Czech University of Life Sciences Prague

Faculty of Forestry and Wood Sciences

Department of Forest Management

Diploma Thesis

Comparing the accuracy of single trees delineation from ALS

Author: Bc. Alejandro Rodríguez Vivancos
Supervisor: Ing. Peter Surový, PhD

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Forestry and Wood Sciences

DIPLOMA THESIS ASSIGNMENT

Bc. Alejandro Rodríguez Vivancos

Forestry, Water and Landscape Management

Thesis title

Comparing the accuracy of single trees delineation from ALS

Objectives of thesis

To know the diametric distribution of a forest is crucial to develop de forest management plans, although it has not been achieved through remote sensing techniques such as LiDAR. This thesis proposes a technique based on single tree extraction from aerial laser scans (ALS) to address the estimation of the diameter distribution. The objectives are:

- Single tree delineation with two different whatershed tecniques: TreeSeg tool (FUSION) and ArcGIS. Study area: Two permanent sample plots of unevenaged Scotch Pine forest(Spain)and three sample plots of evenaged mixture forest (Czech Republic)

- Comparing both techniques, using different kernel distances for smoothing the Canopy Heigh Models (CHM). In addition, 0.2 and 0.5 m pixel resolution for CHM are tested.

- Evaluating the correlation of Diameter at Breast Height (DBH) with Crownsize and Tree Height.

Methodology

- Processing the LiDAR data and extract the CHM.

- Single Tree Detection and delineation using FUSION open tool.

- Single Tree Detection and delineation using ArcGIS.

- Comparing the accuracy of delineation between those techniques.

- Correlating DBH, crown size and tree height.

The proposed extent of the thesis

60 pages

Keywords

laser scan, point cloud, tree characteristics, single-tree detection, diametric distribution

---

Recommended information sources

Edson, C., & Wing, M. G. (2011). Airborne light detection and ranging (LiDAR) for individual tree stem location, height, and biomass measurements. Remote Sensing, 3(11), 2494-2528.

Jakubowski, M. K., Li, W., Guo, Q., & Kelly, M. (2013). Delineating individual trees from LiDAR data: A comparison of vector-and raster-based segmentation approaches. Remote Sensing, 5(9), 4163-4186.

Li, W., Guo, Q., Jakubowski, M. K., & Kelly, M. (2012). A new method for segmenting individual trees from the lidar point cloud. Photogrammetric Engineering & Remote Sensing, 78(1), 75-84.

Panagiotidis, D., Abdollahnejad, A., Surový, P., & Chiteculo, V. (2017). Determining tree height and crown diameter from high-resolution UAV imagery. International journal of remote sensing, 38(8-10), 2392-2410.

Rahman, M. Z. A., & Gorte, B. (2008). Individual tree detection based on densities of high points of high resolution airborne LiDAR. GEOBIA, 350-355.

Reitberger, J., Heurich, M., Krzystek, P., & Stilla, U. (2007). Single tree detection in forest areas with high-density LiDAR data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(3), 139-144.

Shin, J., & Temesgen, H. (2018). Generating tree-lists by fusing individual tree detection and nearest neighbor imputation using airborne LiDAR data. Open Journal of Forestry, 8(04), 500.

Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Computer Architecture Letters, 13(06), 583-598.

Wannasiri, W., Nagai, M., Honda, K., Santitamnont, P., & Miphokasap, P. (2013). Extraction of mangrove biophysical parameters using airborne LiDAR. Remote Sensing, 5(4), 1787-1808.

Zhao, K., & Popescu, S. (2007). Hierarchical watershed segmentation of canopy height model for multi-scale forest inventory. Proceedings of the ISPRS working group'. pp. 436â, 442.

Expected date of thesis defence

2020/21 SS – FFWS

The Diploma Thesis Supervisor

Ing. Peter Surový, PhD.

Supervising department

Department of Forest Management

Electronic approval: 17. 2. 2021

Ing. Peter Surový, PhD.

Head of department

Electronic approval: 17. 2. 2021

prof. Ing. Róbert Marušák, PhD.

Dean

Prague on 31. 03. 2021

Declaration

I declare that I have worked on my diploma thesis titled "Comparing the accuracy of single trees delineation from ALS" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the diploma thesis, I declare that the thesis does not break any copyrights.

In Prague on April 19th, 2021

RODRIGUEZ VIVANCOS ALEJANDRO - 70421948L

Firmado digitalmente por RODRIGUEZ VIVANCOS ALEJANDRO - 70421948L
Fecha: 2021.04.19 15:45:42 +02'00'

Acknowledgement

This project is the result of perseverance and effort to achieve what you want. With special dedication to those people who never stopped believing in me and who did not allow me to give up.

I want to especially thank my supervisor, Peter, who has made this work possible. Also special thanks to the SILVANET research group, not only for the data, but also for the support so that I could finish my master's studies. Finally, to all the people who are in my life, especially one in particular, who still loves me like the first day we met in Prague.

Comparing the accuracy of single trees delineation from ALS

Abstract

This study analyses different Individual Tree Detection (ITD) techniques from airborne LiDAR data (Airborne Laser Scanning, ALS), to increase the knowledge of this potential methodology, since it offers detailed information at the stand scale. Two segmentation techniques of the Canopy Height Models (CHM) are evaluated: First, the watershed segmentation by an elaborating process in ArcMAP. Then, testing the "TreeSeg" algorithm developed in FUSION. The influence of some parameters is also analysed, such as smoothing (different smoothing radii or kernel) and the resolution of these raster layers (0.2 m and 0.5 m). Additionally, these results are compared with those obtained by classifying trees on the point cloud directly, using the LIDR tool in the RStudio program. The vegetation was classified into two typologies based on their height (low and high vegetation). Finally, the relationship between the Diameter at Breast Height (DBH) and other biophysical parameters of the tree is analysed, to obtain a regression model.

The watershed segmentation results in ArcMAP were significantly better than those obtained with FUSION (improvement of 20%, p-value = 0.028 < 5%). Low vegetation trees were not detected by the CHM methods, while few of them were detected by the point cloud classification, which produced estimates as accurate as the CHM methods. Resolution and smoothing had a significant influence on the estimates. CHMs of 0.5 m resolution showed better results with 3-cells radius (K3) smoothing, while CHMs of 0.2 m required the highest smoothing (K7). However, a higher point clouds density allowed to reduce the smoothing of the layers in both resolutions. Finally, a good potential regression model (DBH ~ Height, R^2 = 0.91) was obtained after finding strong correlations between all the variables.

