1949

Advancements in Individual Treetop, Tree and Crown Base Height Location Using Laser Scanning Data of Various Resolutions

Thesis for the Degree of Doctor of Philosophy (PhD)

Author:
Gergő Diószegi

Supervisor:
Szilárd Szabó DSc

UNIVERSITY OF DEBRECEN
Doctoral Council for Natural Sciences and Engineering
Doctoral School of Earth Sciences
Debrecen, 2025

Hereby I declare that I prepared this thesis within the Doctoral Council of Natural Sciences and Engineering, Doctoral School of Earth Sciences, University of Debrecen, to obtain a PhD Degree in Natural Sciences at Debrecen University.

The results published in the thesis are not reported in any other PhD thesis.

Debrecen, 13 February 2025

…….
Gergő Diószegi
Signature of the candidate

Hereby I confirm that the candidate Gergő Diószegi conducted his studies with my supervision within the Doctoral Program of natural and anthropogenic processes of the litho- and hydrosphere of the Doctoral School of Earth Sciences between 2024 and 2025. The independent studies and research of the candidate work significantly contributed to the results published in the thesis.

I also declare that the results published in the thesis are not reported in any other thesis.

I support the acceptance of the thesis.

Debrecen, 13 February 2025

…….
Szilárd Szabó DSc
Signature of the supervisor

Advancements in Individual Treetop, Tree and Crown Base Height Location Using Laser Scanning Data of Various Resolutions

Dissertation submitted in partial fulfilment of the requirements for the doctoral (PhD) degree in Earth Sciences

Written by **Gergő Diószegi** certified forest scientist (MSc)

Prepared in the framework of the Earth Sciences Doctoral School of the University of Debrecen
(doctoral program of natural and anthropogenic processes of the litho- and hydrosphere)

Dissertation advisor: **Szilárd Szabó DSc**

The official opponents of the dissertation:

Dr. ....
Dr. ....

The evaluation committee:

chairperson:	Dr.	.....
members:	Dr.	.....
	Dr.	.....
	Dr.	.....
	Dr.	.....

The date of the dissertation defence: …… 2025

To my Mother,

who shared the light.

To my Wife,

who has been found by the light.

To my Son,

who does not want to be embraced by the light,

To God,

who is the Light.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ALS	Aerial Laser Scanr
BFH	Branch Free Height
CBH	Crown Base Height
CHM	Canopy Height Model
CSF	Cloth Simulation Filtering
DBSCAN	Density-Based Spatial Clustering
FETR	First Emerging Tree Region
FSCT	Forest Structural Complexity Tool
GTR	Growing Tree Region
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
ITD	Individual Treetop Detection
ITS	Individual Tree Segmentation
LiDAR	Light Detection and Ranging
LM	Local Maxima
MA	Matching Algorithm
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
TLS	Terrestial Laser Scanning
TTS	Topology-based Tree Segmentation
ULS	Unmanned Laser Scanning
VWF	Variable Window Filtering

1 INTRODUCTION

Describing forest ecosystem status is a key element in defining forest management strategies. Detailed tree inventories, including information on species composition, height, canopy closure, and biomass, provided these descriptions. Traditional inventory variables are sampled sparsely or systematically at the plot level to calibrate the regional and nationwide forest analyses (Beaudoin et al. 2017; Blackard et al. 2008; Kändler 2006; Tomppo 2006). Active remote sensing technology, Light Detection and Ranging (LiDAR), used by forest managers and researchers, especially in combination with airborne laser scanning (ALS), has established itself in the last two decades (Hyyppä et al., 2008; Lim et al., 2003; White et al., 2016). The rapid development of laser scanning technologies has guided the replacement of labor-intensive and costly forest inventories with remotely laser-scanned data-based forest inventories. LiDAR data can provide forest information for an enhanced understanding of forest structure and ecology (Koch et al., 2006; White et al., 2016; Wulder et al., 2008).

Depending on the scale of interest, inventory attributes are generally modelled/predicted using LiDAR data-derived forest information based on two approaches: the area-based approach (Bouvier et al., 2015; Parkitna et al., 2021) and, accounting for finer-scale variability, the individual tree-based approach (Hardenbol et al., 2023; Kaartinen et al., 2012; Popescu et al., 2003). Individual tree species, particularly stem volume, are the two main parameters used to quantify the vegetation carbon distribution across a vertically stratified forest (Zianis et al., 2005). Individual trees also host insects and pathogens, and the complex interaction between them influences tree mortality, foliage spots, wilting, and dieback. Such stress indicators can only be identified at the tree level, and early detection can facilitate adaptive forest management strategies for their timely implementation (Hantsch et al., 2014). Moreover, crown plasticity, an indicator of inter- and intraspecific competition in temperate and tropical forests, has been quantified at the tree level (Rozendaal et al., 2006; Seidel et al., 2011). Therefore, individual tree-level data collection requires not only a high accuracy of individual tree detection (ITD) and segmentation, but also the capture of quantitatively robust metrics that can describe the natural complexity of trees and forests. In recent years, the latter one has witnessed a substantial increase on the research field of terrestrial

laser scanning (TLS) (Lecigne et al., 2018; Terryn et al., 2023; Xi and Hopkinson, 2022). Regarding aerial laser scanning (ALS), the increase in point density by repeated overlapping scan flies or constant technical advancement of the ALS sensor (meaning a significant rise in the laser pulse canopy penetration ratio) enables the retrieval of high-resolution 3D tree structural parameters. Thus, it is possible to apply the initial TLS metrics to ALS data. Many of these metrics, such as the previously mentioned plasticity, 3D alpha crown volume, and projected crown area (Terryn et al., 2023), depend on accurate crown location and isolation. The tree canopy (i.e., crown) is generally defined (has been used and measured by traditional forest inventory techniques) as part of the tree higher than the Crown Base Height (CBH) (Kelly et al., 2018; Luo et al., 2018; Stefanidou et al., 2020; Sumnall et al., 2017; Vauhkonen, 2008), which is also called Branch Free Height (BFH) in the field of forestry.

This dissertation comprises two interrelated studies, each proposing distinct yet complementary forestry tools, plus a stand-alone tree isolation-centred development. The first study introduces an ITD methodology utilizing a Canopy Height Model (CHM) generated from low-resolution LiDAR data. The second study focuses on a Crown Base Height (CBH) location technique using high-resolution point cloud data, whereas stand-alone development uses terrestrial LiDAR data for point-based tree isolation.

Initially, a novel ITD algorithm was developed based on the Growing Tree Region (GTR) concept, tailored to Central European mixed forest ecosystems, using LiDAR-derived CHM data. Subsequently, an open-source point-cloud-based CBH locating tool, named *treecbh*, was introduced using high-resolution aerial laser-scanned data. *treecbh* is applicable to forests prior to segmentation into individual trees. The GTR-based ITD tool developed in the first study was used for treetop identification. The CBH locator utilizes the hierarchical graph clustering software *treeiso* (Xi and Hopkinson, 2022) for tree trunk isolation and first-branch detection. The novelty of *treecbh's* CBH location algorithm lies in its utilization of a vertical cross-sectional K-means-clustering-based 2D kernel method. In addition, a third research topic, which aims to employ the Discrete Morse Theory for individual tree isolation using very high-resolution point cloud data scanned with TLS, was conducted.

The main objectives were formulated as follows:

1. To develop a simple yet flexible CHM-based treetop locator algorithm.
2. This algorithm was integrated into an open-source tool that seamlessly fits within the coding framework of the *lidR* package in the R computational environment.
3. To test the proposed algorithm in mixed forests with various compositions and canopy densities, and to compare its performance with the most robust and freely available treetop location method that employs an LM search using VWF.
4. To provide an automated, adaptable, and easily optimizable tool for locating CBH using point cloud data of already segmented individual trees,
5. To validate the effectiveness of *treecbh* in locating CBH using leaf-off and leaf-on ALS data against field-measured data obtained from central European forests with diverse compositions, and
6. To determine the optimal point density necessary to accurately represent the structure of the input tree point cloud, particularly focusing on the tree trunk and lower branches.
7. Implementing the fundamental concepts of discrete Morse theory in an R package.
8. To create a framework using existing algorithms to locate tree stems.
9. To enhance this framework, tree crowns are identified and connected to previously located stems using the discrete Morse theory.
10. To implement the fundamental concepts of discrete Morse theory to an application in forestry as an R package. To create a framework using existing algorithms for detecting tree stems.

This dissertation is organized into three research topics to maintain a coherent and structured approach. Each section addresses one of these core topics, ensuring comprehensive exploration of the methodologies and findings central to the overall study.

2 LITERATURE REVIEW

2.1 Individual Treetop Location

Understanding the importance of the high accuracy of LiDAR data-derived ITD and segmentation is inseparably bound to the traditional (i.e., inventorial) individual tree variables. The main focus of forest management is retrieving individual tree parameters such as tree height, tree location, tree species, diameter at breast height or crown diameter (González-Ferreiro et al., 2012; Hu et al., 2014; Koch et al., 2006; Maltamo and Gobakken, 2014) in order to estimate important forest characteristics at stand level: growing stock volume, tree species composition, canopy density and mean basal area (Ene et al., 2012; Lee et al., 2013; Parkitna et al., 2021; Unger et al., 2014). These tree parameters are biased by the errors of various ITD methods; thus, persuading the most accurate algorithm is still an ongoing challenge, and numerous studies have been conducted by the research community, providing a variety of ITD methods for scientific and operational use (Brosofske et al., 2014; Coops et al., 2021). Technically, the ITD and crown delineation methods can be grouped into two main concepts. (i) The point-based concept detects and segments individual trees directly from a point cloud by clustering points into objects (Duncanson et al., 2014; Li et al., 2012; Yao et al., 2014). (ii) The raster-based concept first utilizes the point-to-raster algorithm, which defines a digitalized square unit (pixel) representation of the forest surface in a CHM (Popescu et al., 2002; Popescu and Wynne, 2004). Then, the treetops are located using the local maxima (LM) algorithm with a fixed or variable window filtering (VWF) size (Pitkänen, 2001; Popescu and Wynne, 2004; Wulder et al., 2000). Finally, marker-controlled watershed segmentation or marker-controlled decision tree segmentation using treetops as markers has been applied for tree boundary (i.e., crown) delineation (Beucher and Meyer, 1993; Dalponte and Coomes, 2016; Lamar et al., 2005; Wang et al., 2004). Essentially, the raster-based concept is a CHM-based method. Traditionally, CHM is computed by subtracting the digital terrain model (DTM) from the digital surface model (DSM). However, using different methods for CHM retrieval results in qualitatively different images; thus, CHM processing should be treated with care (Mielcarek et al., 2018). Therefore, applying triangulation-based algorithms to obtain a CHM is recommended (Roussel et al., 2020). Both point-to-raster and

triangulation-based model generation can result in either empty or erroneous pixels (pits), which can be filled by implementing various types of post-processing filters (Roussel et al., 2020; Stereńczak et al., 2020). In addition, to reduce the number of LM a priori, Gaussian filtering with various kernel window sizes (adjusting the smoothing effect) can be applied to the improved model (Stereńczak et al., 2020). The prepared CHM can finally be subjected to ITD.

An extensive set of ITD methods based on CHM was compared by Eysn et al. (2015) in the case of Central-European alpine forests, and an automated matching procedure was proposed with high accuracy evaluation metrics for ITD (Eysn et al., 2015; Stereńczak et al., 2020). The best results were achieved by the LM search with the VWF (i.e., 60% matching rate for single-story coniferous stand, 47% for single-layered mixed forest of 29% coniferous proportion). Moreover, the detection of dominated trees was found challenging for all the considered methods. Benchmarking studies published previously (Kaartinen et al., 2012; Vauhkonen et al., 2012) have reported similar results. A novel ITD and segmentation approach using a CHM-based hierarchical transformation of height levels from highest to lowest presented accuracies between 83 and 86% in various forest stands (Zhao et al., 2017). A complex, self-calibrating segmentation (without ITD) algorithm, tested in a non-mountainous Central-European forest, achieved accuracies of 85% in coniferous and deciduous stands and 75% in mixed forests (Stereńczak et al., 2020). The two previously mentioned well-performing tools are not licensed in the open-source domain. Hence, there is a need for an ITD approach that (i) is free to use, (ii) outperforms the LM with VWF and can be subjected to various segmentation approaches, and (iii) is transparently and easily developable.

2.2 Crown Base Height Location

Crown Base Height is defined as the vertical distance between the ground surface and the lowest live branch marking the base of the tree canopy (Hermosilla et al., 2014; Luo et al., 2018; Maguya et al., 2015; Næsset and Økland, 2002; Popescu and Zhao, 2008; Stefanidou et al., 2020; Xu et al., 2013). It is one of the most important structural parameters used both on individual trees and forest stand levels for

estimating crown volume (Korhonen et al., 2013; Terryn et al., 2023), evaluating crown conditions (Bianchi et al., 2020) and forest health (Zarnoch et al., 2004), and obtaining canopy fuel parameters (Andersen et al., 2005; Engelstad et al., 2019; Erdody and Moskal, 2010; González-Ferreiro et al., 2013b). Moreover, accurate and efficient CBH estimations are crucial for forest fire prediction simulations (Finney, 2006, 1998; Kelly et al., 2018; Riaño et al., 2003).

Despite its significance, obtaining CBH accurately remains a challenging issue. Traditional field surveys provide reliable and accurate CBH measurements; however, these methods are time-consuming, labor-intensive, and costly (Dean et al., 2009; Luo et al., 2018; Stefanidou et al., 2020). The active remote-sensing technique of LiDAR offers a time- and cost-efficient alternative. Indeed, LiDAR systems have been proven to provide CBH measures at both the individual tree and plot levels (Andersen et al., 2005; Erdody and Moskal, 2010; Luo et al., 2018; Maguya et al., 2015; Stefanidou et al., 2020). Therefore, using high-resolution airborne laser scanning data, the creation and development of such algorithms are assumed to be feasible.

Another important aspect of CBH research is that none of the above-mentioned studies provided any hands-on software that could help forest management, and the research community applied CBH location. Using highly developed programming software, such as R, the gap in the lack of an applicable open-source CBH location tool can be bridged.

Previous research on LiDAR-derived CBH estimation methods can be broadly categorized into two main approaches: studies applying regression analysis (Botequim et al., 2019; Erdody and Moskal, 2010; Hermosilla et al., 2014; Kelly et al., 2018), and direct estimation of CBH (Dean et al., 2009; Luo et al., 2018; Popescu and Zhao, 2008; Vauhkonen et al., 2012). Regression-based predictive models require extensive field data, and their estimation accuracy can be significantly influenced by the quantity and quality of sample plots (Stefanidou et al., 2020). Moreover, these models are often site- and species-specific, limiting their broad applicability (Engelstad et al., 2019).

Direct methods, on the other hand, do not rely on field measurements and enable the mapping of inaccessible areas without reference data (Sumnall et al., 2017). However, these methods depend on an adequate pulse penetration capability, which requires sufficient

points in the lower canopy. This capability is influenced by several interacting factors such as sensor characteristics, canopy cover, steep slopes, and relative air moisture content (Hsu et al., 2015; Sumnall et al., 2016).

While most previous studies have focused on LiDAR-based CBH estimations at the plot level (Andersen et al., 2005; Jakubowksi et al., 2013; Maguya et al., 2015; Næsset and Økland, 2002), research conducted at the individual tree level is relatively scarce (Luo et al., 2018; Popescu and Zhao, 2008; Vauhkonen, 2008). Individual tree-based CBH estimations can provide more precise information on forest structure than plot-level analyses (Luo et al., 2018). However, these methods have primarily been studied in boreal coniferous or conifer-dominated forests (Luo et al., 2018; Popescu and Zhao, 2008; Vauhkonen, 2008). Therefore, developing a direct CBH estimator specifically for deciduous forests at the individual tree level is necessary and highly valuable.

2.3 Tree Stem Location

Terrestrial LiDAR, also known as Terrestrial Laser Scanning (TLS), can scan the environment and produce millions of three-dimensional (3D) points. These points can be used to derive tree metrics, such as location, height, and diameter (Liang et al., 2019, 2018), and to create detailed tree models (Hackenberg et al., 2015; Raumonen et al., 2013).

To perform such analyses, it is crucial to identify each tree within the point cloud. However, challenges, such as inconsistent point cloud quality, diverse forest structures, and complex plant morphology, make it difficult to develop a universal, efficient, and fully automated solution (Calders et al., 2020; Liang et al., 2018; Martin-Ducup et al., 2021; Wilkes et al., 2017).

Most existing individual tree segmentation methods aim to isolate single-tree point clouds from the overall forest point cloud by incorporating external information, such as allometric functions (Burt et al., 2019), user-defined parameters (Trochta et al., 2017), or manual inspection and adjustment of segmentation outcomes (Burt et al., 2019; Krisanski et al., 2021; Raumonen et al., 2013). However, reliance on external information can make these algorithms specific to certain datasets, difficult to generalize, and accessible only to experts with

specialized knowledge. A recent promising study proposed a new method called Topology-based Tree Segmentation (TTS), which can provide the most robust results with non-complex user interactions (Xu et al., 2023). TTS identifies distinctive structures within the forest, such as tree bases and tops, and segments individual trees using a topological framework based on the discrete Morse theory (Forman, 1998).

From a computational perspective, segmenting a forest TLS point cloud into individual tree point clouds, known as individual tree segmentation or isolation, is a labor-intensive and time-consuming process, particularly in large areas (Burt et al., 2019; Calders et al., 2020; Krisanski et al., 2021). Although automation is required, few algorithms fully automate this task. Most existing methods use a bottom-up approach, beginning with tree bases, leveraging the fact that TLS point clouds are captured from the ground, providing detailed stem data, and making tree bases reliable starting points for segmentation (Martin-Ducup et al., 2021; Xu et al., 2023).

For example, Trochta et al. (2017) introduced a bottom-up segmentation algorithm, in which points are clustered in horizontal slices parallel to the ground. Tree bases were identified as clusters near the ground and individual trees were constructed by merging these clusters with nearby stems. However, this method requires a significant number of user-defined parameters to guide the process, often necessitating manual fine-tuning to achieve accurate results.

Incorporating ecological knowledge, such as metabolic scaling theory (West et al., 1997), can improve tree segmentation results. In two similar studies, a graph of points was created, and individual tree point clouds were formed by assigning points to the closest located stems based on the shortest path, with distances scaled according to metabolic theory (Tao et al., 2015; Wang, 2020). While this theory offers a universal framework for tree growth, its application requires the consideration of factors such as demographic traits and light conditions. Additionally, these methods involve user-defined parameters, making them time-consuming to learn and apply effectively. Segmentation methods based on the allometric relationships between tree metrics are another approach. Burt et al. (2019) developed a method called *treeseg* that uses a Point Cloud Library (PCL) (Rusu and Cousins, 2011). After identifying tree stems, *treeseg* applies allometric relationships, such as those between stem diameter, tree height, and crown extent, to

segment the tree crowns. However, the method's reliance on hardcoded parameters limits its adaptability to different datasets.

Deep learning techniques have become increasingly popular for analyzing TLS point clouds. Krisanski et al. (2021) created the Forest Structural Complexity Tool (*FSCT*), which uses PointNet++ (Qi et al., 2017) to classify points and clustering algorithms such as Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and Density-Based Spatial Clustering (DBSCAN) to group stem points (Campello et al., 2015). RANSAC was then applied to fit cylinders representing tree structures, with points assigned to the nearest cylinder to form individual tree clouds. While effective, *FSCT*'s deep learning approach demands significant computational resources, making it challenging for large datasets. Other promising methods also use deep learning to extract tree features from point clouds by converting them into 2D images or voxel representations (Chang et al., 2022; Wang et al., 2019; Xi and Hopkinson, 2021). However, these deep learning techniques require extensive annotated data for training, which can be difficult to gather in practice. The Topology-based Tree Segmentation (TTS) method has demonstrated superior robustness in comparative analyses (Xu et al., 2023). The discrete Morse theory-based TTS outperformed both the *FSCT* and *treeseg* across a range of forest environments.

2.4 Decomplexifying the Discrete Morse method

2.4.1 Simplical complexes and the α-complex

We used simplicial complexes to deduce the topological structure of the 3D point cloud based on the connectivity. Formally, a *k-simplex* σ is defined as the convex hull of *k*+1 affinely independent points in Euclidean space. *k* is the dimension of σ, meaning that a *0-simplex* is a point, a *1-simplex* is a line segment, a *2-simplex* is a triangle, and a *3-simplex* is a tetrahedron. We represent a *k-simplex* spanned by the vertices $v_0, v_1, \dots, v_k$ as $\sigma = \{v_0, v_1, \dots, v_k\}$. Any simplex β that is the convex hull of a non-empty subset of the points generating a simplex σ, is considered a face of σ.

A *simplicial complex* Σ is a finite collection of simplexes with the following properties: (i) every face of a simplex in Σ is also included in Σ;

and (ii) for each pair of simplexes σ and β, either $\sigma \cap \beta \neq 0$ or $\sigma \cap \beta$ is a face of both.

An α-*complex* is a specific type of *simplicial complex* constructed from the simplexes of a Delaunay tetrahedralization Σ^T. Let σ be a simplex in Σ^T, and let its circumcircle C_σ have a radius *r*. The α-*complex* (Edelsbrunner, 2012) $\Sigma^\alpha$is a subcomplex of the Σ^T containing all vertices of Σ^T, along with all simplexes σ such that: (i) $r < \alpha$ and C_σ does not contain any points, or (ii) σ is a face of $\beta \in \Sigma^\alpha$.

2.4.2 Discrete Morse Theory

The discrete Morse theory (Forman, 1998) is a combinatorial version of the Morse theory that enables the analysis of the topology of *simplicial complexes*. In the context of a *simplicial complex* Σ, a *discrete vector* is defined as a pair of simplices (σ, β) where one simplex is a face of the other. Pairs are formed by a *1-simplex* (edge) and a *2-simplex* (triangle) and by an edge and a vertex (*0-simplex*). A *discrete vector field V* consists of such pairs, with the condition that each simplex appears in at most one pair. Simplices that do not appear in any pair are referred to as critical ones.

For a triangle mesh:

Critical edges are considered as saddles.

Critical vertices are considered minima.

Critical triangles are considered the maxima.

A V-path is a sequence of simplices (σ_2, σ_1, σ_0) where σ_1 is a face of σ_2 and σ_0 is a face of σ_1. A V-path is termed closed if σ_0 is a facet of σ_2 and σ_0 is different from σ_2. A discrete vector field is known as a Forman gradient if it contains no closed V-paths. In a triangular mesh, there are two types of separatrix path. A separatrix edge path connects a critical edge to a critical vertex, and a separatrix triangle path connects a critical triangle to a critical edge.

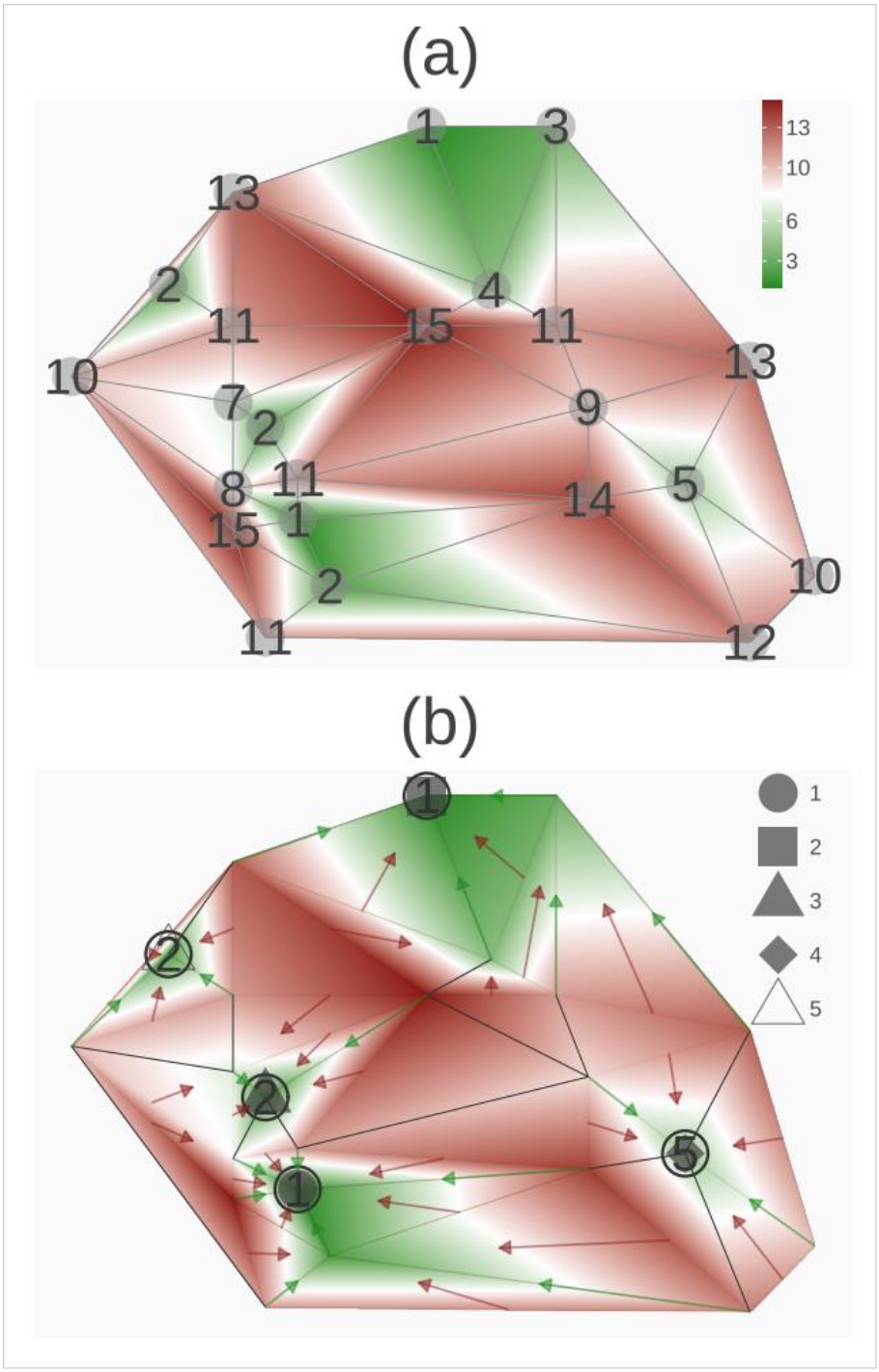


Figure 1. (a) Scalar function, defined on a triangle mesh, colour coded by means of a green-red divergent colour map. Values of vertices are also displayed. (b) Forman gradient computed on the same scalar field. Arrows indicate gradient pairs. Black circles indicate the five unique (represented by distinct shapes) critical minima (vertices); black lines indicate saddles (edges).

When the discrete gradient is derived from scalar function *f*, it acts as a combinatorial representation of the gradient and its critical points. For instance, in Figure 1(a), a scalar function is defined on a triangular mesh. Figure 1(b) illustrates this gradient, where the arrows represent the gradient pairs. The arrows align with the direction of the gradient and indicate where the function value decreases. The Forman gradient naturally segments the simplicial complex into regions influenced by the critical points of the function (De Floriani et al., 2015). Regarding critical simplices, the five unique minima are highlighted by black circles, whereas the saddles are highlighted by black lines. Note that the critical triangles (without arrows) are not highlighted because we planned to use the edge-path separatrix, which considers minima and saddles only.

3 MATERIALS AND METHODS

3.1 Study areas

Interrelated studies were conducted in three study areas: Hardtwald Forest (Karlsruhe), the Bretten municipal forest (both belonging to the federal state of Baden-Württemberg, Germany), and the Nagyerdő recreational forest in Debrecen (federal state of Hajdú-Bihar, Hungary) (Fig. 2). While the two Karlsruhe (KA09 und KA10) and the two Nagyerdő (A and B, Fig. 3) forest sites lie on flat terrain, the two Bretten (BR01 and BR05) forest plots are characterized by hilly landscape.

