, forest type and overall accuracy and the number of citations. We chose to focus on four of the most highly cited algorithms, which are representative of different algorithm structures.

Dalponte2016 and Dalponte2016+. Dalponte2016 is the 2D-raster based ITS algorithm, which is widely used as it is integrated in the lidR package (Dalponte and Coomes., 2016). The algorithm firstly finds treetops with local maxima filtering on a top of canopy height raster and determines the tree crown boundary with a region growing method. During the region growing, TH_CR and TH_SEED are key parameters determining the edges of the crown. TH_SEED controls the height difference between the treetop and neighbouring pixels. A neighbouring pixel would be included in the region if its height is greater than the product of TH_SEED and the height of the treetop. While TH_CR controls the difference between the mean height of the regions and their surrounding pixels, and a neighbouring pixel would be view as part of the region if its height is greater than the product of TH_CR and the mean height of the growing region (see Table 2). Therefore, the two parameters range from 0 to 1. The code for Dalponte2016 used in the study can be found in (Roussel et al., 2020).

We also assessed an improved version of Dalponte2016 (hereafter Dalponte2016 +) (Coomes et al., 2017b), which can find more realistic treetops before the region growing stage. The window size increases

with canopy height so that taller trees have larger crowns. The rate at which the window size increases is defined by the user-defined parameter tau. To find the optimal tau for different forest structures, the global allometry database (Jucker et al., 2022) was used. The crown size and tree height relations were described using quantile regression. The parameter tau refers to the percentile of the quantile regression, which represents different crown size and tree height relations. The remaining parameters (TH_SEED and TH_CR) are defined in the same way in both Dalponte2016 and Dalponte2016 +. Comparing these algorithms is therefore an ideal way to isolate the effect of the variable window size on tree segmentation accuracy.

Li2012 (Li et al., 2012) is a rule-based point ITS algorithm, which has been integrated in the LiDAR processing packages (Roussel and Auty., 2022) and commercial software. This ITS algorithm starts with finding the global highest point in the data space. Then the highest point and a dummy point far away from the highest point are used to create two sets. The rest of points are assigned to the two sets following a top-down order, relative spacing, shape index and point density distribution. In the procedure, DT1 and DT2 are the key parameters (see Table 2), which present the minimal distance to tree set and non-tree set respectively. The range of the two parameters is from 1 to 10. Point clouds of individual trees can be obtained by repeating above procedure till no unlabelled point left. The code for Li2012 used in our experiment can be found in LidR package.

Adaptive Meanshift 3D (AMS3D) was implemented according to the descriptions by (Ferraz et al., 2012; Ferraz et al., 2016). It is based on the general principle of mean shift clustering (Cheng., 1995), applied to the 3-dimensional coordinate space of the lidar point cloud and adapted to the fact that tree crowns in upper canopy strata are larger than in lower strata. For every point in the point cloud the center of point density in a cylindrical search neighborhood (the kernel) is identified. The weight with which each neighbor point contributes to the calculation of the center depends on its relative position inside the kernel. The kernel is then shifted to this center of point density. This procedure is iteratively applied until the kernel reaches a stable position, which usually is closely underneath the apex of a tree crown. After a final kernel position for every point in the point cloud has been found, tree crown clusters are formed from all points for which the kernel positions converge at the same crown apex. Since the final kernel positions do not fully converge, the DBSCAN clustering algorithm (Ester et al., 2001) was applied to the final kernel positions to identify clusters. To enable fast AMS3D computations, the algorithm was implemented in C++ using an R*-tree spatial index structure for efficient spatial queries (Steinmeier., 2022).

The two parameters of the algorithm control the size of the kernel (cylinder height and diameter) as a function of kernel center height above ground. This height dependence ensures large crown clusters in the upper and small crown clusters in the lower canopy. Thus, typical crown length and crown diameter to tree height ratios are good choices for the parameters.

2.6. Assessing ITS algorithms accuracy against benchmark

Grid search over all parameters. We implemented each algorithm varying the key parameters using a grid search method to give a comprehensive overview of the sensitivity to these parameters, see Table 2. The outputs were assessed by how well they matched tree crown polygons from the benchmark data set (see Fig. 4). The accuracies given in Fig. 7 refer to the combination of parameters with the highest accuracy i.e. the best these algorithms can do on the current data. Full results for all combinations of parameters are given in the supplementary materials.

Assessment workflow and index. The assessment was carried out using polygons to represent the individual tree crowns from both the benchmark and predicted trees (Gillies et al., 2007) (see Fig. 4).

The ITS predictions were done on the raw ALS data covering an area

Table 2
Details for ITS algorithms and the parameters assessed in the experiment.

| Algorithm	Parameter	Parameter Range	Definition
Dalponte2016	TH_CR	[0, 0.1, 1]	Height threshold for including neighbouring pixel compared to average crown height
	TH_SEED	[0, 0.1, 1]	Height threshold for including neighbouring pixel compared to treetop height.
Dalponte2016+	Tau	[0, 10%, 100%]	Percentile of quantile regression between crown size and tree height
	TH_CR	[0, 0.1, 1]	TH CR [0, 0.1, 1] As in Dalponte2016.
	TH_SEED	[0, 0.1, 1]	TH SEED [0, 0.1, 1] As in Dalponte2016
Li2012	DT1	[1, 1, 10]	Minimal distance threshold of a point to the corresponding tree subset
	DT2	[1, 1, 10]	Minimal distance threshold of a point to the corresponding non tree subset
AMS3D	H2CD	[0, 0.1, 1]	Ratio of tree height to crown diameter to set kernel width.
	H2CH	[0, 0.1, 1]	Ratio of tree height to crown height to set kernel height.
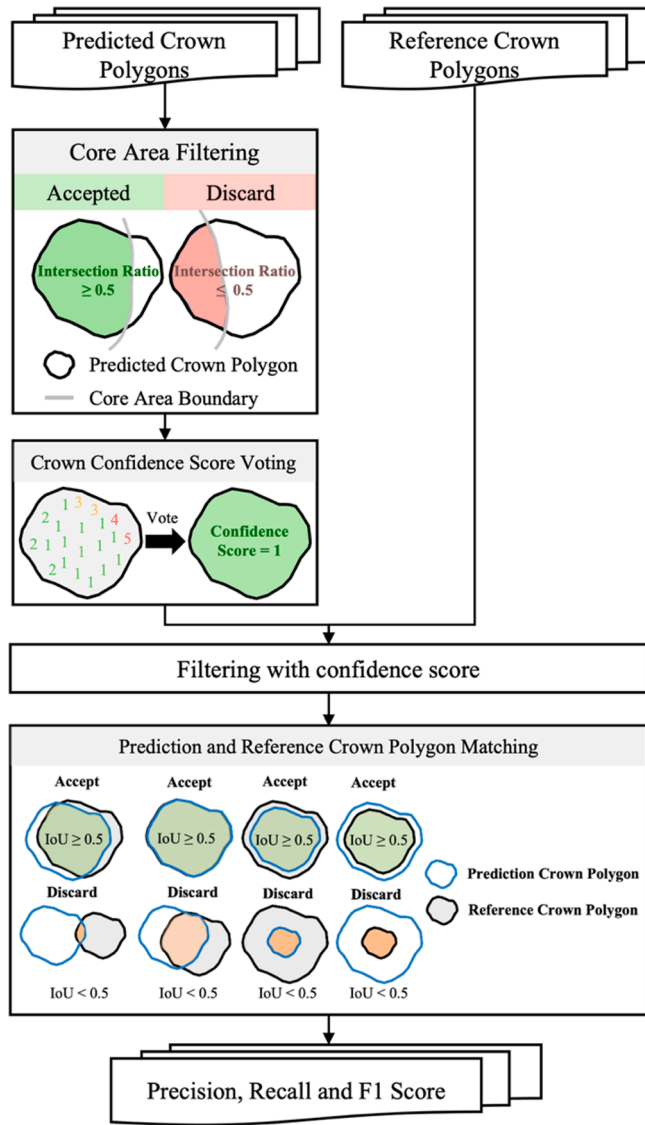
Fig. 4. Crown polygon-based assessment framework for parameter tuning and inter-comparison for ITS algorithms.