Keywords: Remote sensing; Airborne Laser Scanning; Individual Tree Delineation; Watershed Segmentation; Canopy Heigh Model; Resolution; Smoothing; Kernel; point cloud classification; regression model; RMSE; ANOVA.

Porovnání přesnosti vymezení jednotlivých stromů z ALS

Abstrakt

Tato studie analyzuje různé techniky detekce jednotlivých stromů (ITD) od leteckých dat LiDAR (Airborne Laser Scanning, ALS), aby zvýšila znalosti této potenciální metodiky, protože nabízí podrobné informace v měřítku porostu. Vyhodnocují se dvě segmentační techniky modelů výšek porostu (CHM): Nejprve segmentace předělů zpracováním procesu v ArcMAP. Poté testuje algoritmus „TreeSeg“ vyvinutý ve FUSION. Analyzuje se také vliv některých parametrů, jako je vyhlazování (různé poloměry vyhlazování) a rozlišení těchto rastrových vrstev (0,2 m a 0,5 m). Tyto výsledky jsou porovnány s výsledky získanými přímou klasifikací stromů v mračnu bodů pomocí nástroje LIDR v programu RStudio. Vegetace byla rozdělena do dvou typů na základě jeji výšky (nízká a vysoká vegetace). Nakonec je analyzován vztah mezi výčetní tloušťkou (DBH) a dalšími biofyzikálními parametry stromu, aby se získal regresní model.

Výsledky segmentace povodí v ArcMAP byly významně lepší než výsledky získané pomocí FUSION (zlepšení o 20%, p-hodnota = 0,028 <5%). Metody CHM nenašli stromy s nízkou vegetací, i když některé byly detekovány klasifikací mračna bodů, která poskytla odhady stejně přesné jako metody CHM. Rozlišení a vyhlazení mělo významný vliv na odhady. CHM s rozlišením 0,5 m vykázalo lepší výsledky při vyhlazení poloměru 3 buněk (K3), zatímco CHM 0,2 m vyžadovalo nejvyšší vyhlazení (K7). Vyšší hustota mračen bodů však umožnila snížit vyhlazení vrstev v obou rozlišeních. Nakonec byl po nalezení silných korelací mezi všemi proměnnými získán dobrý potenciální regresní model (DBH ~ výška).

Klíčová slova: Dálkový průzkum Země; Airborne Laser Scanning; Detekce jednotlivých stromů; Segmentace povodí; Canopy Heigh Model; Řešení; Vyhlazování; Jádro; klasifikace mračen bodů; regresní model; RMSE; ANOVA.

CONTENTS

LIST OF APPENDICES

FIGURES

TABLES

1 INTRODUCTION

1.1 Forestry and the importance of the environment monitoring

The latest data published by the Food and Agriculture Organization (FAO) of the United Nations confirm that forests occupy 4.06 billion hectares, what is a third of the total surface of the earth (Figure 1). In addition, approximately half of the forest surface (2.05 billion ha) are subject to forest management plans (FAO and UNEP, 2020).

"We depend on forests for our survival, from the air we breathe to the wood we use"

(WWF, 2020)

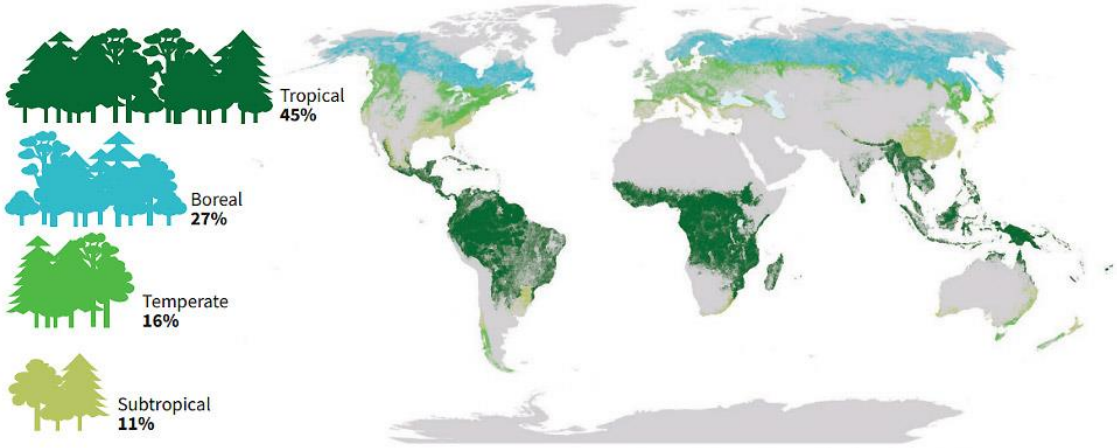


Figure 1. Global distribution of the forest (FAO and UNEP, 2020).

The importance of forests has historically been supported by their diverse productivity of natural resources, not only of timber, but also non-timber products (Führer, 2000), such as resin, cattle pasture, or cork, among many others. However, the functionality that these ecosystems have is what makes them truly valuable. At the environmental level, forests are carbon sinks, that is, they retain this gas that is so harmful to the ozone layer, improving the atmospheric quality and reducing the global warming, which mitigates climate change. The combination of canopies, plant debris and roots build a strong "shield" against intense rainfall, which not only reduces erosion, but also increases water infiltration into the soil, feeding the aquifers and regulating the

hydrological regime. In addition, they are the home of innumerable species (animals, plants, and fungi), forming complex communities that increase biodiversity. Socially, both recreational and professional activities coexist in balance in these natural or semi-natural environments.

Therefore, the work of environmental managers is essential, who must also guarantee the sustainable development of these ecosystems. This concept was coined in the “Our common future” report published in 1987 by the Brundtland Commission in which solidarity is appealed to "meet the needs of the present without compromising the ability of future generations to meet their own needs" (World Commission on Environment and Development, 1987). Ensuring the continuity of these resources begins by knowing their current state, their characteristics, as well as their distribution and organization, aspects that are included within the concept of forest structure (Spies & Franklin, 1991; Smith, Larson, Kelty, & Ashton, 1997; Zimble, et al., 2003).

In particular, the most relevant information concerning the study of forest stands structure are the distribution, height and canopy size of trees, the vertical distribution, the diameter distribution, and the specific composition both in diversity of species and in its spatial distribution (Del Río, Montes, Cañellas, & Montero, 2003).

Not only the current state of the forests is important, but also evaluating their evolution through the growth and mortality of the forest stand is necessary (Rodrigues de Souza, de Azevedo, Brum Rossi, dos Santos, & Higuchi, 2014). Two of the most useful parameters for this purpose are the basal area and the volume, which can be easily calculated knowing the number of trees present on each diameter class. In addition, the change in the frequency of the trees on each this classes makes it possible to evaluate the growth of the forest stands. For these reasons, the diameter distribution in one of the most important parameters to develop the forest management plans.

