RESEARCH ARTICLE

Treegraph: tree architecture from terrestrial laser scanning point clouds

Wanxin Yang1,2, Phil Wilkes1,2,3, Matheus B. Vicari1, Kate Hand4, Kim Calders5 & Mathias Disney1,2

1Department of Geography, University College London, London WC1E 6BT, UK
2NERC National Centre for Earth Observation (NCEO), UK
3Royal Botanic Gardens, Kew, Wakehurst, Ardingly, West Sussex, UK
4Faculty of Science, Technology, Engineering & Mathematics, The Open University, London, UK
5CAVELab – Computational & Applied Vegetation Ecology, Department of Environment, Ghent University, Ghent, Belgium

Keywords
3D tree reconstruction, metabolic scaling theory, quantitative structure model (QSM), terrestrial laser scanning (TLS), tree architecture, tree topology

Abstract
Accurate quantification of tree architecture is critical to interpreting the growth, health and functioning of trees and forests. Terrestrial laser scanning (TLS) offers millimetre-level point cloud data, but current approaches to 3D tree reconstruction from TLS point clouds primarily focus on retrieving total volume at tree scale for aboveground biomass (AGB) estimation. Few methods have been designed specifically to provide tree architectural properties, including branch-level morphology and topology, rather than AGB; derived topological traits have tended to be a compromise, and of secondary importance to volume. We present Treegraph, a new approach explicitly designed to retrieve the architectural traits of trees at multiple scales, from the whole tree scale down to individual branches and internodes, using TLS data with limited assumptions about tree form. It provides morphological traits such as branch length and diameter, alongside topological traits including parent–daughter connections of branches and internodes, furcation (branching) number and branch order. We compare Treegraph-derived morphological and topological traits with manual measurements of branches from eight destructively harvested trees, yielding RMSE values of 0.60 m (5.96%) for branch length, 2.99 cm (33.45%) for branch diameter, 0.46 (19.38%) for furcation number and 0.08 m (33.16%) for internode length, respectively. In a broader application to 603 trees from tropical, temperate and urban forests, we demonstrate that the derived morphological and topological traits support testing of structure-related metabolic scaling theories. Testing branches over 10 cm in diameter across 18 657 branching nodes shows that Treegraph-derived branch-level scaling exponents deviate from WBE predictions, exhibiting area-preserving behaviour while displaying asymmetry in length and diameter of daughter branches. Available as open-source Python software, Treegraph provides fine-level branching network information, promoting improved insights into tree structure and function. This data-driven approach reduces the need for empirical heuristic parameters, which has the potential for advancing large-scale ecological studies on tree architecture.

Introduction
Tree architecture, encompassing both tree morphology and topology, plays a critical role in tree growth, development, reproduction and adaptability to environmental conditions (Barnes et al., 1997; Eloy et al., 2017; Hallé et al., 1978; Osada & Takeda, 2003). Here, we define morphology to be the shapes, sizes and physical
attributes of the trunk and branches, and topology to be the connectivity and hierarchical relationships between the trunk and branches. Laurans et al. (2024) highlight the significance of incorporating plant architecture into trait-based ecology, revealing that multilevel morphological, topological and ontogenetic features can profoundly influence growth and ecological strategies. Morphological diversity closely aligns with functional diversity, with architectural traits providing insights into the interplay of structure with hydraulics (Smith et al., 2014), light capture (MacFarlane et al., 2017), and biomechanical stability (Jackson et al., 2021). Accurately capturing tree morphology and topology enables a detailed exploration of these structural–functional relationships, offering insights into the ecological functions and adaptive mechanisms of trees within forest ecosystems.

Terrestrial laser scanning (TLS) is a ground-based remote sensing technique, capable of mapping 3D surfaces with millimetre accuracy using laser pulses. TLS offers a non-destructive and highly accurate way to characterize tree architecture (Calders et al., 2020; Disney, 2018; Malhi et al., 2018; O’Sullivan et al., 2021; van Leeuwen & Nieuwenhuis, 2010). While TLS holds promise for identifying and quantifying architectural traits of trees, there remain challenges in providing accurate metrics on the morphological and topological properties. In plant architecture studies focusing on similarities in the branching patterns, network models have been constructed to predict scaling of plant form and function within and across species. However, such models often ignore variation in branching traits and rely on simple size-based scaling relationships (Vasseur et al., 2012; West et al., 1997; West et al., 1999). To test the assumptions and predictions of metabolic scaling theory such as West, Brown and Enquist (WBE) model (West et al., 1997, 1999), there is a need for accurate data on branch radii, lengths and the connectivity relationships between daughter and parent branches at node level throughout the tree. Despite the potential of TLS to provide this information, the accuracy and precision in retrieving these parameters remains constrained by current TLS methods.