larger than the core plot to avoid edge effects. Predicted tree crown polygons with more than half their area inside the core plot area were selected as candidates. These predictions cover all trees in the plot, but we want to assess the accuracy only on those trees whose labels have 'good' or 'medium' confidence. We therefore filtered out predictions which mostly cover trees with low confidence labels. Next, the Intersection over Union (IoU) was calculated for each predicted tree crown polygon. the IoU of a predicted tree crown polygon can be determined by selecting the maximum IoU among its corresponding reference tree crown polygons. For those predicted tree crown polygons with its max IoU $\geq$ 0.5, their tree crown score would be 1, which means these tree crown polygons are correctly segmented, otherwise 0.

The true positive (TP) was calculated by accumulating all tree crown scores. Those predictions which failed to match benchmark crown polygons are false positive (FP), while those benchmark tree crowns with no corresponding crown polygons in the prediction were classified into false negative (FN). Finally, TP, FP, FN were used for obtaining Precision (Equation (2)), Recall (Equation (3) and F1 Score (Equation (4)) to represent algorithm performance with specified parameter combination.

$$TP = \sum_{i=1}^{N} TreeCrownScore_i \qquad (1)$$

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

$$F1Score = 2\frac{Precision \bullet Recall}{Precison + Recall} \qquad (4)$$

3. Results

3.1. Novel benchmark data set

We produced a novel benchmark data set for ITS algorithm assessment using TLS to label individual trees in ALS data (Fig. 2). We manually assigned confidence scores to 419 trees in Sepilok and 467 trees in Wytham. Overall, 249 (59.4%) trees in Sepilok and 225 (48.1%) trees in Wytham had a 'high' confidence score. Many canopy trees showed a visually perfect match between the TLS and ALS labels and taller trees generally higher confidence scores, mainly due to better ALS coverage. Importantly, our data set also contains many understory trees with 'good' or 'medium' confidence scores (75 trees in Sepilok Forests and 11 in Wytham Woods, see Fig. B1 in Supplementary B). Understory trees were defined by tree heights ($<$25 m for Sepilok and $<$ 15 m for Wytham). This enables us to assess ITS algorithm performance for understory trees. We tested the sensitivity of our results to the inclusion of trees with different confidence scores and found that our results were robust, as long as the low confidence trees were excluded.

3.2. Segmentation accuracy improves with tree height

ITS algorithms accuracy increased with tree heights across both sites (see Figs. 5, 6 and 7 for statistics).

In Wytham, the precision and recall of all four ITS algorithm increased with tree height. This was expected because the taller trees are more clearly visible in ALS data. The best algorithm overall was AMS3D, with precision 0.5 and recall 0.41 for the tallest trees ($>$25 m). All algorithms tended to slightly underestimate the number of canopy trees and dramatically underestimate the number of understory trees.

In Sepilok, Dalponte2016+, AMS3D and Li2012 increased in precision and recall with tree height. AMS3D had the highest F1 score for the tallest trees (F1 = 0.71, 55–65 m). We note that the $>$ 65 m class only contained 3 trees, so we report accuracies for the penultimate height bin. Dalponte2016 + and AMS3D showed moderate accuracy (mean F1 = 0.3 and 0.49 respectively) for the medium sized trees (25–55 m), while Li2012 performed poorly in this range (mean F1 = 0.11) and was only accurate for trees over 55 m. The performance of Dalponte2016 was low for both understory and canopy trees, and moderate for medium sized trees (F1 = 0.42 for 25–35 m trees). Note that all references to small, medium and tall trees throughout this manuscript are site-specific.