Field information have been historically obtained using classics forest inventories by measuring a relatively small portion of the area in relation to the total extent. The variation of the forest stand in the different strata or ecosystems is evaluated in qualitative terms, as well as the composition of the forest and the characteristics of the registered species (such as the shape of the trunk or the crown). On the other hand, the number of species per unit area and the dasometric variables are determined in quantitative terms

such as diameter, commercial and total height of the inventoried individuals. After collecting that field information, the total basal area and the volume per unit area can be estimated which is useful to organize the forestry activities in the stands. These are the basis of a larger-scale planning (management project).

However, classic inventories are associated with a sizing based on the error to be assumed when sampling. Sometimes, a high accuracy is required, so the field work demanded to keep the maximum admitted error is excessive or unaffordable. In addition, all this work is not always economically compensated, because the monetary value of the forest is getting lower. For these reasons it is necessary to find viable alternatives to the classic inventories and taking advantage of the opportunities offered by the current technologies.

1.2 Remote sensing techniques

Forest attributes have traditionally been mapped using passive airborne and satellite sensors (McRoberts & Tomppo, 2007), but they had important limitations to quantify the characteristics of the vegetation because they only generated information in two dimensions. Moreover, in areas of dense coverage, the highly reflected energy tends to saturate the signal captured by the sensor (Lefsky, Cohen, Parker, & Harding, 2002), which makes it impossible to estimate the different dasometric variables above a certain threshold. In addition, passive remote sensing technology is highly dependent on the atmospheric conditions, which difficult the information gathering in adverse conditions (poor light, clouds, etc.). To solve this, LiDAR technology was developed.

1.2.1 LiDAR technology

Light Detection and Ranging (LiDAR), was an emerging remote sensing technology with promising potential to map, monitor, and assess forest resources (Gatziolis & Andersen, 2008). It is based on the emission of a pulse and the measurement of the time it takes to reach the surface and return to the sensor that emitted it. Since the speed at which it travels (speed of light) is also known, the distance between the sensor and the target surface can be determined.

There are two general types of LiDAR. Terrestrial Laser Scanning (TLS) has the potential to accurately measure Above-Ground Biomass (AGB) on a plot scale, while Airborne Laser Scanning (ALS) from manned aircraft can serve as a tool to apply the plot measurements at the landscape scale. In the case of being mounted on an aircraft, it is necessary to correctly reference the point that has been measured in the field. To accomplish this, the system is based on the combination of three different elements: a laser scanner mounted on an airplane together with a Global Positioning System (GPS) and an Inertial Navigation System (INS) (Baltsavias, 1999). The laser scanner emits a signal that is reflected on the ground surface and then captured by a sensor, which together with the GPS and INS, allow the calculation of its coordinates. Inertial Navigation System (INS) accurately measures the orientation of the sensor, correcting the movements that it may suffer during the flight, while the GPS determines the exact position of the sensor. In summary, the distance between the sensor and the target surface is known for each returned point, together with the angle and the geographical location of the sensor. Considering all the parameters, the location (coordinates) of every returned point is calculated (Figure 2).

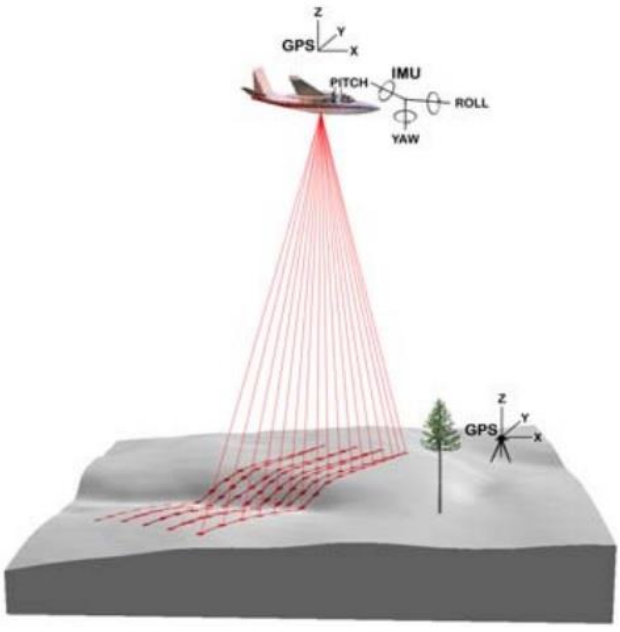


Figure 2. Scheme of an ALS system (McGaughey, 2018).

The main technical specifications of a LiDAR sensor are the wavelength of the laser signal; the power, the pulse duration, and its repetition cycle; the angle of divergence; the specifications of the scanner mechanism; the size of the area illuminated by the laser pulse (footprint) and the information it collects on each reflected pulse (Lefsky, Cohen, Parker, & Harding, 2002).

LiDAR data may be recorded according two different techniques (Wasser, 2020), as illustrated in Figure 3.

- **Discrete Return LiDAR:** This system identifies peaks and record a point (return) at each peak location in the waveform curve. A discrete system usually records from 1 to 4 returns from each laser pulse (sometimes more).
- **Full Waveform LiDAR:** The full distribution of returned light energy is recorded. The system collects how the intensity of this energy varies over time, providing the height distribution of the surfaces illuminated by laser.

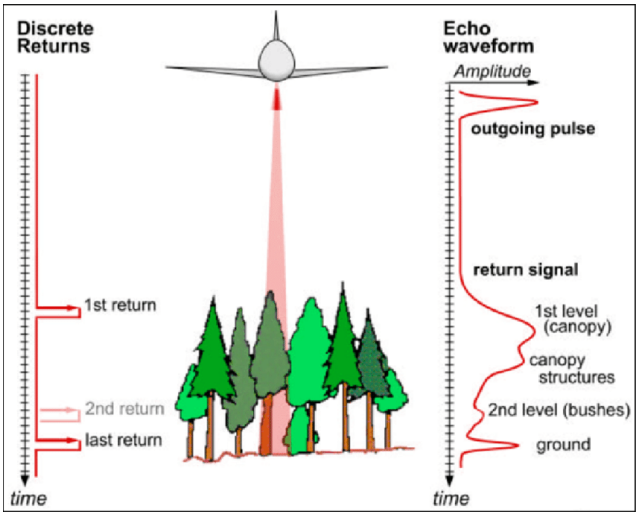


Figure 3. Discrete return LiDAR sensor and a full waveform LiDAR system (Daly, Vuyovich, & Finnegan, 2011).