A wide range of methods have been developed to derive individual tree architectural traits from TLS-derived 3D point cloud data, particularly via so-called quantitative structure models (QSMs). Crucially, many of these methods have been developed to estimate tree volume, rather than topology. TreeQSM (Raumonen et al., 2013) fits patches to segments of the wood points of a tree and then fits cylinders to generate a QSM enabling calculation of branch volume and length. AdQSM (Du et al., 2019; Fan, Nan, Dong, Su, & Chen, 2020; Fan et al., 2020) extracts the skeleton of a tree using a minimum spanning tree algorithm. SimpleTree (Hackenberg et al., 2014, 2015, 2015) uses cylinders to represent trunk and branch sections, fitted recursively from the tree base towards the branch tips. However, these methods are primarily designed to retrieve total tree volume for estimation of aboveground biomass (AGB) (Brede et al., 2022; Burt et al., 2021; Calders et al., 2015; Fan, Nan, Dong, et al., 2020; Gonzalez de Tanago et al., 2017; Hackenberg, Spiecker, et al., 2015). The assumptions made to optimize total volume estimation can impact and even constrain the other architectural properties. Lau et al. (2018) found that TreeQSM (version 2.2.1) underestimates the volume of larger branches (diameter > 20 cm) while greatly overestimating smaller ones. Although these biases may offset each other to provide a total volume approximating the true value, errors in estimating diameter and length at branch and inter-node levels will potentially be large, and the connections between daughter and parent branches may be different from the actual tree. Similarly, AdQSM (Du et al., 2019) uses an allometric ‘rule’ to estimate the branch radius as proportional to the distance from the tree base, but this is arbitrarily fixed. For applications requiring a detailed analysis of branch size and connections, such as testing metabolic scaling theory, reliance on these assumptions could lead to significant inaccuracies.

Achieving the optimal balance between automation and user intervention is another challenge for structure fitting methods. Liu et al. (2021) proposed a deep-learning segmentation method in TreePartNet, but this method heavily relies on labelled synthetic data, and it struggles with larger trees and diverse tree species. There is still a real need for a near automatic approach that works across a range of tree sizes and shapes, while minimizing the need for user-defined structural heuristics.

Lastly, another challenge for deriving detailed tree architecture from TLS data is validation: measuring AGB for whole trees is hard but feasible (Burt et al., 2021; Demol et al., 2022); measuring fine-scale architectural traits is almost impossible for anything other than small trees.

To overcome these challenges, we present Treegraph, an open-source Python software specifically designed to derive tree architecture, rather than volume. Treegraph automatically extracts both morphological and topological traits from individual tree point clouds, with the following benefits:

1. Detailed architectural traits retrieval across scales, throughout an entire tree, at branch level and inter-node level, thereby enabling fine-level branching network analysis.
2. Data-driven with minimal assumptions about tree structure, meaning that the retrieved architectural information is suitable for exploring ecological impacts of tree structure.
3. Automated parameter tuning to adjust to input point cloud characteristics.

An early version of Treegraph was used by Wilkes et al. (2021) to model harvested branch architecture, but the branches were not scanned in situ but under controlled conditions after removal to an indoor scanning location. This paper describes the Treegraph method in full as well as applying it to entire tree data scanned in situ.

Materials and Methods

Data

Terrestrial laser scanning (TLS) data

The TLS data used here were collected from three sites across different forest types, capturing a range of tree species, sizes and shapes (Table 1). Five hundred and fifty-four temperate trees comprising seven species, in Wytham Woods, United Kingdom (Calders et al., 2022b); four tropical forest trees comprising four species, in Caxianã, Brazil (Burt et al., 2021); and forty-five urban street trees comprising six species, in Milton Keynes, United Kingdom. All TLS data were collected using a RIEGL VZ-400 instrument featuring a wavelength of 1550 nm, 0.35 mrad beam divergence and 0.04° angular resolution. In Wytham and Milton Keynes, TLS data were collected using a grid sampling, whereas in Caxianã scans were collected radially around each tree. Individual scans were co-registered using RiSCAN PRO software (RIEGL Laser Measurement Systems GmbH, 2016). Individual trees were manually extracted from each site. The point clouds for Milton Keynes are shown in Figure 1 as they have not previously been published, unlike those for Wytham and Caxianã.

Harvest-derived measurements

Reference data for validating our model includes census data of tree species, height, and DBH for TLS-sampled trees across these three sites (Table 1), and AGB measurements from four harvested tropical trees (Burt et al., 2021).

To validate branch morphology and topology, we manually measured 106 branches from eight harvested trees, spanning four tree species, in Milton Keynes, United Kingdom (Table S0). Using a measuring tape and vernier callipers, we traced from the tree base to selected branch tips. Diameters along the stem were recorded at 50 cm intervals. For each branch, we measured diameters below and above each furcation node (branching point), as well as the length between furcation nodes. We recorded parent–daughter internode connections to enable the reconstruction of the branch topology. Across these eight

Table 1. Summary of the TLS plots and census data of the trees.