3.3. All ITS algorithms fail to segment understory trees

We tested the range of input parameters for each algorithm, and we report the results for the parameter set with the highest accuracy. All ITS algorithms performed poorly for understory trees in both Sepilok ($<$25 m, see Fig. 7) and Wytham ($<$15 m, see Fig. 7). All of the algorithms had precision and recall scores below 0.1 for understory trees, with the exception of Dalponte2016 in Sepilok, which had a moderate precision (0.33) but very low recall (0.01) and F1 score (0.01). Importantly, the Wytham benchmark data set has a low point density, so understory trees have fewer ALS points, meaning that they are very challenging to
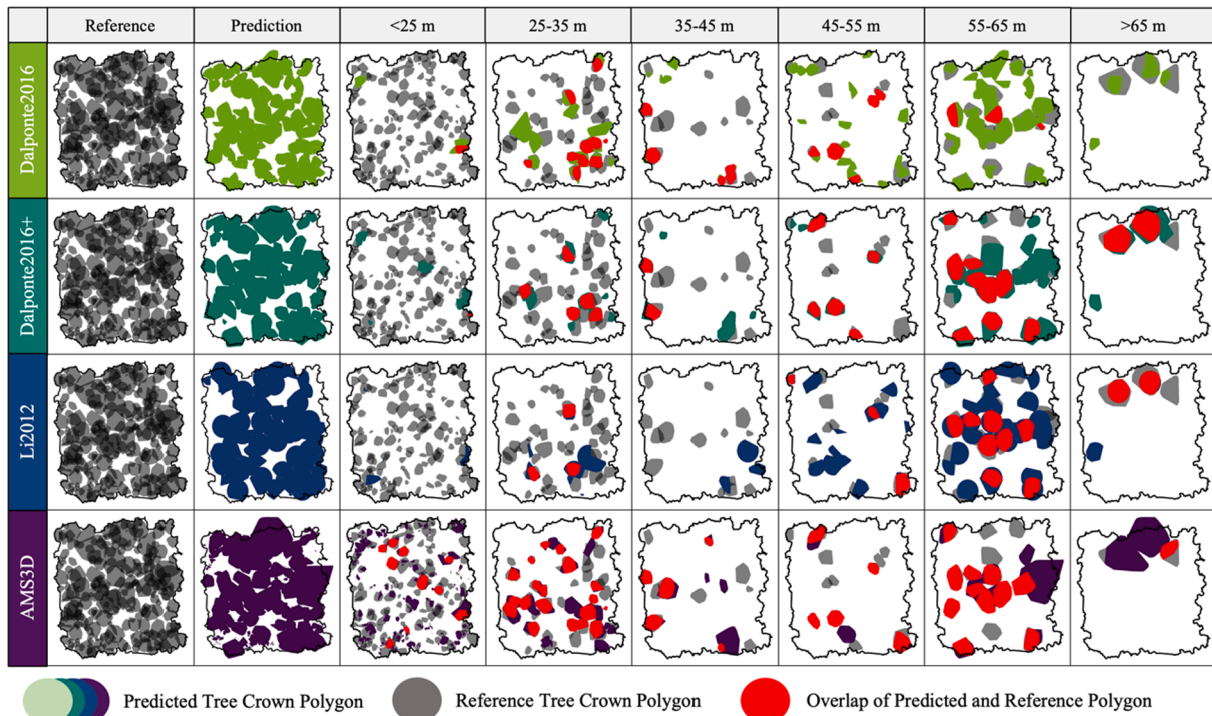
Fig. 5. Crown polygon visualization to display ITS algorithm performance changes with regards to tree heights in Sepilok Forest. The corresponding statistics are in Fig. 7.
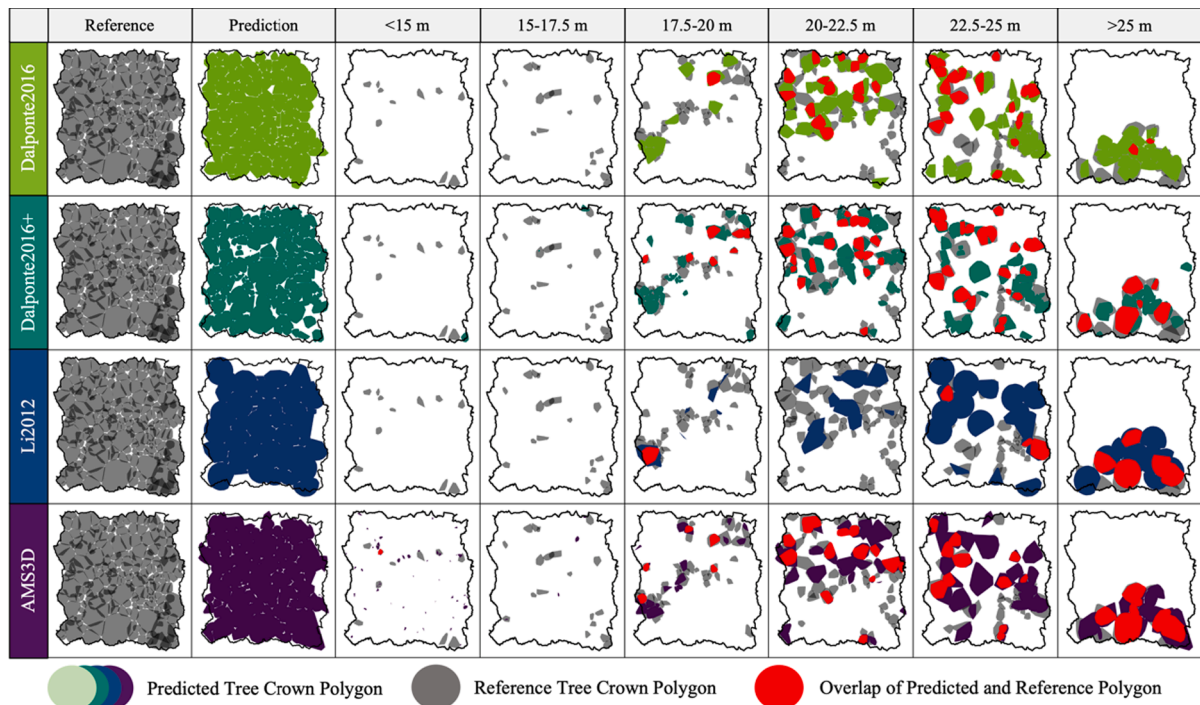

Fig. 6. Crown polygon visualization to display ITS algorithm performance changes with regards to tree heights in Wytham Woods. The corresponding statistics are in Fig. 7.

accurately segment. This is representative of many ALS data sets. The Sepilok benchmark data set has a high point density and many understory trees with good coverage, which were nevertheless poorly segmented.