The managed German forest stands consist of the following main tree species: Norway spruce (*Picea abies* (L.) H. Karst), Scots Pine (*Pinus sylvestris* L.), Douglas fir (*Pseudotsuga menziesii* (Mirb.) Franco), common oak (*Quercus robur* L.), red oak (*Quercus rubra* L.), sessile oak (*Quercus petraea* (Matt.) Liebl.), European beech (*Fagus sylvatica* L.), and European hornbeams (*Carpinus betulus* L.). Scot pine, red oak, and European beech were the dominant species in KA09 and KA10. More diverse tree species compositions with spruce, Douglas fir, European beech, oaks, and European hornbeam characterize the BR01 and BR02 plots. The mixed species composition, dense canopy cover, and multiple layers describe the four forest stands. Detailed information on the plot characteristics can be obtained from Weiser et al. (2022).

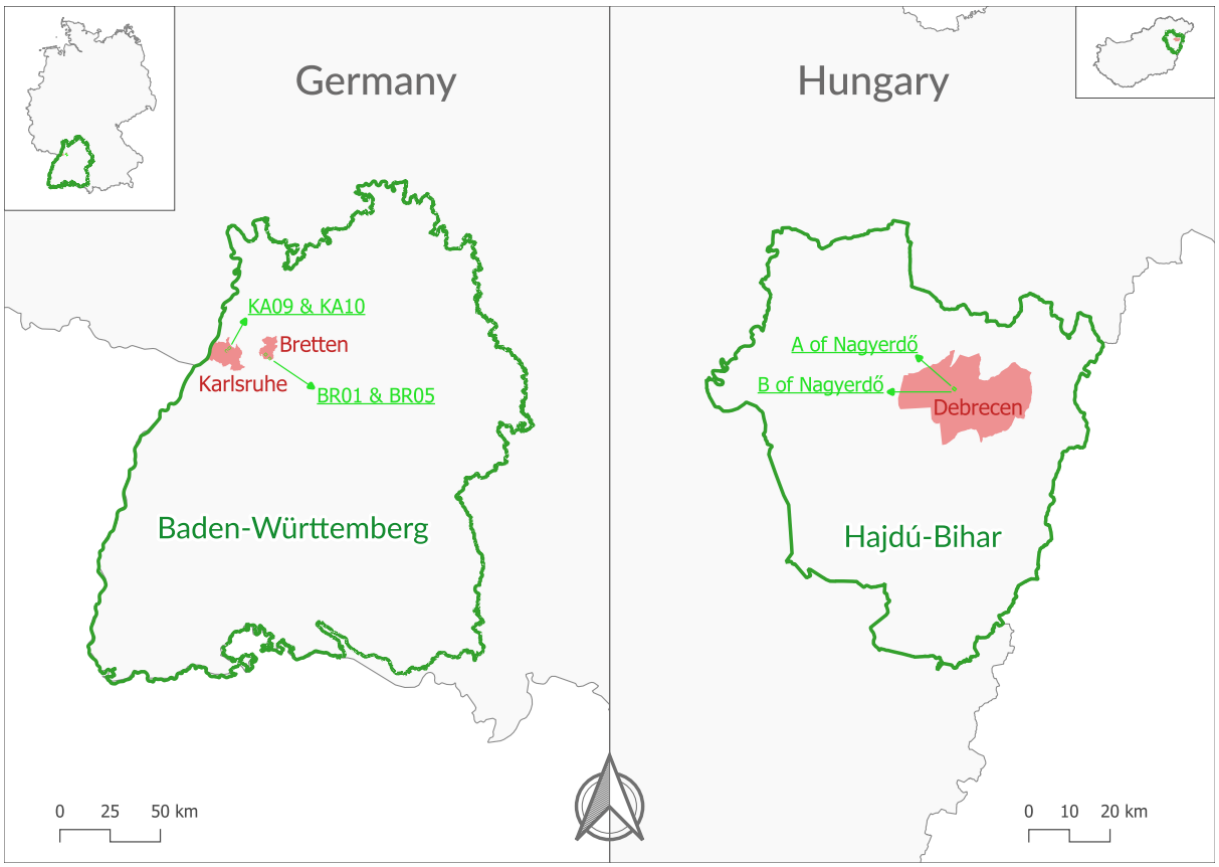


Figure 2. Locations of the four German forest plots (left) in the federal state of Baden-Württemberg (green border, EPSG:25832) and Hungarian sites A and B of the Nagyerdő forest (right) in the federal state of Hajdú-Bihar (green boundary, EPSG:23700).

The two one-hectare study sites of the Nagyerdő forest (Fig. 3) accommodate Austrian pine (*Pinus nigra* J. F. Arnold), Scots pine, common hackberry (*Celtis occidentalis* L.), common oak, red oak, silver poplar (*Populus alba* L.) and eastern American black walnut (*Juglans nigra* L.) tree species. Both sites were very dense and characterized by a mixed species composition. While site A shows mostly single layer canopy structure with deciduous-coniferous ratio of 0.8:0.2, site B is multi layered and more deciduous-dominated (0.9:0.1).

While the development and testing processes of the GTR-based (from here on referred to as GTR) treetop location algorithm was conducted in four German forests (KA09, KA10, BR01, and BR05) and the A forest of Nagyerdő, the CBH locator was created and evaluated on sites A and B of Nagyerdő and in the KA09, KA10, BR01, and BR05

German forests (Fig. 3).

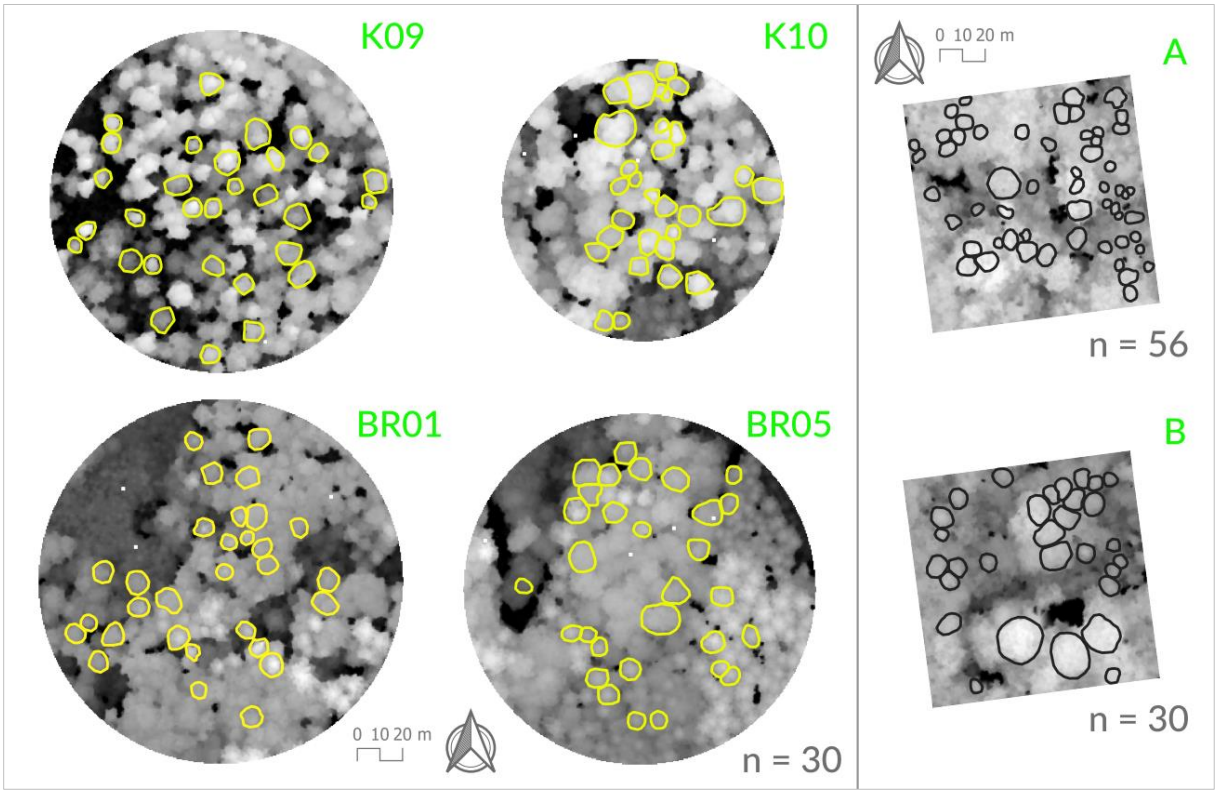


Figure 3. Locations of the four German plots (KA09, KA10, BR01, and BR05) along with the two one-hectare sites (A and B) of the Nagyerdő mixed forest plantation. The number of tree segments utilized as input for *treecbh* is indicated by n=, represented by black (Nagyerdő) and yellow-bordered (German forests) polygons.

Our TLS forest site, called Wytham Woods, is located in Oxfordshire County, United Kingdom (Fig. 4). It is managed by Oxford University (https://www.wythamwoods.ox.ac.uk/home). The dominant tree species in Wytham Woods are European Ash (*Fraxinus excelsior*), sycamor (*Acer pseudoplatanus*), and common hazel (*Corylus avellana*). We selected 30 trees using a random sampling method from the available individual-tree-based TLS data of Wytham Woods (Calders et al., 2022). These data served as inputs for the point-density sensitivity analysis (described in Section 3.5).

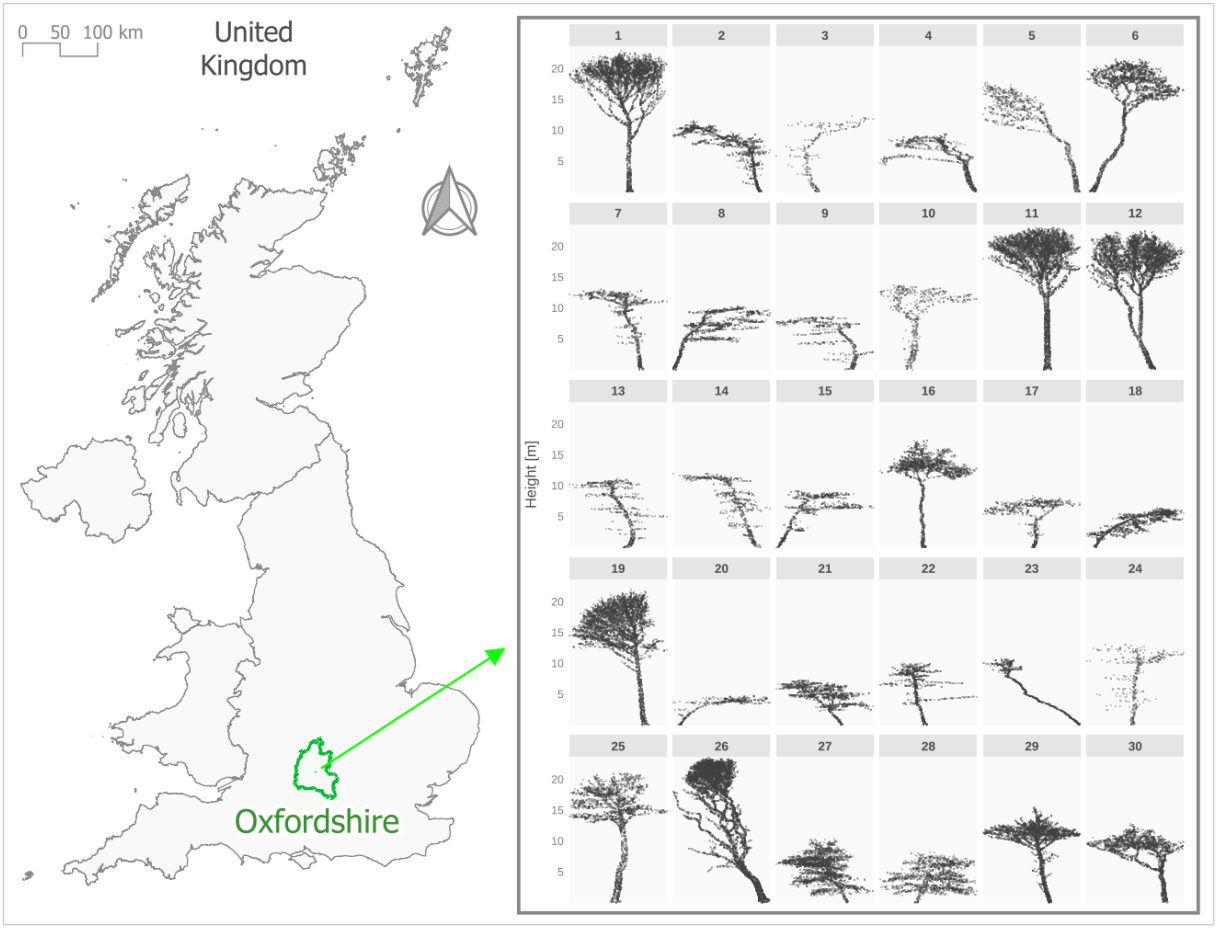


Figure 4. Location of the TLS plot, Wytham Woods in Oxfordshire (green border, EPSG:27700) and the 30 randomly selected trees displayed in 2D.

3.2 LiDAR data

This section describes the three types of LiDAR data used in this study. An overview of the datasets related to the corresponding research topics and areas is presented in Table 1.

Table 1. LiDAR datasets used in this dissertation.

Study area	Study	LiDAR data
KA09	ITD, CBH	ALS*
KA10	ITD	ALS*
BR01	ITD, CBH	ALS*
BR05	ITD, CBH	ALS*

A	ITD, CBH	ALS*, ALS**
B	CBH	ALS**
Wytham Woods	CBH, Tree stem location	TLS***

* Table 2
** Table 3
*** Table 4

Because point cloud characteristics are quantitively different from one another, it is clearer and more understandable to present them according to the two-study approach.

3.2.1 Individual Treetop Location

LiDAR acquisition for KA09, KA10, BR01, and BR05 was performed using the RIEGL VQ-780i (RIEGL Laser Measurements Systems, 2019) sensor. The flight was performed on July 5, 2019, by scanning the respective forests multiple times following the general point density generating flight process of overlapping flight lines, as for Nagyerdő, the LiDAR data was acquired on May 22, 2020, using the RIEGL VQ780ii system with one scanning flight. For a detailed description of German LiDAR acquisition, see Weiser et al. (2022). LiDAR data characteristics of the five study areas are summarized in Table 2.

Table 2. Characteristics of LiDAR data used for developing the GTR treetop locator algorithm.

Characteristics	KA09, KA10, BR01 and BR05	Nagyerdő site A
ALS sensor	RIEGL VQ-780i	RIEGL VQ-780ii
Flying altitude AGL	650 m	1077 m
Off-nadir scan angle	±30°	±30°
Laser beam divergence	0.25 mrad	0.18 mrad
Pulse repetition frequency	1000 kHz	2000 kHz
Flight line distance	175 m	24 m
Flight line overlap	76 %	20 %
Point density	>40 pts/m^2 (max: 103 pts/m^2)	>2 pts/m^2 (mean: 10 pts/m^2)

The most significant difference between the German and Hungarian ALS data is the quantitative comparison of their point densities (Table 2). The German data can be labelled as medium-, whereas the Hungarian data as low-resolution LiDAR data.

3.2.2 Crown Base Height Location

Acquisition flight for the German forests was performed using a RIEGL VQ-780i (RIEGL Laser Measurements Systems, 2019) sensor. The flight was performed on July 5, 2019, as for Nagyerdő, the ALS data was acquired on March 03, 2023, using the RIEGL VQ780ii system. For a detailed description of the German ALS acquisition, see Weiser et al. (2022). ALS data characteristics of the five study areas are summarized in Table 3.

Table 3. ALS data and their derived point-cloud technical characteristics were used in this study.

Characteristics	BR01, BR05 and KA09			A and B	
ALS sensor	RIEGL VQ-780i			RIEGL VQ-780ii	
Flying altitude AGL	650 m			664 m	
Off-nadir scan angle	±30°			±30°	
Laser beam divergence	0.25 mrad			0.18 mrad	
Pulse repetition frequency	1000 kHz			2000 kHz	
Flight line distance	175 m			210 m	
Flight line overlap	76 %			60 %	
Study area	BR01	BR05	KA09	A	B
Input point clouds' mean point density (pts/m^2)	204	172	186	162	217

Consecutive laser scanning was conducted four times over the Nagyerdő forest. Concerning deciduous species, the acquisition date was after the end of the tree dormancy period, depicting a scenario that

can be best described as between leaf-off and leaf-on (with young leaves growing), enabling very high laser penetration rates (through a non-closed canopy layer). Therefore, the achieved point density ranges from 180 to 230 pts/m^2 within the two considered forest sites. Using such high-resolution aerial laser scanning data facilitated the creation of the CBH location tool.

TLS data were collected under leaf-off conditions from late November 2015 to January 2016. All scans were performed on windless days. A summary of the main TLS characteristics is provided in Table 4. For more information, please refer to the study by Calders et al. (2022).

Table 4. TLS data characteristics.

Characteristics	
TLS	RIEGL VZ-400
Minimum range	0.5 m
Maximum range	350 m
Laser beam divergence	0.35 mrad
Pulse repetition frequency	300 kHz
Angular sampling resolution	0.04°
Outgoing pulses per scan	22500000
Beam diameter	2.45 cm
Input point clouds' mean point density (Fig. 4)	4000 pts/m^2 -> 111 pts/m^2*

* 111 pts/m^2 was the mean point density of the finalised input point clouds

3.3 LiDAR data preprocessing

This section presents the previously implemented three-study approach.

3.3.1 Individual Treetop Location

Three major pre-processing steps were conducted using the lidR package (Roussel et al., 2020) in the R computational environment (R Core Team, 2022). First, raw LiDAR point data were classified into ground and surface points using cloth simulation filtering (CSF) (Zhang et al., 2016). Next, the CHMs were derived directly from the classified

above-ground and ground points using a pit-free algorithm (Khosravipour et al., 2014). Test sites BR01, BR05, and KA10 showed two, one, and three pits, respectively. These were non-influential on the ITD and were therefore not subjected to further processing. Finally, Gaussian filtering with a window size of 3 × 3 pixels was applied for image smoothing and noise removal.

3.3.2 Crown Base Height Location

3.3.2.1 ALS

In each of the German forest plots, 30 tree segments were manually drawn using ALS-derived vertical profiles (Fig. 2) after treetop location using the GTR-based treetop locator. For Hungarian forest site A, 56 tree segments were drawn (Fig. 2). These segments were carefully created by applying the same GTR-based treetop location, combined with the manual tree delineation method described for the German plots. Once the reference trees were segmented, the ground-classified point cloud of the forest was extracted for each segment. To normalize the heights within each segment, a normalization process was applied individually to each point within the segments using following formula (Eq. 1):

$$\text{norm\_height}_{i:s} = \text{height}_{i:s} - \text{mean}(\sum \text{height}_{i:g});$$

$$i = 1 \ldots 56,\ s = 1 \ldots \infty,\ g = 1 \ldots \infty, \qquad \text{(Eq. 1)}$$

whose application is justified by the low mean error (Fig. 5), which expresses the mean height deviation of the ground points for the 56 tree segments computed by Eq. 2:

$$\mu_{error} = \sum((\max(\text{height}_{i:g}) - \min(\text{height}_{i:g})) / 2) / \max(i) \qquad \text{(Eq. 2)}$$

where *i* denotes the *i*-th ITS containing the binary classified points *s*, *surface*, and *g*, *ground* labelled.

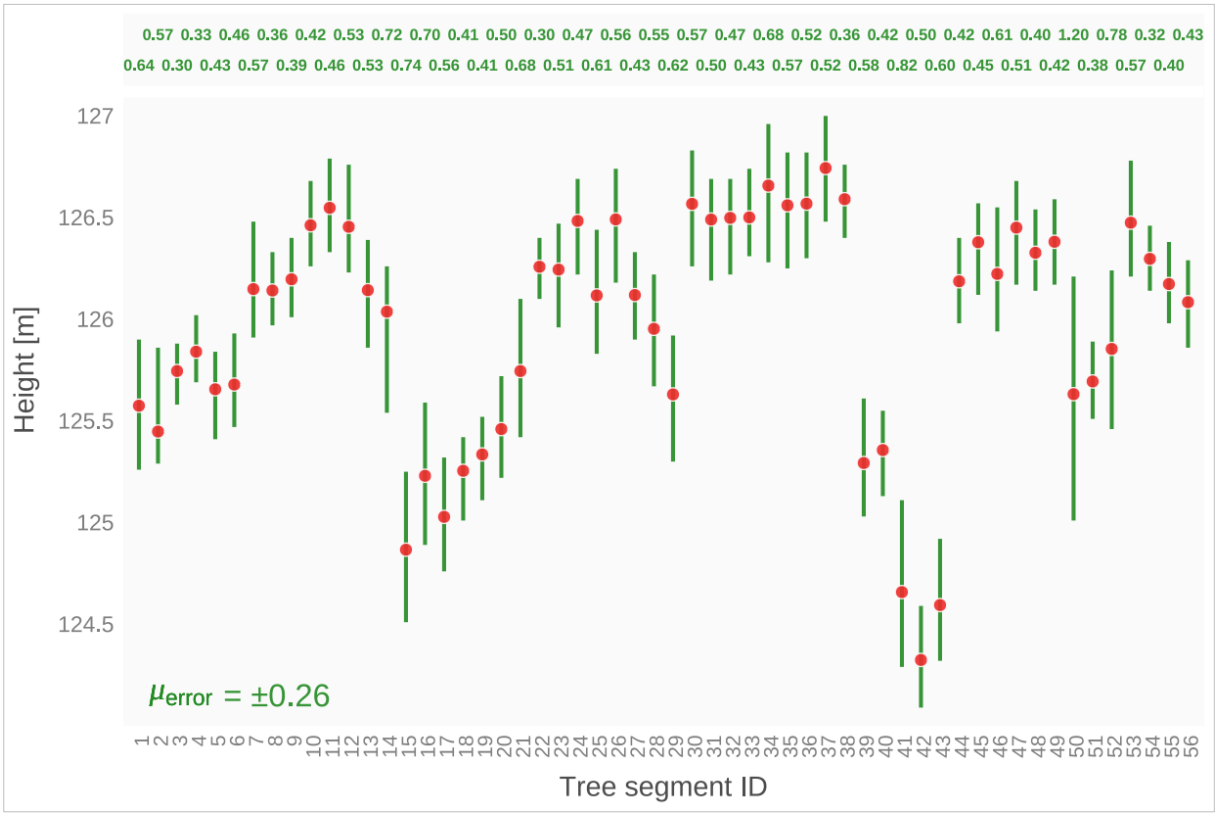


Figure 5. Ground point ranges within the 56 individual tree segments of Nagyerdő forest A indicated by green vertical lines and corresponding green labels, while the respective means are shown by red dots.

Regarding the structurally more complex forest B, individual reference tree segments were drawn around the 30 highest GTR-based located coordinative treetops (Fig. 3 and Fig. 6). The point cloud extraction process (for the 30 segments) involved visual validation of the extracted vertical profiles derived from the LiDAR data. Expert biologists checked and confirmed the extracted point clouds in situ for each point cloud (i.e., individual tree).

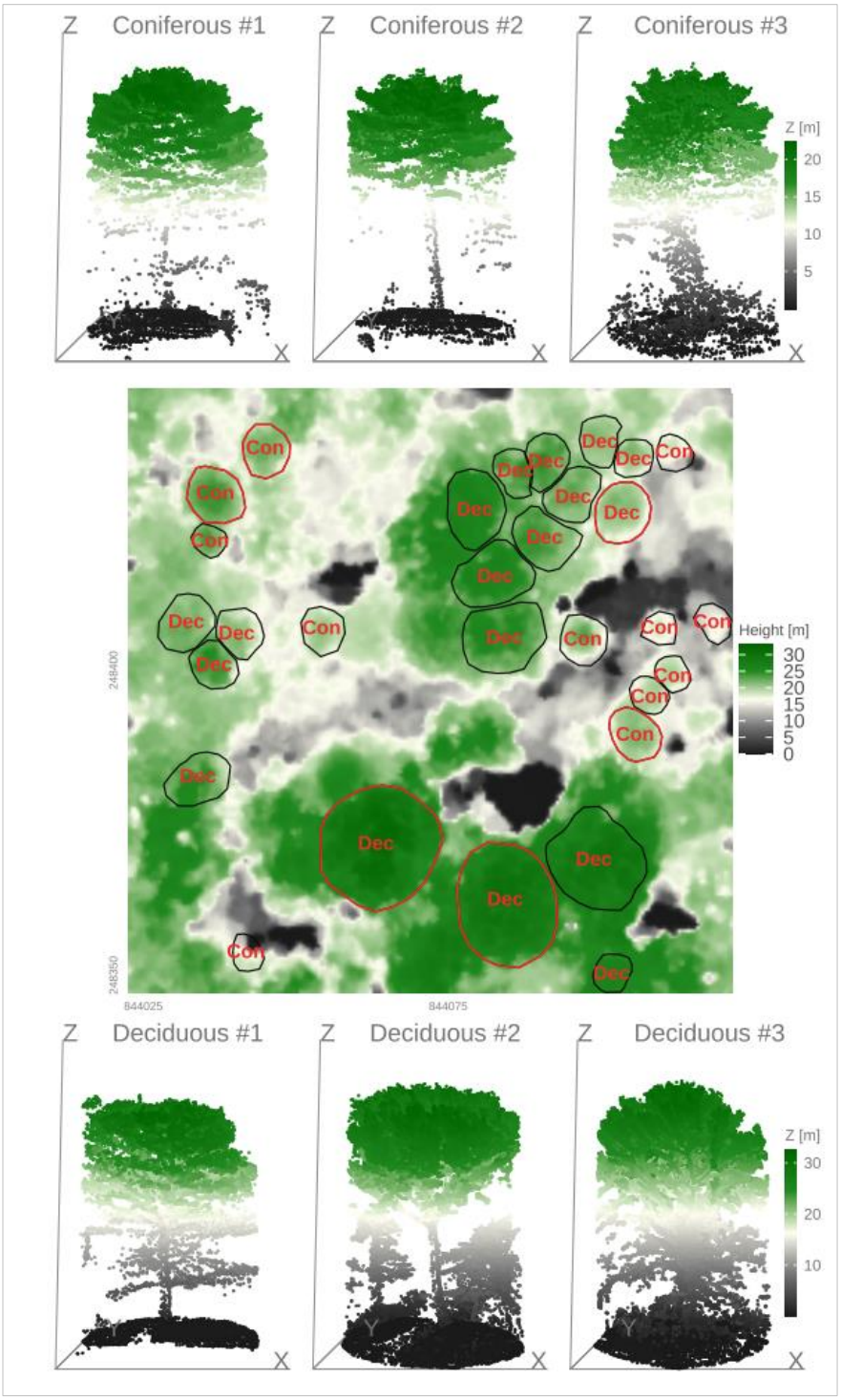


Figure 6. Reference individual tree segments in site B (black). The 3-3 deciduous and coniferous example segments are marked in red (examples derived from the 3D point clouds of the LiDAR data.

After point cloud extraction from the ALS data (Bretten, Karlsruhe, and Debrecen) to the corresponding tree segments, the obtained tree point clouds were used as input data for *treecbh*. Accordingly, we term this data as "input point cloud." The characteristics of the input point cloud are listed in Table 5.

Table 5. Input point cloud characteristics.

Study area	Number of input point cloud	Mean height (m)	Conifer-broadleaf ratio*
A	56	23	0.2:0.8
B	30	24	0.1:0.9
BR01	30	33	0.4:0.6
BR05	30	35	0.6:0.4
KA09	30	28	0.6:0.4

* calculated from inventory data

3.3.2.2 TLS and the point density sensitivity

Thirty trees were selected using a random sampling method from available individual-tree-based TLS data from Wytham Woods (Calders et al., 2022). These data served as inputs for the point density sensitivity analysis.

The point density of the 30 input tree point clouds was randomly decimated using the *decimate_points()* function from the lidR package (Roussel et al., 2020). This function employs a sampling algorithm that randomly samples or removes points from the area of a tree point cloud to obtain the desired point density (pts/m^2). The following six point density classes were defined: 20, 53, 111, 156, 205 and 253 pts/m^2.