<table>
<thead>
<tr>
<th>Site</th>
<th>Location (forest type)</th>
<th>Scan pattern (citation)</th>
<th>Species</th>
<th>Common name</th>
<th>No. of trees</th>
<th>Avg. tree height (m)</th>
<th>Avg. DBH (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW Wytham Woods, UK (Temperate mixed deciduous)</td>
<td>20 m grid (Calders et al., 2018, Calders et al., 2022a, 2022b)</td>
<td>Acer campestre</td>
<td>Field maple</td>
<td>2</td>
<td>7.46 ± 1.10</td>
<td>7.60 ± 0.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acer pseudoplatanus</td>
<td>Sycamore</td>
<td>360</td>
<td>16.59 ± 6.06</td>
<td>24.27 ± 15.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Corylus avellana</td>
<td>Common hazel</td>
<td>34</td>
<td>8.53 ± 2.40</td>
<td>9.16 ± 2.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Crateriphytum moluccanum</td>
<td>Hawthorne</td>
<td>17</td>
<td>8.06 ± 2.53</td>
<td>9.97 ± 4.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fraxinus excelsior</td>
<td>European ash</td>
<td>50</td>
<td>12.51 ± 7.42</td>
<td>22.56 ± 17.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quercus robur</td>
<td>English oak</td>
<td>24</td>
<td>20.19 ± 3.35</td>
<td>60.92 ± 14.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unknown</td>
<td>–</td>
<td>67</td>
<td>7.63 ± 3.12</td>
<td>9.03 ± 5.57</td>
<td></td>
</tr>
<tr>
<td>MK Milton Keynes, UK (Urban street trees)</td>
<td>10 m grid (unpublished)</td>
<td>Populus tremula</td>
<td>Aspen</td>
<td>6</td>
<td>13.00 ± 2.02</td>
<td>26.82 ± 2.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acer campestre</td>
<td>Field maple</td>
<td>4</td>
<td>6.90 ± 0.68</td>
<td>15.20 ± 1.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Platanus x acerifolia</td>
<td>London plane</td>
<td>3</td>
<td>12.77 ± 0.80</td>
<td>24.47 ± 3.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quercus rubra</td>
<td>Red oak</td>
<td>7</td>
<td>8.50 ± 1.09</td>
<td>18.64 ± 3.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fraxinus angustifolia</td>
<td>Raywood ash</td>
<td>23</td>
<td>9.70 ± 1.07</td>
<td>17.42 ± 3.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sorbus aria</td>
<td>Whitebeam</td>
<td>2</td>
<td>8.60 ± 0.00</td>
<td>30.35 ± 2.47</td>
<td></td>
</tr>
<tr>
<td>CAX Caxianã, Brazil (Tropical rainforest)</td>
<td>Circle around the tree (Burt et al., 2020b, Burt et al., 2021)</td>
<td>Hymenaea courbarill L.</td>
<td>Brazilian Cherry</td>
<td>1</td>
<td>29.80</td>
<td>64.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inga alba (Sw.) Willd.</td>
<td>White Inga</td>
<td>1</td>
<td>46.20</td>
<td>117.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tachigali paniculata var. alba (Ducke)</td>
<td>Dwyer</td>
<td>-</td>
<td>34.90</td>
<td>90.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trattinnickia burserifolia</td>
<td>Mart.</td>
<td>-</td>
<td>35.20</td>
<td>69.70</td>
<td></td>
</tr>
</tbody>
</table>
harvested trees, we obtained 3557 diameter measurements in total, covering an aggregate path length of 315 m.

**Treegraph: algorithm description**

The Treegraph method consists of six main steps, each detailed in the subsequent subsections.

1. Building a distance graph from the point cloud to determine each node’s shortest path (Fig. 2B).
2. Identifying skeleton nodes that outline the primary branch structure (Fig. 2C).
3. Generating a skeleton graph that characterizes the topology of the tree’s branching network (Fig. 2D).
4. Segmenting individual branches based on the reconstructed topology attributes (Fig. 2E).
5. Estimating radii for individual branches from the point cloud (Fig. 2E).
6. Generating a QSM based on the extracted skeleton and the radius estimates (Fig. 2F).

**Build distance graph**

In the first step, we convert the input leaf-off point cloud into a weighted bidirectional network graph based on Vicari et al. (2018). To reduce the computing time, we apply a voxel grid downsampling (default voxel length of 0.04 m, but customisable). In this graph, each point becomes a node and connections between neighbouring points are represented as edges with weights determined by the Euclidean distances between these points. Then, a shortest path analysis using the Dijkstra algorithm (Dijkstra, 1959) is performed from all nodes to the base node, obtaining the attribute of ‘distance from base’ for each node.

To refine the location of the graph’s base node for shortest path analysis, we first extract the tree’s lower stem section using the normal vector method from Zhang et al. (2022), and then fit a cylinder to this section. The base node is repositioned at the cylinder’s bottom circle centre, with the cylinder’s radius indicating the tree’s basal...
Figure 2. Illustration of Treegraph method. (A) Input point cloud of an individual tree. (B) Build distance graph. (C) Identify skeleton nodes. (D) Generate skeleton graph. (E) Segment individual branches and estimate branch radii. (F) Generate a QSM model.
radius. Finally, the updated base node is integrated into the distance graph, connected to the remaining basal nodes (Figure S1.2). Shortest path information is subsequently derived from this updated graph (see Appendix S1 for comparison before and after refinement).