The reason for this poor performance in Dalponte2016, Dalponte2016 + and Li2012 is that they predicted too few trees in these low height classes. In Sepilok, the benchmark data contained 153 trees < 25 m, while Dalponte2016 predicts 17, Dalponte2016 + predicted 45 and Li2012 predicts 16 respectively. In Wytham, the benchmark data contained 11 trees < 15 m, while Dalponte2015 predicted 0, Dalponte2016 + predicts 1 and Li2012 predicts 0. AMS3D fails for the opposite reason: it predicted large numbers of very small trees in the low height classes,
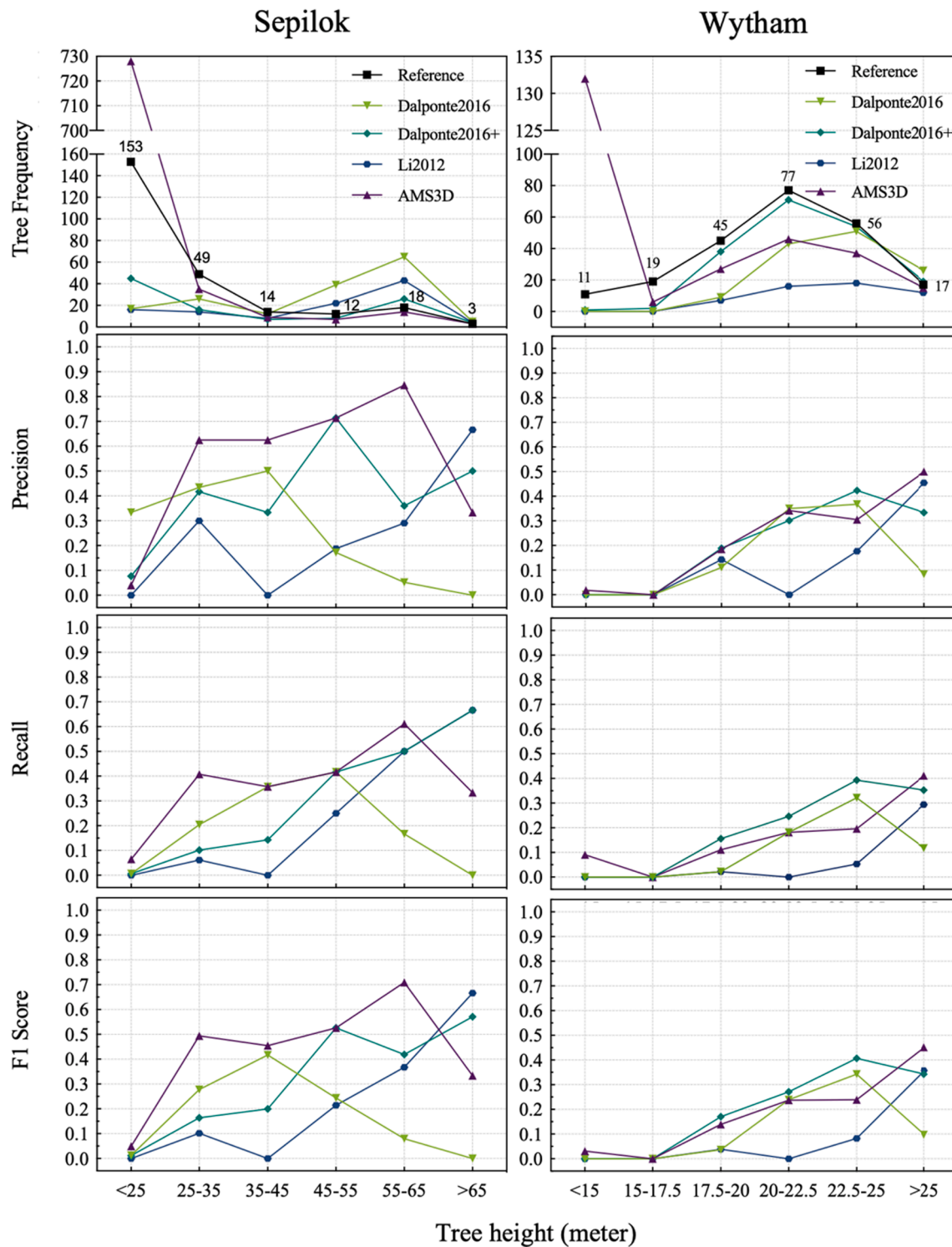
Fig. 7. Statistics in terms of tree frequency, precision, recall and F1 score to show ITS algorithm performances in Wytham Woods (left panel) and Sepilok Forest (right panel).

which do not match any of the trees in the benchmark data. Specifically, the most accurate predictions from AMS3D predicted 728 trees < 25 m in Sepilok and 132 trees < 15 m in Wytham.

3.4. Sensitivity of segmentation accuracy to allometric parameters

The two best algorithms, AMS3D and Dalponte2016 + both have an input parameter related to the tree height to crown diameter allometry

(H2CD and tau, respectively). This enables them to look for trees with larger crowns in the higher parts of the canopy. Unsurprisingly, the segmentation accuracy was highly sensitive to these parameters.

The accuracy of Dalponte2016 + was highest with tau = 80, where the search window size varies according to the the 80th percentile of the tree height to crown diameter allometry. The accuracy increased steadily from tau = 10 to tau = 80 and then dropped dramatically after this point (see Fig. 9). A similar pattern was observed in Wytham, with

precision, recall and F1 score increasing with tau within from 10% to 70%. The accuracies then plateaued until 90% and finally decreased at 99%. This drop in accuracy for large tau was particularly noticeable in the medium size trees in both sites. Dalponte2016 + failed to detect < 25 m trees in Sepilok and < 15 m trees in Wytham regardless of tau, while it had decent precision, recall and F1 score when faced > 55 m and > 25 m with F1 score can reach 0.57(Sepilok) and 0.42(Wytham) respectively.

The accuracy of AMS3D varied dramatically with the H2CD parameter, which controls the ratio between tree height and crown diameter. The accuracy was highest in with H2CD at 0.5 for Sepilok and 0.3 for Wytham. Variations in H2CD reduced this accuracy (see Fig. 8).

3.5. The number of trees predicted varies with allometric parameters

The number of trees predicted by each algorithm was highly sensitive to the allometric parameter (tau or H2CD). These algorithms can therefore be tuned to predict a realistic number and size distribution of trees, but this will reduce the overall accuracy of tree segmentation.

In Dalponte2016+, a higher tau led to fewer predictions throughout height ranges in both plots. Underestimation of the number of under-story trees occurred for all allometry percentiles (closest estimate was 132 out of 146 < 25 m trees in Sepilok and 4 out of 11 < 15 m trees in Wytham). Dalponte2016 + over-estimated the number of medium and large size trees when choosing < 60% allometry percentiles in two sites.

