Exploring the Scientific Mechanism of Tree Structure Network based on LiDAR Point Cloud Data

Haoliang Chen 1, Yi Lin 1,*

1 Institute of RS and GIS, School of Earth and Space Sciences, Peking University, Beijing, China - 2301110689@pku.edu.cn
* Correspondence - yi.lin@pku.edu.cn

Keywords: Tree Structure Network, LiDAR, Quantitative Structure Model, Pareto Optimality, Point Cloud.

Abstract

To explore how trees optimize their structure, we developed a method based on Pareto optimality theory. This method consists of the following operations. Firstly, we utilize Quantitative Structure Models for Single Trees from Laser Scanner Data (TreeQSM) to extract tree structures from point clouds acquired through Light Detection and Ranging (LiDAR). Subsequently, we utilize a graph-theoretical model to characterize the natural tree structure networks and implement a greedy algorithm to generate Pareto optimal tree structure networks. Finally, based on the Pareto optimality theory, we explore whether tree structures are multi-objective optimized. This paper demonstrates that tree structures lie along the Pareto front between minimizing "transport distance" and minimizing "total length". The growth pattern of trees, which produces multi-objective optimized structures, is likely an intrinsic mechanism in the generation of tree structure networks. The location of tree structures along the Pareto front varies under different environmental conditions, reflecting their diverse survival strategies.

1. Introduction

Trees play an indispensable role in regulating the climate, slowing down the greenhouse effect (Cannell, 1996), conserving water and maintaining soil quality, etc. In the process of energy conversion, trees need to transfer, generate, and convert nutrients, as well as maintain their mechanical stability. This places specific demands on the structure formed by their trunks and branches. Robust structure performance involves intricate trade-offs, considering factors such as cost, transmission efficiency, and fault tolerance (Tero et al., 2010). It is valuable to investigate how trees optimize their structure to obtain and allocate resources effectively while simultaneously reducing the cost of building their structure.

The exploration of function-structure modeling in plant growth commenced in the mid-1990s. Sachs et al. (1995) presented an alternative perspective on tree morphology, contending that the interaction between the environment and the tree goes beyond a mere modifying role; rather, it serves as a pivotal determinant of tree form. According to their viewpoint, genetic and molecular mechanisms do not outright dictate the tree's form but establish rules for self-organization through the competitive interaction among branches. Méch et al. (1996) suggested that interactions with the environment play a crucial role in shaping the development of plants and plant ecosystems. They proposed a model that explores the bidirectional exchange of information between plants and the environment, aiming to investigate the growth of trees through the study of this information exchange.

In recent years, many scholars have conducted extensive studies on the feedback mechanism of plant structure and established various in-depth models, such as the ADEL-maize model (Fournier et al., 1999), LIGNUM model (Perttunen et al., 2001), L-Peach model (Allen et al., 2005) and the maize growth virtual model (Ma et al., 2008). These models can be broadly categorized into two types: the first type integrates the mature plant functional model with the structural model, enabling mutual feedback between structure and function throughout the plant growth process; the second type involves considering both physiological and structural functions within the same model, simulating the feedback mechanisms of these two aspects in parallel.

The function-structure modeling approaches abstracting genetic mechanisms have driven plant development into a computational framework. This framework can be utilized to investigate the genetic mechanisms of canopy morphology development (Palubicki, 2013), the impacts of root development on nutrient and water uptake (Dunbabin et al., 2013), etc. Conn et al. (2017) have selected tobacco, tomato, and other crops as research subjects to investigate the correlation between the growth of crops and two structural characteristics: the total length and the transport distance. They also examined the benefits of selective fitness in the process of plant transportation.

To understand plant functions through mathematical modeling, quantitative measurements of plant morphology are important. Advances in remote sensing technology, especially terrestrial laser scanning (TLS), have provided a highly accurate data source for extracting tree structures. TLS allows for highly detailed 3D mapping of smaller areas, resulting in dense 3D point clouds of tree surfaces (Lin and Hyppä, 2022), which enables quantitative analysis of trees and the reconstruction of their 3D models.