Identify skeleton nodes

Point clouds are binned based on the ‘distance from base’ attribute. We reduce bin widths towards branch tips using an exponential function mapped onto the ‘distance from base’ attribute. The bin width of the \( k \)th bin out of \( n \) bins is computed by Eq. 1.

\[
b_k = (b_{\text{max}} - b_{\text{min}}) \times \frac{e^{-k-1}}{e-1} + b_{\text{min}}
\]  

(1)

where \( b_{\text{max}} \) and \( b_{\text{min}} \) indicate the maximum and minimum bin width to slice the point clouds. The selection of these two parameters and the sensitivity analysis is demonstrated in Appendix S2. The total bin count \( n \) starts from 50 and iteratively updates until the criteria in Eq. 2 is met.

\[
d_f b_{\text{max}} < \sum_{k=0}^{n} b_k < 1.05 \times df b_{\text{max}}
\]

(2)

where \( df b \) is the shortest path distance from the base to individual nodes.

Within each bin, we extract initial skeleton nodes using DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Ester et al., 1996). Given the variable point density across the tree, we determine an adaptive search radius by calculating the \( k \)th nearest neighbour distance for each point and identifying the ‘knee’ point, where the rate of distance growth peaks. Next, we apply a cylindrical prior constraint (CPC) optimization method (Fu et al., 2020) to improve the centredness of cluster centroid.

At a furcation node (where a parent branch separates into two or more daughter branches), daughter branches often emerge in close proximity, making it difficult for DBSCAN to distinguish sub-branches as separate clusters. Additionally, intertwined, overlapping or closely spaced branches pose additional challenges for DBSCAN in accurate separation of individual branches.

To further separate individual branches and refine the skeleton nodes, we apply a self-tuning spectral clustering approach (Zelnik-Manor & Perona, 2004). This method uses eigenvectors of a point-to-point similarity matrix (Eq. 3, ibid.) to transform the data into a lower dimensional space, capturing the underlying manifold structure of the data. The optimal number of clusters is automatically determined by the maximum gap between eigenvalues derived from the normalized similarity matrix (Eq. 5, ibid.). This process is iteratively applied to each cluster obtained from DBSCAN until the current cluster cannot be further segmented (Fig. 3).

\[
\tilde{A}_{ij} = \exp \left( -\frac{d^2(s_i, s_j)}{\sigma_i \sigma_j} \right)
\]

(3)

where \( d \) is the Euclidean distance between two points \( s_i, s_j \). The local scale \( \sigma_i \) for a given point \( s_i \) corresponds to the Euclidean distance to its \( k \)th nearest neighbour. Here, we determine the value of \( k \) based on the total number of points within a cluster.

\[
D_n = \sum_{j=1}^{n} \tilde{A}_{ij}
\]

(4)

\[
L = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}
\]

(5)

where \( D \) is a diagonal matrix used to construct the normalized similarity matrix \( L \).

Generate skeleton graph

Defining connections between skeleton nodes is essential for outlining the tree topology. We incorporate proximity information of cluster edges in generating a skeleton graph. First, we calculate the convex hull vertices for each cluster identified in the previous step. Then, a k-nearest neighbours search is performed to identify neighbouring vertices from adjacent clusters. Edge weight between any two neighbouring clusters is computed using Eq. 6, prioritizing clusters closer in Euclidean distance with more neighbouring hull vertices.

\[
W_{c_i,c_j} = \frac{\min(\text{dist} (V_{c_i}, V_{c_j}))}{\log(n)}
\]

(6)

where \( V_{c_i} \) denotes the convex hull vertices of cluster \( c_i \), and \( V_{c_j} \) refers to the vertices of cluster \( c_j \) among \( V_{c_i} \)’s k-nearest neighbours. \( n \) is the count of neighbour pairs between \( V_{c_i} \) and \( V_{c_j} \).

Next, we construct an undirected weighted graph from the edge list, ensuring that all clusters are connected. Dijkstra’s algorithm (Dijkstra, 1959) is then used to extract the shortest path and distance from each cluster to the base node. Finally, we generate a skeleton graph from the shortest path information, capturing the topology of individual branches extending from the tree base to branch tips.

Segment individual branches

We first identify the branch tip nodes that appear only once across all paths, representing the endpoints of individual branches. For each tip node, we calculate the shortest path to the tree base and sort these paths by length. We begin branch segmentation with the longest path, designating the tree base.
node as the branch base. For each subsequent path, we define the branch base at the furcation node, where it intersects with any previously segmented branch path, ensuring all branches are distinct and non-overlapping, and each node is uniquely assigned to a specific branch.

Branch lengths are calculated as the path distance from the branch base to the tip. Branch IDs are then indexed in descending order based on length. The longest branch, characterized by the maximum tree pathlength, is identified as ‘branch 0’ (Fig. 4).