Some clarifications are needed in reference to the laser pulses emitted by LiDAR sensors (Tordesillas Torres, 2014):

- The pulse is neither instantaneous nor completely discreet. This pulse travels at the speed of light and its energy have a Gaussian distribution with a pulse duration

from 6 to 12 nanoseconds (that is, between 1.8 and 3.6 meters at the speed of light).

- The laser beam describes a cone shape (and not a line) covering an area (footprint). Therefore, there is a high probability that the same pulse hits several elements, thus obtaining several returns at different heights. Small footprint sensors (0.1 - 2 m), which usually correspond to discrete return sensors, are distinguished from large footprint sensors (typically 10 - 100 m), usually used in full wave sensors.

Therefore, what differentiates LiDAR technology from other remote sensing methods is that it is an active system based on the emission of a laser beam and its reception when it intersects with different objects. The emitted laser is characterized by its frequency in number of pulses per unit of time, which, depending on certain characteristics such as distance (height for ALS) from the emitter to the object or the speed of LiDAR sensor, together with footprint, are subsequently translated into a point density per unit area. ALS density is typically ranged from 1 to 10 points·m^{-2}, while TLS can map precisely canopy elements like stems and branches, since it produces high density point clouds. (Brede, Lau, Bartholomeus, & Kooistra, 2017). For forestry purposes, LiDAR technology is able to collect information from the forest cover in the three-dimensions with the advantage of not presenting signal saturation, which allows to evaluate the three-dimensional patterns of the tree canopy and to estimate the vertical structure from plant communities (Lefsky, Cohen, Parker, & Harding, 2002).

1.2.2 ALS approaches

There are two main approaches to estimate forest variables through a LiDAR inventory (Hyyppä, Hyyppä, Leckie, Gougeon, & Yu X., 2008): (i) Area Based Approach (ABA) and (ii) single-tree detection.

The ABA methods are based on the study of small areas of forest relaying measurements in the field and information extracted from the LiDAR point cloud using statistical regression models. It starts from a dasometric inventory in which the necessary measurements are made for the subsequent calculation of the variables of interest such as Tree Cover Density (TCD), Basal Area, Volume, Mean Diameter, Dominant Height or Above-Ground Biomass (AGB). Subsequently, the statistics of the point cloud in those

sample plots (percentiles, density, dispersion) are extracted and related to the field information. Finally, the variables of interest are estimated on a larger scale (stand or landscape scale).

In the other hand, Individual Tree Delineation (ITD) methods consist of the location and delimitation of each single tree and the direct calculation of its biophysical parameters, such as tree height or crown size. Other variables of interest can be estimated from them by using allometric equations or regression models, such as DBH, basal area or volume. The aggregation of the individual values provides the information relative to the stand scale. In contrast with ABA method, the identification of individual trees from LiDAR data requires a high density of LiDAR points, at least 4-5 points·m^{-2} (Wulder, Bater, Coops, Hilker, & White, 2008; Reutebuch & Andersen, 2005).

The veracity of this methods is more influenced by the forest structure than by the algorithms used. The main challenge with this type of assessment lies in dense forests with overlapping tree crowns. Furthermore, many algorithms fail to identify understory and dominated trees or to identify trees under high density conditions or groups of trees. (Goodwin, Coops, & Culvenor, 2006; Zhao, Popescu, & Nelson, 2009; Vauhkonen, Korpela, Maltamo, & Tokola, 2010).

2 LITERATURE REVIEW

2.1 Laser scanning background

Several studies have achieved accurate estimations for the main forest attributes using the ABA procedure (Hyyppä, Hyyppä, Leckie, Gougeon, & Yu X., 2008; Wulder, Bater, Coops, Hilker, & White, 2008), obtaining greater precision than through classical inventories or other approaches with optical sensors (Maltamo, Malinen, Packalén, Suvanto, & Kangas, 2006). However, some variables, as TCD or Quadratic Mean Dimeter (QMD) usually produce poor fits with LiDAR data, especially in stands with high regeneration or many dominated trees (Tordesillas Torres, 2014; Næsset, 2002; Maltamo, Eerikäinen, Pitkänen, Hyyppä, & Vehmas, 2004; Treitz, et al., 2012). These trees, being dominated by other individuals would be unnoticed in the LiDAR point cloud. The precision in estimating the diameter distribution using the ABA method varies enormously (Gobakken & Næsset, 2004; Maltamo, Suvanto, & Packalén, 2007; Breidenbach, Gläser, & Schmidt, 2008). On the other hand, LiDAR has been also proved to be a reliable method to assess wildlife habitat (Hinsley, Hill, Bellamy, & Balzter, 2006) and to quantify stand susceptibility to fire (Andersen, McGaughey, & Reutebuch, 2005). In addition, the evaluation of forest dynamics is being carried out successfully by the repetition of LiDAR flights, both for the growth of stands (Hopkinson, Chasmer, & Hall, 2008) and for the wood extraction (Andersen, Reutebuch, McGaughey,, d'Oliveira, & Keller, 2014).

Recently, LiDAR sensors have been mounted on Unmanned Aerial Vehicles (UAV) to combine the advantages of LiDAR and UAV technology. It could produce greater density of points than a conventional ALS since it flies closer to the stand and at a lower speed, although it does not have as much detail as TLS, since the latter is in a fixed position within the forest stand. One of the first UAV LiDAR systems reached point cloud densities from 100 to 1500 points·m^{-2} with the objective of single trees detection (Jaakkola, et al., 2010). High point cloud densities are efficient to conduct surveys of forest plots in which terrain and understory height, tree location, tree height, crown area and volume could be derived (Wallace, Musk, & Lucieer, 2014). However, UAV LiDAR can also be used to obtain lower point cloud density. For instance, a density of 0.5

points·m^{-2} captured with UAV was sufficient to perform a good vegetation filtering and DTM generation (Wei, Yang, Jiang, Cao, & Wu, 2017).

2.1.1 Individual Tree Delineation

Two main methodologies have been developed for the tree segmentation. Initially, the studies focused their efforts on the study of aerial images (passive remote sensing), but LiDAR emerged as an alternative at the beginning of the 21st century. Since then, many studies have been followed and the knowledge in this area has increased. (Figure 4).