The analysis was conducted by executing *treecbh* in its CBH-locate-only mode (*cbh_ONLY* = 3 in Table 6, as described in Section 4.3.2). Accordingly, the selected input point clouds of the Wytham

Woods TLS data were processed using the CBH locator considering six-point density classes (Fig. 7).

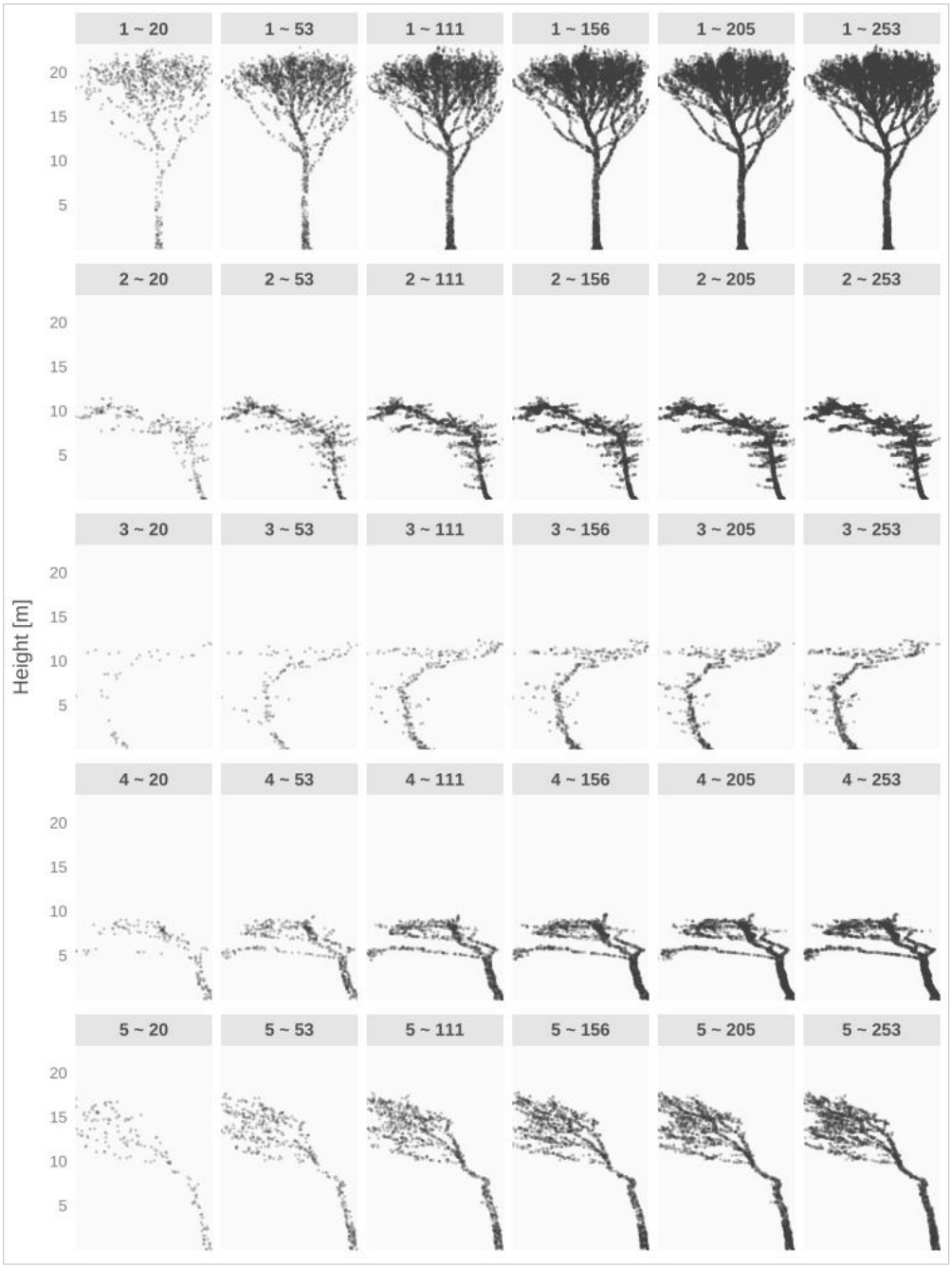


Figure 7. The first five tree point clouds (Fig. 4) from the selected Wytham Woods TLS data are displayed with respect to the six point density classes used for the point density sensitivity analysis: 20, 53, 111, 156, 205 and 253 pts/m^2.

3.3.2.3 TLS data for the Tree Stem Location

From the Wytham TLS data (Table 1 and 4), 12 individual closely spaced adjacent trees were selected, forming a clustered point cloud (Fig. 8). Although understorey vegetation was removed by default, the close proximity of these trees allowed the thin stems to be considered as understory. This factor plays a crucial role in the development of stem isolation.

Figure 8. The 12 selected trees from the Wytham Forest TLS dataset, forming the test TLS point cloud. Seven adjacent trunks, considered understory vegetation, are highlighted within a red box.

3.4 Reference data

The major aim of these interrelated studies was to develop a robust CBH locator using a treetop locator for deciduous- and deciduous-dominated forests. Nagyerdő forest site B was specifically chosen because of its more pronounced vertical structural complexity compared to Nagyerdő

site A and the four German forests. This consideration was also reflected in the scientific analyses of the study sites.

For the individual treetop location, forest A was designated as the “target” forest, meaning that the treetop locator algorithm was specifically developed for Nagyerdő forest. The four German forests were used to evaluate the performance of the treetop locator, and are therefore referred to as “test” forests. To develop the Crown Base Height (CBH) identification algorithm (second study), both Nagyerdő sites (A and B) were utilized. Forest plots A and B as well as three German sites were used for validation purposes, and the preparation of the reference data was significantly influenced by fundamental technical differences between the two algorithms. While the GTR algorithm required prior point interpolation, as it is entirely CHM-based, the crown base height locator algorithm worked directly with the point cloud. The dependence of the CBH locator on any form of individual tree segmentation (ITS) does not alter this fundamental difference. Consequently, this section outlines the processes of reference data preparation according to the two-study method.

3.4.1 Individual Treetop Location

Although thorough field measurements were conducted in German forests, tree heights could not be measured in dense forest stands because the treetops were barely visible (Weiser et al., 2022). Tree stem positions were determined using DTMs derived from ALS point clouds combined with TLS or unmanned laser-scanning (ULS) point clouds (Weiser et al., 2022). Considering the Nagyerdő forest site A, 60 reference stem coordinates were collected using Stonex S9i RTK GNSS (https://www.stonex.it/project/s9i-gnss-receiver/) system. The acquired tree positions included 15 repeated measurements at each tree, whose coordinates were finally calculated by averaging the results of the repeated measurements.

As described previously, forest A was considered the target site, and the four German forests were considered test sites. Because German forests consist of multiple layers and raster-based ITD are not capable of capturing the treetops of suppressed or under-canopy trees (Eysn et al., 2015), CHM-derived forest height distributions were visualized to understand the similarities and differences between the

target and the four test sites (Fig. 9). The differences between the two highest peaks of these distributions were calculated to define CHM-based layering of the respective forests. While Nagyerdő presented a height difference of 1.93 meters, suggesting a single-layer forest, the German sites revealed height differences ranging from 3.42 meters to 12.17 meters with multiple peaks (Fig. 9). Overall, tree heights in the Nagyerdő forest were considerably lower. Moreover, the distributional line segments of BR01 and BR05 showed similarities over their 20-meter ranges in the Nagyerdő height profile (Fig. 9). This phenomenon suggests that the CHM-induced one-to-one comparability between the trees of BR01 and BR05 and the trees of Nagyerdő is only meaningful for trees, especially of the BR01 and BR05 forests, which are higher than 20 m. Therefore, the reference trees of the four German forests were filtered, and only trees higher than 20 m were considered in this study.

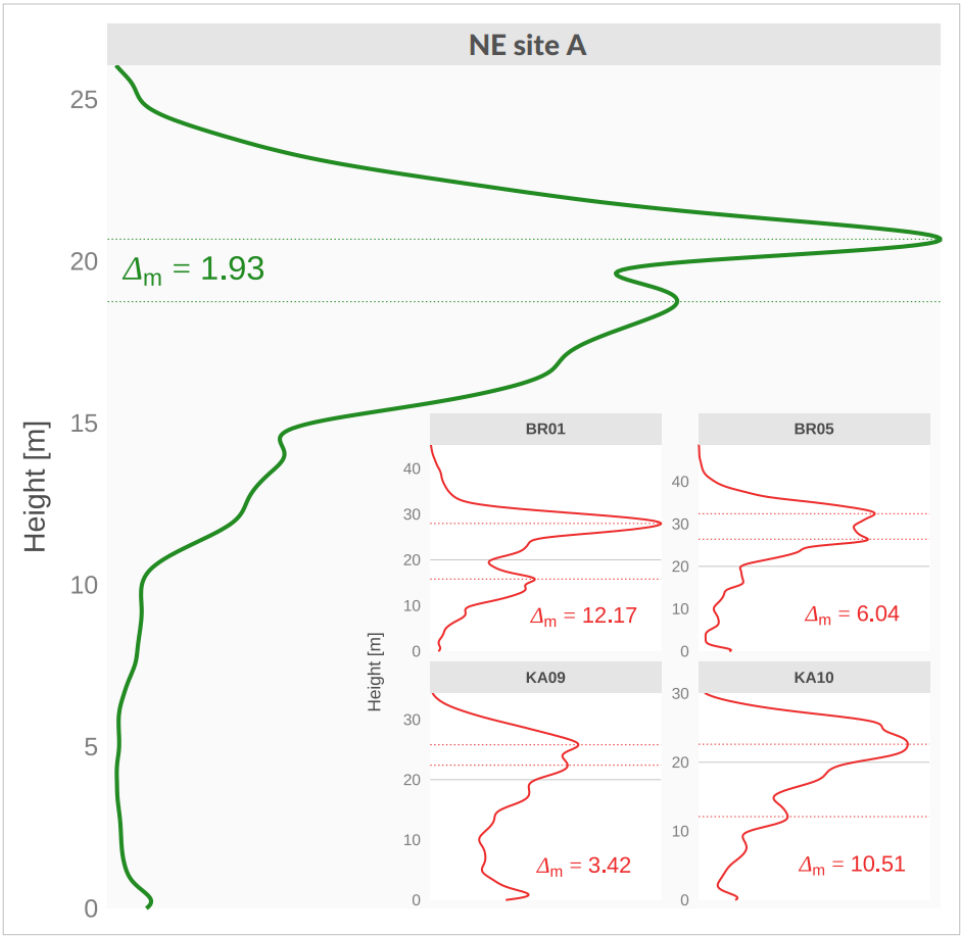
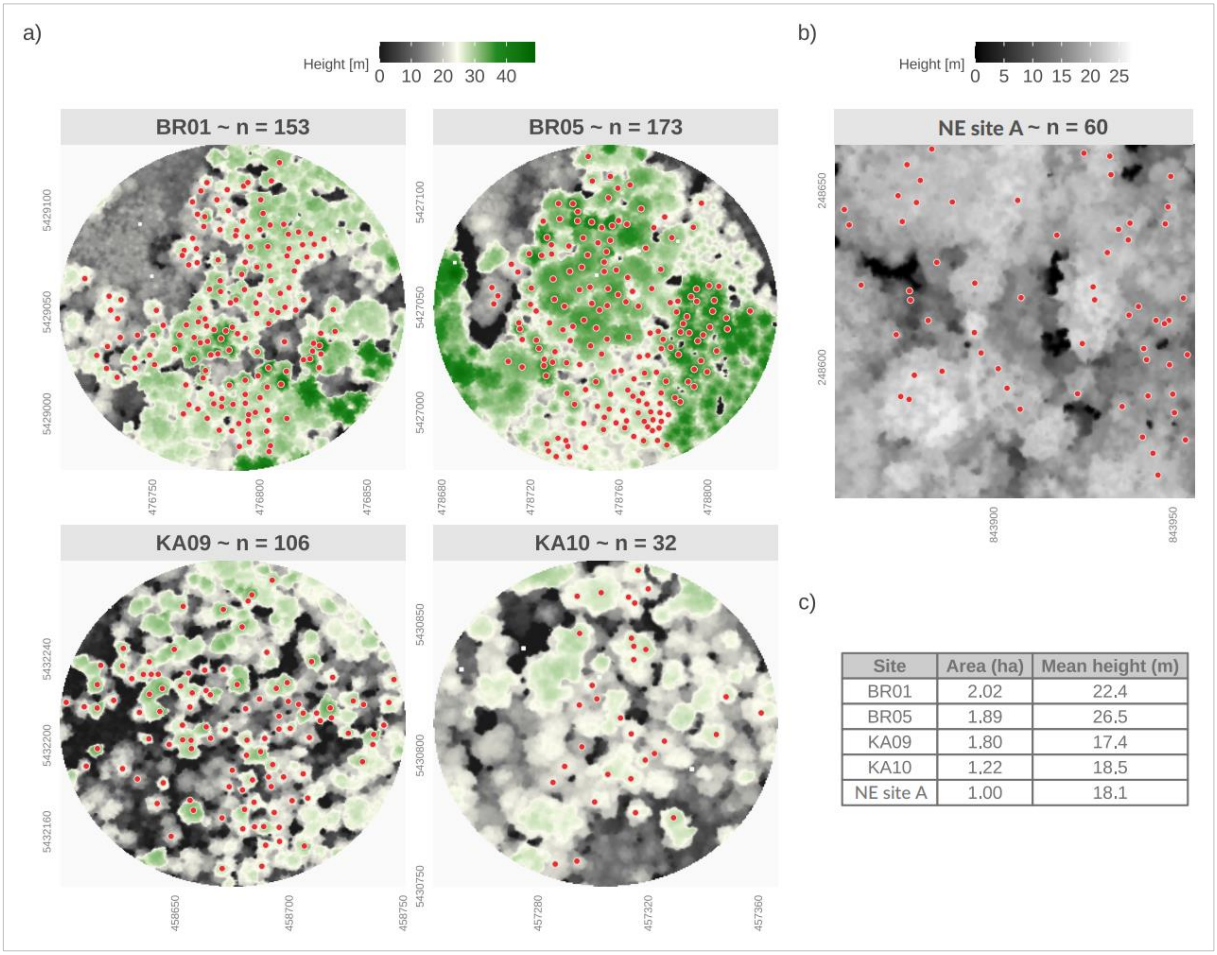


Figure 9. CHM derived the height distribution of the target forest site (Nagyerdő site A, abbreviated as NE site A). Height distributions of the Bretten and the Karlsruhe test sites are shown in the inset graphs. Horizontal dashed lines indicate the highest two peaks of each sites, differences between them are symbolized with Δ_m. The grey horizontal lines display the 20 m threshold in the BR01, BR05, KA09, and KA10 inset graphs.

After subset analysis, the reference trees of the five study sites, BR01, BR05, KA09, KA10, and A, had 153, 173, 106, 32 (Fig. 10a), and 60 samples (Fig. 10b), respectively. Regarding area of the test forest, the BR01 has 2.02 and the BR05 1.89 ha, while the KA09 and KA10 are left with 1.80 and 1.22 ha (Fig. 10c). Our target site, A is the only one ha square area because (i) it (along with the other one ha square Nagyerdő B) was considered for future research (i.e., in the second study) and was used accordingly in the second study, and (ii) circular cuts from CHMs around the German reference trees ensured that the test area

boundaries are near to the test reference trees, making the areas of the test forests as small as possible. The size of the test forest is relevant in the computational speed of ITD.



Site	Area (ha)	Mean height (m)
BR01	2.02	22.4
BR05	1.89	26.5
KA09	1.80	17.4
KA10	1.22	18.5
NE site A	1.00	18.1

Figure 10. Study sites and reference tree coordinates of the four German forests (a), of the A site Nagyerdő forest (b). Site-specific characteristics in tabular form (c). The mean heights were calculated from the respective CHMs.

The selection of these German forests for testing was also motivated by the similarity between their tree species composition and that of Nagyerdő forest. Therefore, the ratios of the reference tree species and coniferous-to-deciduous ratios at each site were plotted and compared visually (Fig. 11). From the ALS perspective, the ratio of coniferous to deciduous trees turned to be of primary interest (Eysn et al., 2015; Stereńczak et al., 2020; Zhao et al., 2017). The reference trees of BR01

and KA10 showed high similarity, those of BR05 showed low similarity, and those of KA09 forest were dissimilar to the Nagyerdő reference trees (Fig. 11c).

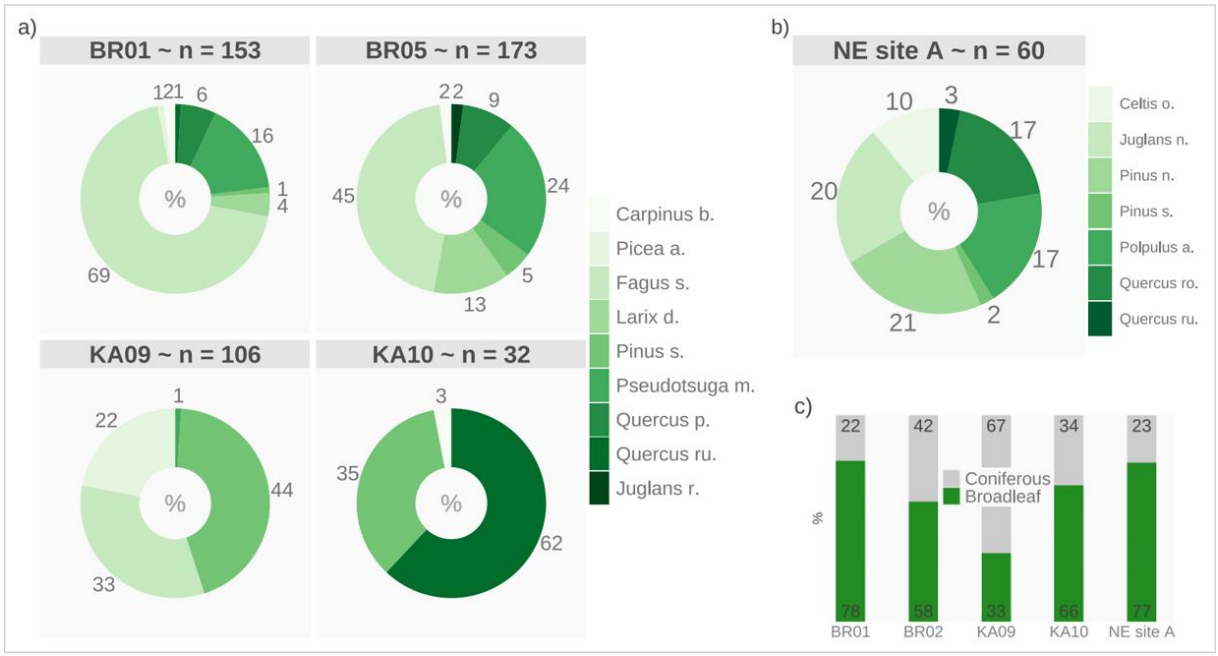


Figure 11. Three species distributions (%) for the reference tree locations of the four German forests (a), of the Nagyerdő forest site A (b); (c) coniferous and broadleaf tree species distribution of the respective forest plots.

3.4.2 Crown Base Height Location

CBHs of the trees were measured in Bretten and Karlsruhe-Hardtwald forests using a Haglöf Vertex-IV hypsometer during exhaustive field measurements (Weiser et al., 2022). The CBH was defined as the height of the lowest branch, with a minimum length of 1 m.

Regarding the Nagyerdő, reference CBHs were collected through a field survey of 86 input point clouds (outlined in Section 3.3.2) in forests A and B. The heights of the first leaved branches were measured in the field using the Haglöf EC II D-R electronic tool (https://haglofsweden.com/project/ec-ii-d-r). Trees were marked at 2 m with a measuring tape to verify the accuracy of the tool and ensure the precision and reliability of the measurements.

CBH was defined as the height of the first leaved branch protruding from the stem. The measured CBH value was determined

based on the reading displayed by the EC II D-R tool when aimed at a precise spot on the tree stem where the first leaved branch initiated its protrusion. It is important to note that the measurement of the reference CBH did not account for the width or direction (downward and upward) of the respective tree branch, and the reference CBHs for the input point clouds of the TLS data were created through visual interpretation because the very high point density of the data (Table 4 in Section 3.2.2) enabled obtaining CBHs accurate field-measured data was available for this dataset.

3.5 The "LowerStar" algorithm

Various approaches have been developed to compute discrete gradients on quadratic and triangular grids and simplicial complexes (De Floriani et al., 2015; Gyulassy et al., 2008; Robins et al., 2011; Shivashankar et al., 2012; Weiss et al., 2013). One of the most proficient and easily understandable computational algorithm to implement is termed as "Lower star" (Robins et al., 2011).

The algorithm processes the lower star of vertices within a *simplicial complex* Σ. The lower star of a vertex denoted as σ (i.e., *0-simlpex*) in Σ, with a scalar function *f* defined on the vertices of Σ, is the set of all simplices ($\beta_1, \beta_2, \ldots, \beta_n$) containing σ, where the scalar value (*v*) at σ is greater than or equal to the scalar value at any other vertex in Σ. Formally, the lower star set of σ is defined by Equation 3:

$$\textit{LowerStar}(\sigma) = \{\beta_n \in \Sigma \mid \sigma \in \beta_n \text{ and } max_{v \in \beta_n} f(v) \leq f(\sigma)\} \qquad \text{(Eq. 3)}$$

Figure 12(a) illustrates a triangular mesh with a scalar function coloured in green-red. A randomly selected vertex (with a value of 11) and its simplicial complex containing its neighbouring simplices are shown in Figure 12(b). The lower star calculation eliminates simplices (vertices, edges, and triangles) that are smaller than or equal to our selected vertex with a value of 11. The lower star of the vertex with value 11 consists of vertices with values 1,2, 8, and 9, their connective edges, and triangles 11-1-8 and 2-11-8 (Fig. 12c).

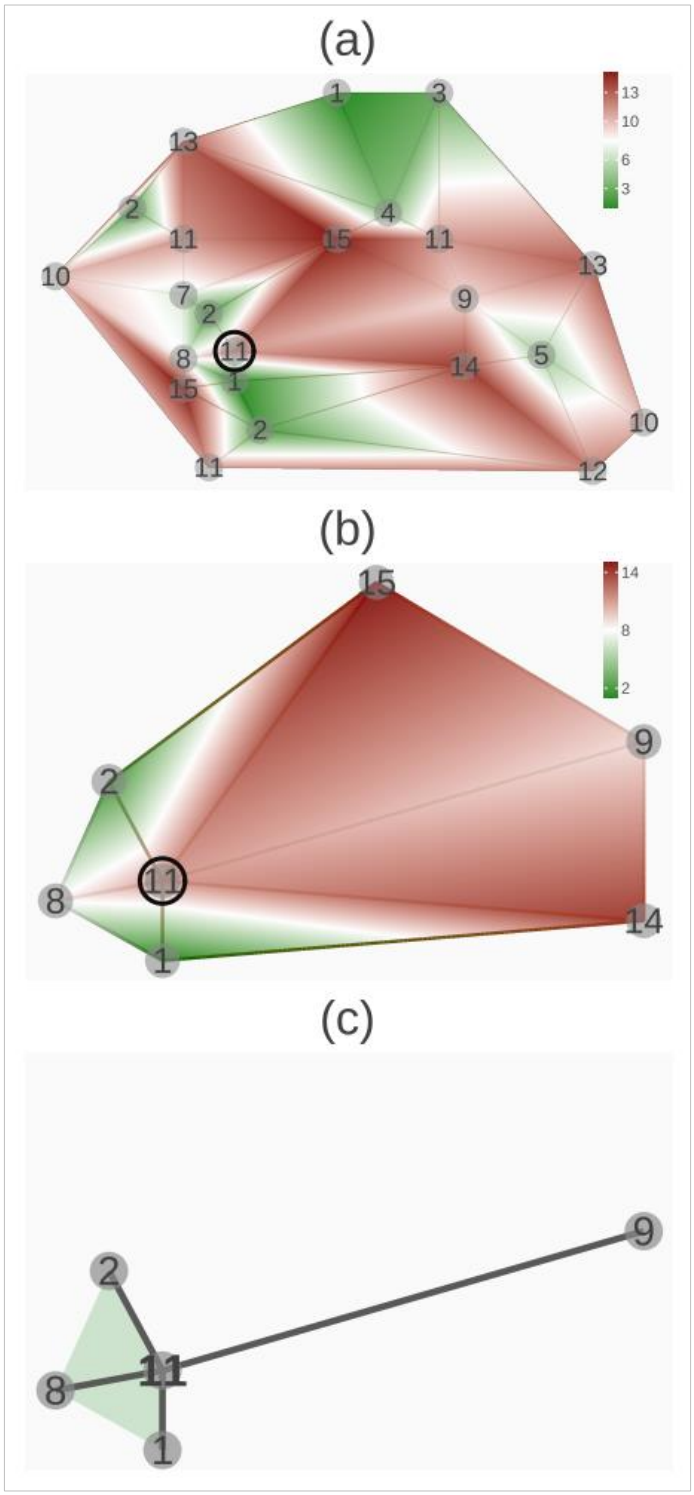


Figure 12. (a) Scalar function, defined on a triangle mesh, colour coded by means of a green-red divergent colour map. Values of vertices are also displayed. Vertex with value 11 is highlighted. (b) Vertex with value 11 and its simplicial complex consisting of neighbouring simplices. (c) Simplices of the Lower star from (b): four vertices, four edges and two triangles.

The algorithm operates by iterating over the vertices of the simplicial complex with no specific requirement for the order of processing. The algorithm identifies the lower star for each vertex. If the lower star set is empty, then the vertex is classified as a local minimum. If the lower star contains other simplices, the vertex is paired with the lowest incident edge, which is determined using lexicographic ordering. Next, the algorithm identifies the lowest unpaired simplex, denoted as $\sigma_{unpaired(k)}$ (remember: *k* indicates the dimension of σ), and attempts to pair it with the lowest unpaired higher-dimensional simplex, $\beta_{unpaired(k+1)}$. This pairing process is repeated to ensure that every simplex in the lower star is paired with a higher dimensional cell. If a cell cannot be paired, it is classified as being critical. For a more detailed mathematical description of the algorithm, please refer to Robins et al. (2011).

3.6 Accuracy assessment

TREETOPS and *treecbh* were developed using the R programming language to integrate with the C++-based forestry-focused application, the lidR package (Fig. 16). In the first study, the LM combined with the optimized VWF technique (described in the next section), both implemented in lidR, were compared with the treetops identified by the treetop locator TREETOPS. Various combined metrics were used in the GTR assessment, whereas *treecbh* was evaluated using two simple metrics for the CBH accuracy assessment. The methods used to assess the accuracy of these tools differ. Each method is described in the following subsection.

3.6.1 TREETOPS

3.6.1.1 LM with VWF versus the GTR

To ensure logical comparability between the two methods, we proposed an equation based on our target study site to adjust the window size of the VWF depending on the tree height embedded in the CHM. The function is defined as follows (Eq. 4 and Fig. 13):

$$y = 5.7 \cdot \left(-\left(e^{-0.08 \cdot (x-5)} - 1\right) + 1 \right. \quad \text{(Eq. 4)}$$

where y represents the window size for the VWF and the distance parameter of *finalize_TREETOPS()* and x represents the height values stored by the CHM. A sensible minimum height threshold of 5 m was chosen, below which no trees were found in the single-layered Nagyerdő forest. The delinearized green curve became linear at 20 m (Fig. 13, x-axis), indicating that the window size was fixed at 5 m for the B01, B05, KA09, and KA10 test forests. Throughout this study, LM with VWF was optimized and computed using the lidR package.