In Sepilok, no clear trend was found as most allometry percentiles tended to underestimate the number of understory trees while over-estimate medium and large trees. Among these allometry percentiles, 90% could be a better one if precision, recall and F1 score were also taken into consideration. A more complete visualization can be found in Fig. I1. of Supplementary I. In Wytham, Dalponte2016 + with 90% allometry percentile had predicted the number of trees > 35 m most accurately (182/193), though it still underestimated the < 35 m trees. A more complete visualization can be found in Fig. I2. of Supplementary I.
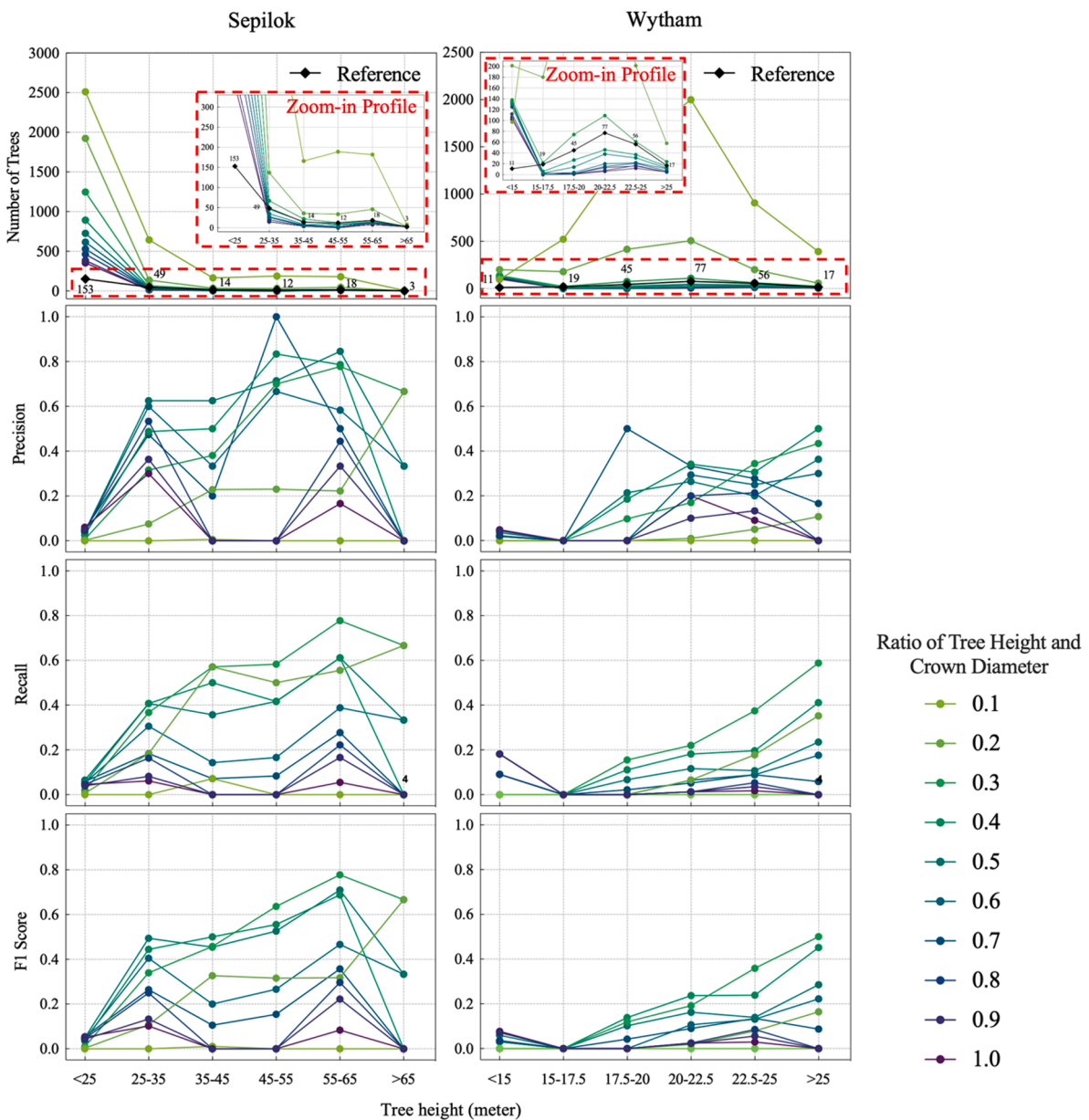

Fig. 8. Allometric relation's contributions to AMS3D; The left Panel and the right panel shows the effects of tree height and crown diameter (H2CD) on Dalponte2016 + in Sepilok and Wytham plots.
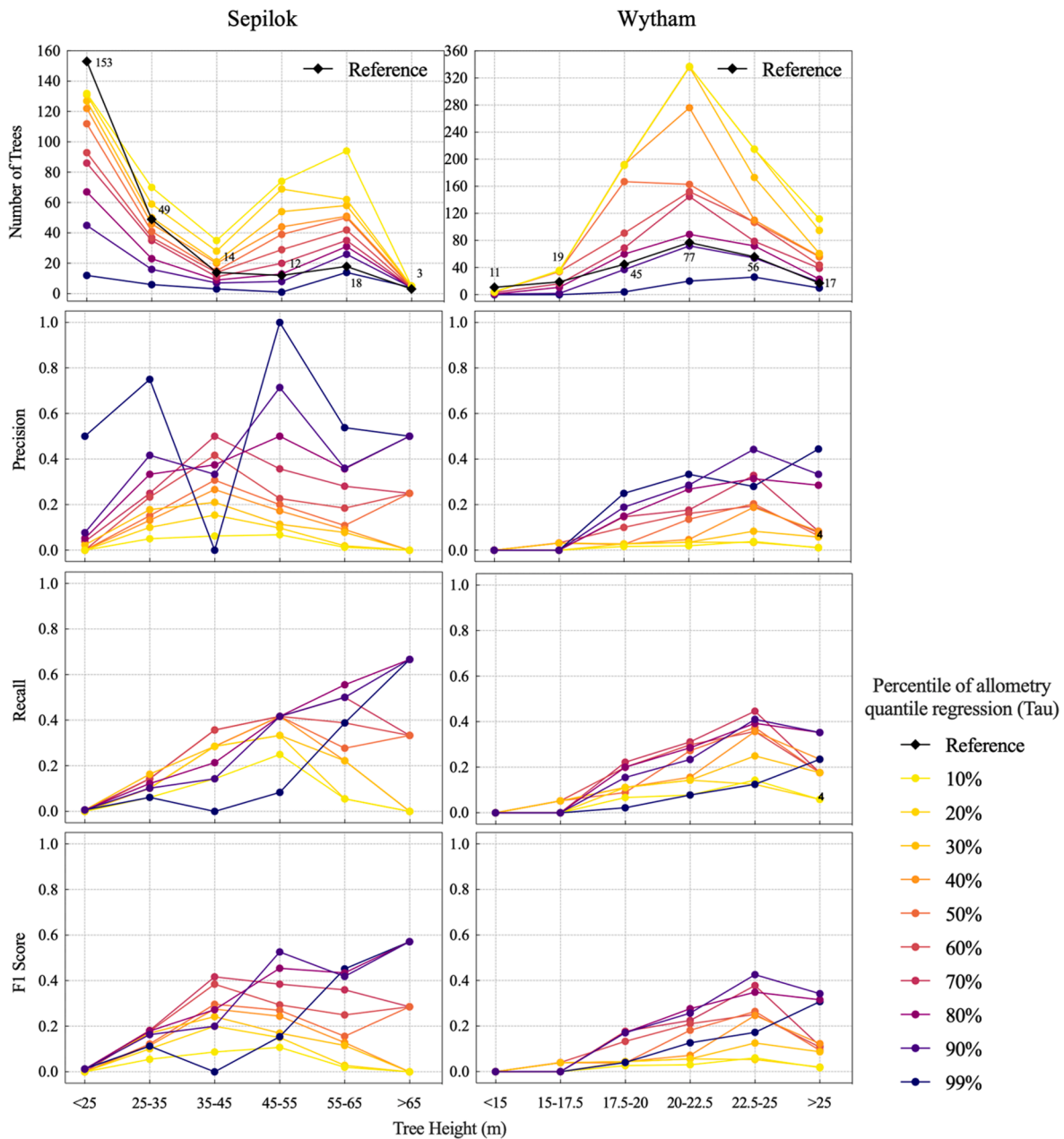
Fig. 9. Allometric relation's contributions to Dalponte2016+; The left Panel and the right panel shows the effects of tau on Dalponte2016 + in Sepilok and Wytham plots.

4. Discussion

4.1. Tree segmentation algorithms are only accurate for tall trees

We compared the accuracy of four widely used ITS algorithms for ALS data in 1 ha of temperate and 1 ha of tropical forest. Our most striking result was that all four ITS algorithms, including 3D-point-cloud algorithms, failed to accurately segment understory trees. Tree segmentation accuracy increased dramatically with tree height across both sites. This is likely because the top of the forest canopy is clearly visible to aerial surveys, while understory trees are often obscured. Note that we tested a wide range of input parameter combinations for all algorithms (see Supplementary E and F), and reported results from the parameter combination with highest accuracy. Also, we defined understory trees purely by their height, so some of the understory trees with better ALS coverage are found in canopy gaps.