Iteratively, we establish the parent–daughter relationships for each segmented branch and internode (a point where a branch splits or terminates as a tip). This enables us to quantify the furcation number at every internode and the branch order for each branch segment. The trunk is assigned the lowest order, and the order increases as branches extend towards the tips.

**Estimate branch radii**

Here, we avoid *a priori* constraints on branch size relationships. We estimate branch radius by ‘straightening’ the branch and calculating the mean point-to-axis distance of each segment. For each branch, we apply spline interpolation on the X, Y, Z coordinates as a function of distance from base, and straighten the branch by subtracting the interpolated coordinates from the original ones (Eq. 7). The straightened branch is segmented at 30 cm intervals, and mean point-to-axis distance is computed for each segmented section. We calculate the coefficient of variation (CV) via Eq. 8 to quantify distance variation within each section. The branch radius for each section is then estimated using Eq. 9. This approach results in a more conservative branch radius estimation when there is high variability in the point-to-axis distances, mitigating the impact of noise or local variations in the data.

\[
\begin{align*}
    tx_i &= x_i - f_x(d_i) \\
    ty_i &= y_i - f_y(d_i) \\
    tz_i &= z_i - f_z(d_i)
\end{align*}
\]  

where \((x_i, y_i, z_i)\) and \((tx_i, ty_i, tz_i)\) are the original and transformed coordinates of a point in a branch point cloud, respectively, while \(d_i\) denotes its distance from base.

\[
\begin{align*}
    CV &= \frac{\sigma}{\mu} \\
    r_i &= \mu - CV \times \sigma
\end{align*}
\]  

where \(\mu\) and \(\sigma\) are the mean and standard deviation of point-to-axis distance in each section.
Modify the radius estimates and generate cylinder models

Segmentation of individual branches can still incorporate unexpected components like points from other branches, leaves or lianas, leading to erroneous fluctuations in radius estimates. To mitigate this, we apply a radius correction that limits the initial radius of a daughter branch to be no more than twice the radius of its parent branch. This correction is grounded in empirical observations from our manual measurements, where we noted instances of daughter branches having larger radii than their parent branches, but not exceeding twice the parent’s radius.

For each branch, the path from the stem furcation node to the branch tip is divided into equal distance sections. For each section, we calculate upper bound candidate points based on the mean radius, providing an upper limit to constrain a reduction in branch radius towards the branch tip (Figure S3.1). This modification reduces the impact of noise in overestimating branch radius (Figure S3.2).

Finally, we generate a cylinder between successive skeleton nodes, with the vector between two skeleton nodes serving as the axis, and the modified radius as the cylinder radius, from which we can calculate the volume and surface area.

Testing Treegraph

To quantitatively assess the Treegraph algorithm, we compare the reconstructed model with the original point clouds by computing the mean Euclidean distance of the points to the model surface (Eq. 10). We use the standard deviation of this distance to express variability in the residuals of the model (Eq. 11).

\[
\text{Acc}_{\text{model}} = \frac{\sum_{i=1}^{n} |d_i - R_i|}{n} \quad (10)
\]

\[
\text{SD}_{\text{model}} = \sqrt{\frac{\sum_{i=1}^{n} (|d_i - R_i| - \text{Acc}_{\text{model}})^2}{n-1}} \quad (11)
\]

where \(n\) denotes the total number of raw points, \(d_i\) is the distance between the \(i^{th}\) point and its corresponding axis of the modelled cylinder and \(R_i\) is the radius of that cylinder.

We compute the error between the model estimates and the harvest-derived measurements introduced in Figure 4.
Section 2.1.2. We use the standard deviation of the error to express precision and the root mean square error to quantify accuracy.

\[
e = \text{est} - \text{ref} \quad (12)
\]

\[
\text{Rel} = \frac{|e|}{\text{ref}} \quad (13)
\]

\[
\text{ME} = \frac{1}{n} \sum_{i=1}^{n} e_i \quad (14)
\]

\[
\text{SD} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (e_i - \bar{e})^2} \quad (15)
\]

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2} \quad (16)
\]

It is generally not possible to identify individual manually measured branches in the point cloud following tree felling due to the resulting crown damage. As a result, we consider the path from the tree base to the farthest branch tip as the tree’s longest branch path. The total length of this path is used as a reference for validating Treegraph-derived branch length estimation of the longest detected branch path. To validate branch diameter, furcation number and internode length, we divide the distance from base into intervals, and compare manual measurements of the mean values of all branch sections at each interval with Treegraph estimates.

**Testing metabolic scaling theory**

As Treegraph models allow us to extract internode-level structural information, we explore the scaling of tree architecture for all 603 trees. For each daughter-to-parent pairing, branch (node) level scaling exponents of radius (anode), length (bnode) and estimated metabolic rate (θnode) are computed using equations listed in Table 1 in Bentley et al. (2013).