Pfeifer et al. (2004) implemented a method to reconstruct conifer tree structures by tracking along branches and generating cylinders. Binney et al. (2009) reconstructed tree branch structures using LiDAR data. Hackenberg et al. (2015) utilized a sphere centered on the tree skeleton to track the branching structure of the tree. Raunonen et al. (2015) proposed a method to automatically reconstruct a Quantitative Structural Model (QSM) using fixed-size large surface patches to segment the point cloud into trunks and branches.

Based on the aforementioned studies, this paper focuses on examining the scientific mechanism of tree structure by analyzing the trade-off between the total length and the transport distance. We extract tree structure information from LiDAR point clouds using TreeQSM and apply the Pareto optimality theory to explain the formation of the tree structure.
2. Materials

The data used for this study were obtained from two publicly available LiDAR point cloud datasets: one collected in Rushworth (Calders et al., 2015) and the other collected in Ndélélé (Momo et al., 2018).

2.1 Study Area

Rushworth (36.58985°S, 145.01331°E) is located in southeastern Australia, in the subtropical region of southern hemisphere. This region has a temperate maritime climate. It is warm and comfortable in the summer, while cool and wet in winter, with an average annual rainfall ranging between 750-1000 mm. The forests in the region are evergreen, featuring main tree species such as Eucalyptus leucoxylon, E. microcarpa, and E. tricarpa (Calders et al., 2015).

Ndélélé (4.03907°N, 14.93055°E) is located in eastern Cameroon, in the tropical region of northern hemisphere. This region is characterized by a regional average annual precipitation of between 1,500 and 2,000 millimeters, two distinct dry seasons, an average annual temperature of 24°C, and an elevation ranging between 600 and 700 meters above sea level. The forests in this region are semi-deciduous, with dominant species belonging to the mallow and cannabis families (Momo et al., 2018).

2.2 Individual Tree Point Cloud Data

Rushworth: In May 2012, TLS data were collected using a RIEGL VZ-400 terrestrial laser scanner. A C++ library, based on the open-source Point Cloud Library (Rusu & Cousins, 2011), was used to extract trees from the global point clouds. Only LiDAR data with reflectance values above −4 dB were used to identify individual trees (Calders et al., 2015). In this paper, the tree data provided by this dataset were labeled with names starting with “PT”.

Ndélélé: Between July 2015 and August 2016, point cloud data of several individual trees were collected using a Leica C10 Scanstation terrestrial laser scanner in the Ndélélé area. Co-registration of point clouds from different scanning positions and extraction of individual trees were conducted using Leica cyclone software (v. 9.1). Additionally, leaves were manually removed from the point cloud for each tree (Momo et al., 2018). In this paper, the tree data provided by this dataset were labeled with names starting with “MT”.

3. Methods

This study utilizes TreeQSM (Raumonen et al., 2015) to extract tree structures from LiDAR point clouds. To characterize the natural structure network of each tree, we utilize a graph-theoretical model to abstract the tree structure network into a graph consisting of vertices and edges. To generate Pareto optimal structure networks, we have developed an algorithm that integrates Pareto improvement with a greedy approach to effectively balance the trade-off relationship between two key features of tree structure networks: Transport Distance and Total Length. Finally, we fit the Pareto front for each tree and observed the location of tree structures along the Pareto front (Giagkiozis et al., 2014).

3.1 Extraction of Tree Structure

To quickly and automatically build accurate tree structure models from single-tree point cloud data collected by LiDAR, we utilize TreeQSM. This tree modeling algorithm was publicly introduced and has undergone continuous enhancements in recent years (Raumonen et al., 2013).

During the execution of the TreeQSM algorithm, two coverage sets are successively generated. The first set utilizes the parameter PatchDiam1 to create an initial point cloud segmentation, removing individual points that are not part of the trees. The second set utilizes two parameters, PatchDiam2Min and PatchDiam2Max, which determine the minimum and maximum diameters of the point cloud chunks in the coverage set. The chosen values for the three parameters are empirical.