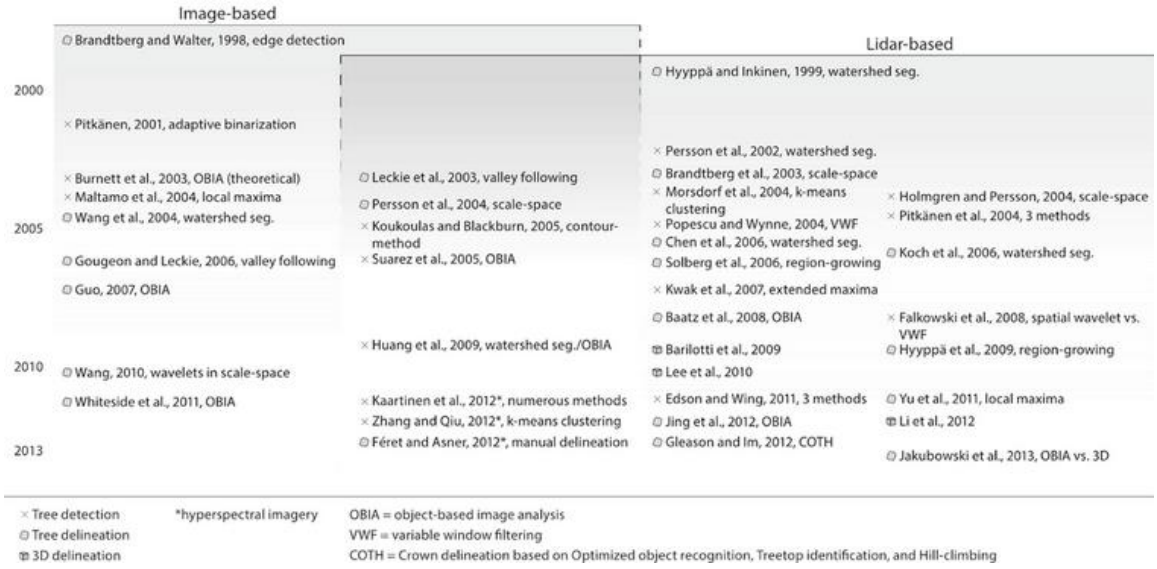


Figure 4. Scientific ITD studies both with passive and active remote sensing data (Jakubowski, Guo, & Kelly, 2013).

Transforming the laser point cloud into the Canopy Height Model (CHM) is the most common method to face the tree detection. CHM represents the difference in height between the top of the canopy surface and the underlying ground surface, that is, the height layer of the vegetation (Panagiotidis, Abdollahnejad, Surový, & Chiteculo, 2017). With this layer, individual trees can be detected using the watershed segmentation (Hyyppä & Inkinen, 1999; Chen, Baldocchi, Gong, & Kelly, 2006), which is based on the inversion of the CHM, so that the treetops appear as the lowest points and each tree can be considered a catchment basin. The highest point of each tree can be detected by the

evaluation of the local maxima. However, the crowns can be very complex and there may be more than one local maxima, especially in broadleaves species. (Panagiotidis, Abdollahnejad, Surový, & Chiteculo, 2017). To solve this, different distances (kernel distance) must be evaluated depending on the characteristics of the vegetation, within which there can only be one maximum point that represents an individual tree. This depends mainly on the shape of the tree canopy and the distance between the trees (Figure 5). Afterwards, the local maximas are used as the pour points for the basins of each tree.

Recent studies try to avoid transforming the LiDAR point cloud into a CHM to perform the tree segmentation. To do this, they develop different algorithms to evaluate the point cloud directly. For example, some algorithms apply watershed segmentation directly on the point cloud (Lee, Slatton, Roth, & Cropper, 2010), while others use different criteria. One of them is based on the natural distance between the trees (Li, Guo, Jakubowski, & Kelly, 2012). Even in the closest tree canopies, the treetops always have some separation.

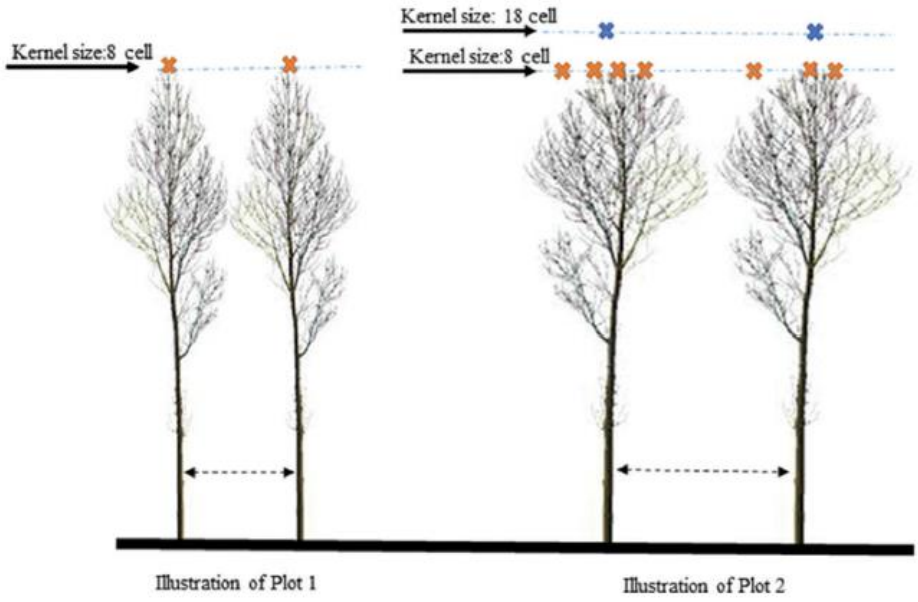


Figure 5. Different kernel distance in relation with the vegetation properties, as crown size and separation between trees (Panagiotidis, Abdollahnejad, Surový, & Chiteculo, 2017).

3 OBJECTIVES AND METHODOLOGY

3.1 Objectives

LiDAR technology has demonstrated its efficacy and precision in estimating most of the forest variables. At present, the challenges lie in estimating the diameter distribution, as well as in the possibility of predicting the evolution of the forest variables.

To know the diameter distribution of a forest is crucial to develop de forest management plans, although it has not been achieved through remote sensing techniques such as LiDAR. This thesis proposes a technique based on single tree extraction from aerial laser scans (ALS) to address the estimation of the diameter distribution. The objectives are:

1) **Single tree delineation with two different watershed techniques**: TreeSeg tool (FUSION) and ArcGIS. Study area: Two permanent sample plots of uneven aged Scotch Pine forest (Spain) and three sample plots of even aged mixture forest (Czech Republic)

2) **Comparing both techniques**, using different kernel distances for smoothing the Canopy Heigh Models (CHM). In addition, 0.2 and 0.5 m pixel resolution for CHM are tested.