In creating our benchmark data set, we found that many understory trees were poorly sampled by the ALS data and therefore had lower confidence scores (see supplementary B). This was particularly important in Wytham, which had lower ALS point density, making segmentation more challenging. The understory trees in Sepilok also had far fewer points and lower confidence scores, however, there remained over 100 understory trees with high confidence scores on which we assess the ITS algorithms.

4.2. Tree segmentation algorithms are more accurate when tuned to the local allometry

The most accurate ITS algorithm was AMS3D, followed by Dalponte2016 +. The key similarity between these two algorithms is that they both contain a parameter which describes the expected relationship between tree height and crown size. This allometric information is

widely available (Jucker et al., 2022) and we suggest that users choose ITS algorithms which incorporate this information.

The value of including allometric information is clearly demonstrated by the fact that Dalponte2016 + dramatically outperformed Dalponte2016. These two algorithms both initially detect treetops before 'growing' the tree crowns until they reach certain stopping conditions. The only difference between them is that Dalponte2016 + detects treetops using a searching window whose width increases with canopy height. This enables it to look for trees with large crowns in tall areas of forest, and trees with small crowns in short areas of forest.

Although allometric information can increase accuracy, it may also give false confidence if segmentation accuracy is not assessed directly. In Sepilok, we found that AMS3D could be tuned to predict roughly the correct number and size distribution of trees (with H2CD parameter = 0.9). However, majority of these predicted trees did not correspond to real trees in the ground truth data set. The best performance was achieved with (H2CD = 0.5), and the understory trees were not well segmented by any combination of parameters. We therefore advise caution when assessing ITS algorithm accuracy, especially if the algorithm has been tuned to a local allometry.

4.3. Value of ALS benchmark data sets

It is critical that tree segmentation accuracy is assessed directly, rather than by comparing with ecological indices such as biomass or the number and size distribution of trees. Benchmark data are needed to achieve this, but they are rare in broadleaf forests (see (Weiser et al., 2022)), especially in the tropics. Our Sepilok benchmark dataset is, to the best of our knowledge, the only available ITS benchmark data set in a tropical forest. The benchmark data sets used in this study (and provided online) are particularly valuable as they include many understory trees. This was possible because we labelled the individual trees in the ALS data using TLS data (Xin, 2021), which contains detailed information on understory trees. We controlled for data quality issues by manually assigning a confidence score to each tree. This enables direct accuracy assessment for the understory trees, whereas most manually interpreted benchmark data sets focus only on the easily visible canopy trees. Moreover, this also encourages more 3D quantitative metrics in terms of tree shape and crown to describe the performance of ITS algorithms, which allow more comprehensive investigation on the algorithm applicability. We hope other researchers use these benchmark data and follow our approach to produce additional benchmark sites, to help assess the accuracy of ITS algorithms across the tropics.

4.4. What next for tree segmentation algorithms?

Understory trees are not visible in most remote sensing data as they only detect the top of the canopy. Laser scanning penetrates through the canopy and so can potentially be used to segmented understory trees. However, we show that the accuracy of understory tree segmentation is currently very low for typical airborne laser scanning data sets in broadleaf forests. This challenge may be overcome using extremely high point density laser scanning data collected using Unoccupied Aerial Vehicles (Hamraz et al., 2017). The processing power required to apply existing ALS segmentation algorithms to these massive data sets may be prohibitive, so we expect efficient algorithms to be adapted for this specific purpose.

Machine learning methods are proving helpful for distinguishing trees with studies ranging from object detection (Sun Y et al., 2022; Weinstein et al., 2019) to segmentation (Ball et al., 2023) and even species classification (Veras et al., 2022). One key challenge in this area is collecting sufficient reference data to train the machine learning models.

Recent studies have demonstrated the exciting potential of combining the structural information from ALS with spectral information to improve segmentation accuracy for canopy trees (Aubry-Keintz et al., 2021). RGB or hyperspectral data (Shi et al., 2021) can help distinguish tree crowns by their colour or texture, which are often species specific (Williams et al., 2022). This may therefore prove particularly useful in highly diverse tropical forests. We expect that combining techniques in this way will result in more reliable segmentation for the visible canopy trees.

5. Conclusions

This study compared the accuracy of four representative individual tree segmentation (ITS) algorithms against benchmark data sets in temperate and tropical forests. We found that all four ITS algorithms were able to segment canopy trees accurately but performed poorly for the understory trees. AMS3D was the most accurate ITS algorithm, closely followed by Dalponte2016 +. Both algorithms benefited from using allometric information to improve their predictions. However, their accuracy was highly sensitive to these parameters. Crucially, we found that these parameters could be used to tune the algorithms to predict a realistic number and size distribution of trees, but that this reduced segmentation accuracy. This highlights the importance of robustly assessing segmentation results using labelled benchmark data sets, such as the openly available ones generated in this study.

Data and code availability.

- Benchmark data for Sepilok Forest and Wytham Woods:

  https://zenodo.org/record/7181101#.Y6W4XuzP0ea.

- Raw ALS data collected in Sepilok in 2020:

  http://dx.https://doi.org/10.5285/dd4d20c8626f4b9d99bc143 58b1b50fe.

- Raw ALS data collected in Sepilok in 2014:

  https://data.ceda.ac.uk/neodc/arsf/2014/GB13_10.

- Segmented TLS data collected in Wytham in 2015:

  https://doi.org/10.5281/zenodo.7307956.

- Segmented TLS data collected in Sepilok in 2017 will be available on acceptance.
- Labelling ALS with TLS:

  https://github.com/Elephant-C/ALS-labelling.

- Tree crown polygon-based assessment:

  https://github.com/Elephant-C/tree-crown-based-assessment.

- AMS3D:

  https://opus4.kobv.de/opus4-hnee/frontdoor/index/index/docId/ 372.

Author contributions.

TJ, **YC** and **JB** conceived the idea and designed the study. **YC** conducted all the analysis and prepared code used in the study. **LS** and **NK** reproduced the AMS3D software. **KC**, **AB**, TJ, PW and MD collected the TLS data in both Wytham and Sepilok. PW and KC processed the TLS data in Sepilok and Wytham respectively. **TJ** and **DC** organized the ALS data collection in Sepilok. **DC** organized the ALS data collection in Wytham. **TJ** and **YC** led the writing of the manuscript, in close collaboration with **JB** and **YL**. All authors contributed critically to the drafts and gave final approval for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my data and code at the Attached File stpe

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2023.103490.