Although empirical parameters are suitable for modeling most normal-sized single trees, achieving a better fit for trees of specific sizes and structures, as well as different tree species, can be challenging (Calders et al., 2015). To enhance the extraction process, we modeled 16 groups of parameters (unit: m) for each tree, as shown below:

\[
\text{PatchDiam1} = 0.12; \\
\text{PatchDiam2Min} = 0.02 \times \text{factor}; \\
\text{PatchDiam2Max} = 0.07 \times \text{factor},
\]

where \( \text{factor} = 0.5 + 0.1k \) (\( k = 0, 1, 2, 3, \ldots, 15 \)).

For each cylinder in quantitative structure models (QSM) extraction results, we calculated the average distance from all points closest to it. Subsequently, we averaged these values across all cylinders to determine the criterion for fitting accuracy. Among the 16 groups of parameters for each tree, the one that yields the highest accuracy is selected for QSM extraction.

The output results of TreeQSM for extracting single-tree point clouds encompass six data types: Cylinder, Branch, Treedata, Rundata, Pmdistance, and Triangulation. The Cylinder represents multiple cylinders obtained through fitting, while the Branch identifies the tree branches determined by the algorithm. Both Cylinder and Branch encapsulate the fundamental composition and topological properties of the tree structure network, serving as the primary data structures employed in this study.

3.2 Structure Network Based on Graph Theory

To facilitate the analysis of tree structures, we employ a concise graph-theoretic approach to describe an individual tree.

Initially, we simplify the cylinders in TreeQSM results into line segments by using the centers of their upper and lower surfaces as endpoints. This enables us to represent the tree structure using the graph \( G = (V, E) \). As nutrients can flow bidirectionally, from leaves to roots and vice versa, all edges in \( E \) are undirected, with their weights denoting the lengths of the corresponding cylinders.

\[
V_{c_0} = \text{tree root}; \\
V_{c_i} = \text{end point of } C_i; \\
|E_{c_i}| = \text{length of } C_i,
\]

where \( C_i \) is the \( i \)-th cylinder in TreeQSM results.

In this study, the fundamental units constituting the tree structure networks are branches. By leveraging the branch data structure extracted from TreeQSM, we can further refine and optimize the graph \( G_k \) into \( G_k = (V_k, E_k) \). The set of vertices in the graph is refined to include only the root and tip points of each branch, while the edges connecting the root and tip of a branch are simplified into single edges.
\[ V_{b,0} = V_{c,0}; \]
\[ V_{b,\text{start}} = V_{c,\text{min}(\text{CyList}_i)}; \]
\[ V_{b,\text{end}} = V_{c,\text{max}(\text{CyList}_i)}; \]
\[ |E_{b,i}| = \sum_{j:\text{CyList}_i} |E_{c,j}| , \]

where \( B_i \) is the \( i \)th branch in TreeQSM results; \( \text{CyList}_i \) is the index list of cylinders in \( B_i \); \( P_i \) is index of the branch on which the root of \( B_i \) lies.

### 3.3 Pareto Optimality: Transport Distance vs. Total Length

Numerous problems involve multiple objectives that frequently conflict with and impact each other. Achieving the optimal state for all objectives simultaneously is challenging, but efforts are made to bring them within an optimal range in practice. Pareto Optimal Improvement is a common method used to achieve multi-objective optimization. Simply put, in the bi-objective case, Pareto optimization involves optimizing one objective without degrading the other.

In the context of a solution, if no other solution can outperform it in all objectives, it is considered a Pareto-optimal solution for the multi-objective optimization problem. Using Figure 1 as an example, targets 1 and 2 represent the optimization effects of the two objectives. Assuming the feasible decision space is the area above the black curve in the graph, the blue points represent Pareto-optimal solutions, each outperforming the black points in terms of optimization. These blue points collectively form the Pareto front curve when mapped by the objective function, illustrating the trade-offs and optimal solutions in the bi-objective optimization problem (Giagkiozis et al., 2014).