3) **Evaluating the correlation of Diameter at Breast Height (DBH) with Crown size and Tree Height**.

In addition to these main objectives, a third Individual Tree Delineation (ITD) methodology (Li, Guo, Jakubowski, & Kelly, 2012) is performed in order to compare the potential of the point cloud segmentation directly.

To sum up, this study is focused in: (i) comparing different ITD techniques to evaluate their potential in extracting the individual trees from ALS; (ii) Evaluating the relation between different biophysical tree parameters (DBH, height and crown size) to obtain a regression model. In future studies, diameter distribution could be obtained by extracting the individual trees from ALS data and then calculate their DBH.

3.2 Material and methods

3.2.1 Study area

Two study areas will be assessed. The first one is in Cercedilla, a small town located 60 kilometers northwest from Madrid (central Spain), characterized by the mountain range that crosses the central Spanish zone from southwest to northeast (Central System). Part of this area is protected by the most recent (2013) Spanish National Park figure (Sierra de Guadarrama). In addition, it is the closest National Park to the Spanish capital, where approximately 6 million people live. This makes it a very particular and visited place. Specifically, in 2015 it was the most visited Spanish National Park after the Teide National Park (Canary Islands).

Within the National Park is the Fuenfría valley forest and develops a horseshoe shape with N-S orientation. The altitudinal interval ranges between 1250 and 2025 meters above sea level, and it is crossed by the Venta river, tributary of the Guadarrama river.

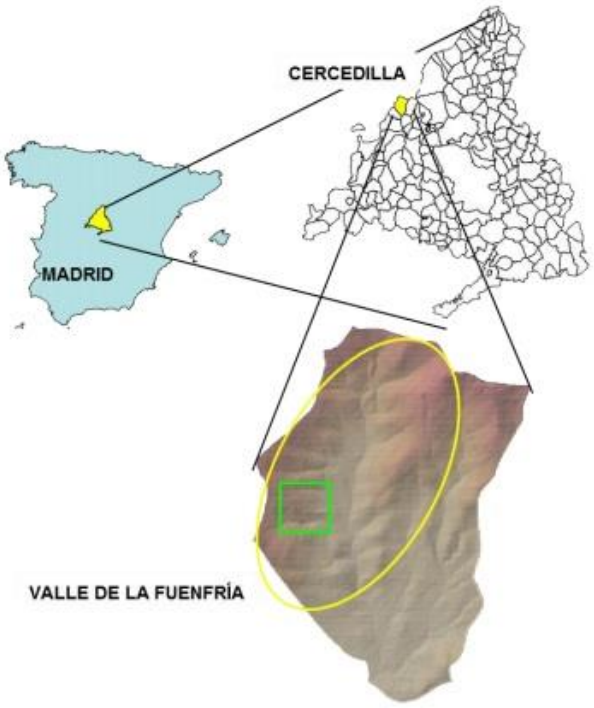


Figure 6. Spanish study area location
(Pascual, García-Abril, García-Montero, & Martín-Fernández, 2008).

The Fuenfría valley is within the Spanish public protection figure "Monte de Utilidad Pública" (M.U.P. number 32), called "Pinar y Agregados" which is owned by the City Council of Cercedilla. It has a total area of 2520 hectares, of which 2274 hectares are forested.

Due to its proximity and characteristics, this forest has been a regular place of practice for Forest Engineering students at the Technical University of Madrid, as well as a research place for faculty staff. For this reason, there are numerous forest inventories, generally circular plots with diameter ranges between 30 and 36 meters, in which only the diameters of all the trees and the height of some of them are usually measured. However, there are two permanent rectangular plots (40x60 meters) in which the most important dasometric variables of all the trees in the plot are measured. For this reason, these plots were chosen for this study.

The mean altitude of the plots is 1408 and 1470 m above sea level, respectively. This altitude range is within the supramediterranean zone of the Mediterranean region and corresponds to an uneven-aged forest and monospecific forest of Scots pine (*Pinus sylvestris* L.). The slope in the plots is 23 and 25 degrees, respectively.

In the other hand, the second study area is located 30 km east from Prague, in the forests managed by the School Forest Enterprise ($ŠLP_{[CZ]}$, $SFE_{[EN]}$), close to Kostelec nad Černými lesy (Figure 7). The main objective of School Forest Enterprise of CULS is to provide internships and scientific support to university students. However, the difference with the Spanish area is that this one is from municipal property, while that of the Czech Republic is managed by the university.

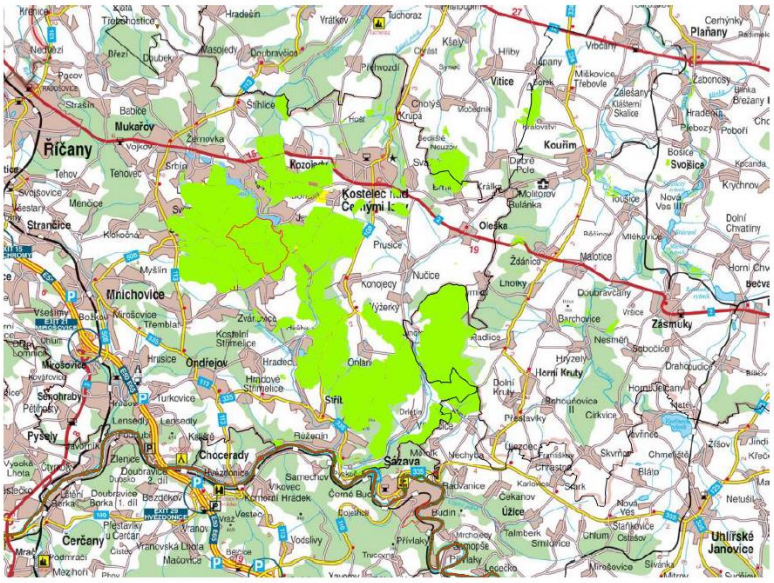


Figure 7. General map of ŠLP Kostelec nad Černými lesy (LHP, 2011).

Almost half of the area corresponds to the 3rd stage of vegetation (oak-beech). Less surface area (20%) is occupied by the 2nd (beech-oak) and the 4th stage of vegetation (beech). The lower areas are represented by the 1st stage of vegetation (oak), while in the higher areas there is the 5th stage of vegetation (fir-beech). The vegetation stage (pines) is represented here only exceptionally, for example, in rocky outcrops (LHP, 2011).