References

Amiri, N., Polewski, P., Heurich, M., Krzystek, P., Skidmore, A.K., 2018. Adaptive stopping criterion for top-down segmentation of als point clouds in temperate coniferous forests. ISPRS J. Photogramm. Remote Sens. 141, 265–274.
Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., Coomes, D., Piboule, A., Vincent, G., 2019. A comparative assessment of the performance of individual tree crowns delineation algorithms from als data in tropical forests. Remote Sens. (Basel) 11, 1086.
Aubry-Kientz, M., Laybros, A., Weinstein, B., Ball, J.G., Jackson, T., Coomes, D., Vincent, G., 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.
Ayrey, E., Fraver, S., Kershaw Jr, J.A., Kenefic, L.S., Hayes, D., Weiskittel, A.R., Roth, B. E., 2017. Layer stacking: A novel algorithm for individual forest tree segmentation from lidar point clouds. Can. J. Remote. Sens. 43, 16–27.
Ball, J.G., Hickman, S.H., Jackson, T.D., Koay, X.J., Hirst, J., Jay, W., Archer, M., Aubry-Kientz, M., Vincent, G., Coomes, D.A., 2023. Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN. Remote Sensing in Ecology and. Conservation.
Burt, A., Disney, M., Calders, K., 2019. Extracting individual trees from lidar point clouds using treeseg. Methods Ecol. Evol. 10, 438–445.
Butt, N., Campbell, G., Malhi, Y., Morecroft, M., Fenn, K., Thomas, M., 2009. Initial results from establishment of a long-term broadleaf monitoring plot at wytham woods, oxford, uk. University Oxford, Oxford, UK, Rep.
Calders, K., Origo, N., Burt, A., Disney, M., Nightingale, J., Raumonen, P., Akerblom, M., Malhi, Y., Lewis, P., 2018. Realistic forest stand reconstruction from terrestrial lidar for radiative transfer modelling. Remote Sens. (Basel) 10, 933.
Chen, Q., Baldocchi, D., Gong, P., Kelly, M., 2006. Isolating individual trees in a savanna woodland using small footprint lidar data. Photogramm. Eng. Remote Sens. 72, 923–932.
Cheng, Y., 1995. Mean shift, mode seeking, and clustering. IEEE Trans. Pattern Anal. Mach. Intell. 17, 790–799.

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F., Lewis, S. L., Nilus, R., Phillips, O.L., Phua, M.H., et al., 2017. Area-based vs tree-centric approaches to mapping forest carbon in southeast asian forests from airborne laser scanning data. Remote Sens. Environ. 194, 77–88.
Dai, W., Yang, B., Dong, Z., Shaker, A., 2018. A new method for 3d individual tree extraction using multispectral airborne lidar point clouds. ISPRS J. Photogramm. Remote Sens. 144, 400–411.
Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. 7, 1236–1245.
Duncanson, L., Cook, B., Hurtt, G., Dubayah, R., 2014. An efficient, multilayered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. Remote Sens. Environ. 154, 378–386.
Ester, M., Kriegel, H.P., Sander, J., 2001. Algorithms and applications for spatial data mining. Geographic Data Mining and Knowledge Discovery 5, 600.
Ferraz, A., Bretar, F., Jacquemoud, S., Gon calves, G., Pereira, L., Tom′e, M., Soares, P., 2012. 3-d mapping of a multi-layered mediterranean forest using als data. Remote Sens. Environ. 121, 210–223.
Ferraz, A., Saatchi, S., Mallet, C., Meyer, V., 2016. Lidar detection of individual tree size in tropical forests. Remote Sens. Environ. 183, 318–333.
Ferraz, A., Saatchi, S.S., Longo, M., Clark, D.B., 2020. Tropical tree size frequency distributions from airborne lidar. Ecol. Appl. 30, e02154.
Gillies, S., et al., 2007–. Shapely: manipulation and analysis of geometric objects. URL: https://github.com/shapely/shapely.
Hamraz, H., Contreras, M.A., Zhang, J., 2016. A robust approach for tree segmentation in deciduous forests using small-footprint airborne lidar data. International journal of applied earth observation and geoinformation 52, 532–541.
Hamraz, H., Contreras, M.A., Zhang, J., 2017. Forest understory trees can be segmented accurately within sufficiently dense airborne laser scanning point clouds. Sci. Rep. 7, 1–9.
Hui, Z., Cheng, P., Yang, B., Zhou, G., 2022. Multi-level self-adaptive individual tree detection for coniferous forest using airborne LiDAR. International Journal of Applied Earth Observation and Geoinformation. 1 (114), 103028.
Jakubowski, M.K., Li, W., Guo, Q., Kelly, M., 2013. Delineating individual trees from lidar data: A comparison of vector-and raster-based segmentation approaches. Remote Sens. (Basel) 5, 4163–4186.
Jucker, T., Bongalov, B., Burslem, D.F., Nilus, R., Dalponte, M., Lewis, S.L., Phillips, O.L., Qie, L., Coomes, D.A., 2018. Topography shapes the structure, composition and function of tropical forest landscapes. Ecol. Lett. 21, 989–1000.
Jucker, T., Fischer, F.J., Chave, J., Coomes, D.A., Caspersen, J., Ali, A., Panzou, G.J.L., Feldpausch, T.R., Falster, D., Usoltsev, V.A., et al., 2022.
Junttila, S., Nasi, R., Koivumaki, N., Imangholiloo, M., Saarinen, N., Raisio, J., Holopainen, M., Hyyppa, H., Hyyppa, J., Lyytikainen-Saarenmaa, P., et al., 2022. Multispectral imagery provides benefits for mapping spruce tree decline due to bark beetle infestation when acquired late in the season. Remote Sens. (Basel) 14, 909.
Ke, Y., Quackenbush, L.J., 2008. Comparison of individual tree crown detection and delineation methods, in: Proceedings of 2008 ASPRS annual conference.
Ke, Y., Quackenbush, L.J., 2011. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. Int. J. Remote Sens. 32, 4725–4747.
Lee, H., Slatton, K.C., Roth, B.E., Cropper Jr, W., 2010. Adaptive clustering of airborne lidar data to segment individual tree crowns in managed pine forests. Int. J. Remote Sens. 31, 117–139.
Li, W., Guo, Q., Jakubowski, M.K., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud. Photogramm. Eng. Remote Sens. 78, 75–84.
Liang, X., Hyyppa, J., Kaartinen, H., Lehtomaki, M., Pyorala, J., Pfeifer, N., Holopainen, M., Brolly, G., Francesco, P., Hackenberg, J., et al., 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. ISPRS J. Photogramm. Remote Sens. 144, 137–179.
Liu, T., Im, J., Quackenbush, L.J., 2015. A novel transferable individual tree crown delineation model based on fishing net dragging and boundary classification. ISPRS J. Photogramm. Remote Sens. 110, 34–47.
Ma, K., Chen, Z., Fu, L., Tian, W., Jiang, F., Yi, J., Du, Z., Sun, H., 2022. Performance and sensitivity of individual tree segmentation methods for uav-lidar in multiple forest types. Remote Sens. (Basel) 14, 298.
Minarik, R., Langhammer, J., Lendzioch, T., 2020. Automatic tree crown extraction from uas multispectral imagery for the detection of bark beetle disturbance in mixed forests. Remote Sens. (Basel) 12, 4081.
Mongus, D., Zalik, B., 2015. An efficient approach to 3d single tree-crown delineation in lidar data. ISPRS J. Photogramm. Remote Sens. 108, 219–233.
Morsdorf, F., Meier, E., Allg¨owe, B., N¨uesch, D., 2003. Clustering in airborne laser scanning raw data for segmentation of single trees. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 34, W13.
Qin, H., Zhou, W., Yao, Y., Wang, W., 2022. Individual tree segmentation and tree species classification in subtropical broadleaf forests using uav-based lidar, hyperspectral, and ultrahigh-resolution rgb data. Remote Sens. Environ. 280, 113143.
Roussel, J.R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R., Meador, A.S., Bourdon, J.F., De Boissieu, F., Achim, A., 2020. lidr: An r package for analysis of airborne laser scanning (als) data. Remote Sens. Environ. 251, 112061.
Shi, Y., Wang, T., Skidmore, A.K., Holzwarth, S., Heiden, U., Heurich, M., 2021. Mapping individual silver fir trees using hyperspectral and LiDAR data in a Central European mixed forest. International Journal of Applied Earth Observation and Geoinformation. 1 (98), 102311.
Silva, C.A., Hudak, A.T., Vierling, L.A., Loudermilk, E.L., O'Brien, J.J., Hiers, J.K., Jack, S.B., Gonzalez-Benecke, C., Lee, H., Falkowski, M.J., et al., 2016. Imputation of