**Figure 1.** Pareto front (black curve) in a bi-objective

In tree structures, the primary sites for nutrient absorption and synthesis are located at the terminations of the structural network, namely the leaves and the roots. Trees continuously optimize their structure to thrive in the intense competition for survival. Within a tree structure network, performance can be evaluated based on **Transport Distance**, while cost can be assessed in terms of **Total Length**, as defined below:

\[
\text{Total}(G_0) = \sum_{E_{b,i} \in E_b} |E_{b,i}| ;
\]
\[
\text{Transport}(G_0) = \sum_{V_{b,0} \in V_b} \text{dist}(V_{b,0}, V_{b,i}) ,
\]

where \( \text{dist}(V_{b,0}, V_{b,i}) \) is the shortest distance between a tree root and a branch tip along the structure network.

Transport distance is related to performance metrics of plant structures, including resistance to and consumption of nutrients transported between the leaf and root systems (Peel, 2013), while total length serves as a measure of the resources required to construct a branch and the rate at which the tree consumes resources to sustain itself (Bloom et al., 1985). Given that minimizing both total length and transport distance can provide plants with a stronger survival advantage (Conn et al., 2017), we need to consider the joint optimization of these two objectives.

Since trees are rich in species and grow in different environments, we consider the transport distance and the total length as two basic objective functions. We construct a linear joint objective \( \text{JointT} \) with a variable parameter \( \alpha \) to quantify how plants collectively minimize the transport distance and total length as follows.

\[
\text{JointT} = \alpha \cdot \text{Transport} + (1 - \alpha) \cdot \text{Total} .
\]

To test whether the Pareto Optimality can reasonably explain the generation mechanism of tree structure networks, we need to compare the relationship between the natural structure and the Pareto front for trees in the Transport-Total coordinate system.

#### 3.3.1 Natural Structure

To get the transport distance and the total length of a tree’s natural structure network, we calculate them directly using Equation 4.

#### 3.3.2 Random Structure

To demonstrate that a tree structure network located along the Pareto front is unlikely to occur by chance, in this study, we also compare the location of a randomly generated structure network with the Pareto front. Taking the point set of tree root and all branch tips as input, the root is considered the starting link-point. For each link-point, a point is randomly selected from the points that are not already part of the structure network to establish a connection with it. Then, we randomly select a point from the points that have already been incorporated into the structure network to serve as a new link-point. The process of connecting is repeated until all the points have joined the structure network.

#### 3.3.3 Pareto Front

Constructing a tree structure generation algorithm for a specific \( \text{JointT} \) is similar to the generalized Steiner tree generation problem (Conn et al., 2017). Since computing a 3D minimum Steiner tree is a typical NP-hard problem in combinatorial optimization that cannot be solved in polynomial time, it is nearly impossible for us to find a minimum Steiner tree within a reasonable time frame when the number of points exceeds ten. We resort to an iterative greedy algorithm to construct the structure network, aiming to make the joint objective approach optimality.

Firstly, by considering the point set consisting of the tree root and all branch tips as the input point set, denoted as \( P_{\text{input}} = \{P_0, P_1, ..., P_n\} \) where \( P_i = V_{b,i} \), we establish the initial growth edge \( L_0 \). This edge connects the root point \( P_0 \) with the center point \( P_{\text{center}} = \frac{\sum_{i=1}^{n} P_i}{n} \) and serves as the trunk for the tree.

Secondly, the parameter \( D \) (e.g., \( D = 20 \)) is chosen to divide the newly added grow-edge uniformly into \( D \)-sub-segments of uniform length. We take the \( D \)-equipartition points and the endpoint of the grow-edge as grow-position set \( S \).
Thirdly, we proceed with an iteration over all remaining vertices and grow-position combinations \( \{P, S_1\} \). To get the new grow-edge, we connect points in the combination that minimizes the increment \( \Delta \text{JointT} \).

Finally, the process of adding new grow-edges is reiterated until all the points are integrated into the structural network. The last step involves removing any unused portions of the trunk.