According to the Forest management plan of ŠLP Kostelec nad Černými lesy, the specific composition in the area is mostly woody conifers. Spruce (SM) is the most represented species with approximately 50%. The other conifers are pine (BO) and larch (MD), with 16% and 4% respectively. Broadleaves species are represented mainly by the beech (BK, 14%), followed by oak (DB, 9%) and hornbeam (HB, 1%).

3.2.2 Plot characteristics

The field plots in Spain are located on a steep slope (> 50% slope) and have a rectangular shape (40 x 60 m^2). Within the plots, the height and DBH of all the trees, which are all Scots pines, are known. Both plots show un-even aged forest stands where the coexistence between adult trees and regeneration can be seen in the following frequency histograms. This also causes a high stem density (> 500 trees·ha^{-1}).

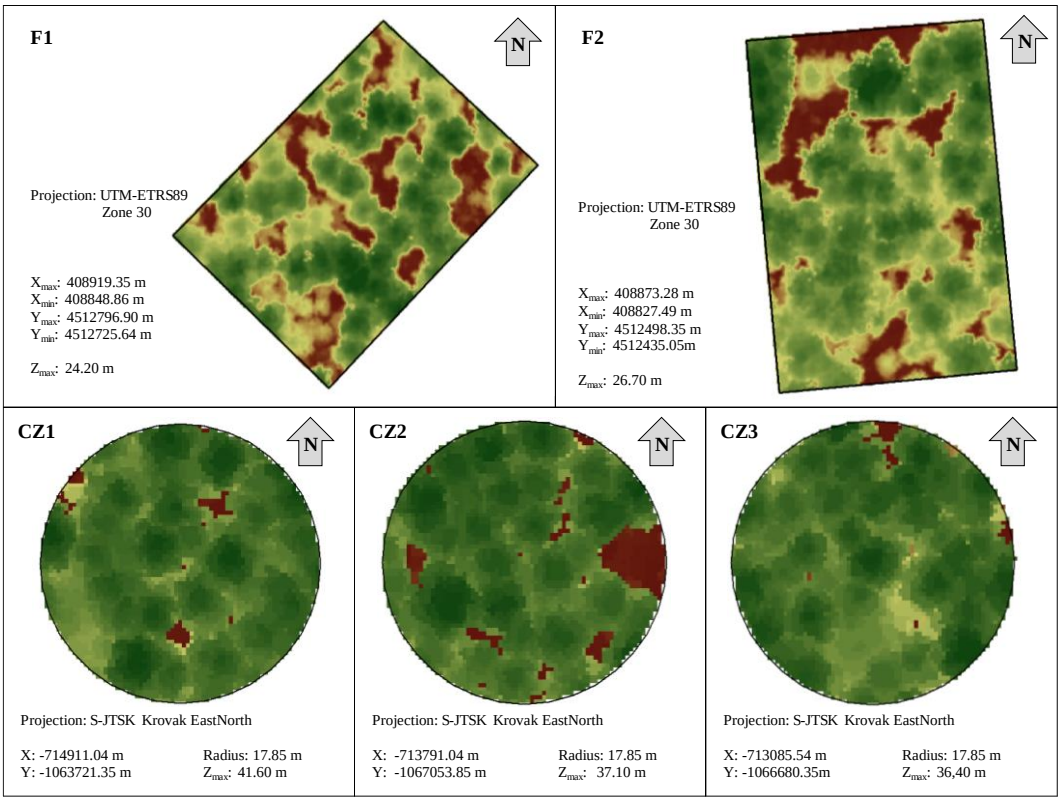


Figure 8. Summary of field plots properties and location.

In Czech Republic, 3 circular plots (1000 sq. m) have been chosen, distributed throughout the forest and whose characteristics differ in terms of the specific composition. All of them have a regular distribution regarding the size of the trees. The stand is composed of several forest species, with heights exceeding 30 meters in general and the regeneration is scarce. Plot 2 (CZ2) contains only spruce trees (*Picea abies*), while plots 1 and 3 (CZ1 and CZ3) are mixed. CZ1 is mostly composed by spruce (89%), followed by larch (Larix decidua, 6%), pine (Pinus sylvestris, 3%) and oak (Quercus patraea, 3%). CZ3 is different from the other, because its composition is mainly pine (67%), followed by spruce (31%) and oak (2%). The stem density range is between 340 and 420 trees·ha^{-1}, and the slope ranges from 10 to 25%.

General information about the location and dimension for each field plot is shown in Figure 8, while the DBH and height distributions are shown in Figure 9.