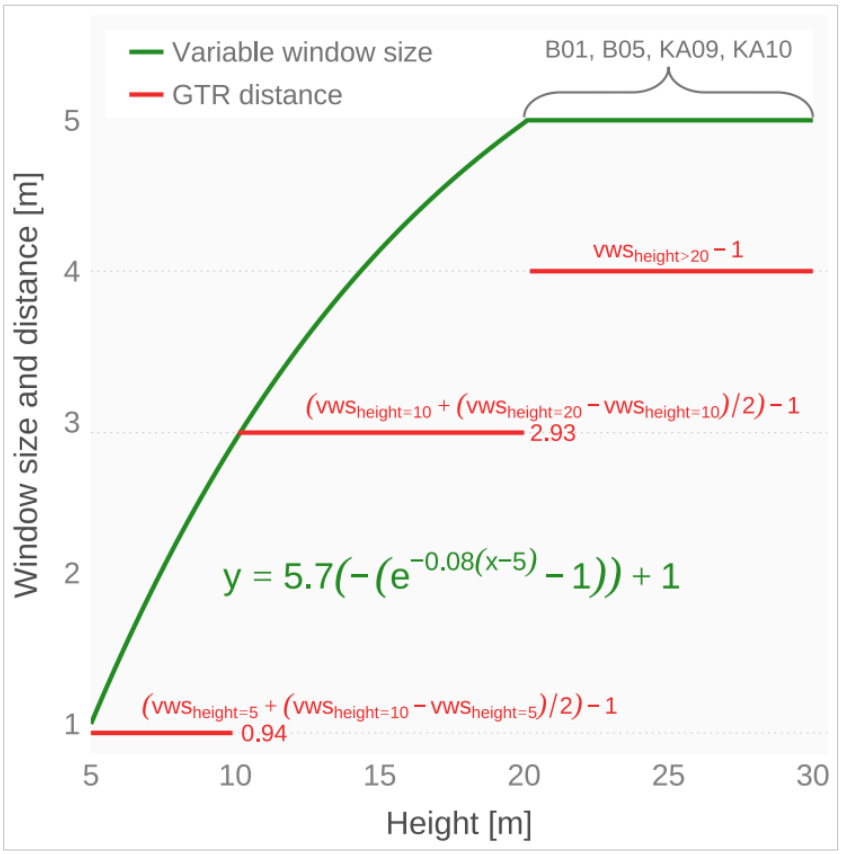


Figure 13. Calibrated variable window size for the Nagyerdő forest (depicted by the green line) and adjusted GTR distances (represented by the red line) in relation to height (x-axis). The main equation is displayed in green, while equations and values of distances are written in red; $\mathrm{VWS}_{\mathrm{height}>\text{ or }=}$ signifies a variable window size greater than or equal to a given height (y-axis).

The GTR distances (i.e., the window sizes of VWF) were calculated for the following thresholds: 5.00 to 9.99 m, 10.00 to 20.00 m, and above 20.01 m, using the following formulae (Eq. 5-7 and Fig. 13):

$$4 = VWS_{height>20} - 1 \quad \text{(Eq. 5)}$$

$$2.93 = (VWS_{height=10} + (VWS_{height=20} - VWS_{height=10})/2) - 1 \quad \text{(Eq. 6)}$$

$$0.94 = (VWS_{height=5} + (VWS_{height=10} - VWS_{height=5})/2) - 1 \quad \text{(Eq. 7)}$$

where $VWS_{height>}$ and $VWS_{height=}$ denote the variable window size greater than and equal to a given height (y-axis in Fig. 13), respectively. The distance for treetop reduction for treetops taller than 20 m corresponds to 4 m, whereas for treetops between 10 and 20 m, it is 2.93 m. For treetops smaller than 10 m, this distance was negligible (Fig. 13).

3.6.1.2 Accuracy assessment of Individual Treetop Location

A verification method developed by Eysn, et. al (2015) has established itself in studies focussing on Central-European alpine and lowland forests (Eysn et al., 2015; Stereńczak et al., 2020). The key algorithm of this approach is called the Matching Algorithm (MA), which links the location results to the reference data. During the matching process, MA produces various qualitative and quantitative statistical parameters (Eysn et al., 2015). A modified version of MA was applied to assess the accuracy of the ITD results. This had to be altered because our particular interest lay in a comparative analysis between the LM with VWF and the GTR on a one-to-one basis. More importantly, our reference data sets (i.e., German and Nagyerdő reference trees) did not provide reference information for every tree within the study site boundaries. Thus, the number of trees extracted using the two methods was not reliably linked to the number of reference trees. This phenomenon technically hindered obtaining the correct extraction and commission rates, respectively. To address this issue, both the extraction and commission rates were mathematically adjusted. The following parameters were returned by the MA for validation:

N_{Test} → Number of Extracted trees,

N_{Ref} → Number of Reference trees,

R_{Ref} → Rate of Reference trees ~ N_{Ref} / (N_{Ref} + N_{Test}),

N_{MTest} → Modified N_{Test} ~ $if\ R_{Ref} < 0.3 \longrightarrow$

$N_{MTest} = N_{Test} \cdot (1 - R_{Ref}) - N_{Ref}\ else \longrightarrow N_{MTest} = N_{Test}$,

N_{Match} → Number of Matched trees,

N_{Com} → Number of Extracted trees that could not be matched ~

$N_{MTest} - N_{Match}$,

N_{Om} → Number of Reference trees that could not be matched ~

$N_{Ref} - N_{Match}$,

H_{Test} → Height of Matched Test trees,

H_{Ref} → Height of Reference trees,

D_{Hor} → Horizontal distance between Matched Test and Reference trees,

N_{Match} / N_{Ref} → Matching rate,

N_{MTest} / $N_{Ref} - log_{10}(N_{MTest}) \cdot 20$ → Modified Extraction rate,

N_{Com} / $N_{MTest} - log_{10}(N_{MTest}) \cdot 20$ → Modified Commisson rate, Extracted trees that could not be matched,

N_{Om} / N_{Ref} → Omission rate, Reference trees that could not be matched.

Using the values of H_{Test} and H_{Ref}, the mean absolute error (MAE) was determined for the height differences between Matched Test and Reference trees by Equation 8:

$$MAE_{height\,[m]} = (1 / N_{Ref}) \cdot \sum |H_{Ref_i} - H_{Test_i}| \qquad \text{(Eq. 8)}$$

where *i* indicates the i-th value regarding Reference and Test heights and the modulus is denoted by the | | symbol. Furthermore, the parameters described above were used to define the following root mean square (RMS) parameters:

RMS_{Match} → Root Mean Square of Matching rates,

RMS_{Extr} → Root Mean Square of Extraction rates,

RMS_{Com} → Root Mean Square of Commission rates,

RMS_{Om} → Root Mean Square of Omission rates.

3.6.2 Accuracy assessment of Crown Base Height Location

The located CBHs were overlaid onto the vertical 2D cross-sectional plots of the input point clouds (Section 3.1). CBHs were buffered by 1 m in both the positive and negative directions along the z-axis (Fig. 14, Crown Base Height). A matching criterion was applied to assess the match between the estimated and the buffered reference CBHs. A match (+) was recorded if the located CBH fell within the buffered CBH range; otherwise, it was considered a nonmatch (–). The matching rate was calculated using Equation 9:

$$Matching\ rate_{\,[\%]} = (1 - (N_{(+)} \,/\, N_{(-)})) \cdot 100 \qquad \text{(Eq. 9)}$$

where the two *N*s represent the counts (i.e., sums) of matches (+) and non-matches (-), respectively. The mean absolute error (MAE) was applied to quantify the located CBH deviation from the reference CBH and was computed using Equation 10:

$$MAE_{CBH\,[m]} = (1 \,/\, N) \cdot \sum |Ref_{CBH_i} - \ Test_{CBH_i}| \qquad \text{(Eq. 10)}$$

where *i* indicates the *i*-th individual tree segment, Ref_{CBH} and $Test_{CBH}$ are the reference and the located CBHs, respectively. Parameter *N* represents the number of tree segments and the modulus is denoted by the | | symbol. The unit of MAE is given in meters, as indicated by the small *m* in square brackets.

The accuracy assessment consists of three sequential steps. First, the CBH locator was executed using the input point clouds (Section 3.1) employing the deactivated *treeiso* mode (*No treeiso*). Next, *treecbh* estimated the CBHs for the input point clouds in conjunction with the activated *treeiso* mode. The parameters of the tree isolator were maintained at their default settings (Section 2.2.2) and were denoted as *Default treeiso* throughout this study. Finally, the input point clouds were

processed using *treecbh*'s interactive visual CBH adjustment mode, herein referred to as *visual* mode.

4 IMPLEMENTATION AND DEVELOPMENT OF THE ALGORITHMS

Both algorithms were implemented at the programming level, providing forest managers and the research community with two open-source tools intended to be used in conjunction. The processing workflows of the algorithms can be integrated and generalized, thereby simplifying and considering them as a single workflow. Before this integration step, each of the two algorithms is described.

4.1 Individual treetop location

GTR, which is based on the LiDAR-derived CHM, is a treetop localizer algorithm. The algorithm was written in the R programming language and relies heavily on the C++ based terra (Hijmans et al., 2023) and sf (Pebesma et al., 2023) packages. Therefore, the terms used in this dissertation were adjusted to the terms used by those packages. Similar to the LM algorithm using VWF, this algorithm considers tree crowns as mountains protruding from a three-dimensional canopy surface (i.e., CHM). The apex of a tree is equal to the top of a mountain. Height values decreased continuously starting from the treetop, running along the slope of the mountain which is the tree crown. This descent stops in the valley, which can either mark the boundary between the trees or the edge of the crown. The algorithm consists of three major steps.

1) CHM cutting and storage.

2) Treetop location.

3) Reduction in the number of treetops.

4.1.1 CHM cutting and storage

The GTR algorithm employs a horizontal plane to vertically slice the CHM from top to bottom into multiple horizontal raster layers, where

non-no data values are labelled as 1 (Fig. 14). The files were stored in a stacked format. The vertical cutting distance, referred to as the height increment, was set to 0.2 m. Consequently, each raster represents the CHM value (1) at a specific height, with the height information (Z) stored in the filename of the corresponding raster file within the stack. During the tree-growth process, sections of growth intersect, forming a new patch at a certain height. In our example, the first new patch (i.e., intersection) occurred between the growing sections that grew at height from 31.8 m to 30.0 m (Fig. 14).

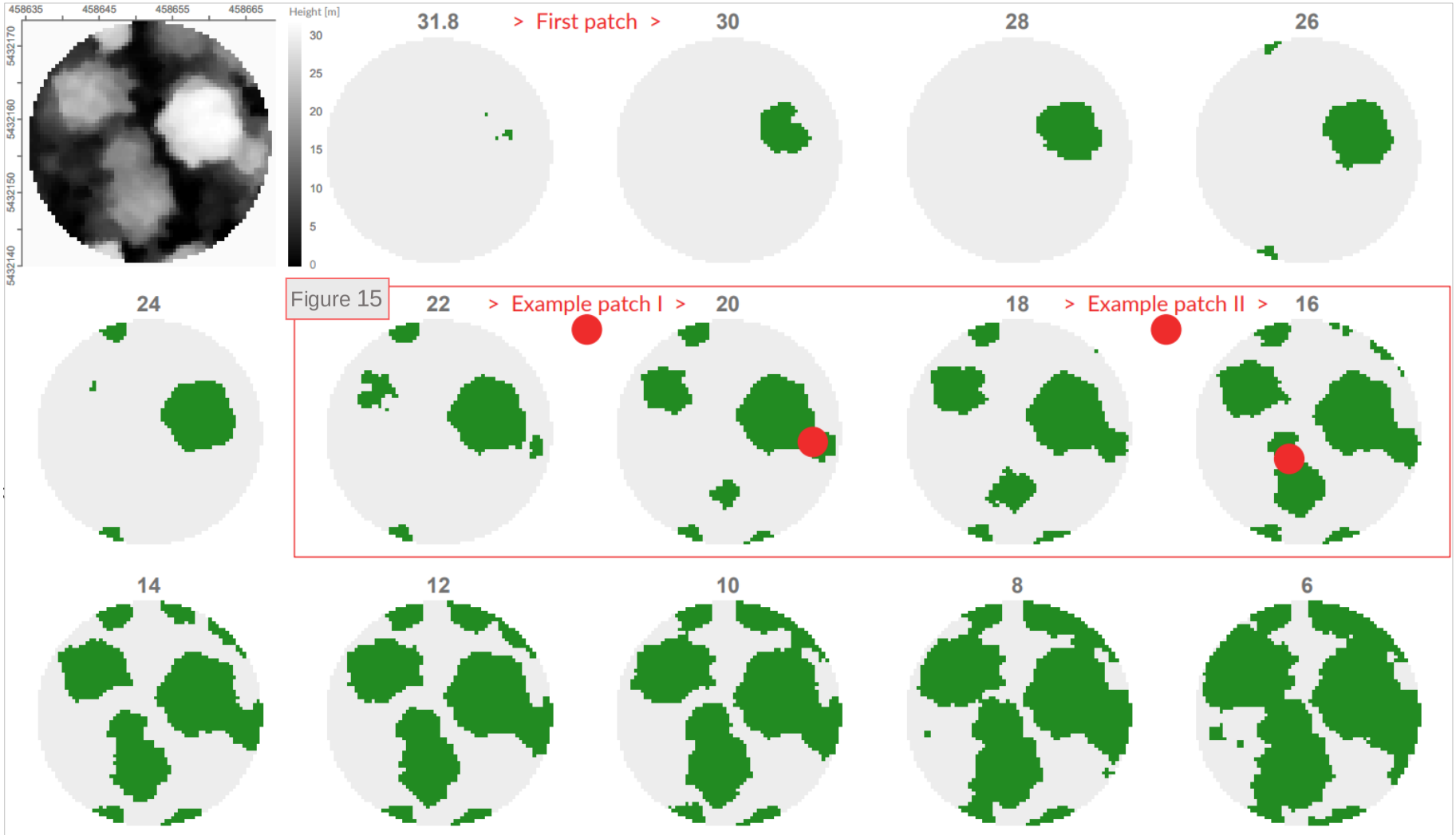


Figure 14. Circular example of a CHM with a pixel size of 0.5·0.5 m, displayed in decreasing order between 31.8 m (maximum height) and 6 m. The decrement interval is 0.2 m (shown is a 2 m decrease). In the top row on the left, the initial example CHM is depicted, with green representing value 1 and grey indicating no data. The locations of the two example patches (intersections) are highlighted by filled red circles with transparency.

4.1.2 Treetop location

GTR identification is a crucial step for locating treetops. The process is iterative, and involves the selection of three consecutive layers (from top to bottom) at each iteration (Fig. 15). During one iteration, the algorithm searches for new intersections and identifies GTR. Treetops are extracted by determining the centroids of the GTR calculated from layers 2 and 3, and then filtering them based on the criterion of overlapping with the GTR from layers 1 and 2 (Fig. 15). The three-layer concept is explained as follows.

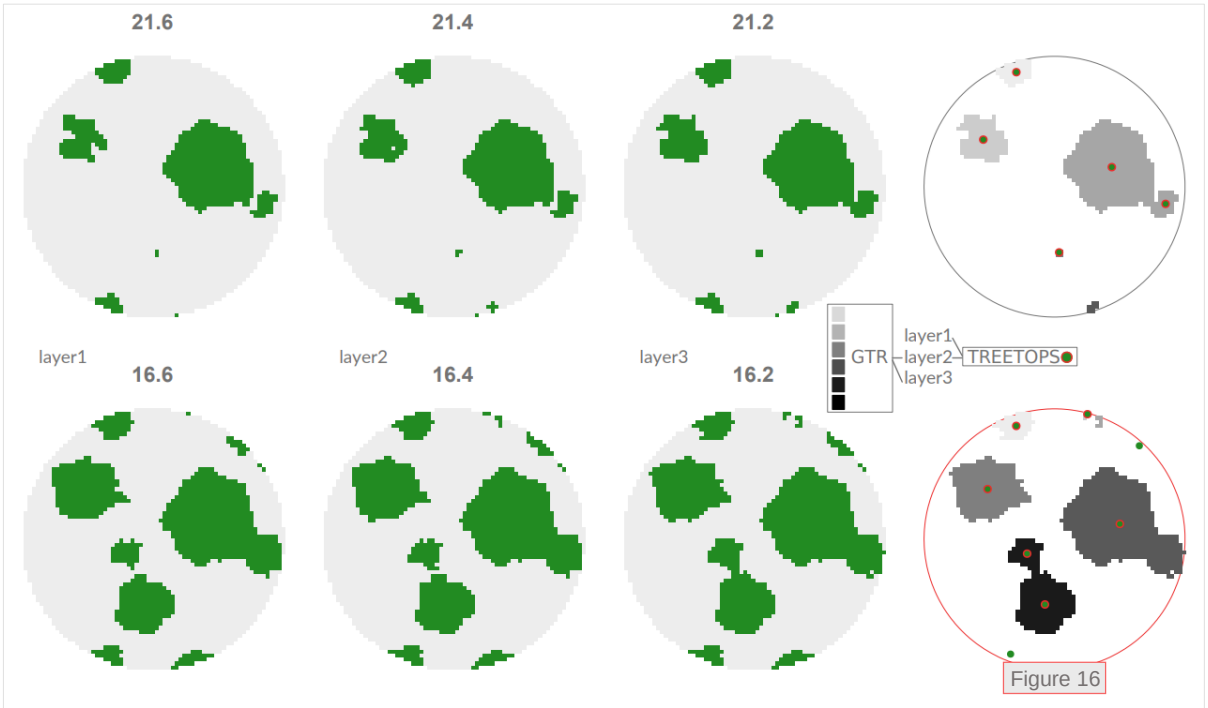


Figure 15. Example patches (Figure 14) provide an overview of the three-layer concept. Green denotes a value of 1, whereas gray represents no data.

First, the first emerging tree region (FETR) is determined by subtracting layer1 from layer2 (FETR21), and the operation of layer3 - layer2 results in FETR32 (Fig. 16). Subsequently, GTR21 and GTR32 were computed. GTR21 was calculated by utilizing layer2 and FETR21 under the following conditions (Eq. 11, Fig. 16):

$$layer2patch_i \supseteq FETR21patch_n \,\&\, N_{cell}\, layer2patch_i \neq N_{cell}\, FETR21patch_n \quad \text{(Eq. 11)}$$

where *i* in $layer2patch_i$ denotes the *i*-th patch of layer2, ⊇ represents improper superset and *n* in $FETR21patch_n$ expresses the *n*-th patch of FETR21 (Fig. 16). Thus, this double condition can be expressed as follows: if the *i*-th patch of layer2 is a superset of the n-th patch of the FETR21 patch whose pixels are element of layer2 AND the number of pixels of the i-th patch of layer2 is not equal to the number of pixels of the *n*-th patch of FETR21. The output GTR21 (and GTR32) is a binarized raster with GTR and non-GTR values. Next, individual regions (i.e., patches) of both the GTR21 and the GTR32 were identified. Finally, the treetops were located by using the following formula (Eq. 12, Fig. 16):

$$TREETOPS_{layer2} = Centroids_{GTR21} \in GTR32 \qquad \text{(Eq. 12)}$$

where $TREETOPS_{layer2}$ indicates that the treetops identified in a given iteration are associated with the height value of layer2; ∈ represents the element-of symbol; centroids of GTR21 overlapping GTR32 are considered as the treetops.

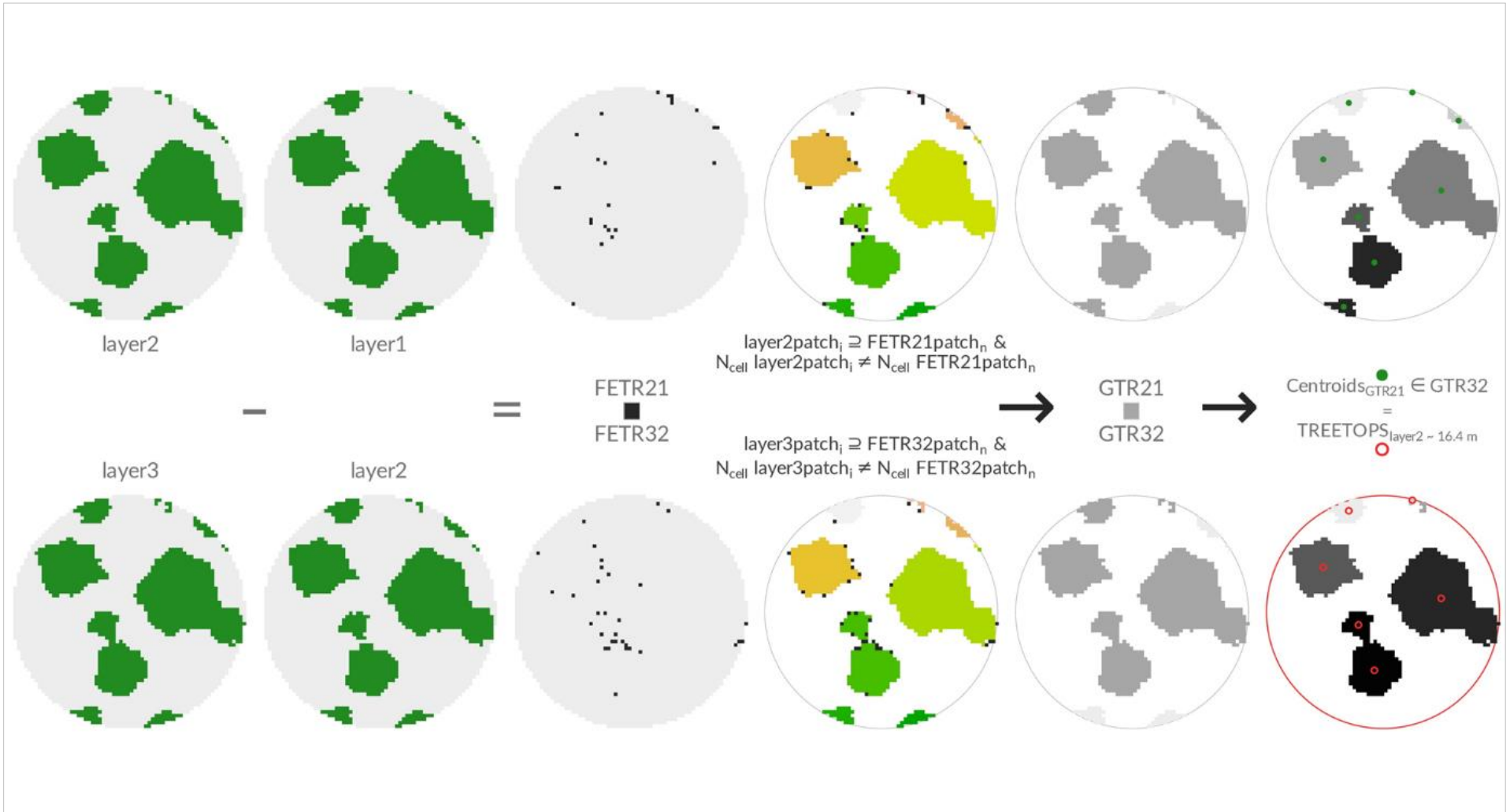


Figure 16. The three-layer concept (Figure 15), which is at the core of the GTR algorithm, is visualized and mathematically explained. In the first three columns, green represents value 1, whereas grey indicates no data. The distinct colours in the fourth column and shades of grey in the sixth column represent the patch IDs. The red circle around the bottom-right raster highlights the ITD output (red unfilled dots) of the algorithm.

4.1.3 *Reduction in the number of treetops*

Because treetops are stored in a self-updating process after every iteration (Fig. 17a), the number of treetops must be reduced. Our solution is a highly flexible three-parameter, distance-based treetops reduction method. Technically, it finds all the neighbours within the defined distance (first parameter) and maintains the highest value (Fig. 17a and b). The two other parameters, minimum height and maximum height, facilitate threshold settings, meaning that lower trees (unless under canopy) of multilayered forests can be identified (Fig. 17b).

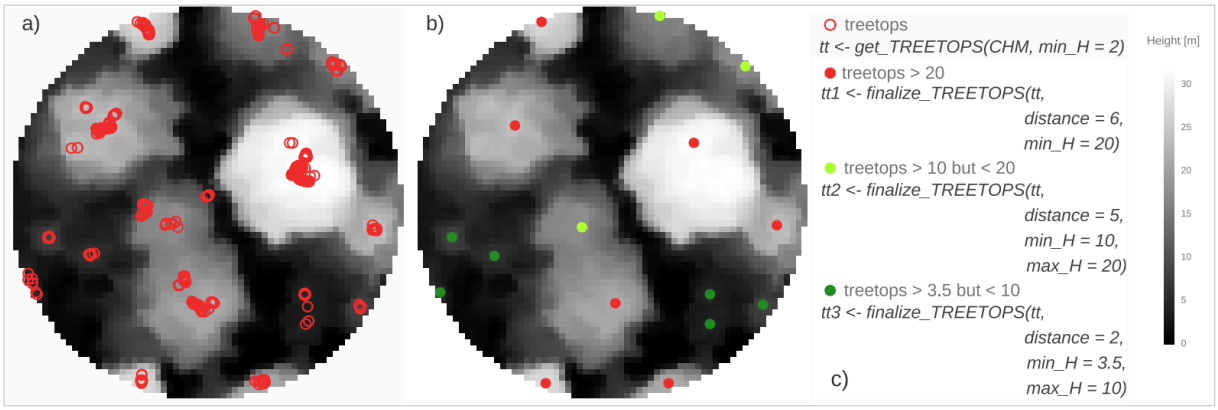


Figure 17. Example CHM (0.5·0.5 m pixel size) with treetops (a); example CHM with treetops after applying the distance-based treetop filter (b); (c) the two main functions, *get_TREETOPS()* and *finalize_TREETOPS()* from the R package TREETOPS, and the implementation of threshold settings using the *min_H* and *max_H* parameters.

4.1.4 *TREETOPS*

From the user's perspective, the GTR algorithm is bundled into an R package called TREETOPS, which provides two main functions (Fig. 17c): *get_TREETOPS()* and *finalize_TREETOPS()*. The *get_TREETOPS()* function executes the entire ITD process, including CHM cutting and storage and the treetop location. The *finalize_TREETOPS()* function applies the distance-based treetop reduction method. Both functions are user friendly and require a maximum of four parameters. To set the threshold, the user must set the *min_H* and *max_H* (height-defining) parameters within the *finalized _TREETOPS()* function (Fig. 17c). The outputs of both functions were simple features, as defined by the sf R package (Pebesma and Bivand, 2023).

4.2 Crown Base Height locator

The *treecbh* framework operates based on the following three-step concept:

1) Tree trunk isolation: In this initial phase, the framework focused on isolating the tree trunk with the first leaved branches of the dominant tree within the input point cloud. The input point cloud was obtained through manual segmentation, as detailed in Section 3.3.2.

2) Vertical cross-sectional k-means clustering: After tree isolation, this stage addresses the separation of understory elements from the tree trunk and its lower branches.

3) 2D kernel method for CBH location: The framework employs a 2D kernel density method specifically designed for locating the Crown Base Height (CBH) based on the extracted tree stem and its lower branches.

4.2.1 Tree trunk isolation

Assuming the presence of smaller non-dominant trees and understory vegetation in Central European Forest stands, a point cloud-based segmentation approach was implemented to precisely identify and extract the trunk of the dominant tree along with its first-leaved branches (Fig. 18a). This methodology requires a point cloud of a segmented individual tree as input and leverages the *treeiso* software, which employs two successive cut-pursuit clusterings (Xi and Hopkinson, 2022). Conceptually, the point cloud was treated as a graph, where each point represents a node and the connections between points were denoted as edges. Using the cut-pursuit algorithm, the point-cloud graph first undergoes partitioning into clusters. The boundaries between these clusters were determined by minimizing the total variation within the initial cloud (Landrieu and Obozinski, 2016), which corresponds to the input point cloud for *treecbh*. Subsequently, a secondary cut-pursuit clustering function was employed to merge small 3D clusters into larger clusters, thereby improving the coherence of segmentation (Xi and Hopkinson, 2022). The resulting point cloud is an

isolated tree cloud (Fig. 18a, Tree trunk with first leaved branches), which encompasses the trunk along with its lower branches of the dominant tree isolated from the input point cloud (Fig. 18a, Input point cloud). It only contains points positioned above 0.2 meters (Fig. 18a). The shape of an isolated tree point cloud can vary depending on the geometry of the input point cloud, which is influenced by the forest structure. Although rare, the resulting point cloud may occasionally include multiple isolated tree trunks.