To approximate the Pareto front, we select with the parameters \( \alpha = 0, 0.25, 0.50, 0.75, \ldots, 1 \) for 41 cases. In these cases, the total length and transport distance values were calculated, and the points corresponding to these 41 groups of values were connected.

Due to the individual differences between trees, the total length and transport distance calculated for each tree exhibit significant variations in values. To facilitate the comparative analysis of all results, it is essential to normalize the calculation results. In this study, normalization has been applied to both the total length and the transport distance, mapping them to the range of 0-1, as shown below:

\[
\text{Transport}_{\text{nor}} = \frac{\text{Transport}}{\text{Transport}_{\text{max}}}, \quad \text{Total}_{\text{nor}} = \frac{\text{Total}}{\text{Total}_{\text{max}}}
\]  

(6)

### 4. Results

#### 4.1 Structure Network of Single Tree: Confliction between Total Length and Transport Distance

Based on the algorithm mentioned above, we analyze whether there is a contradiction between minimizing total length and minimizing transport distance in a tree structure network. Taking the root and all branch tips as input nodes (as shown in Figure 2A), the structure network diagrams were drawn for the natural structure (Figure 2B), the Steiner-tree structure (Figure 2C, minimized total length), and the Satellite-tree structure (Figure 2D, minimized transport distance). The total length and transport distance of three structure types were calculated, as shown in Table 1.

<table>
<thead>
<tr>
<th>Structure Type</th>
<th>Total length (m)</th>
<th>Transport distance (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>43.28</td>
<td>550.54</td>
</tr>
<tr>
<td>Steiner</td>
<td>32.11</td>
<td>591.35</td>
</tr>
<tr>
<td>Satellite</td>
<td>521.29</td>
<td>521.29</td>
</tr>
</tbody>
</table>

Table 1. Total Length and Transport Distance of a tree.
In the case of minimizing transport distance, the total length is larger compared to the natural case because each vertex is independently connected to the root of the tree. Conversely, in the case of minimizing total length, the transport distance of some vertices may be large.

### 4.2 Tree Structure is Close to the Pareto front

We extracted Pareto fronts for a total of 43 trees in the dataset and compared them with the randomly generated structures and the natural structure of trees. Figure 3 illustrates the Pareto fronts for some trees in the dataset and plots the location of randomly generated structures alongside the natural structure in the same coordinate system.

From Figure 3, we can see that the Total Length and Transport Distance corresponding to the natural structure are located in the inner Pareto front, i.e., in the feasible domain of the Pareto optimality theory. These locations initially reflect that the Pareto optimality theory has no theoretical error in analyzing the formation of the tree structure network (Giagkiozis et al., 2014). Further, we need to calculate the distance of the corresponding points of the natural structure from the Pareto curve to determine if the Pareto optimality theory can elucidate the self-optimization of tree structures. Since the Pareto front in this study is a folded line connecting 41 points, we use the minimum distance of the points from all the line segments on the folded line to characterize the distance from the Pareto front.

While calculating the distance, it is also possible to calculate the value of the parameter $\alpha$ that is closest to the structure of the tree itself through interpolation.

Combined with the tree height information, we can observe that trees with heights ranging from 8-40 meters tend to be close to the Pareto front. Considering a certain positive correlation between tree heights and the length of tree growth time, we can assume that the structure network construction pattern close to the Pareto front persists for an extended period during the tree growth process. Achieving a good trade-off between total length and transport distance may represent an important survival strategy, ensuring a balance between resource transportation and consumption. This enables trees with stronger survival abilities to pass on their genetic material through generations, enduring the challenges of natural selection.

### 4.3 Similarities and Differences among Species

In this study, we measured the cost of the tree structure network in terms of the **Total Length** of the tree used to connect the branch ends to the roots and the **Transport Distance** of the tree used to deliver resources to the branch ends.

As shown in Figure 3, both the Ndélélé and Rushworth trees achieve a high degree of "cost-effectiveness" in their structures compared to those that minimize transport distances. In the construction of structure networks, significant cost savings are
achieved by reducing the total length significantly, even if it results in a slight increase in transport distance.