To assess whether trees exhibit area-preserving branching, we calculate the branch area ratio for each furcation node, which measures the ratio of the total cross-sectional area of daughter branches to their parent branch. For evaluating self-similarity in branching patterns, we compute the branch length ratio of each daughter-to-parent pair.

**Results and Discussion**

**Assessment of model reconstruction**

For all 603 trees across the three sites, we found a mean point-to-model distance of 0.027 m, with a corresponding standard deviation per tree of 0.035 m. The results listed in Table 2 demonstrated that the mean accuracy of the model remains at the centimeter level across all trees.

In Figure 5, tree height does not significantly impact the modelling accuracy, though lower point density is usually observed at the upper crown of taller trees. For example, the four tallest trees here are the tropical trees at CAX site, ranging in height from 29.8 to 46.2 m (Figure S4.1). Notably, these tropical trees display higher modelling accuracy than the Quercus robur trees at WW site, even though the latter have a lower mean height of 20.19 ± 3.35 m (Table 2). Higher variation of modelling accuracy appears in trees with larger size, characterized by > 10^6 points and/or volume > 4 m^3. Moreover, the complexity of tree architecture also influences the modelling accuracy. For example, strong epicormic presence (twiggy growth from the trunk or branches, Figure S4.2) can affect the clustering results when generating a skeleton graph. This can potentially lead to erroneous branch segmentation and radius estimates, subsequently reducing the modelling accuracy.

Figure 6 shows an example of point clouds of branches ≤ 0.3 m in diameter at the upper crown, for an English Oak from Wytham site, as well as the resulting cylinder models generated by TreeQSM v2.0 and Treegraph.

**Tree-level attributes QSM**

**Tree height and DBH estimation**

Figure 7 compares Treegraph-derived estimates and manual measurements of tree height and DBH. The RMSE is...
0.17 m for tree height and 5.14 cm for DBH. For DBH, a bias of −1.82 cm is noted, suggesting that Treegraph tends to slightly underestimate DBH here, partly due to the non-circular cross sections of lower trunks.

**Aboveground biomass estimation**

The total AGB estimated from Treegraph compared with the harvest-derived results, with RMSE values of 226.9 kg (3.1%) (Table S5.1). Furthermore, Treegraph demonstrated high precision in estimating AGB distribution between the stem and crown, with a cumulative relative error of 2.2% for the stem and 7.5% for the crown (Table S5.2). Treegraph can also provide volume distribution across different branch diameter and branch order categories. Figure S5.1 indicates that small branches (with diameter under 10 cm) predominate the upper canopy and collectively contribute substantially to the tree’s volume.

**Branch-level attributes retrieval**

**Branch length estimation**

Treegraph estimates the longest branch with an RMSE of 0.60 m (5.96%) compared to manual measurements (Table 3). On average, the model overestimates the longest branch by 0.46 m, with a variation of 0.39 m.

**Branch diameter estimation**

In terms of branch diameter, Figure 8 shows that Treegraph estimations reveal no statistically significant deviation from the manual measurements based on the 95% confidence intervals. Originating at the tree base, a general trend of decreasing diameter is observed, as would be expected. Notably, a slight increase is observed at the main branching points, presumably to support the ramification into multiple daughter branches. Following these branching points, the diameter rapidly decreases and then follows a more gradual decline.

Treegraph estimates of mean branch diameters give an overall RMSE of 2.99 cm (33.45%) and a bias of −1.08 cm (−12.11%) across 3557 diameter measurements. The point at which diameter overestimation starts to occur is a strong indication of the limits of the TLS data. There is currently no agreed method for quantifying TLS point cloud quality for tree structural applications. However, this quality will fundamentally constrain the reliability of resulting structural attributes—particularly the smallest branches for which we can ‘trust’ derived information. Given that generally we will not know when branch diameter overestimation occurs, the rapid growth in Treegraph diameter uncertainty, indicated by the CI starting to grow again after reaching a minimum, is a potentially useful indicator of this reliability: Beyond this point, we should probably avoid using branch size information if possible.

The uncertainty of branch radius estimation comes not only from poorer point clouds but also the deviation of the real branch shape and the idealized cylinder used to represent branch segments. Most studies use a cylinder model because it has been demonstrated to be the most robust shape in terms of volume (Åkerblom et al., 2015) (also shown in Fig. 9D). However, using cylinders (or any simple geometric shape for that matter) to represent all tree components will not accurately capture the true geometry of trunks and branches when they deviate from a cylindrical shape. For example, the cross section of a
buttress (Fig. 9A) and a branching node (Fig. 9C) is not circular.

The accuracy of branch diameter estimation can be further impacted by co-registration errors, wind effects, occlusion and foliage points. As a result, uncertainty of branch diameter estimated from TLS-QSM method increases gradually with distance from the tree base, but the onset of the crown and the advent of smaller branches increase the variation in the uncertainty (Figure S3.3). Several other studies have also shown that the uncertainty of branch volume increases with decreasing branch diameter (Abegg et al., 2023; Demol et al., 2022; Lau et al., 2018), and the inflation of small branch diameter estimates can lead to an overestimation in crown volume. While this may be mitigated by aiming to scan in windless conditions as far as possible, occlusion is harder to
deal with. It is a function not just of the sampling regime (Wilkes et al., 2017), but also of the canopy itself.