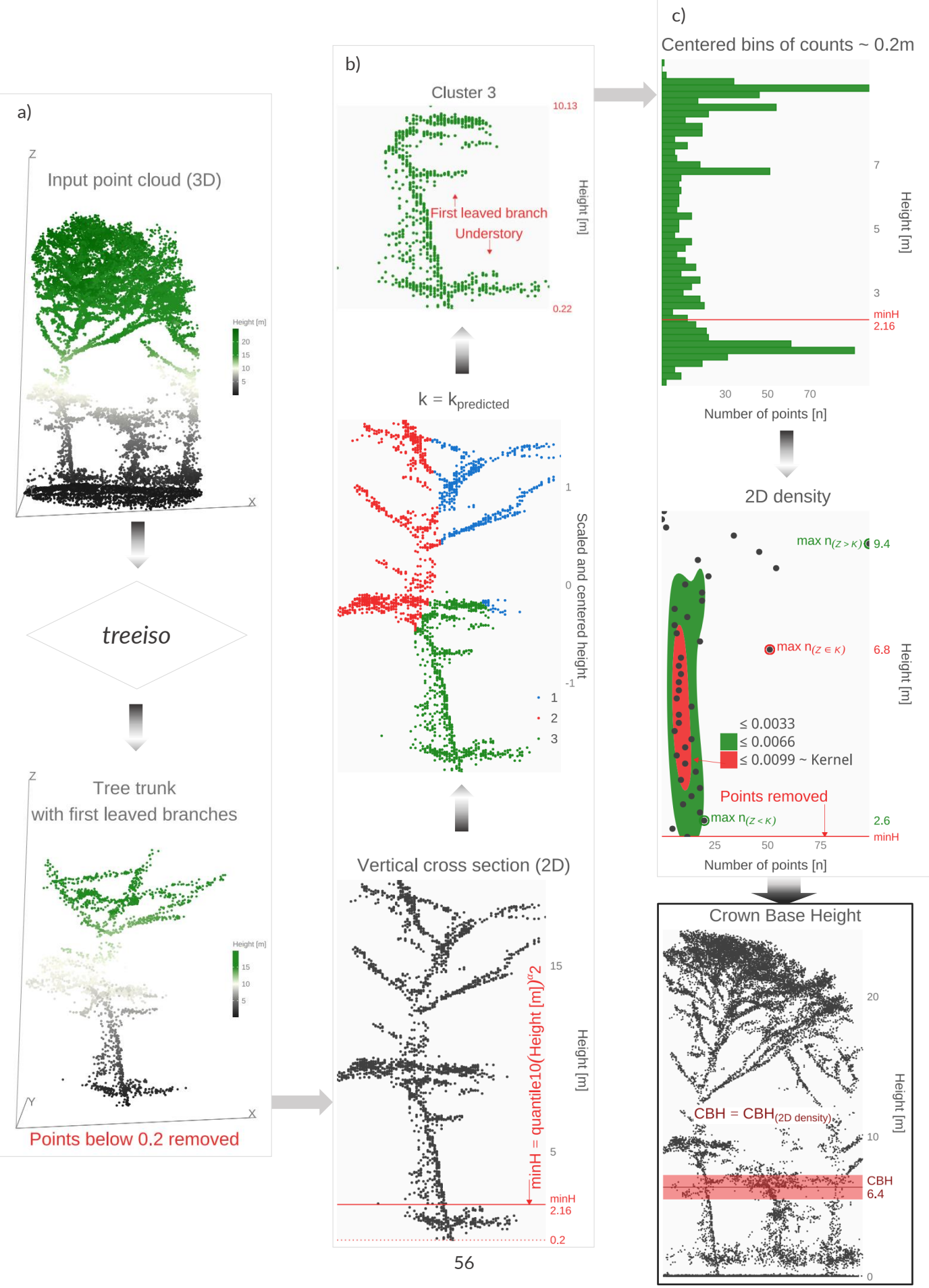

a)
Input point cloud (3D)
Z
X
Height [m]
20
15
10
5
treeiso
Tree trunk
with first leaved branches
Y
X
Points below 0.2 removed
b)
Cluster 3
10.13
First leaved branch
Understory
0.22
Height [m]
k = k predicted
1
0
-1
Scaled and centered height
1
2
3
Vertical cross section (2D)
15
5
minH = quantile10(Height [m])^2
minH
2.16
0.2
Height [m]
c)
Centered bins of counts ~ 0.2m
7
5
3
Height [m]
minH
2.16
30
50
70
Number of points [n]
2D density
max n(Z > K)
9.4
max n(Z ∈ K)
6.8
Height [m]
≤ 0.0033
≤ 0.0066
≤ 0.0099 ~ Kernel
Points removed
max n(Z < K)
2.6
minH
25
50
75
Number of points [n]
Crown Base Height
20
10
0
Height [m]
CBH = CBH(2D density)
CBH
6.4

Figure 18. Framework of *treecbh*: tree isolation (a), vertical cross-sectional K-means clustering (b), and the 2D kernel method (c). The located Crown Base Height is illustrated in the graph at the bottom right.

4.2.2 Vertical cross-sectional K-means clustering

In the second step, a vertical cross-sectional cuboid, 5 m wide, was applied to the xy attributes of the isolated tree cloud. This 5-meter width adequately encompasses any tree cloud in the y dimension, capturing points in both the x and z dimensions. This operation creates a transect through the tree cloud, defining a specific region of interest within the xy plane of the point cloud. This allowed for the extraction of a vertical cross-section from the tree cloud, resulting in a 2D profile that displayed the xz attributes of the points within the specified region of interest (Fig. 18b).

Next, the *minH* parameter was introduced to serve as a minimal height threshold, accounting for the probable presence of understory vegetation. Understory vegetation was represented by a higher point count than that of the tree trunk, resulting in a higher peak in the point distribution of the 2D vertical cross-section. We assumed that understory vegetation was present below the 10^{th} quantile in the distribution of the input point cloud and its 2D vertical cross-section (xz). The computation is given by Equation 13:

$$minH_i = \text{quantile10}(height_i)^{\alpha} \cdot 2 \qquad \text{(Eq. 13)}$$

where *i* signifies the *i*-th tree cloud, and quantile10 represents the function to capture the sample quantile at the 10th percentile of the height (z) of the *i*-th tree cloud ($height_i$). The scaling parameter α facilitates a user-defined height threshold related to the height of the tree cloud, under which understory vegetation is assumed and removed internally (Fig. 18b). α is denoted by the *min_H_scale* parameter in *treecbh*, and has a default value of 0.13, which equates to 13% of the tree height.

Subsequently, the height attribute (z) of the 2D data underwent scaling and centering processes, as required for the K-means clustering methodology. The estimation of the number of clusters defining the *k* parameter was accomplished using the prediction strength of the clustering technique (Tibshirani and Walther, 2005) (Fig. 18b). Finally,

the cluster characterized by the lowest height value was identified and extracted from the processed data (Fig. 18b). The extracted dataset then served as an input for the subsequent 2D kernel density analysis.

4.2.3 Crown Base Height locator, the 2D kernel density method

The 2D kernel method analyses a histogram of the points from the cluster identified in the second step. In *treecbh*, the *branch_WIDTH* parameter defines the bin width of the histogram (default value of 0.2 m) with which the input points are binned (Fig. 18c, Centered bins of counts ~ 0-2m). Bins with heights equal to or below the *minH* threshold (Fig. 18c, Centred bins of counts ~ 0.2m). Bins with heights equal to or below the *minH* threshold (Eq. 5) were systematically removed. The filtered data were then used for the 2D kernel density estimation. The goal of this analysis was to find the region with the highest bin density, referred to as the "Kernel" (Fig. 18c, 2D density). The 2D kernel method applies three kernel density classes: ≤ 0.0033, ≤ 0.0066 and ≤ 0.0099 on the binned points (Fig. 18c, 2D density). The analysis involved three steps: determining the number of kernels (in cases of more than one isolated tree trunk), defining the range of these kernels, and establishing the maximum count of points vertically within each kernel range. For multiple kernels, the one with the highest point count within its range was selected for further analysis. Additionally, the maximum count of points both below and above the identified kernel was determined (Fig. 18c, 2D density) along with the range of the cluster (i.e., the minimum and maximum heights of the input data). This results in five candidate heights: the heights with the maximum counts below the kernel (max $n_{(z < K)}$), within the kernel range (max $n_{(z \in K)}$), and above the kernel (max $n_{(z > K)}$) (Fig. 18c, 2D density), as well as the cluster height (max(Z_rC)) and kernel height (max(Z_rK)) (Fig. 2).

In the final stage of the CBH identification, one of these candidate heights was selected as the CBH. This decision-making process starts with the number of kernels identified through the automated 2D kernel density analysis and follows a systematic if-yes-no strategy, as illustrated in the workflow diagram (Fig. 19).

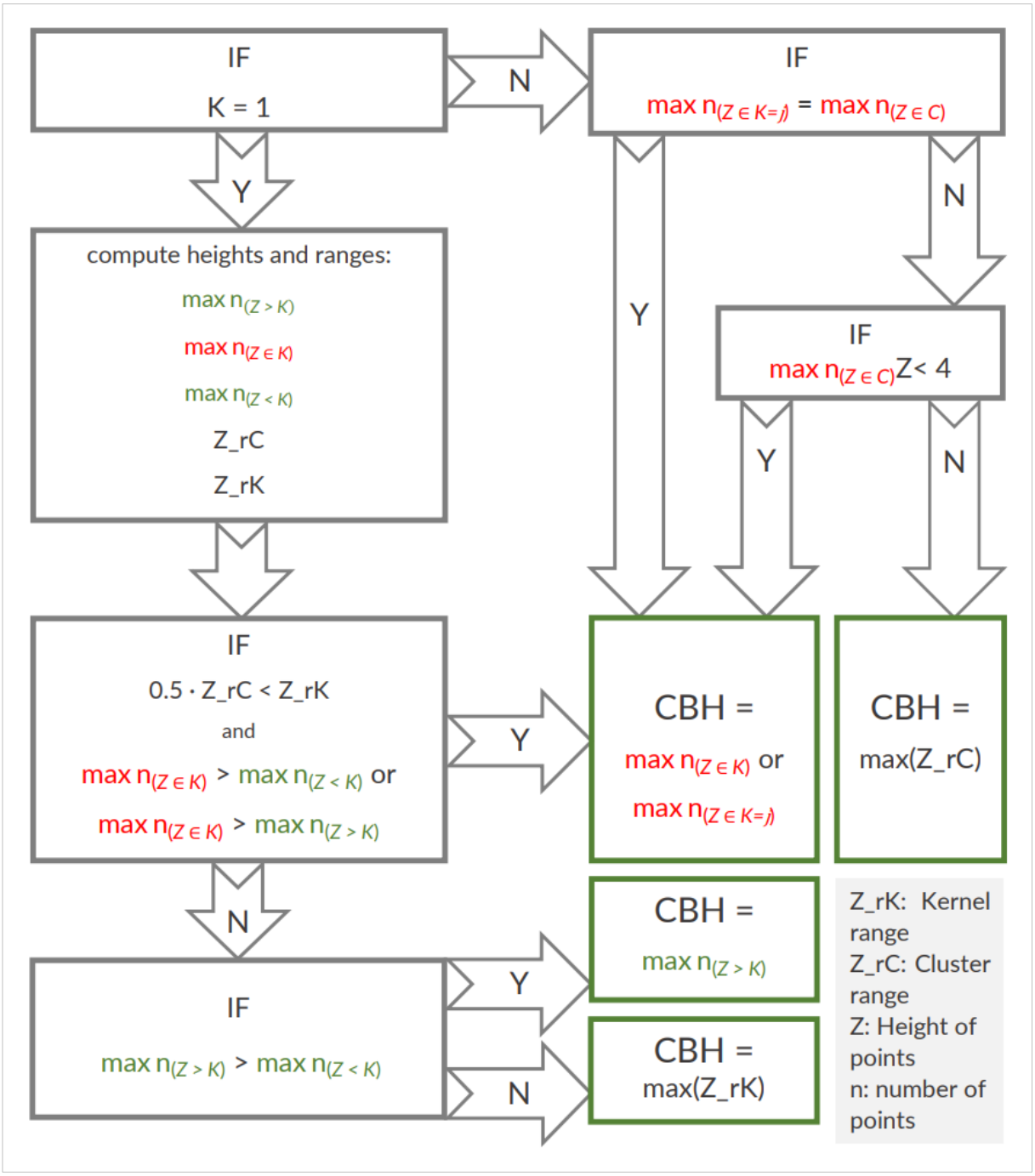


Figure 19. The workflow of the final stage of CBH identification, the decision scheme (Y: yes, N: no; green and red colored text serve transparency and distinguishability, while green rectangular frames indicate identified CBHs; K = 1 means one kernel; Z_rK and Z_rC contain two values; Z represents a vector of heights).

Because the LiDAR data were obtained at the beginning of the vegetative period and the branches and twigs of deciduous tree species were covered with young leaves, it is safe to assume that the CBH of a tree with fully grown leaves might be lower. Furthermore, coniferous species tend to hang their first leaved branches downward. Considering these characteristics, the final CBH was obtained by subtracting the two bin heights from the decision scheme assigned to the CBH.

4.3 TREETOPS and treecbh as a user's tool

4.3.1 User's workflow

As mentioned previously (Sections 2 and 3.3), ITS is required before applying *treecbh* because the CBH locator operates at the individual tree level. Therefore, using the TREETOPS-outputted treetop-controlled ITS combined with *treecbh* facilitates a generalized workflow that aims at a robust CBH location. Thus, embedding the two main functions of TREETOPS (Section 4.1.4), namely *get_TREETOPS()* and *finalize_TREETOPS()*, within the R computational workflow is straightforward (Fig. 20). Generally, the use of a treetop locator coupled with the CBH locator is user-friendly and easy. The user is recommended to follow the computational steps shown in Figure 20.

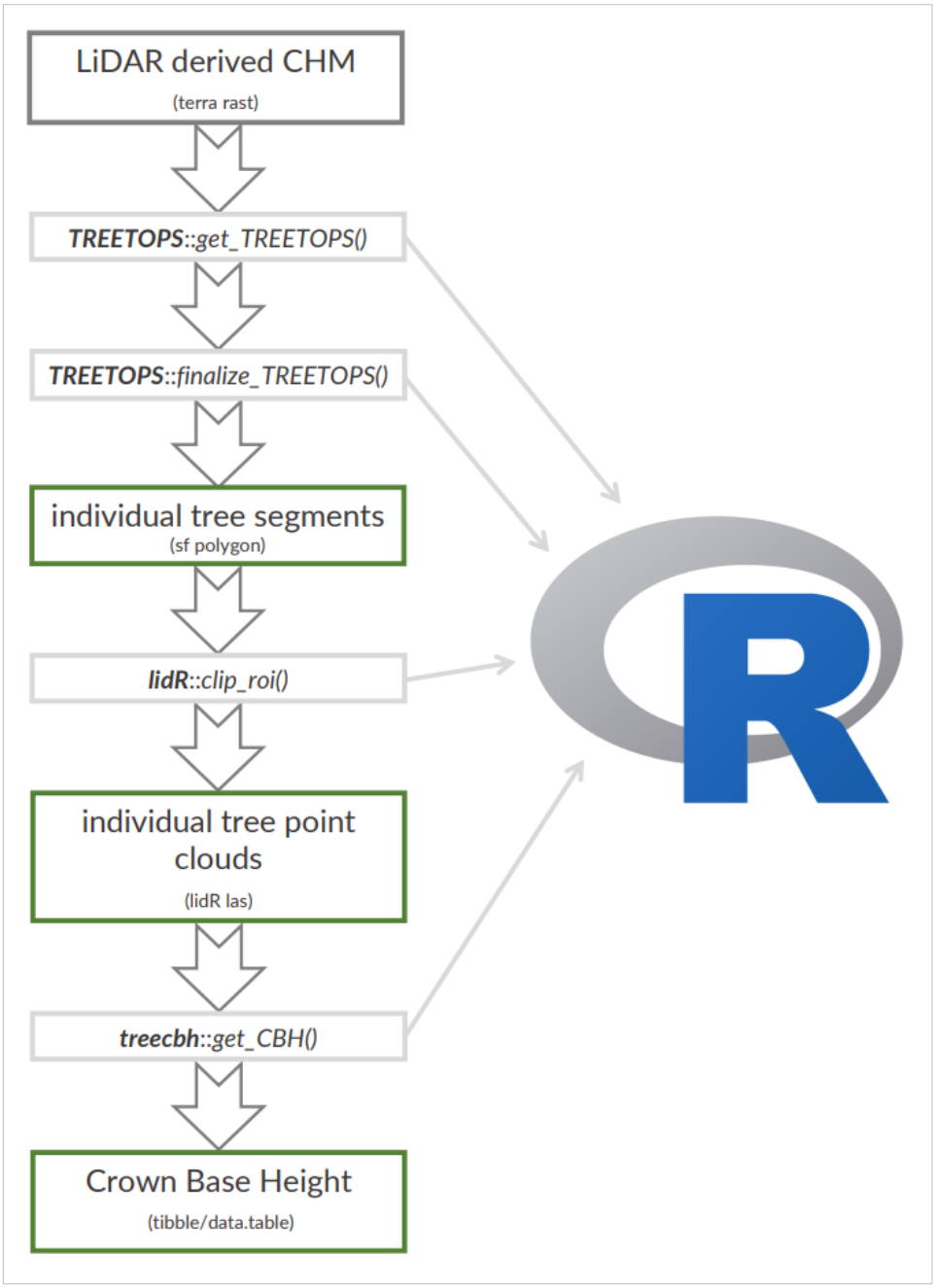


Figure 20. Concept of the computational workflow from the user's perspective regarding GTR treetop locator (R package TREETOPS) and CBH locator (R package *treecbh*). While the dark grey outlined square represents the initial input R object (terra package raster, named rast), the green framed squares indicate the resulting spatial and non-spatial R objects.

4.3.2 Parameters of treecbh

Throughout this study, two parameters, *min_H_scale* and *branch_WIDTH*, were set consistently with fixed default values and remained unchanged. However, *treecbh* incorporates two additional parameters: *cbh_ONLY* and *kM*. An in-depth overview detailing the *treecbh* parameter annotations, their implications, and the corresponding thresholds are presented in Table 6.

Table 6. Parameter characteristics of *treecbh*.

Name	Default value	Implication	Range or values
*min_H_scale***	0.13	Minimum height scaler (m), controlling understory removal	[0.13-0.25]
*branch_WIDTH***	0.2	Assumed CBH branch width (m), controlling bin width for counting points	[0.01-0.5]
cbh_ONLY	1	Options for executing: 1~*treeiso* and cbh, 2~only *treeiso*, 3~only cbh location	[1,2,3]
*kM**	TRUE	K-means clustering (TRUE) or CBH tuning (FALSE)	[-]

* parameter that triggers visual CBH location
** parameters used with their default values

4.3.3 treecbh parameters as user's tools

By setting the parameter *cbh_ONLY*, users can specify whether *treecbh* executes the complete 3D tree isolation followed by the CBH location, performs tree isolation only, or focuses solely on the CBH location. This flexibility is particularly useful when using the tool in the visual CBH adjustment mode because it allows users to disable 3D tree segmentation (i.e., *treeiso*). When the *kM* parameter is deactivated by setting *kM* = FALSE (*treecbh* skips K-means clustering and employs only the 2D kernel technique), an interactive CBH adjustment process is initiated. *treecbh* plots the 2D profile of the input point cloud and

proposes a CBH in the R Console awaiting user approval. Users can either accept or reject the proposed value while inspecting a two-dimensional (2D) plot (Fig. 21). If the proposed value is rejected, *treecbh* prompts the user to suggest a CBH value. It then adjusts the CBH location to the proposed height and searches for CBH within a 0.5 m distance in both the negative and positive directions of the user-defined CBH value.

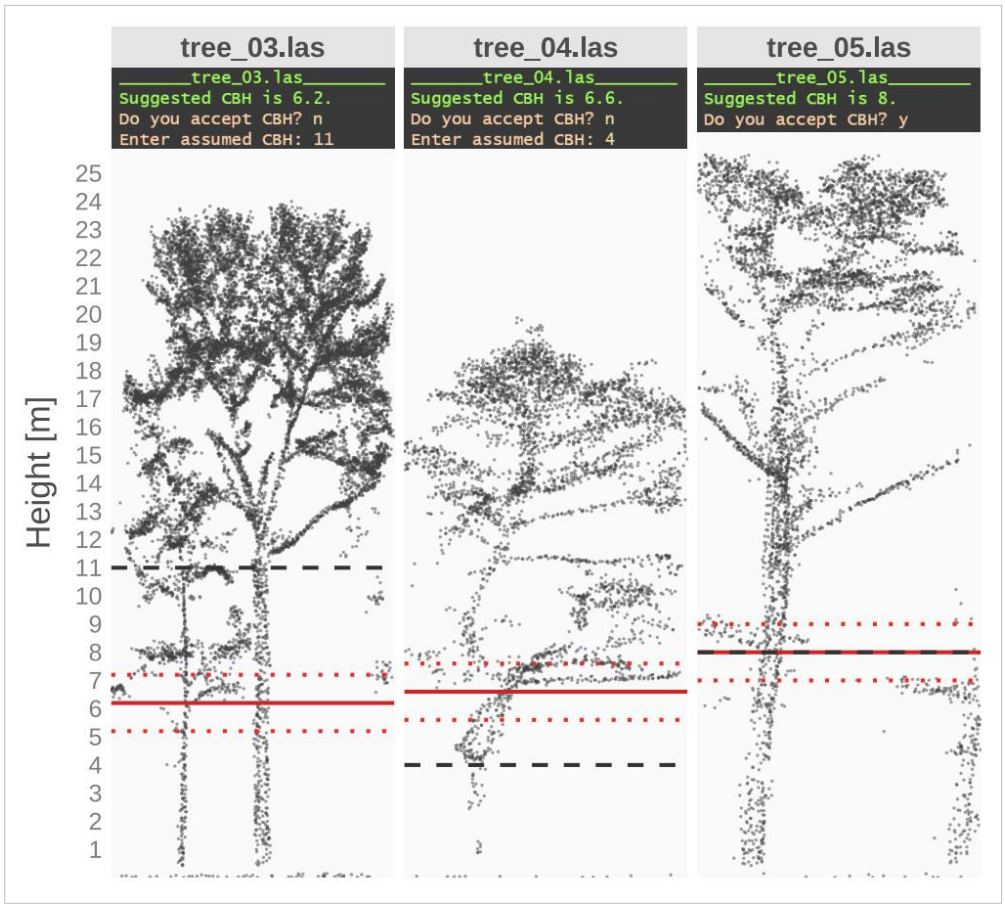


Figure 21. Example of CBH optimization process triggered by *kM* parameter deactivation. The solid red line indicates the suggested CBH by *treecbh*, while the dotted red lines represent the suggested CBH with a range of ± 1 m. The dashed black line indicates the user-entered assumed CBH (not shown by the tool). Dark-grey R soles display interactive processes with user inputs.

4.3.4 Parameters of treeiso

If the parameter *cbh_ONLY* is set to 1 or 2 (as outlined in Table 5), it allows the adjustment of nine additional parameters within the *treeiso*

(Xi and Hopkinson, 2022). It is crucial to note that the default values for these parameters remained unchanged for the purposes of this study. For a comprehensive understanding of the annotations, implications, and thresholds associated with the *treeiso* parameters detailed in Table 7, please refer to Xi and Hopkinson (2022).

Table 7. Parameter characteristics of *treeiso*.

Name	Default value	Implication	Range or values
K1	10	Number of nearest neighbours, controlling unit size of 3D cluster	[3-50]
K2	20		[5-50]
L1	1	Regularizing parameters, a greater number producing more edge cuts	[0.1-40]
L2	20		[5-40]
DEC_R1	0.1	Customizable node weight, value is the inverse of K1	-
DEC_R2	0.1		-
MAX_GAP	0.5	Maximally allowed threshold distance to consider an edge	[0.5-5]
VER_O_W	0.3	Ratio of elevation difference from neighbours to segment length	[0-1.1]
RHO	0.5	Importance of the horizontal overlapping ratio over the vertical	[0-2]

4.3.5 General R-usage of treecbh

In the context of prioritizing the CBH location, *treecbh* relies on the user-defined *kM* parameter (*kM* = FALSE), which triggers an interactive and visual CBH location process. Notably, the tool defaults to automated full tree isolation, completing the secondary cut-pursuit clustering of *treeiso* by linking cloud segments to isolate trees (as described in Section 4.3.3). This default functionality accommodates users interested in isolating

trees from terrestrial laser scanning data and offers flexibility in adjusting the *treeiso* parameters (Xi and Hopkinson, 2022) to suit their specific needs.

4.4 Tree Stem Location

The tree stem location consists of three steps:

1) Triangulation and Discrete Morse computation: in this initial phase, Delaunay triangulation (i.e., α-shaping) is applied to the input TLS point cloud to generate critical simplices, as detailed in Section 2.4.2.

2) Tree Stem Location: This stage involves locating and labelling each tree trunk using the HDBSCAN algorithm.

3) Label-Passing: In the final step, a novel growing radius technique is employed to pass the tree ID labels obtained from HDBSCAN in step 2 to the critical vertices (points) computed in step 1.

4.4.1 Triangulation and Discrete Morse Computation

Initially, the clustered 12-member tree point cloud contained over two million points (Fig. 22a). For α-shaping, after testing various candidate α values, a value of 0.1 was selected (Fig. 22b). The resulting triangulated shape forms a simplicial complex on which the topological surface of the critical simplices is identified. Using the "Lower Star" algorithm (as detailed in Section 3.5), all minima (critical vertices), saddles (critical edges), and maxima (critical triangles), along with other non-critical simplices, were outputted in a text file. The simplices were listed in lexicographical order. The number of critical vertices (points) was reduced to 157311 (Fig. 22c), indicating a significant reduction in the point count compared with the initial point cloud. From this point onward, the point cloud in Figure 22(c) will be referred to as the “critical point cloud”.

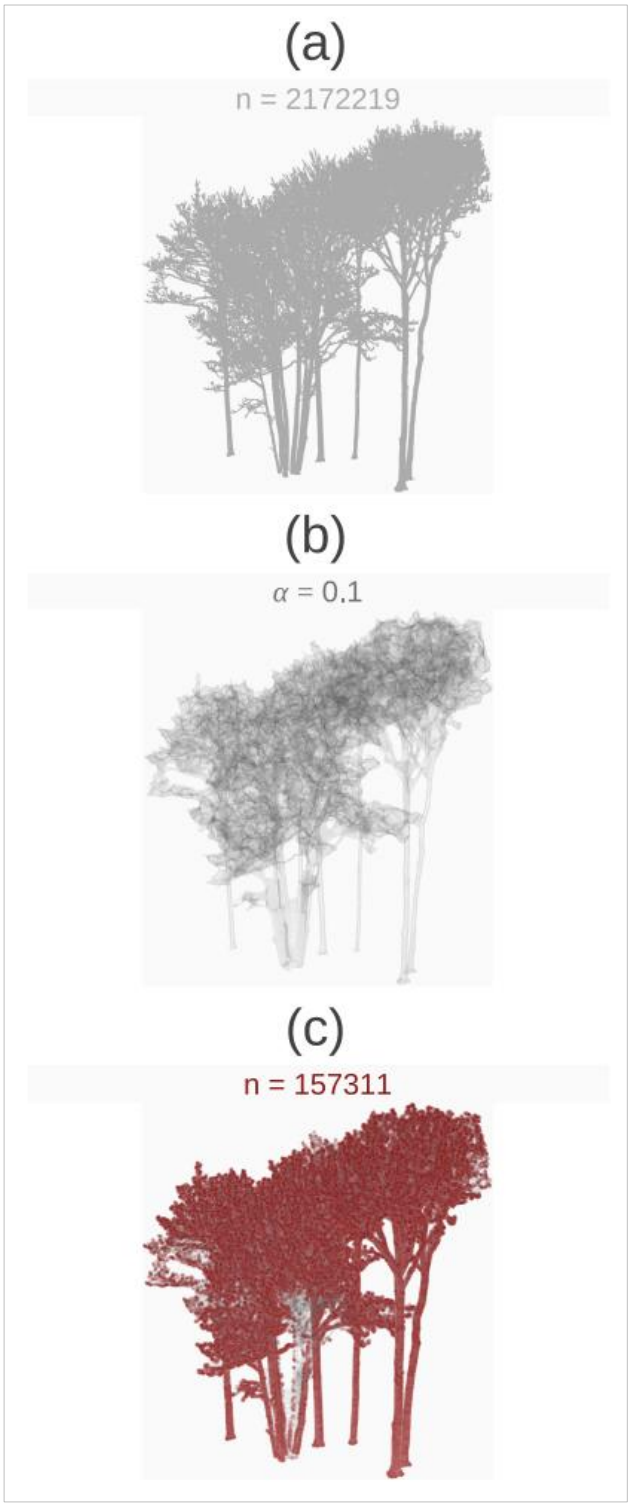


Figure 22. (a) Input 12-member initial tree point cloud. (b) Triangulated tree point cloud using an α value of 0.1. (c) Triangulated tree point cloud overlaid with the computed minima (i.e., critical vertices, points).