For the joint target $\text{JointT} = \alpha \cdot \text{Transport} + (1 - \alpha) \cdot \text{Total}$, a smaller value of $\alpha$ implies that the reduction of total length is more crucial for trees in constructing the structural network. In other words, the locations of tree structures along the Pareto front may vary due to the constraints of their growing environment.

From the left subplot of Figure 4, we can observe that the location of the tree structures along the Pareto front varies between the two regions. The closest $\alpha$ values of tree structures along the Pareto front in the Ndélélé region are distributed between 0 and 0.61, with a mean value of 0.22. In contrast and the distribution of $\alpha$ values of tree structures in the Rushworth region ranges from 0 to 0.07 with a mean value of 0.02. Differences in the location of trees along the Pareto front may reflect the varying effects of their growing environment on survival strategies.

Geographically, the Ndélélé region is situated in the northern hemisphere tropics, experiencing abundant sunshine throughout the year with two distinct dry seasons. This climatic characteristic is crucial for trees in this region to minimize the depletion of nutrients, especially valuable water, during transport. This strategy helps maintain normal tree life activities under the limited precipitation conditions during the dry season. On the other hand, the Rushworth region is located at the subtropical and temperate border of the southern hemisphere, on the southeast coast of Oceania, where precipitation is evenly distributed, and thermal conditions are favorable. As a result, the transport distance has less influence on the structural network of trees in this region.

5. Discussion

Optimization in nature occurs when non-optimized species are outcompeted (Mattheck and Bethge, 1998). According to the theory of Pareto optimality, approaching to the Pareto front is a trend in the evolutionary process of trees (Giagkiozis et al., 2014). Trees that are closer to the Pareto front exhibit a superior ability to balance the total length with the transport distance, making them more likely to survive natural selection.

However, there are more factors influencing the generation of tree structures such as maturation stress (Alméras and Clair, 2016) and environmental factors (Creber and Chaloner, 1984), which cause the natural structure of a tree to not exactly fall on the Pareto front (Conn et al., 2017). Tree’s physiological functions working within a 3D vascular structure engender 3D apparent patterns of metabolic scaling, biomechanical coordinating, and allometric scaling (Lin and Hyppä, 2022). Understanding trees’ 3D architecture needs to focus on tree-branching architecture, biomechanics, metabolism, respiration, reproduction (Lin et al., 2023).

The experimental data in this study encompasses trees located in only two regions, each characterized by significantly different environments. This diversity makes it challenging to accurately analyze the effects of factors such as precipitation, temperature, and sunshine on the structure networks of trees. To gain deeper insights, a more detailed analysis might involve subjecting identical tree seedlings to diverse growth environments.

When modeling tree structure, we disregarded the bends in the branches. In reality, there are several bends from the root of the branch to the end (Ojo and Shoele, 2022), which results in the distance values of the tree itself being larger than those deduced by the algorithm. At the same time, the diameter of the branches is related to the amount of resources required for the construction of the tree per unit length, nutrient transportation, etc. In this study, the thickness of the tree branches and trunks was not taken into account, which will lead to deviations when calculating the total length.

In addition, TreeQSM itself has a certain degree of randomness, which causes the extraction results to have some differences from the real tree structure. Since TreeQSM cannot extract leaf location information effectively, in this paper, we settle for the second best. In TreeQSM results, only the tips of the branches were extracted as leaf locations, while leaves growing directly from trunks and branches were ignored.

6. Conclusion

This study introduces a graph-theoretic algorithm to evaluate the design strategies of trees. We find that the natural structures of trees lie along the Pareto front between two objectives: minimizing the total length and minimizing the transport distance. The growth pattern, producing structures close to the Pareto front, is probably an intrinsic mechanism in the generation of tree structure networks. Additionally, varying survival conditions will lead to different effects of this mechanism on tree growth.
Acknowledgements

The work was financially supported by the National Key Research and Development Program of China (No. 2022YFE0112700) and the National Natural Science Foundation of China (No.32171782).

References