Internode-level attributes retrieval

Figure 10 shows that the majority of branching nodes in the harvested London Plane (Ave_B, Ave_C), English Oak (Ove_O1, Ove_O2, Ove_O3) and Whitebeam (Ove_W2) trees have two furcations, that is, a parent branch divides into two daughter branches; the Raywood Ash (Dan_R5, Dan_R9) predominantly feature branching nodes with three or more furcations. Treegraph tends to overestimate the number of furcations by 11.18%, particularly in the upper crown where point cloud quality degrades, causing errors in sub-branch segmentation through clustering methods. As a consequence of identifying more skeleton nodes as furcation nodes, the internode length—measured as the length of branch section between two furcations—is underestimated with a bias of −0.05 m.

Scaling exponents estimated from Treegraph models

The majority of radius scaling exponents for parent branches with diameters larger than 15 cm exceed the WBE (West et al., 1999) predicted value of 0.5, while the smaller branches exhibit a lower radius scaling rate (Fig. 11). Across diameter classes, the scaling of branch

Table 3. Comparison of the branch length of the longest path (from the tree base to the farthest branch tip) between manual measurements and Treegraph model estimates.

<table>
<thead>
<tr>
<th>TreeID</th>
<th>Species</th>
<th>Ref. (m)</th>
<th>Model (m)</th>
<th>Error (m)</th>
<th>Rel. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ove_W2</td>
<td>Sorbus aria</td>
<td>9.69</td>
<td>9.96</td>
<td>0.27</td>
<td>2.79%</td>
</tr>
<tr>
<td>Ove_O1</td>
<td>Quercus rubra</td>
<td>9.72</td>
<td>9.54</td>
<td>−0.18</td>
<td>1.85%</td>
</tr>
<tr>
<td>Ove_O2</td>
<td>Quercus rubra</td>
<td>8.91</td>
<td>8.92</td>
<td>0.01</td>
<td>0.11%</td>
</tr>
<tr>
<td>Ove_O3</td>
<td>Quercus rubra</td>
<td>10.94</td>
<td>11.67</td>
<td>0.73</td>
<td>6.67%</td>
</tr>
<tr>
<td>Dan_R5</td>
<td>Fraxinus angustifolia</td>
<td>7.90</td>
<td>8.80</td>
<td>0.90</td>
<td>11.39%</td>
</tr>
<tr>
<td>Dan_R9</td>
<td>Fraxinus angustifolia</td>
<td>9.20</td>
<td>9.58</td>
<td>0.38</td>
<td>4.13%</td>
</tr>
<tr>
<td>Ave_C</td>
<td>Platanus x acerifolia</td>
<td>12.21</td>
<td>12.83</td>
<td>0.62</td>
<td>5.08%</td>
</tr>
<tr>
<td>Ave_B</td>
<td>Platanus x acerifolia</td>
<td>12.96</td>
<td>13.93</td>
<td>0.97</td>
<td>7.48%</td>
</tr>
<tr>
<td>Mean error (m)</td>
<td></td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD of the error (m)</td>
<td></td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (m)</td>
<td></td>
<td>0.60 (5.96%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
length derived from Treegraph is less variable than branch radii, but their median values generally fall below WBE model prediction (Fig. 11, Figure 12). The metabolic rate estimated from the Treegraph-derived radius and length scaling exponents varies significantly for branch segments < 10 cm in diameter (Fig. 11). For branch segments over 10 cm in diameter across 18,657 branching nodes, deviations from WBE prediction are more significant for
length than for radius scaling exponents, with a median estimated metabolic rate of 0.58 falling below the theoretical value of 0.75 predicted by WBE model (Fig. 12). This aligns with studies from Lau et al. (2019), Bentley et al. (2013), Savage et al. (2010), which also found greater variability in the distribution of length ratios compared to radii ratios and observed exponents lower than WBE prediction. There are various reasons these values may be lower than WBE predicts, including environmental factors such as asymmetric competition for light (Coomes et al., 2011) and departures from the assumption of constant sapwood area across all branching levels (Sopp & Valbuena, 2023).

In Figure 13, the logarithm of mean branch area ratio approaches zero, suggesting that for branches with diameters over 10 cm, the combined cross-sectional areas of daughter branches are approximately equal to that of their parent branch, which indicates area-preserving

Figure 9. Diagram of trunk slices of the tropical tree T2 at CAX site (left panel) showing (A) slice from lower tree trunk with an irregular buttress; (B) slice from the middle of the trunk; (C) slice capturing a branching node; and (D) slice from the upper trunk encountering occlusion. Top view of point cloud at each trunk slice with a generated cylinder for comparison (red circles with radius determined by the mean point-to-axis distance (middle panel). Distribution of point-to-axis distance of all points within the corresponding slice (right panel).
behaviour. Similar results were presented by Bentley et al. (2013). However, the deviations of Treegraph-observed scaling rates and WBE model predictions (Fig. 12) suggest asymmetric branching in the evaluated species, characterized by daughter branches having varied lengths and diameters at each branching node.