4.4.2 Tree Stem Location

Assuming that understory vegetation is removed (as in our TLS dataset), tree trunk location and labeling can be performed by applying HDBSCAN. HDBSCAN is an advanced clustering algorithm derived from DBSCAN and designed for datasets with varying densities. It combines hierarchical clustering with density-based methods to identify clusters and noise effectively (Campello et al., 2013).The algorithm was applied to the xy plane, cutting through both the initial and critical point clouds at a suitable height (z attribute of the point cloud), where the tree trunks were clearly separated from one another (Fig. 23b). In our datasets (both the initial and critical point clouds), the height was set to 3 m (Fig. 23a).

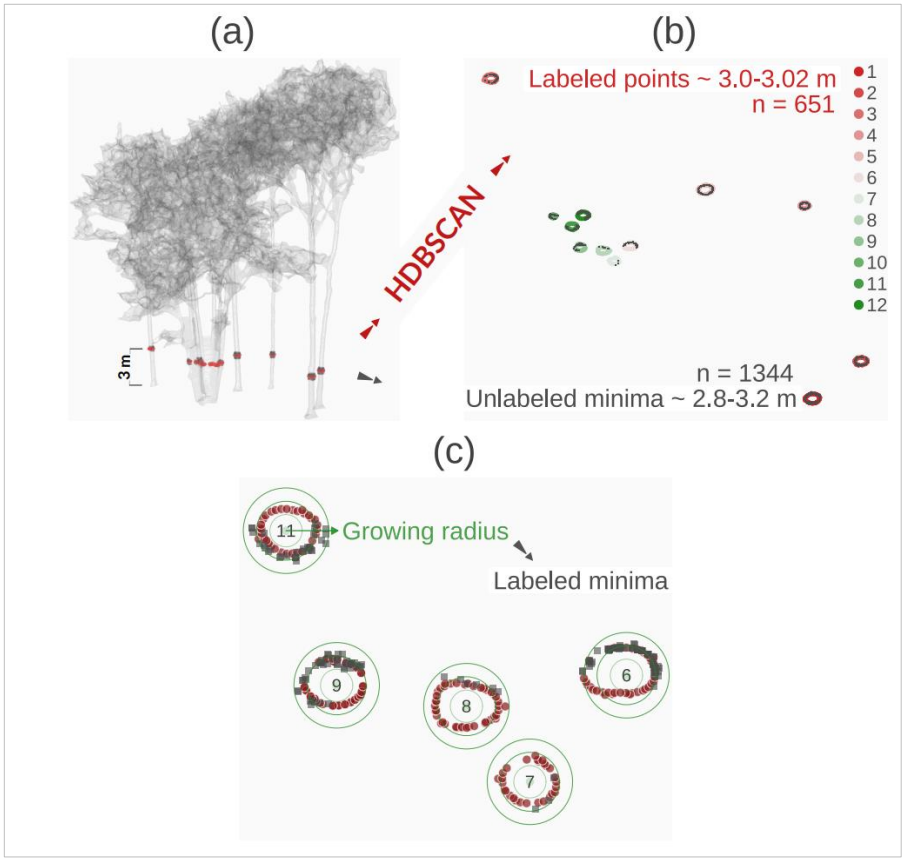


Figure 23. (a) Triangulated 12-member initial tree point cloud overlaid with the minima at height between 2.8 and 3.2 m in grey, and with the points extracted from the initial point cloud between 3.00 and 3.02 m height. (b) Unlabeled minima and its count in grey, HDBSCAN-detected and -labeled (tree ID) points of the initial point cloud scale coloured after a red-green colour palette, point count is expressed in red. (c) Tree IDs 6,7,8,9 and 11 from (b), displaying label-passing, which utilises the growing radius technique (concentric circles).

After applying HDBSCAN, the points from the initial point cloud between the heights of 3.00 and 3.02 meters were tagged with the HDBSCAN-identified tree ID. However, the critical points between 2.8 and 3.2 m remain unlabeled (Fig. 23b). The Label-Passing method is used to transfer the tree ID tags from the labelled initial points (3.00-3.02 m) to the unlabelled critical points (2.8-3.2 m).

4.4.3 Label-Passing

This simple yet robust method captures the unlabelled critical points defined at a height of 2.8–3.2 m during an incremental radius-growing process, which generates increment-defined concentric circle, and passes the tree IDs identified by HDBSCAN from the labelled initial points (3.00–3.02 m) to the critical points (Fig. 23c). The growth of the concentric circles stopped automatically once no additional points were captured within the next annulus (Fig. 23c).

5 RESULTS

5.1 Individual Treetop Location (TREETOPS)

5.1.1 ITD results at site per method

The highest matching rate was found in the KA10 forest, where both methods showed 81% matching. Interestingly, the lowest rate (55 %) was produced by LM with VWF in BR01 forest (Fig. 24a). Regarding our target forest, the GTR method (80 %) slightly outperformed the LM with VWF (77 %). Overall, a minor performance advantage was observed in favour of the GTR method.

Regarding the accuracy of height matching, the GTR method displayed Mean Absolute Error (MAE) values between 0.73 and 2.04 m, while the LM with VWF evidenced MAE values between 1.01 and 2.50 m (Fig. 24a). Across all forest sites, the GTR achieved slightly better MAE results.

It is noteworthy that the two types of matched trees (those located by the LM with VWF and GTR methods matching a particular reference tree) were situated within a distance of less than 4 m from

each other in every case, where both methods matched the same reference tree (Fig. 24b).

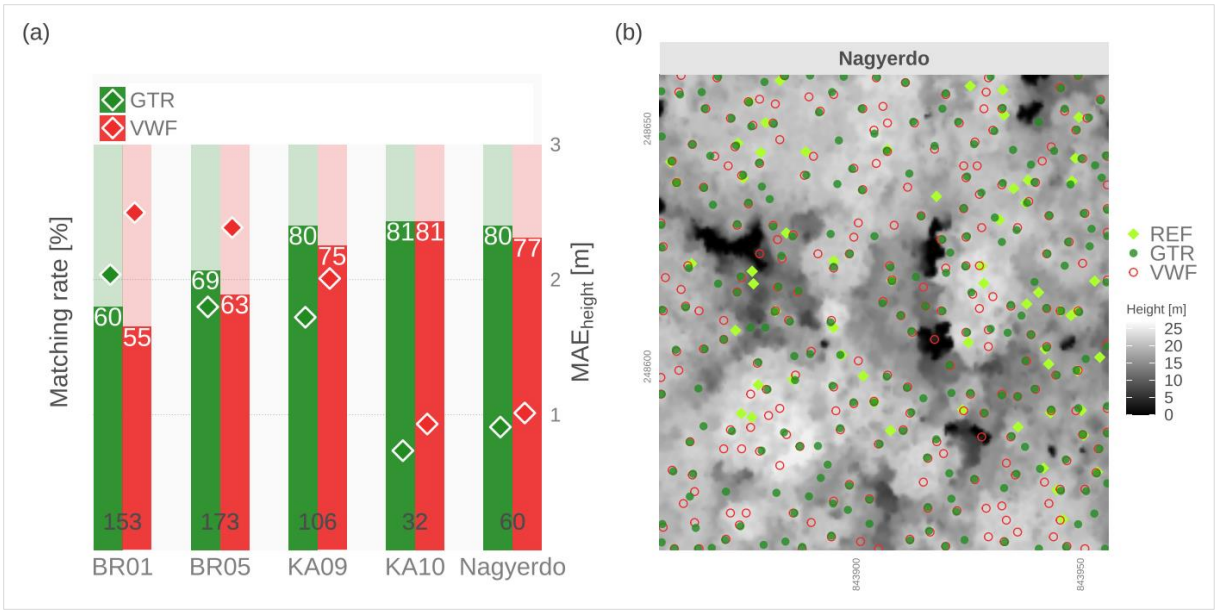


Figure 24. Matching rates concerning forest sites and the dual methods (a; diamonds indicate MAE values, grey numbers on bars represent the number of reference trees for each forest site, white numbers show the matching error rates for each forest depending on the method: GTR, Growing Tree Region, and VWF, Variable Window Filtering). The spatial locations of the 60 reference (REF) trees, GTR-extracted treetops, and Local Maxima with VWF-extracted treetops of the Nagyerdő forest are shown in (b).

Regarding commission errors, no differences were observed between the LM with VWF and the GTR methods at the KA09 site (Fig. 25). While the LM with VWF obtained a 2% lower value in KA10, the three remaining forests - BR01, BR05, and Nagyerdő - witnessed the GTR having 2%, 1%, and 2% lower rates, respectively (Fig. 25).

Both the higher matching rates and lower MAE values achieved by the GTR technique suggest that the GTR approach can locate treetops more accurately than the LM with VWF. This can be explained by sequential raster cutting, which ensures that the maximum protrusion of the CHM is independent of the crown base area. The VWF can be made variable by considering the correlation between tree height and crown base area of the tree, which can be difficult to tune in structurally complex mixed forests (with different species).

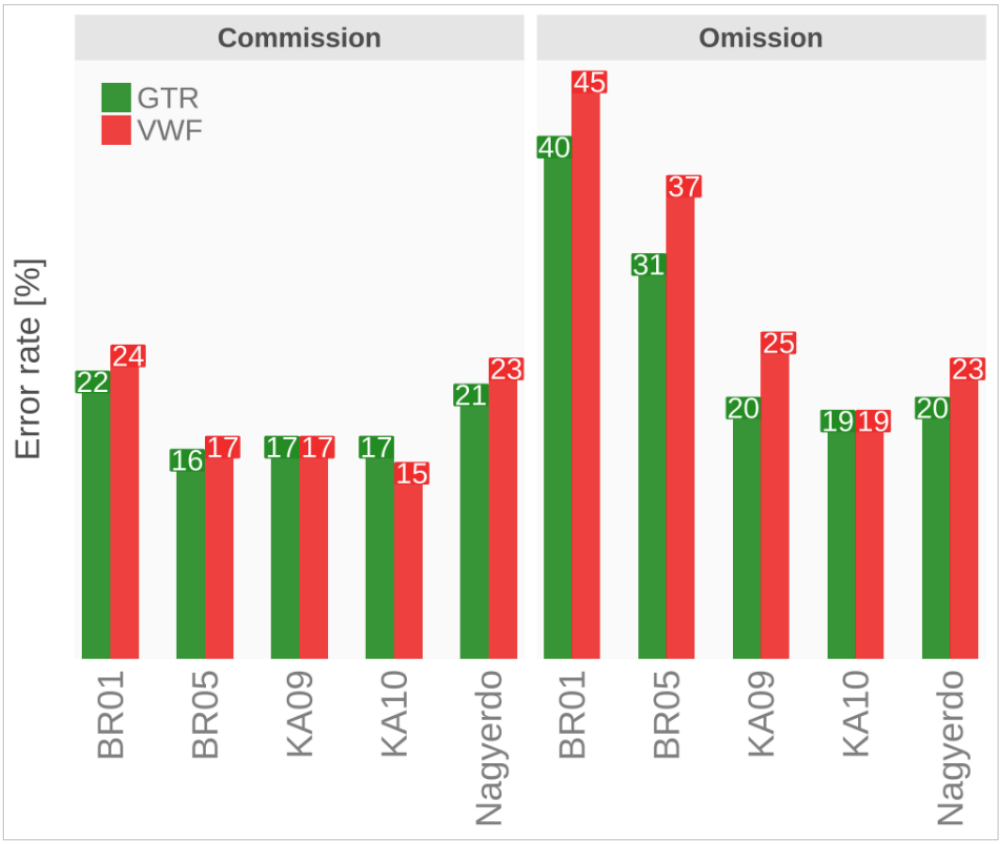


Figure 25. Modified commission and omission error rates for the studied forest sites with regard to the two methods (GTR: Growing Tree Region; VWF: Variable Window Filtering).

The highest omission rate of 45% was found with LM using VWF in BR01, whereas the lowest rate (19%) was observed with both algorithms in the KA10 forest (Fig. 25). For BR01, BR05, KA09, and Nagyerdő, the GTR method exhibited lower omission error rates of 5, 6, 5, and 3%, respectively.

5.1.2 ITD results per method

Quantitatively, both ITD approaches yielded high RMS values: 74% for GTR and 71% for LM with VWF. The RMS differences in the modified commission and omission rates were 1% and 4%, respectively, providing an edge to the GTR (Fig. 26). The modified extraction rates were high for both methods, 140% for GTR and 134% for LM with VWF (Fig. 26).

Figure 26. The overall performance of the compared ITD methods, Root Mean Square (RMS) of matching, modified extraction, modified commission, and omission error rates for the two methods (GTR: Growing Tree Region; VWF: Variable Window Filtering) were computed across all study sites.

5.2 Crown Base Height Location (treecbh)

5.2.1 Matching rate

5.2.1.1 At site per treecbh mode

At sites A and B (Hungary), we achieved matching rates of 41% and 20% in *No treeiso* mode, which were augmented to 45% and 60%, respectively, with the *Default treeiso*. Notably, the interactive visual CBH adjustment further increased these rates to 71% and 77%, respectively (Fig. 27).

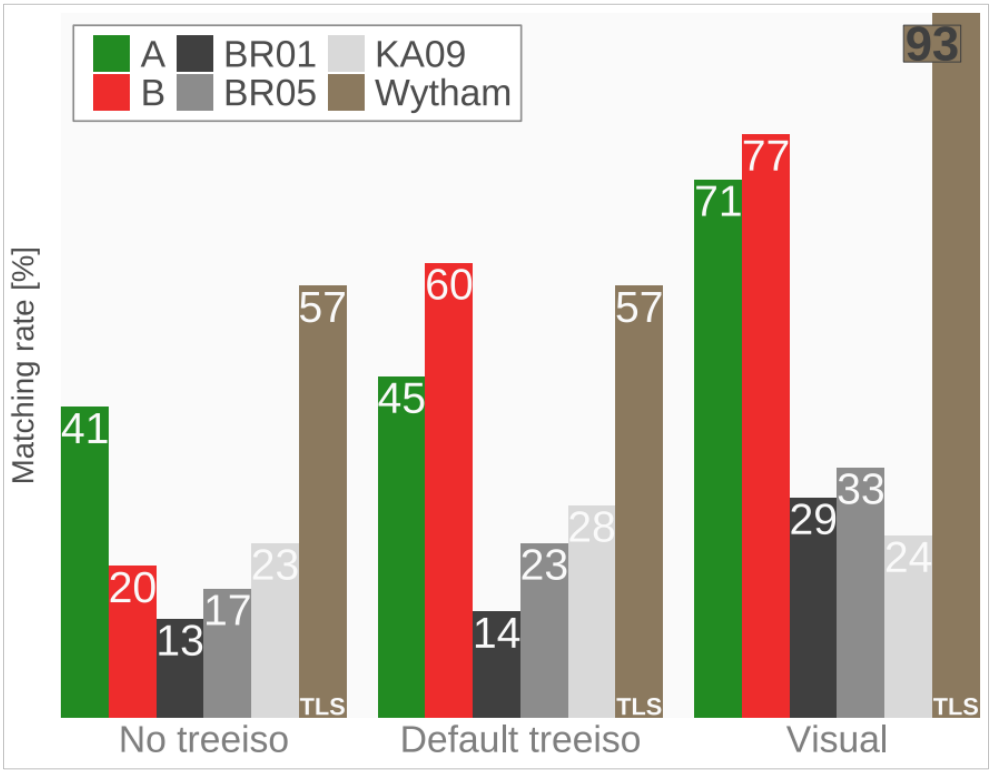


Figure 27. Comparison of Crown Base Height location results from ALS sites A, B, BR01, BR05, KA09, and TLS site Wytham Woods across three operating modes of the *treecbh* tool: deactivated *treeiso* (*No treeiso*), default parameter treeiso (*Default treeiso*), and interactive visual (*Visual*) CBH adjustment mode.

For the three German forests, *treecbh* produced low matching rates of 13–23% in *No treeiso*, 14–28% in *Default treeiso*, and 24–33% in *Visual* mode (Fig. 27). However, what stands out is the consistent enhancement observed across *treecbh*'s operating modes in the BR05 forest. Notably, in the *Visual* mode, BR05 achieved a matching rate of 33%, which was the highest among German forests (Fig. 27).

As expected, the CBH location of the TLS input point clouds was not influenced by the activation of *treeiso* because the input TLS trees were isolated in their default state. Both *No treeiso* and *Default treeiso* modes yielded a 57% matching rate (Fig. 27).

5.2.1.2 *Comparison of leaf-off and leaf-on data*

In the *Default treeiso* mode, the CBH locator attained a matching rate of 60% in forest B, identifying 18 matches and 12 non-matches (Fig. 28). The distinct panels illustrate that the structural information of the tree was effectively captured by the leaf-off input point clouds.

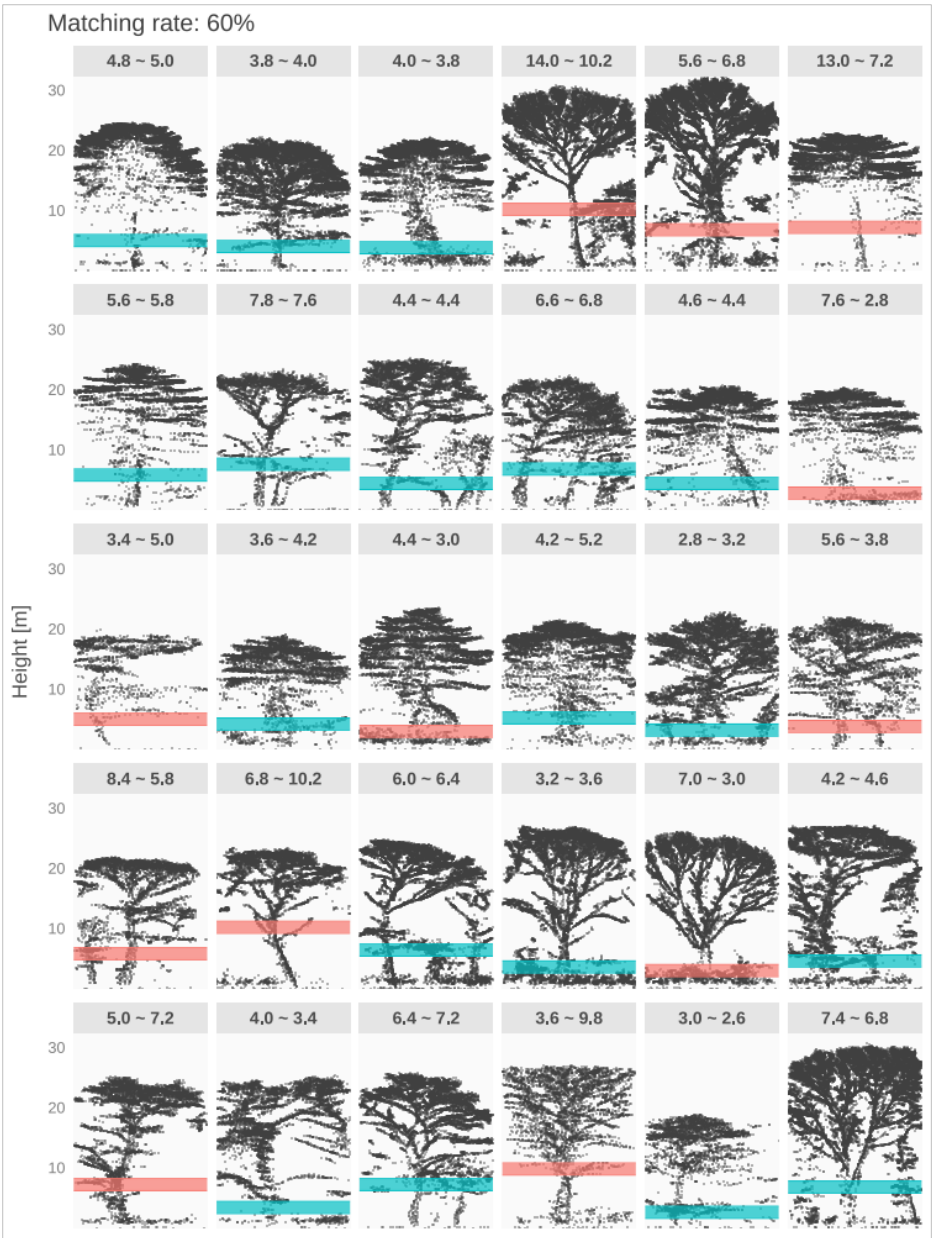


Figure 28. Crown Base Height location results of *treecbh*: matching rate of the leaf-off ALS site B in *Default treeiso* mode. Horizontal transparent red bands (2 m wide) indicate non-matches, whereas transparent green bands (2 m wide) represent matches. Each panel is labelled with the reference CBH and its corresponding estimated CBH separated by the ~ symbol.

In the BR05 German forest, the *Default treeiso* mode of *treecbh* achieved a matching rate of 23%, with 7 matches and 23 non-matches (Fig. 29). It is important to emphasize the insufficient representation of forest structure by the input point clouds, as illustrated in Figure 23. Laser scanning during the leaf-on season resulted in the majority of the laser pulses being reflected by the canopy leaves, hindering effective penetration. As a result, tree structures under and within the canopy were not captured, as was clearly observed in Forest B (Fig. 28).

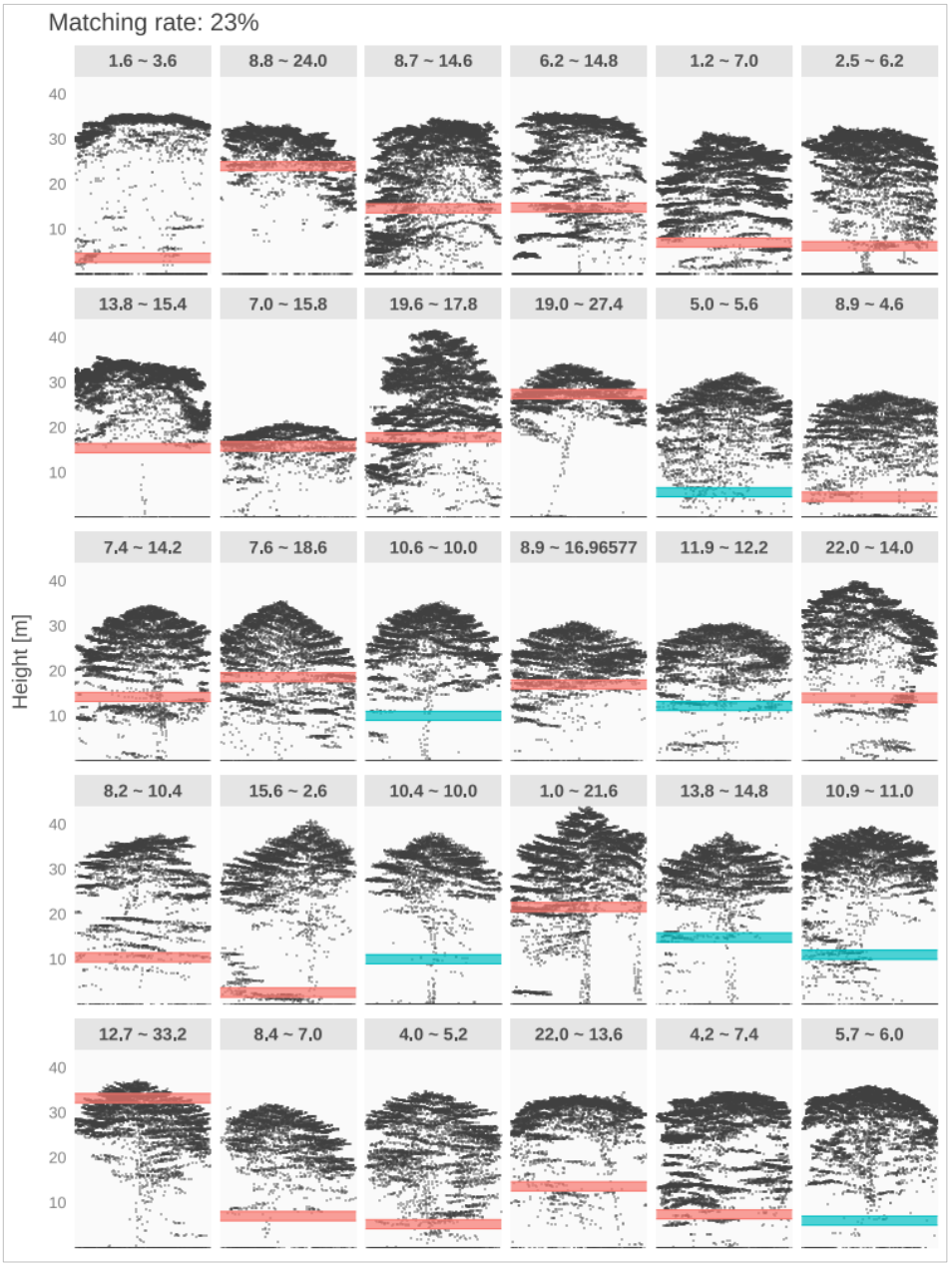


Figure 29. Crown Base Height location results of *treecbh*: matching rate of the leaf-on ALS site BR05 in *Default treeiso* mode. Horizontal transparent red bands (2 m wide) indicate non-matches, whereas transparent green bands (2 m wide) represent matches. Each panel is labelled with the reference CBH and its corresponding estimated CBH separated by the symbol.

5.2.2 MAE at site per treecbh mode

In Nagyerdő forests A and B, MAEs of 2.6 m and 3.7 m were revealed for the *No treeiso* mode. Upon implementing tree isolation, these errors were reduced to 2.1 m and 1.5 m (Fig. 30), respectively. Adjustment of CBH in the *Visual* mode resulted in a significant reduction in errors, with MAEs decreasing to 1.1 m and 0.8 m (Fig. 30).

In comparison, the three German forests exhibited high MAEs across all three CBH locator operating modes (Fig. 30). It is evident that *treecbh* cannot handle the airborne laser scanning data scanned in the German forests.