individual longleaf pine (pinus palustris mill.) tree attributes from field and lidar data. Can. J. Remote. Sens. 42, 554–573.

Steinmeier, L., 2022. Finding trees in lidar point clouds: towards a fast and scalable segmentation workflow. masterthesis. Hochschule f˙ur nachhaltige Entwicklung Eberswalde.

Strimbu, V.F., Strimbu, B.M., 2015. A graph-based segmentation algorithm for tree crown extraction using airborne lidar data. ISPRS J. Photogramm. Remote Sens. 104, 30–43.

Sun, Y., Li, Z., He, H., Guo, L., Zhang, X., Xin, Q., 2022. Counting trees in a subtropical mega city using the instance segmentation method. International Journal of Applied Earth Observation and Geoinformation. 1 (106), 102662.

Vega, C., Hamrouni, A., El Mokhtari, S., Morel, J., Bock, J., Renaud, J.P., Bouvier, M., Durrieu, S., 2014. Ptrees: A point-based approach to forest tree extraction from lidar data. International Journal of Applied Earth Observation and Geoinformation 33, 98–108.

Veras, H.F.P., Ferreira, M.P., da Cunha Neto, E.M., Figueiredo, E.O., Dalla Corte, A.P., Sanquetta, C.R., 2022. Fusing multi-season uas images with convolutional neural networks to map tree species in Amazonian forests. Eco. Inform. 101815.

Wang, D., 2020. Unsupervised semantic and instance segmentation of forest point clouds. ISPRS J. Photogramm. Remote Sens. 165, 86–97.

Wang, Y., Hyyppa, J., Liang, X., Kaartinen, H., Yu, X., Lindberg, E., Holmgren, J., Qin, Y., Mallet, C., Ferraz, A., et al., 2016. International benchmarking of the individual tree detection methods for modeling 3-d canopy structure for silviculture and forest ecology using airborne laser scanning. IEEE Trans. Geosci. Remote Sens. 54, 5011–5027.

Weinacker, H., Koch, B., Heyder, U., Weinacker, R., 2004. Development of filtering, segmentation and modelling modules for lidar and multispectral data as a fundament of an automatic forest inventory system. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36, W2.

Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. Remote Sens. (Basel) 11, 1309.

Weiser, H., Schafer, J., Winiwarter, L., Kräsovec, N., Fassnacht, F.E., Hofle, B., 2022. Individual tree point clouds and tree measurements from multi-platform laser scanning in german forests. URL: Earth Syst. Sci. Data 14, 2989–3012 https://essd.copernicus.org/articles/14/2989/2022/, https://doi.org/10.5194/essd-14-2989-2022.

Wilkes, P., Disney, M.I., Armston, J., Bartholomeus, H., Bentley, L.P., Brede, B., Burt, A., Calders, K., Chavana-Bryant,C., Clewley, D., Duncanson, L., Forbes, B., Krisanski, S., Malhi, Y., Moffat, D., Origo, N., Shenkin, A., Yang, W., 2022.Tls2trees: a scalable tree segmentation pipeline for tls data.

Wilkes, P., Lau, A., Disney, M., Calders, K., Burt, A., de Tanago, J.G., Bartholomeus, H., Brede, B., Herold, M., 2017. Data acquisition considerations for terrestrial laser scanning of forest plots. Remote Sens. Environ. 196, 140–153.

Williams, J., Sch¨onlieb, C.B., Swinfield, T., Lee, J., Cai, X., Qie, L., Coomes, D.A., 2019. 3d segmentation of trees through a flexible multiclass graph cut algorithm. IEEE Trans. Geosci. Remote Sens. 58, 754–776.

Williams, J., Jackson, T.D., Sch¨onlieb, C.B., Swinfield, T., Irawan, B., Achmad, E., Zudhi, M., Habibi, H., Gemita, E., Coomes, D.A., 2022. Monitoring early-successional trees for tropical forest restoration using low-cost uav-based species classification. Frontiers in Forests and Global Change 5. URL: https://www.frontiersin.org/articles/10.3389/ffgc.2022.876448,. https://doi.org/10.3389/ffgc.2022.876448.

Winiwarter, L., Pena, A.M.E., Weiser, H., Anders, K., Sanchez, J.M., Searle, M., Hofle, B., 2022. Virtual laser scanning with helios++: A novel take on ray tracing-based simulation of topographic full-waveform 3d laser scanning. Remote Sens. Environ. 269, 112772.

Xiao, W., Zaforemska, A., Smigaj, M., Wang, Y., Gaulton, R., 2019. Meanshift segmentation assessment for individual forest tree delineation from airborne lidar data. Remote Sens. (Basel) 11, 1263.

Xin, Honglu, et al., 2021. Individual tree detection and crown segmentation based on metabolic theory from airborne laser scanning data. Journal of Applied Remote Sensing 15 (3). https://doi.org/10.1117/1.JRS.15.034504.

Xu, S., Ye, N., Xu, S., Zhu, F., 2018. A supervoxel approach to the segmentation of individual trees from lidar point clouds. Remote Sensing Letters 9, 515–523.