Conclusions

In this study, we present Treegraph, a new approach aimed explicitly at automatically deriving morphological and topological traits of trees across all levels of branching from TLS data. We minimize a priori assumptions about tree architecture to prevent circularity in applications using derived structure properties.

We test Treegraph on 603 trees of widely varying size, structure and type, finding an average modelling accuracy of 0.027 m with a standard deviation of 0.035 m. Comparing with manual measurements from whole-tree destructive harvest, Treegraph shows higher accuracy in estimating branch length than diameter. Uncertainty in diameter estimates stems from point cloud quality and deviation between actual branch shapes and the idealized circular representation. Treegraph effectively characterizes tree architecture but can overestimate furcation numbers and underestimate internode lengths, particularly in the upper crown. We apply Treegraph to test metabolic scaling theory (MST) across 17 species and find that branch-level scaling exponents of branch radius, length and estimated metabolic rate deviate from WBE predictions for optimized symmetrically branching trees.

Treegraph, as an open-source Python tool, is recommended for estimating tree branch sizes and connectivity. While caution is advised when interpreting tree volume

Figure 10. Analysis of mean furcation number (upper row) and mean internode length (bottom row) as a function of position along the tree, using manual measurements (left column) and Treegraph estimates (right column). Mean values are derived at 1 m interval from the first branching node for eight trees in Milton Keynes site.

© 2024 The Author(s). Remote Sensing in Ecology and Conservation published by John Wiley & Sons Ltd on behalf of Zoological Society of London.
Figure 11. Boxplots of branch-level scaling exponents calculated on a node basis for radius (top), length (middle) and estimated metabolic rate (bottom) across parent diameter classes of 5 cm intervals. Boxes indicate the interquartile range, with central lines showing medians. Whiskers extend to data within 1.5 times the IQR. Red dashed lines indicate WBE predicted values for an idealized tree.
Figure 12. Distribution of branch-level scaling and estimated metabolic rate calculated on a node basis for individual species. Only branch segments with a diameter larger than 10 cm are considered here. Dark blue dash lines represent WBE model prediction, while the light blue dash lines indicate the median values derived from Treegraph model estimates.

Figure 13. Scatter plot illustrating the distribution of branch area ratio (left) and branch length ratio (right) in relation to parent branch area, including data from 17 species across three sites. The data covers all branch segments with a diameter larger than 10 cm.
estimates, as these are calculated as a by-product of the architecture. With tests conducted on 17 species, Treegraph suggests a capacity for adaptation to other species for MST testing, though additional data are needed for thorough verification. Treegraph provides the TLS ecology community with a solution for automating the reconstruction of architectural traits, with potential for integration in large-scale ecological studies of tree architecture.

Acknowledgements

We would like to express our sincere gratitude to The Milton Keynes Parks Trust, Philip Wheeler and Kadmiel Maseyk from The Open University, and all volunteers who contributed to the data collection in Milton Keynes site. We also acknowledge the collaborators who made data collection possible for the other sites (see previous publications for full details). We acknowledge capital funding and travel support from NERC via the National Centre for Earth Observation (NCEO) and UCL Geography; NERC Standard grants NE/N00373X/1 and NE/P011780/1; European Research Council grant no. 757526. Tropical Forest Degradation Experiment (FODEX); Metrology for Earth Observation and Climate Project (MetEOC-2) within the European Metrology Research Programme (EMRP) funded by the European Union under grant no. ENV55.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The source code for Treegraph is available at https://github.com/wanxinyang/treegraph. The TLS data and Treegraph model outputs of urban trees in Milton Keynes sites are archived at Zenodo https://zenodo.org/record/8412120. The TLS data and QSM for trees from the Wytham Woods plot were published in Calders et al. (2022b), and are free to download at Zenodo https://zenodo.org/record/7307956 (Calders et al., 2022a). Data of the four tropical trees at the CAX site were published in Burt et al. (2021). The TLS measurements can be downloaded from Zenodo https://zenodo.org/record/4056903 (Burt et al., 2020b), and their corresponding destructive measurements are available at Zenodo https://zenodo.org/record/4056899 (Burt et al., 2020a).

References


of Sciences of the United States of America, 107(52), 22722–22727.
Zhang, J., Wang, J., Dong, P., Ma, W., Liu, Y., Liu, Q. et al. (2022) Tree stem extraction from TLS point-cloud data of natural forests based on geometric features and DBSCAN. Geocarto International, 37, 1–11.

Supporting Information

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

Appendix S0. Manual measurements on branches from harvested trees.
Appendix S1. Refine base node in initial distance graph.
Appendix S2. Parameter selection for determining bin width.
Appendix S3. Branch radius estimates.
Appendix S4. Examples of Treegraph QSM.
Appendix S5. AGB estimation from Treegraph.