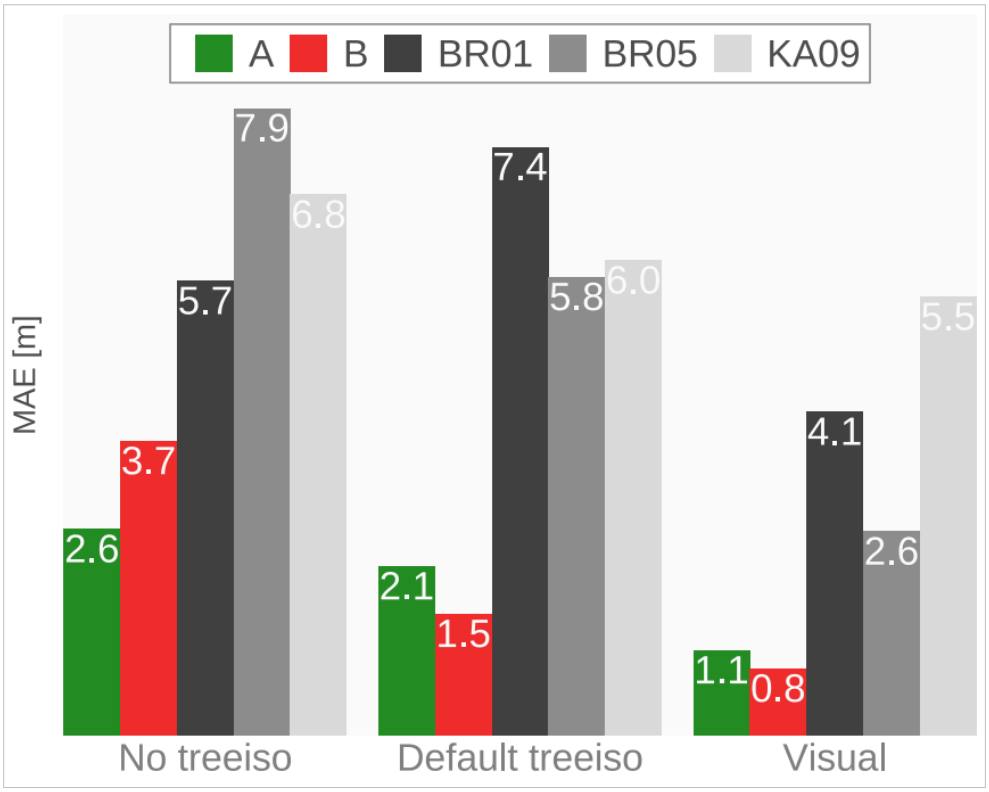


Figure 30. Crown Base Height location results of *treecbh*: Mean Absolute Errors of ALS sites A, B, BR01, BR05, and KA09 regarding *treecbh*'s three operating modes: deactivated *treeiso* (*No treeiso*), default parameter *treeiso* (*Default treeiso*) and interactive visual (*Visual*) CBH adjustment mode, respectively.

5.2.3 Point density sensitivity

The performance evaluation of *treecbh*, focusing solely on the CBH location, revealed the influence of the input point cloud densities on the

results. Contrary to expectations, increasing the point densities above 111 pts/m^2 did not enhance *treecbh*'s performance (Fig. 31). The corresponding performance remained at a 57% matching rate, as shown in Figure 7. Indeed, the 111 pts/m^2 point density aligns with that employed for CBH location on the TLS input point clouds (Table 5).

While *treecbh* demonstrated a 20% matching rate when processing input point clouds with a density of 20 pts/m^2, when presented with higher point densities of 156, 205, and 253 pts/m^2, matching rates ranging from 40% to 43% were achieved (Fig. 31).

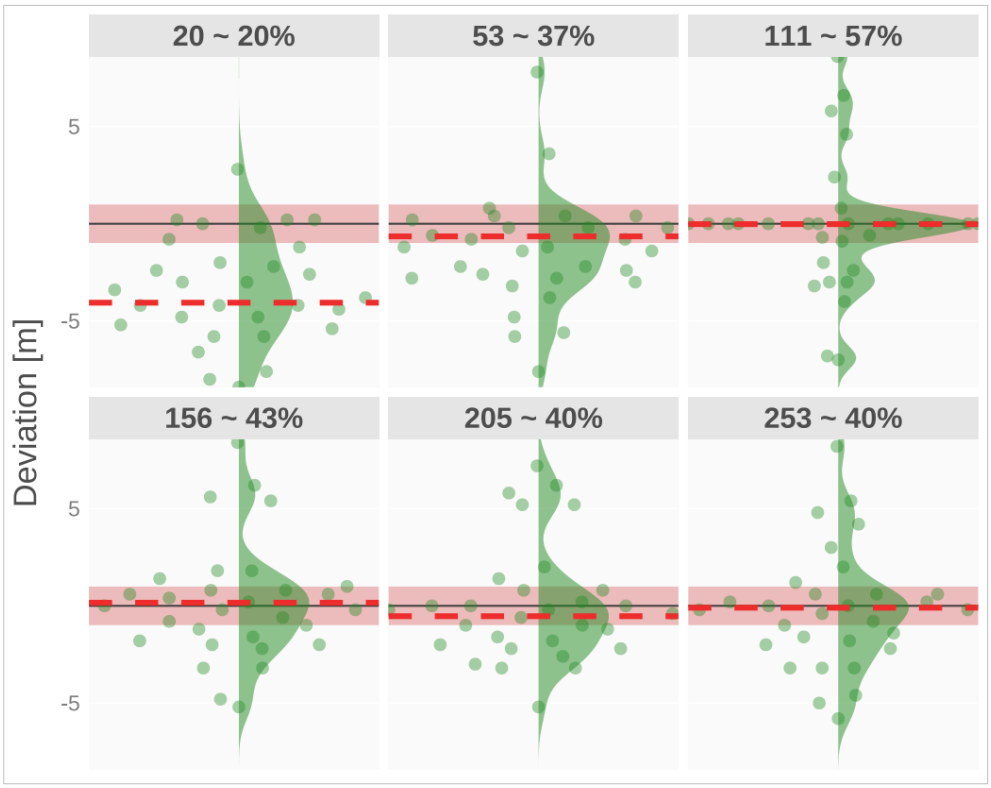


Figure 31. Crown Base Height location results of point sensitivity analysis (*No treesio* mode) for point density classes. The horizontal dark line represents the reference CBHs, and the red horizontal dashed lines mark the maxima of the deviation densities displayed by the vertical green densities, which represent the distributions of the 30 input-point clouds (green dots). The horizontal red transparent band (2 m wide) marked the ±1 m acceptance region. Each panel is labelled with its point density class along with its respective matching rate separated by the symbol.

5.3 Benchmarking results

The development was carried out on an HP ProBook laptop featuring an Intel® Core™ i7-6500U processor with 8 GB of RAM and 4 processing units, each running at 2.50 GHz. The processor, based on a 64-bit architecture, has 2 cores and is supported by an Intel® HD Graphics 520 GPU. The system was operated on Windows® 10 Pro. Forest A in the Nagyerdő was selected for benchmarking purposes, taking advantage of the availability of the two ALS datasets with distinct resolutions. The TREETOPS was applied to the low-resolution dataset, while the treecbh tool utilized the high-resolution dataset. Additionally, the time required for computing the Forman Gradient conducted on the 12-tree TLS data was included (Tables 1 and 4). A comprehensive comparison of the benchmarking results for each tool, along with the computational time, is presented in Table 8.

Table 8. Benchmarking results.

Method	Time	LiDAR data
LM with VWF (lidR)	15 sec	ALS*
TREETOPS	205 sec	ALS*
Treecbh (default treeiso)	1680 sec (28 mins)	ALS**
Forman Gradient	5673 sec (95 mins)	TLS***

* Table 2
** Table 3
*** Table 4

6 DISCUSSION

6.1 Individual Treetop Location

The GTR Method described in this study is a simple CHM-based tree region-growing algorithm that is provided in a ready-to-use form for forest managers as well as the research community. Its open access availability makes it indisputably useful, which is a crucial issue that other researchers tend to ignore (Stereńczak et al., 2020; Zhao et al., 2017). We offer a flexible user-friendly tool that outperforms the widely

used LM with VWF and can be adjusted to various forest ecosystems depending on their vertical structure. The five tested Central-European forests were mixed, deciduous dominated. According to previous studies, similar forest communities make ITD highly demanding (Eysn et al., 2015; Vauhkonen et al., 2012). Based on prior knowledge (Eysn et al., 2015; Stereńczak et al., 2020), we selected possibly visible trees from the ALS perspective using the canopy layer of multi-story forests. Moreover, our approach of accuracy assessment is derived from a method established itself in ITD performance evaluation focussing on Central-European alpine and lowland forests (as described in Section 5.1.2). This enabled our findings to be directly linked to those presented by Eysn et al. (2015). A segmentation-centred study that conducted ITD before delineation in mixed and deciduous forests, similar to the ones in this research, also provided results comparable to our findings (Zhao et al., 2017).

An important aspect of the comparative analysis between the GTR and the LM with VWF is that the VWF was calibrated for the forest A of Nagyerdő (Section 5.1.1). Notably, GTR was mathematically adopted to make it logically comparable to VWF, meaning that the threshold setting capability (described in Section 4.1.3) of the GTR tool was not thoroughly facilitated.

Eysn et al. (2015) analysed the performance of eight different ITD methods and reported RMS matching rates between 66% and 82% for trees above 20 m (Eysn et al., 2015). One of their best performers (considering extraction, commission, and omission error rates), LM with VWF, achieved 72%. Although their research areas were located in the Alpine region, our results (74% revealed by GTR and 71% by VWF) fell in line with theirs. This phenomenon suggests, unlike statements from other studies (González-Ferreiro et al., 2013a; Khosravipour et al., 2014), that dominant and co-dominant treetops can be equivalently extracted independently of the topographic characteristics of the forest. In general, the declared overall performance was comparable. While an article published by Eysn et al. (2015) described 47% as the best matching rate (Eysn et al., 2015), our study showed 52% for both methods. These correspondences of results underline the plausibility of our proposed method concerning commission and extraction adjustments.

At Nagyerdő site A, the GTR matching rate evidenced 80%,

similar ITD results with percentages between 80 and 87 were reported in temperate less dense boreal forests of deciduous nature (Zhao et al., 2017). This analogy can be explained by the fact that both algorithms applied the theoretically similar top-down CHM cutting approach.

Although GTR outperformed the calibrated LM with VWF, the obtained RMS differences concerning matching, modified commission and omission rates are marginal. While the VWF was used throughout this study in its calibrated form based on Nagyerdő site A, the GTR was adjusted to conduct a logically founded comparative analysis. Thus, it can be stated that the GTR is a competitive alternative to the LM with VWF, and it should also be optimized by using its threshold-setting feature without limitations. However, a calibrated VWF can match the robustness of the non-optimized GTR method in lowland mixed and deciduous forests.

Three additional issues need to be addressed. First, considering the CHM as a topographic surface makes it difficult for the GTR to detect suppressed trees and trees with flatly shaped crowns because such objects lack the necessary protrusions emerging from the CHM. This issue becomes relevant if treetop distances in the same height class (as defined by threshold setting) of the studied forest vary, and the output of the *finalize_TREETOPS()* function would be erroneous. Second, the lack of abundant reference trees covering the target (Nagyerdő forest) study area led us to develop a strategy for modified commission and extraction rates, which were used to evaluate the performance of both algorithms at the test sites (BR01, BR05, KA09, and KA10). Although plausible, this strategy is a prone-to-error replacement for commission and extraction rates obtained using forest inventory data covering the entire study area. Third, the GTR was tested in temperate Central European mixed- and deciduous-dominated forests; thus, the robustness of the algorithm in other forested ecosystems is unknown. It is reasonable to assume that the algorithm would perform better in a coniferous environment, as ITD in coniferous forests has been reported to achieve higher accuracy. The robustness of the GTR method should be evaluated in the future using other types of LiDAR forest data, combined with area-covering forest inventory data from different forested ecosystems.

6.2 Crown Base Height Location

6.2.1 Performance of treecbh

Our findings highlight that the performance of our proposed CBH locator is primarily influenced by the quality of the input data, specifically the ability of the input point clouds to accurately describe the structure of trees under or within the canopy (Fig. 28 and 29). Before utilizing *treecbh*, we recommend that users assess the input point cloud and verify whether the input point cloud sufficiently represents the tree structure, particularly the tree trunk and the lowest branches. This is a basic requirement for the algorithm to achieve meaningful results.

Given that leaf-on scanning tends to produce high point densities within the canopy (Fig. 29), but limited densities for the lower parts of the canopy, leaf-on data is less suitable as input to *treecbh*. *treecbh*'s CBH location technique relies on points representing the tree trunk and lower branches (as detailed in Sections 4.2.1–4.2.3) and it is therefore imperative that the structure of trees beneath the canopy is adequately captured by the input point data. Therefore, the preferred input to *treecbh* is leaf-off ALS data (Fig. 28). Moreover, to effectively utilize *Treecbh* 's *visual* mode, ensuring the “readability” of the tree structure (i.e., the tree trunk and its lowest branches) is essential. Users must be able to reliably identify the CBH in order to provide the assumed CBH during the interactive visual CBH adjustment process (Fig. 21).

treecbh demonstrates optimal performance using the leaf-off ALS point cloud of Forest B (Fig. 27), which had a conifer-broadleaf ratio of 0.1:0.9 (Table 4). In contrast, for the point cloud of forest A, which was characterized by a conifer-broadleaf ratio of 0.4:0.6, *treecbh* showed reduced robustness (Fig. 27). The lowest performance was observed using the point cloud from the leaf-off BR01 forest, which had a conifer-broadleaf ratio of 0.6:0.4 (Table 4). These findings underscore the robustness of our CBH locator in broadleaf forests compared with mixed conifer forests. Thus, our CBH locator is limited to performing well in broadleaf forests, and we recommend using the current version of the locator in such a setting.

Interestingly, the performance of *treecbh* in the *No treeiso* mode improved only when the point density increased from 20 pts/m^2 to 111 pts/m^2 (Fig. 31). This suggests that a minimum point density is required for the sound operation of *treecbh*. A point density of around 100

pts/m^2 appears to be adequate for effectively representing tree structure. To evaluate *treecbh*'s CBH location, we used a mean point density of 111 pts/m^2 for the TLS point cloud (Table 3 and Fig. 27). This density was chosen because it proofed to visually well represent the tree structure. However, it is important to note that the TLS data collected during leaf-off conditions may be perceived as less realistic when assuming an operational scenario compared to ALS datasets from Nagyerdő, Bretten, and Karlsruhe forests, as the TLS data provided perfectly delineated individual tree point clouds.

6.2.2 Future development of treecbh

Addressing the inherent structural intricacies within point clouds is crucial. The vertical arrangement of a forest (single-, double-, or multilayered) significantly influences the outcomes of treetop location and crown segmentation (Eysn et al., 2015). Consequently, *treecbh* faces specific challenges in under-canopy environments, particularly in isolating tree stems in the lower canopy and removing understory vegetation.

To isolate tree stems effectively in the lower canopy, an additional algorithm is necessary, for example the two-stage cut-pursuit clustering algorithm (Xi and Hopkinson, 2022), which was implemented in this study. Enhancing the performance of our CBH locator should start with the implementation of a more robust tree isolator, such as 3DFin, which is available as a CloudCompare plugin, as well as a Python module (Laino et al., 2024).

Another important aspect for the future development of *treecbh involves the identification and effective* removal of understory vegetation (Luo et al., 2018). This task must be carried out at the individual tree level, especially in multispecies heterogeneous forest stands.

From an algorithmic perspective (Section 4.2.1, Figure 18), vertical cross-sectional K-means clustering can be replaced with faster algorithms that analyse horizontal expansion (on the xy plane). Additionally, combining the current 2D density kernel method with a replacement for the K-Means clustering technique could improve computational efficiency and accuracy.

In its current implementation, *treecbh* utilizes the *minH* formula (Eq. 5) to remove understory vegetation. Its scaling parameter,

min_H_scale, was consistently set at 0.13 throughout this study (Sections 4.2.2 and 4.3.2). Although adjusting *min_H_scale* has demonstrated potential for enhancing CBH location accuracy, determining its optimal value necessitates iterative testing. We refrained from doing so in this study, but future work could focus on this as well as refining the *minH* formula to enhance the accuracy of the CBH location.

6.2.3 Operational use of treecbh

Compared with previous studies that focused on CBH estimation at the individual tree level, our CBH locator was competitive only when its *visual* mode was utilized on leaf-off data, achieving MAE values of 0.8 and 1.1 m. Vauhkonen (2010) reported a Root Mean Square Error (RMSE) between 1.54 m to 3.56 m for Scots pine (Vauhkonen, 2010). In a similar study that estimated CBH for both broadleaf and conifer species, an average RMSE of 2 m was obtained (Popescu and Zhao, 2008). Similarly, an RMSE of approximately 1.6 m has been achieved in mixed-species forests (Luo et al., 2018). These studies used ALS data of 4 to 10 pts/m^2 point densities, scanned in structurally simpler forests compared to our study areas. Tree isolation and understory removal have not been reported to be particularly challenging in these forests (Luo et al., 2018; Popescu and Zhao, 2008). Addressing these challenges (Section 4.2) was a key driver in the development of *treecbh*.

In our study, we used manually extracted tree segments to generate individual tree input point clouds (Section 3.3.2.1). For the operational implementation of our CBH locator, this segmentation step must precede the use of *treecbh*. The second crucial step involves employing a tree isolator that is capable of effectively identifying the tree trunk and lower branches within the input point cloud. To address potential inaccuracies in CBH location due to imperfect tree isolation, *treecbh* offers a visual CBH adjustment mode. However, relying solely on the workflow as outlined in this study is suboptimal for operational viability, partly because of the manual segmentation process and partly because of reliance on visual adjustment; alternatively, achieving high CBH location accuracy in the *Default treeiso* mode is feasible if the tree isolator used is sufficiently robust (as described in Section 7.2.2) to handle inaccuracies such as the presence of other trees or parts

neighbouring trees in the input point cloud. Implementing a more robust tree isolator would facilitate automation of the initial segmentation process. Automating this segmentation process using any suitable segmentation algorithm would significantly enhance the operational viability of *treecbh*.

6.3 Future development of the Tree Stem Locator to Individual Tree Isolator

The core concept of tree isolation in our study is rooted in the principle of vertical connectivity, which is characterized by the detection of minima at heights of 3 m ± 0.2 m. By following the V-path separatrix edge-path, which links saddles (critical edges) to minima (critical vertices), it becomes possible to group these minima into distinct clusters (Xu et al., 2023). This connection of minima is inherently bi-directional: descending towards the base of the trunk and ascending towards the tree crown. Consequently, our tree-stem location method provides a pathway for fully isolating tree stems. However, a key question arises: Is the intermediate step of stem isolation truly necessary if direct tree isolation, bypassing this step, is achievable?

In order to develop an effective tree isolation technique using Terrestrial Laser Scanning (TLS) data, three significant challenges must be addressed within a bottom-to-top approach:

1. Understory removal (might be unnecessary)
2. Tree stem location and isolation (3D solution)
3. Tree isolation: Connecting tree crowns to the previously located stems

In our ongoing work, "Individual Tree Isolation," we have successfully tackled the initial step of tree stem location and isolation. However, a more robust 3D tree stem detection method could be developed by computing and thresholding the linearity values of the input point cloud (Hackel et al., 2016). It is hypothesized that tree stem isolation may also be achievable without the need for understory removal. This remains a subject for future research.

Although the presented tree stem isolator method has yet to be fully validated, it is designed to be highly robust and could serve as a

solid foundation for the development of a comprehensive tree isolation technique. Addressing the aforementioned challenges will further refine the process, with tree stem location forming the cornerstone of a promising approach to isolate individual trees in complex forest environments.

7 SUMMARY AND CONCLUSIONS

7.1 Treetop Location (TREETOPS)

In the first part of this dissertation, the development of a simple yet robust GTR-based treetop location method was presented. The method was based on LiDAR-derived raster data (CHM). The GTR algorithm consists of three steps: CHM incremental cutting and storing to stack all height layers, a three-layer concept to locate individual treetops, and the number of treetop reductions to filter treetops based on distance. Two functions, *get_TREETOPS()* and *finalize_TREETOPS()* were realized in the R package named TREETOPS for applying the algorithm. While the former performs CHM incremental cutting and storing plus individual treetop location (applying the three-layer concept) steps, the latter facilitates the definition of various distances for different tree height bins (as set by threshold setting). Additionally, an adapted strategy using incomplete field-based data was applied for the comparative evaluation of our method with calibrated local maxima with variable window filtering. The retrieved results (55-81% matching rate, RMS of 19 vs. 20% for commission, and 27 vs. 31% for omission) were accurate, revealing the robustness of both methods, giving the GTR a marginal lead over the LM with VWF.

Future studies should focus on evaluating the GTR method for various types of forests worldwide. Because the tool is freely available, GTR combined with different segmentation techniques can be tested within the programming framework of the lidR package. A two-step procedure for tree detection in multilayered forests could be feasible: CHM-based treetop location and tree crown delineation, followed by point-cloud-based tree segmentation within the CHM-based previously delineated boundary.

7.2 Crown Base Height Location (treecbh)

The subsequent study in this dissertation focused on the description of a proposed robust CBH location method using high-resolution ALS data at the individual tree level. The CBH locator was realized in a user-friendly R package intended for use within the framework of the lidR package. The locator *treecbh* is the only available open-access tool to determine the canopy base height. The tool is equipped with fixed default parameters, enabling straightforward execution, while providing users with an option for visual CBH location adjustment. An innovative aspect of this adjustment process is its interactive nature, where *treecbh* engages with the user. Our investigations revealed that when evaluated against field-measured reference data, the default parameter set yielded CBH matching rates of 45-60% for leaf-off ALS data. However, these rates were outperformed by visual CBH adjustment, which achieved matching rates of 71-77%.

Users are advised to use leaf-off ALS data and become familiar with the quality of point clouds and their information content before utilizing *treecbh*. This familiarity helps in deciding whether to use the interactive visual CBH adjustment feature. Additionally, because tree-isolator *treeiso* is integrated within *treecbh*, users can perform only tree isolation by deactivating the CBH location. Notably, users can also expedite the visual CBH adjustment process by deactivating the tree isolator.

As we look ahead, our future endeavours involve enhancing *treecbh*'s performance by replacing *treesio* with a more robust tree isolator.

7.3 Tree Stem Location

Our research demonstrated the effectiveness of tree location identification through tree stem location by utilizing the HDBSCAN algorithm on TLS point clouds at a height where tree stems are distinctly separated. This method, combined with a novel label-passing technique, allows for the successful transfer of tree IDs identified by HDBSCAN to the critical point cloud, representing the critical vertices of the Discrete Morse-induced Forman gradient, derived from the triangulated input tree point cloud. Futurere development will involve further refining this

technique through the application of V-path clustering (Xu et al., 2023). This process entails two phases: (1) V-path-defined point clustering and (2) cluster connectivity. These steps aim to achieve complete tree isolation, which is crucial for accurate forest management and analysis. This approach presents both algorithmic and methodological challenges, particularly as we aim to develop this system as an accessible, open-source tool. The end goal is to create a plugin for drag-and-drop software such as QGIS, which will empower users, including forest managers and researchers, to implement this tree isolation process seamlessly.

8 TOOL AVAILABILTY

TREETOPS and *treecbh* packages are available at https://github.com/DijoG/TREETOPS and https://github.com/DijoG/treecbh, can be installed by executing the following lines:

devtools::install_github("DijoG/TREETOPS")

devtools::install_github("DijoG/treecbh").

9 ACKNOWLEDGEMENT

The research was supported by the NKFI K138079 and the KKP 144068 projects during manuscript preparation. We are grateful to Envirosense Hungary Ltd. for providing the ALS data used in this study.

10 REFERENCES

Andersen, H.-E., McGaughey, R.J., Reutebuch, S.E., 2005. Estimating forest canopy fuel parameters using LIDAR data. Remote Sens. Environ. 94, 441–449. https://doi.org/10.1016/j.rse.2004.10.013

Beaudoin, A., Bernier, P., Villemaire, P., Guindon, L., Guo, X., 2017. Tracking forest attributes across Canada between 2001 and 2011 using a kNN mapping approach applied to MODIS imagery. Can. J. For. Res. 48. https://doi.org/10.1139/cjfr-2017-0184

Beucher, S., Meyer, F., 1993. Segmentation: The Watershed Transformation. Mathematical Morphology in Image Processing. Opt. Eng. 34, 433–481.

Bianchi, S., Siipilehto, J., Hynynen, J., 2020. How structural diversity affects Norway spruce crown characteristics. For. Ecol. Manag. 461, 117932. https://doi.org/10.1016/j.foreco.2020.117932

Blackard, J.A., Finco, M.V., Helmer, E.H., Holden, G.R., Hoppus, M.L., Jacobs, D.M., Lister, A.J., Moisen, G.G., Nelson, M.D., Riemann, R., Ruefenacht, B., Salajanu, D., Weyermann, D.L., Winterberger, K.C., Brandeis, T.J., Ph.D..Czaplewski, R.L., McRoberts, R.E., Patterson, P.L., Tymcio, R.P., 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sens. Environ. 112 1658-1677 12, 1658–1677. https://doi.org/10.1016/j.rse.2007.08.021

Botequim, B., Fernandes, P., Borges, J., González-Ferreiro, E., Guerra, J., 2019. Improving silvicultural practices for Mediterranean forests through fire behaviour modelling using LiDAR-derived canopy fuel characteristics. Int. J. Wildland Fire. https://doi.org/10.1071/WF19001

Bouvier, M., Durrieu, S., Fournier, R.A., Renaud, J.-P., 2015. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sens. Environ. 156, 322–334. https://doi.org/10.1016/j.rse.2014.10.004

Brosofske, K.D., Froese, R.E., Falkowski, M.J., Banskota, A., 2014. A Review of Methods for Mapping and Prediction of Inventory Attributes for Operational Forest Management. For. Sci. 60, 733–756. https://doi.org/10.5849/forsci.12-134

Burt, A., Disney, M., Calders, K., 2019. Extracting individual trees from lidar point clouds using treeseg. Methods Ecol. Evol. 10, 438–445. https://doi.org/10.1111/2041-210X.13121

Calders, K., Adams, J., Armston, J., Bartholomeus, H., Bauwens, S., Bentley, L.P., Chave, J., Danson, F.M., Demol, M., Disney, M., Gaulton, R., Krishna Moorthy, S.M., Levick, S.R., Saarinen, N., Schaaf, C., Stovall, A., Terryn, L., Wilkes, P., Verbeeck, H., 2020. Terrestrial laser scanning in forest ecology: Expanding the horizon. Remote Sens. Environ. 251, 112102. https://doi.org/10.1016/j.rse.2020.112102

Calders, K., Verbeeck, H., Burt, A., Origo, N., Nightingale, J., Malhi, Y., Wilkes, P., Raumonen, P., Bunce, R.G.H., Disney, M., 2022. Laser scanning reveals potential underestimation of biomass carbon in

temperate forest. Ecol. Solut. Evid. 3, e12197. https://doi.org/10.1002/2688-8319.12197
Campello, R.J.G.B., Moulavi, D., Sander, J., 2013. Density-Based Clustering Based on Hierarchical Density Estimates, in: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (Eds.), Advances in Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, pp. 160–172. https://doi.org/10.1007/978-3-642-37456-2_14
Campello, R.J.G.B., Moulavi, D., Zimek, A., Sander, J., 2015. Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. ACM Trans Knowl Discov Data 10, 5:1-5:51. https://doi.org/10.1145/2733381
Chang, L., Fan, H., Zhu, N., Dong, Z., 2022. A Two-Stage Approach for Individual Tree Segmentation From TLS Point Clouds. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 8682–8693. https://doi.org/10.1109/JSTARS.2022.3212445
Coops, N.C., Tompalski, P., Goodbody, T.R.H., Queinnec, M., Luther, J.E., Bolton, D.K., White, J.C., Wulder, M.A., van Lier, O.R., Hermosilla, T., 2021. Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. Remote Sens. Environ. 260, 112477. https://doi.org/10.1016/j.rse.2021.112477
Dalponte, M., Coomes, D., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. 7. https://doi.org/10.1111/2041-210X.12575
De Floriani, L., Fugacci, U., Iuricich, F., Magillo, P., 2015. Morse complexes for shape segmentation and homological analysis: discrete models and algorithms. Comput. Graph. Forum 34, 761–785. https://doi.org/10.1111/cgf.12596
Dean, T.J., Cao, Q.V., Roberts, S.D., Evans, D.L., 2009. Measuring heights to crown base and crown median with LiDAR in a mature, even-aged loblolly pine stand. For. Ecol. Manag. 257, 126–133. https://doi.org/10.1016/j.foreco.2008.08.024
Duncanson, L.I., Cook, B.D., Hurtt, G.C., Dubayah, R.O., 2014. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. Remote Sens. Environ. 154, 378–386. https://doi.org/10.1016/j.rse.2013.07.044
Edelsbrunner, H., 2012. Herbert Edelsbrunner. Wiad. Mat. 48, 47. https://doi.org/10.14708/wm.v48i2.316

Ene, L., Næsset, E., Gobakken, T., 2012. Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. Int. J. Remote Sens. 33, 5171–5193. https://doi.org/10.1080/01431161.2012.657363

Engelstad, P.S., Falkowski, M., Wolter, P., Poznanovic, A., Johnson, P., 2019. Estimating Canopy Fuel Attributes from Low-Density LiDAR. Fire 2, 38. https://doi.org/10.3390/fire2030038

Erdody, T.L., Moskal, L.M., 2010. Fusion of LiDAR and imagery for estimating forest canopy fuels. Remote Sens. Environ. 114, 725–737. https://doi.org/10.1016/j.rse.2009.11.002

Eysn, L., Hollaus, M., Lindberg, E., Berger, F., Monnet, J.-M., Dalponte, M., Kobal, M., Pellegrini, M., Lingua, E., Mongus, D., Pfeifer, N., 2015. A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space. Forests 6, 1721–1747. https://doi.org/10.3390/f6051721

Finney, M.A., 2006. An Overview of FlamMap Fire Modeling Capabilities. Andrews Patricia Butl. Bret W Comps 2006 Fuels Manag.- Meas. Success Conf. Proc. 28-30 March 2006 Portland Proc. RMRS-P-41 Fort Collins CO US Dep. Agric. For. Serv. Rocky Mt. Res. Stn. P 213-220 041.

Finney, M.A., 1998. FARSITE: Fire Area Simulator-model development and evaluation. Res Pap RMRS-RP-4 Revis. 2004 Ogden UT US Dep. Agric. For. Serv. Rocky Mt. Res. Stn. 47 P 4. https://doi.org/10.2737/RMRS-RP-4

Forman, R., 1998. Morse Theory for Cell Complexes. Adv. Math. 134, 90–145. https://doi.org/10.1006/aima.1997.1650

González-Ferreiro, E., Diéguez-Aranda, U., Barreiro-Fernández, L., Buján, S., Barbosa, M., Suárez, J.C., Bye, I.J., Miranda, D., 2013a. A mixed pixel- and region-based approach for using airborne laser scanning data for individual tree crown delineation in Pinus radiata D. Don plantations. Int. J. Remote Sens. 34, 7671–7690. https://doi.org/10.1080/01431161.2013.823523

González-Ferreiro, E., Diéguez-Aranda, U., Crecente-Campo, F., Barreiro-Fernández, L., Miranda, D., Castedo-Dorado, F., 2013b. Modelling canopy fuel variables for Pinus radiata D. Don in NW Spain with low density LiDAR data. Int. J. Wildland Fire 23. https://doi.org/10.1071/WF13054

González-Ferreiro, E., Diéguez-Aranda, U., Miranda, D., 2012. Estimation of stand variables in Pinus radiata D. Don plantations

using different LiDAR pulse densities. For. Int. J. For. Res. 85, 281–292. https://doi.org/10.1093/forestry/cps002
Gyulassy, A., Bremer, P.-T., Hamann, B., Pascucci, V., 2008. A Practical Approach to Morse-Smale Complex Computation: Scalability and Generality. IEEE Trans. Vis. Comput. Graph. 14, 1619–1626. https://doi.org/10.1109/TVCG.2008.110
Hackel, T., Wegner, J.D., Schindler, K., 2016. Contour Detection in Unstructured 3D Point Clouds, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 1610–1618. https://doi.org/10.1109/CVPR.2016.178
Hackenberg, J., Spiecker, H., Calders, K., Disney, M., Raumonen, P., 2015. SimpleTree —An Efficient Open Source Tool to Build Tree Models from TLS Clouds. Forests 6, 4245–4294. https://doi.org/10.3390/f6114245
Hantsch, L., Bien, S., Radatz, S., Braun, U., Auge, H., Bruelheide, H., 2014. Tree diversity and the role of non-host neighbour tree species in reducing fungal pathogen infestation. J. Ecol. 102, 1673–1687. https://doi.org/10.1111/1365-2745.12317
Hardenbol, A.A., Korhonen, L., Kukkonen, M., Maltamo, M., 2023. Detection of standing retention trees in boreal forests with airborne laser scanning point clouds and multispectral imagery. Methods Ecol. Evol. 14, 1610–1622. https://doi.org/10.1111/2041-210X.13995
Hermosilla, T., Ruiz, L.A., Kazakova, A.N., Coops, N.C., Moskal, L.M., 2014. Estimation of forest structure and canopy fuel parameters from small-footprint full-waveform LiDAR data. Int. J. Wildland Fire 23, 224–233. https://doi.org/10.1071/WF13086
Hijmans, R.J., Bivand, R., Pebesma, E., Sumner, M.D., 2023. terra: Spatial Data Analysis.
Hsu, W.-C., Shih, P.T.-Y., Chang, H.-C., Liu, J.-K., 2015. A Study on Factors Affecting Airborne LiDAR Penetration. Terr. Atmospheric Ocean. Sci. 26, 241. https://doi.org/10.3319/TAO.2014.12.02.08(EOSI)
Hu, B., Li, J., Jing, L., Judah, A., 2014. Improving the efficiency and accuracy of individual tree crown delineation from high-density LiDAR data. Int. J. Appl. Earth Obs. Geoinformation 26, 145–155. https://doi.org/10.1016/j.jag.2013.06.003
Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser

scanning for extracting forest inventory data in boreal forests. Int. J. Remote Sens. - INT J REMOTE SENS 29, 1339–1366. https://doi.org/10.1080/01431160701736489
Jakubowksi, M.K., Guo, Q., Collins, B., Stephens, S., Kelly, M., 2013. Predicting Surface Fuel Models and Fuel Metrics Using Lidar and CIR Imagery in a Dense, Mountainous Forest. Photogramm. Eng. Remote Sens. 79, 37–49. https://doi.org/10.14358/PERS.79.1.37
Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B.M., Wu, J.-C., 2012. An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning. Remote Sens. 4, 950–974. https://doi.org/10.3390/rs4040950
Kändler, G., 2006. The design of the second German national forest inventory.
Kelly, M., Su, Y., Di Tommaso, S., Fry, D.L., Collins, B.M., Stephens, S.L., Guo, Q., 2018. Impact of Error in Lidar-Derived Canopy Height and Canopy Base Height on Modeled Wildfire Behavior in the Sierra Nevada, California, USA. Remote Sens. 10, 10. https://doi.org/10.3390/rs10010010
Khosravipour, A., Skidmore, A.K., Isenburg, M., Wang, T., Hussin, Y.A., 2014. Generating Pit-free Canopy Height Models from Airborne Lidar. Photogramm. Eng. Remote Sens. 80, 863–872. https://doi.org/10.14358/PERS.80.9.863
Koch, B., Heyder, U., Weinacker, H., 2006. Detection of Individual Tree Crowns in Airborne Lidar Data. Photogramm. Eng. Remote Sens. 72, 357–363. https://doi.org/10.14358/PERS.72.4.357
Korhonen, L., Vauhkonen, J., Virolainen, A., Hovi, A., Korpela, I., 2013. Estimation of tree crown volume from airborne lidar data using computational geometry. Int. J. Remote Sens. 34, 7236–7248. https://doi.org/10.1080/01431161.2013.817715
Krisanski, S., Taskhiri, M.S., Gonzalez Aracil, S., Herries, D., Muneri, A., Gurung, M.B., Montgomery, J., Turner, P., 2021. Forest Structural Complexity Tool—An Open Source, Fully-Automated Tool for Measuring Forest Point Clouds. Remote Sens. 13, 4677. https://doi.org/10.3390/rs13224677
Laino, D., Cabo, C., Prendes, C., Janvier, R., Ordonez, C., Nikonovas, T., Doerr, S., Santin, C., 2024. 3DFin: a software for automated 3D forest inventories from terrestrial point clouds. For. Int. J. For. Res. cpae020. https://doi.org/10.1093/forestry/cpae020

Lamar, W.R., McGraw, J.B., Warner, T.A., 2005. Multitemporal censusing of a population of eastern hemlock (Tsuga canadensis L.) from remotely sensed imagery using an automated segmentation and reconciliation procedure. Remote Sens. Environ. 94, 133–143. https://doi.org/10.1016/j.rse.2004.09.003

Landrieu, L., Obozinski, G., 2016. Cut Pursuit: Fast Algorithms to Learn Piecewise Constant Functions, in: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics. Presented at the Artificial Intelligence and Statistics, PMLR, pp. 1384–1393.

Lecigne, B., Delagrange, S., Messier, C., 2018. Exploring trees in three dimensions: VoxR, a novel voxel-based R package dedicated to analysing the complex arrangement of tree crowns. Ann. Bot. 121, 589–601. https://doi.org/10.1093/aob/mcx095

Lee, S.J., Kim, J.R., Choi, Y.S., 2013. The extraction of forest CO2 storage capacity using high-resolution airborne lidar data. GIScience Remote Sens. 50, 154–171. https://doi.org/10.1080/15481603.2013.786957

Li, W., Guo, Q., Jakubowski, M., Kelly, M., 2012. A New Method for Segmenting Individual Trees from the Lidar Point Cloud. Photogramm. Eng. Remote Sens. 78, 75–84. https://doi.org/10.14358/PERS.78.1.75

Liang, X., Hyyppä, J., Kaartinen, H., Lehtomäki, M., Pyörälä, J., Pfeifer, N., Holopainen, M., Brolly, G., Francesco, P., Hackenberg, J., Huang, H., Jo, H.-W., Katoh, M., Liu, L., Mokroš, M., Morel, J., Olofsson, K., Poveda-Lopez, J., Trochta, J., Wang, D., Wang, J., Xi, Z., Yang, B., Zheng, G., Kankare, V., Luoma, V., Yu, X., Chen, L., Vastaranta, M., Saarinen, N., Wang, Y., 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. ISPRS J. Photogramm. Remote Sens. 144, 137–179. https://doi.org/10.1016/j.isprsjprs.2018.06.021

Liang, X., Wang, Y., Pyörälä, J., Lehtomäki, M., Yu, X., Kaartinen, H., Kukko, A., Honkavaara, E., Issaoui, A.E.I., Nevalainen, O., Vaaja, M., Virtanen, J.-P., Katoh, M., Deng, S., 2019. Forest in situ observations using unmanned aerial vehicle as an alternative of terrestrial measurements. For. Ecosyst. 6. https://doi.org/10.1186/s40663-019-0173-3

Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. Prog. Phys. Geogr. Earth

Environ. 27, 88–106. https://doi.org/10.1191/0309133303pp360ra
Luo, L., Zhai, Q., Su, Y., Ma, Q., Kelly, M., Guo, Q., 2018. Simple method for direct crown base height estimation of individual conifer trees using airborne LiDAR data. Opt. Express 26, A562–A578. https://doi.org/10.1364/OE.26.00A562
Maguya, A.S., Tegel, K., Junttila, V., Kauranne, T., Korhonen, M., Burns, J., Leppanen, V., Sanz, B., 2015. Moving Voxel Method for Estimating Canopy Base Height from Airborne Laser Scanner Data. Remote Sens. 7, 8950–8972. https://doi.org/10.3390/rs70708950
Maltamo, M., Gobakken, T., 2014. Predicting Tree Diameter Distributions, in: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, Managing Forest Ecosystems. Springer Netherlands, Dordrecht, pp. 177–191. https://doi.org/10.1007/978-94-017-8663-8_9
Martin-Ducup, O., Mofack, G., Wang, D., Raumonen, P., Ploton, P., Sonké, B., Barbier, N., Couteron, P., Pélissier, R., 2021. Evaluation of automated pipelines for tree and plot metric estimation from TLS data in tropical forest areas. Ann. Bot. 128, 753–766. https://doi.org/10.1093/aob/mcab051
Mielcarek, M., Stereńczak, K., Khosravipour, A., 2018. Testing and evaluating different LiDAR-derived canopy height model generation methods for tree height estimation. Int. J. Appl. Earth Obs. Geoinformation 71, 132–143. https://doi.org/10.1016/j.jag.2018.05.002
Næsset, E., Økland, T., 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sens. Environ. 79, 105–115. https://doi.org/10.1016/S0034-4257(01)00243-7
Parkitna, K., Krok, G., Miścicki, S., Ukalski, K., Lisańczuk, M., Mitelsztedt, K., Magnussen, S., Markiewicz, A., Stereńczak, K., 2021. Modelling growing stock volume of forest stands with various ALS area-based approaches. For. Int. J. For. Res. 94, 630–650. https://doi.org/10.1093/forestry/cpab011
Pebesma, E., Bivand, R., 2023. Spatial Data Science: With Applications in R. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9780429459016

Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Müller, K., Pedersen, T.L., Baston, D., Dunnington, D., 2023. sf: Simple Features for R.

Pitkänen, J., 2001. Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. Can. J. For. Res. 31, 832–844. https://doi.org/10.1139/x01-013

Popescu, S., Wynne, R., 2004. Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable Window Size for Estimating Tree Height. Photogramm. Eng. Remote Sens. 70, 589–604. https://doi.org/10.14358/PERS.70.5.589

Popescu, S., Wynne, R., Nelson, R., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can. J. Remote Sens. 29, 564–577. https://doi.org/10.5589/m03-027

Popescu, S., Wynne, R., Nelson, R., 2002. Estimating plot-level tree heights with LiDAR: local filtering with a canopy-height based variable window size. Comput. Electron. Agric. 37, 71–95. https://doi.org/10.1016/S0168-1699(02)00121-7

Popescu, S.C., Zhao, K., 2008. A voxel-based lidar method for estimating crown base height for deciduous and pine trees. Remote Sens. Environ. 112, 767–781. https://doi.org/10.1016/j.rse.2007.06.011

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., Vastaranta, M., Holopainen, M., Disney, M., Lewis, P., 2013. Fast Automatic Precision Tree Models from Terrestrial Laser Scanner Data. Remote Sens. 5, 491–520. https://doi.org/10.3390/rs5020491

Riaño, D., Meier, E., Allgöwer, B., Chuvieco, E., Ustin, S., 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. Remote Sens. Environ. 86, 177–186. https://doi.org/10.1016/S0034-4257(03)00098-1

Robins, V., Wood, P.J., Sheppard, A.P., 2011. Theory and Algorithms for Constructing Discrete Morse Complexes from Grayscale Digital Images. IEEE Trans. Pattern Anal. Mach. Intell. 33, 1646–1658. https://doi.org/10.1109/TPAMI.2011.95

Roussel, J.-R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R.H., 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. https://doi.org/10.1016/j.rse.2020.112061

Rozendaal, D.M.A., Hurtado, V.H., Poorter, L., 2006. Plasticity in leaf traits of 38 tropical tree species in response to light; relationships with light demand and adult stature. Funct. Ecol. 20, 207–216. https://doi.org/10.1111/j.1365-2435.2006.01105.x

Rusu, R.B., Cousins, S., 2011. 3D is here: Point Cloud Library (PCL), in: 2011 IEEE International Conference on Robotics and Automation. Presented at the 2011 IEEE International Conference on Robotics and Automation, pp. 1–4. https://doi.org/10.1109/ICRA.2011.5980567

Seidel, D., Leuschner, C., Müller, A., Krause, B., 2011. Crown plasticity in mixed forests—Quantifying asymmetry as a measure of competition using terrestrial laser scanning. For. Ecol. Manag. 261, 2123–2132. https://doi.org/10.1016/j.foreco.2011.03.008

Shivashankar, N., M, S., Natarajan, V., 2012. Parallel Computation of 2D Morse-Smale Complexes. IEEE Trans. Vis. Comput. Graph. 18, 1757–1770. https://doi.org/10.1109/TVCG.2011.284

Stefanidou, A., Gitas, I.Z., Korhonen, L., Stavrakoudis, D., Georgopoulos, N., 2020. LiDAR-Based Estimates of Canopy Base Height for a Dense Uneven-Aged Structured Forest. Remote Sens. 12, 1565. https://doi.org/10.3390/rs12101565

Stereńczak, K., Kraszewski, B., Mielcarek, M., Piasecka, Ż., Lisiewicz, M., Heurich, M., 2020. Mapping individual trees with airborne laser scanning data in an European lowland forest using a self-calibration algorithm. Int. J. Appl. Earth Obs. Geoinformation 93, 102191. https://doi.org/10.1016/j.jag.2020.102191

Sumnall, M., Fox, T., Wynne, R., Thomas, V., 2017. Mapping the height and spatial cover of features beneath the forest canopy at small-scales using airborne scanning discrete return Lidar. ISPRS J. Photogramm. Remote Sens. 133, 186–200. https://doi.org/10.1016/j.isprsjprs.2017.10.002

Sumnall, M., Peduzzi, A., Fox, T.R., Wynne, R.H., Thomas, V.A., 2016. Analysis of a lidar voxel-derived vertical profile at the plot and

individual tree scales for the estimation of forest canopy layer characteristics.
Tao, S., Wu, F., Guo, Q., Wang, Y., Li, W., Xue, B., Hu, X., Li, P., Tian, D., Li, C., Yao, H., Li, Y., Xu, G., Fang, J., 2015. Segmenting tree crowns from terrestrial and mobile LiDAR data by exploring ecological theories. ISPRS J. Photogramm. Remote Sens. 110, 66–76. https://doi.org/10.1016/j.isprsjprs.2015.10.007
Terryn, L., Calders, K., Åkerblom, M., Bartholomeus, H., Disney, M., Levick, S., Origo, N., Raumonen, P., Verbeeck, H., 2023. Analysing individual 3D tree structure using the R package ITSMe. Methods Ecol. Evol. 14, 231–241. https://doi.org/10.1111/2041-210X.14026
Tibshirani, R., Walther, G., 2005. Cluster Validation by Prediction Strength. J. Comput. Graph. Stat. 14, 511–528. https://doi.org/10.1198/106186005X59243
Tomppo, E., 2006. The Finnish National Forest Inventory. pp. 179–194. https://doi.org/10.1007/1-4020-4381-3_11
Trochta, J., Krůček, M., Vrška, T., Král, K., 2017. 3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR. PLOS ONE 12, e0176871. https://doi.org/10.1371/journal.pone.0176871
Unger, D.R., Hung, I.-K., Brooks, R., Williams, H., 2014. Estimating number of trees, tree height and crown width using Lidar data. GIScience Remote Sens. 51, 227–238. https://doi.org/10.1080/15481603.2014.909107
Vauhkonen, J., 2010. Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. Int. J. Remote Sens. 31, 1213–1226. https://doi.org/10.1080/01431160903380615
Vauhkonen, J., 2008. Estimating crown base height for Scots pine by means of the 3 D geometry of airborne laser scanning data 17–19.
Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkänen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalén, P., Gobakken, T., Koch, B., Næsset, E., Tokola, T., Maltamo, M., 2012. Comparative testing of single-tree detection algorithms under different types of forest. For. Int. J. For. Res. 85, 27–40. https://doi.org/10.1093/forestry/cpr051
Wang, D., 2020. Unsupervised semantic and instance segmentation of forest point clouds. ISPRS J. Photogramm. Remote Sens. 165, 86–97. https://doi.org/10.1016/j.isprsjprs.2020.04.020

Wang, J., Chen, X., Cao, L., An, F., Chen, B., Xue, L., Yun, T., 2019. Individual Rubber Tree Segmentation Based on Ground-Based LiDAR Data and Faster R-CNN of Deep Learning. Forests 10, 793. https://doi.org/10.3390/f10090793

Wang, L., Gong, P., Biging, G.S., 2004. Individual Tree-Crown Delineation and Treetop Detection in High-Spatial-Resolution Aerial Imagery. Photogramm. Eng. Remote Sens. 70, 351–357. https://doi.org/10.14358/PERS.70.3.351

Weiser, H., Schäfer, J., Winiwarter, L., Krašovec, N., Fassnacht, F.E., Höfle, B., 2022. Individual tree point clouds and tree measurements from multi-platform laser scanning in German forests. Earth Syst. Sci. Data 14, 2989–3012. https://doi.org/10.5194/essd-14-2989-2022

Weiss, K., Iuricich, F., Fellegara, R., De Floriani, L., 2013. A primal/dual representation for discrete Morse complexes on tetrahedral meshes. Comput. Graph. Forum 32, 361–370. https://doi.org/10.1111/cgf.12123

West, G.B., Brown, J.H., Enquist, B.J., 1997. A General Model for the Origin of Allometric Scaling Laws in Biology. Science 276, 122–126. https://doi.org/10.1126/science.276.5309.122

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. Can. J. Remote Sens. 42, 619–641. https://doi.org/10.1080/07038992.2016.1207484

Wilkes, P., Lau, A., Disney, M., Calders, K., Burt, A., Gonzalez de Tanago, J., Bartholomeus, H., Brede, B., Herold, M., 2017. Data acquisition considerations for Terrestrial Laser Scanning of forest plots. Remote Sens. Environ. 196, 140–153. https://doi.org/10.1016/j.rse.2017.04.030

Wulder, M., Niemann, K.O., Goodenough, D.G., 2000. Local Maximum Filtering for the Extraction of Tree Locations and Basal Area from High Spatial Resolution Imagery. Remote Sens. Environ. 73, 103–114. https://doi.org/10.1016/S0034-4257(00)00101-2

Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., White, J.C., 2008. The role of LiDAR in sustainable forest management. For. Chron. 84, 807–826. https://doi.org/10.5558/tfc84807-6

Xi, Z., Hopkinson, C., 2022. 3D Graph-Based Individual-Tree Isolation (Treeiso) from Terrestrial Laser Scanning Point Clouds. Remote Sens. 14, 6116. https://doi.org/10.3390/rs14236116

Xi, Z., Hopkinson, C., 2021. Detecting Individual-Tree Crown Regions from Terrestrial Laser Scans with an Anchor-Free Deep Learning

Model. Can. J. Remote Sens. 47, 228–242. https://doi.org/10.1080/07038992.2020.1861541
Xu, W., Su, Z., Feng, Z., Xu, H., Jiao, Y., Yan, F., 2013. Comparison of conventional measurement and LiDAR-based measurement for crown structures. Comput. Electron. Agric. 98, 242–251. https://doi.org/10.1016/j.compag.2013.08.015
Xu, X., Iuricich, F., Calders, K., Armston, J., De Floriani, L., 2023. Topology-based individual tree segmentation for automated processing of terrestrial laser scanning point clouds. Int. J. Appl. Earth Obs. Geoinformation 116, 103145. https://doi.org/10.1016/j.jag.2022.103145
Yao, W., Krull, J., Krzystek, P., Heurich, M., 2014. Sensitivity Analysis of 3D Individual Tree Detection from LiDAR Point Clouds of Temperate Forests. Forests 5, 1122–1142. https://doi.org/10.3390/f5061122
Zarnoch, S., Bechtold, W., Stolte, K., 2004. Using crown condition variables as indicators of forest health. Can. J. For. Res.-Rev. Can. Rech. For. - CAN J For. RES 34, 1057–1070. https://doi.org/10.1139/x03-277
Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., Yan, G., 2016. An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. Remote Sens. 8, 501. https://doi.org/10.3390/rs8060501
Zhao, Y., Hao, Y., Zhen, Z., Quan, Y., 2017. A Region-Based Hierarchical Cross-Section Analysis for Individual Tree Crown Delineation Using ALS Data. Remote Sens. 9, 1084. https://doi.org/10.3390/rs9101084
Zianis, D., Muukkonen, P., Mäkipää, R., Mencuccini, M., 2005. Biomass and stem volume equations of tree species in Europe. Silva Fenn. 4. https://doi.org/10.14214/sf.sfm4

11 LINKS

Stonex (accessed on 01.11.2024) https://www.stonex.it/project/s9i-gnss-receiver/

Haglöf (accessed on 01.11.2024)
https://haglofsweden.com/project/ec-ii-d-r

Wytham Woods (accessed on 01.11.2024)
https://www.wythamwoods.ox.ac.uk/home

12 ÖSSZEFOGLALÓ

Ez a disszertáció három szoros összefüggésben lévő témakört ölel fel, amelyek célja új erdészeti eszközök kifejlesztése. Ezek a különböző légi és földi lézer-szkennelt adatok feldolgozását, illetve elemzését segítik. Az első téma egy új egyedfacsúcs detektáló (angolul Individual Treetop Detection) módszert mutat be, amely alacsony pontsűrűségű LiDAR adatokból származó raszter-alapú Canopy Height Modellt használ. Az algoritmus magja a Growing Tree Region koncepció, és célja a fakorona csúcsának pontos lokalizálása különböző erdőtípusokban. A módszer különböző szerkezetű és sűrűségű erdőkben volt tesztelve (a német Hardtwald és Bretten, valamint a debreceni Nagyerdő erdeiben). Az algoritmus hatékonyságát a legelterjedtebb raszter-alapú egyedfacsúcs technikával, a lokálos maxima (angolul Local Maxima) keresővel kombinált variálható ablak szürő (angolul Variable Window Filtering) módszerrel hasonlítottam össze és számos különböző kiértékelési metrikát alkalmaztam (matching rate, extraction rate, commission rate, omission rate, mean absolute error). Az egyedfacsúcs detektáló módszer egy R csomagban, a TREETOPS-ban lett megvalósítva.

A második téma egy a fa ágtiszta törzsmagasság (angolul Crown Base Height) lokalizálására szolgáló módszert kínál, amely nagy pontsűrűségű lézer-szkennelt-pontfelhő adatokat dolgoz be. Az előzőleg az egyedfacsúcs-detektáló által szegmentált egyedfa-pontfelhők képezik az beviteli alapot az ágtiszta törzsmagasság-lokalizáló eszközhöz. Ezt szintén egy R csomagban, *treecbh* néven valósítottam meg. Ez az eszköz automatikusan végzi el a fatörzs és a hozzá tartozó alsó ágak detektálását, amelyhez a *treeiso* nevű hierarchikus-grafikus egyedfa-szegmentáló eszközt használja. A *treecbh* a lézer fapontfelhők automatikus feldolgozását teszi lehetővé, és a TREETOPS-hoz hasonlóan Hardtwald, Bretten és Nagyerdő erdeiben voltak tesztelve. A fa ágtiszta törzsmagasság meghatározásának pontossága erdőleltározási adatokkal való összehasonlítással is alátámasztásra került. A törzsmagasság-lokalizátor három beállítási lehetőségét: az egyedfa szegmentálás nélkülit, az alapértelmezettet és az interaktív-vizuálisat teszteltem és kiértékeltem. Ezen kívül, egy pontsűrűség érzékenységi vizsgálat keretén belül a *treecbh*-ba való beviteli fapontfelhő optimális pontsűrűségét is meghatároztam.

A harmadik téma a diszkrét Morse elmélet alkalmazására épül és jelenleg fejlesztés alatt áll. Ez egy új módszert kínál egyedfa elkülönítésére nagy sűrűségű földi lézer-szkennelt adatok alapján. Ennek a módszernek az első lépése az egyedfatörzs lokalizálása és elkülönítése. Ehhez a Wytham Woods erdejében (Oxfordshire, Egyesült Királyság) gyűjtött földi lézer-szkennelt adatokat használtam, és az egyedfatörzs felismerésére egy új algoritmust fejlesztettem ki.

Ez disszertáció új, hatékony eszközöket és módszereket mutat be a lézer-alapú erdészeti adatfeldolgozás területén, különös tekintettel a raszter-alapú egyedfatörzs pontos lokalizálására, a fa ágtiszta törzsmagasság meghatározására, valamint az egyedfatörzs lokalizálására és elkülönítésére. Az alkalmazott algoritmusok és eszközök nemcsak a kutatás, hanem az erdőgazdálkodás, erdőleltározás és az erdészeti megfigyelés területén is jelentős előrelépést jelenthetnek, különösen a különböző pontsűrűségű légi és földi lézer-szkennelt adatok kombinált vagy egyedi hatékony feldolgozásában.