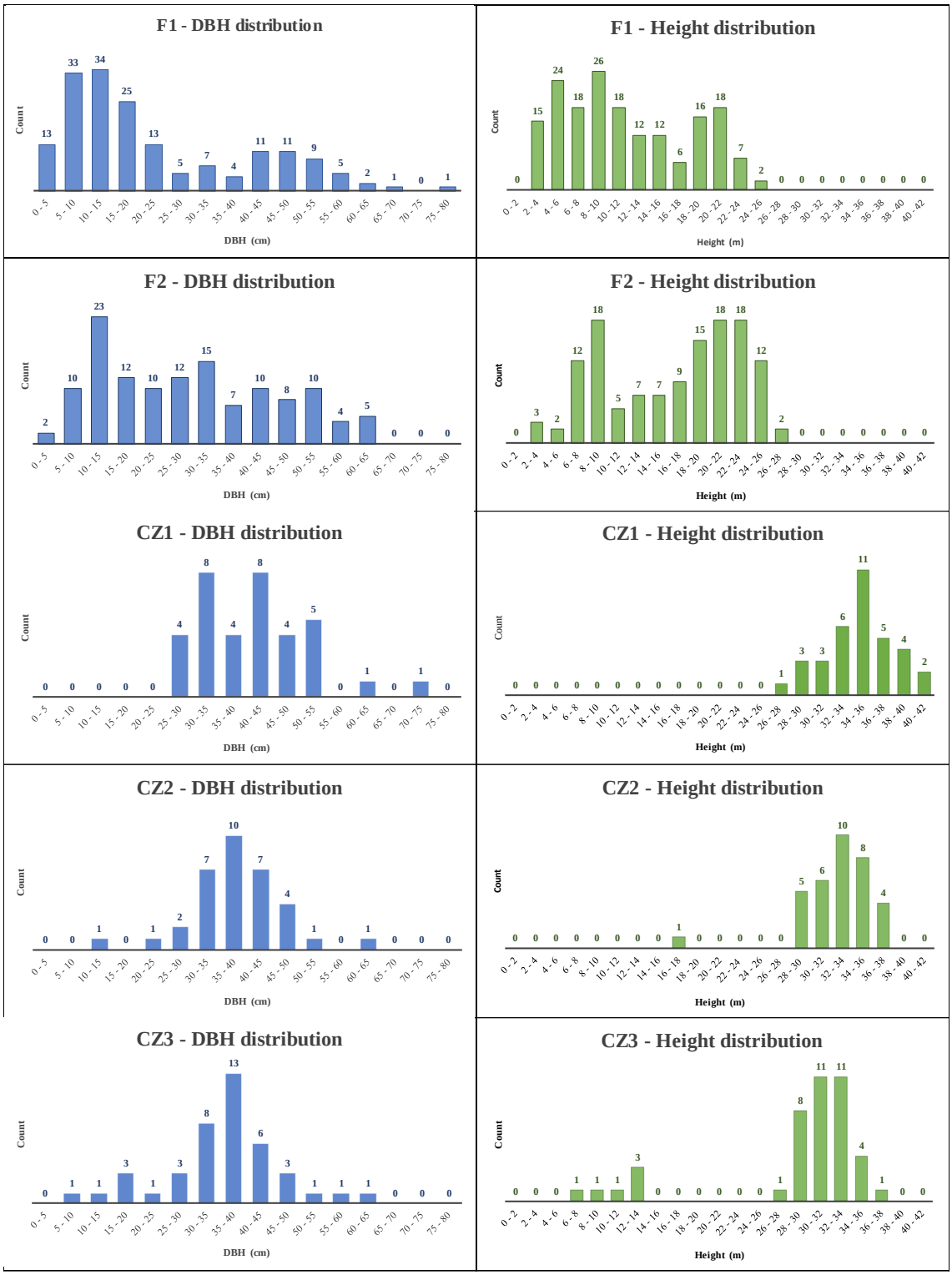


Figure 9. Real DBH and height distribution on each field plot.

3.2.3 Field work

3.2.3.1 Spanish field plots

The Spanish sampling plots were located in areas of thick vegetation, where georeferencing with GPS is imprecise. Therefore, the location of the vertices of the plots were located with topographic measurements creating an oriented topographic (polygonal) itinerary with 18 stations. These points were permanently marked with the installation of resin landmarks. Topographic measurements were made with the Nikon DTM 330 total station. The coordinates of the initial station of the itinerary and the reference point for its orientation were measured with the installation of GPS receivers model Trimble Geoexplorer 3 with external antenna. These points were located on an open meadow (50 meters separated from the nearest trees). The DBH and the total height of all the trees in the plots were measured. Two diameter measurements have been made, one according to the direction of maximum slope and the other according to the level curve. These measurements were taken with a Haglöf Mantax caliper. On the other hand, the total height measurements were made with a Haglöf Vertex III hypsometer. This field work has been carried out during the months of June and July 2003. Later, in 2015, the DHB and the heights of all the trees in the plots were measured again. These are the data used for this study due to proximity to the LiDAR inventory (2011).

3.2.3.2 Czech field plots

The field data collection in the Czech plots took place in the school enterprise Kostelec nad Černými lesy during June and July of 2020. An inventory net (hexagonal cells) based on processed laser data was created. Both the map of the area and the hexagons were displayed in the ArcGIS Collector application, which was used for field work orientation and for field plot establishment. The information about the area and trees values like type of species, DBH, height, height of the crown was taken. Different size circular plots were established on each hexagon (300, 500 and 1000 m^2). The height was measured with Vertex Laser GEO and diameter by caliper and measuring tape.

3.2.4 LiDAR data

3.2.4.1 Spanish ALS flight

The Spanish study area has been flown over with a *LiDAR* sensor twice. The first time was in 2002, when an area of 171 ha was covered with a point density of 4.5 per square meter. The second and last one was in 2011. During this flight both the area studied, and the density of points were increased. For these reasons, in addition to being more recent, the 2011 LiDAR inventory will be used for this case study. Its characteristics are described (Table 1).

For the capture of *LiDAR* data during 2011, the *ALS70-HP* sensor of the Leica company was used, with a scanning angle (*FOV*) of 14 degrees, a repetition rate of the laser pulse of 200 kHz and a scanning rate of 73.7 Hz. The average distance between laser pulses is 0.29 m and the average area of each pulse is 0.5 m^2. All this configuration has resulted in an approximate density of 24 points·m^2.

Table 1. Spanish LiDAR flight characteristics

Characteristics	Cercedilla, Spain
Date of *LiDAR* acquisition	July, 2011
LiDAR sensor	Leica *ALS70-HP*
Sensor type	Multi-return (5 returns/pulse max)
Elevation range	625 - 1200 m
Total coverage area	220.3 ha
Number of passes	7 passes (*Fig. 8*)
Average pass width	400 m
Vertical precision	0.08 m
Horizontal precision	0.15 m
Point density	24 points·m^{-2}
Reference system	*UTM-H*30 *ETRS*89

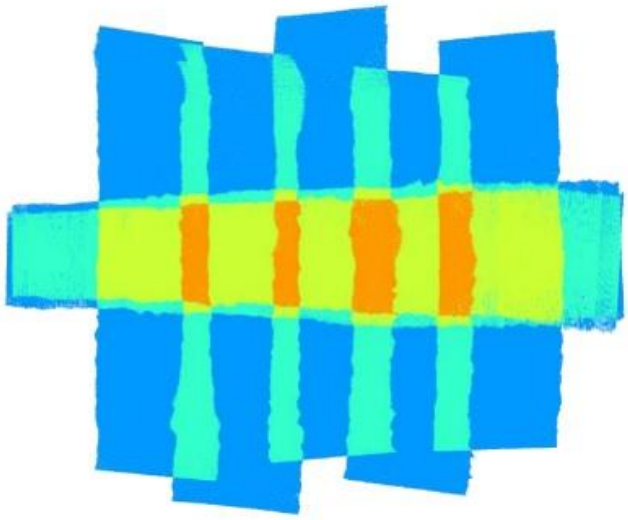


Figure 10. Graphic scheme of the 7 passes made by the plane in the LiDAR inventory of 2011 (Tordesillas Torres, 2014). The colors show where more information has been collected: Dark blue: 1-time scanned, Light blue: 2-times scanned; Yellow: 3-times scanned; Orange: 4-times scanned.

3.2.4.2 Czech ALS flight

The Czech study area has been flown over with a *LiDAR* sensor in April 2020, with the next characteristics (Table 2). For the capture of *LiDAR* data, the *ALS70-CM* sensor of the Leica company was used, with a scanning angle (*FOV*) of 50 degrees and a scanning rate of 47 Hz.

The average density was 3 pulses·m^{-2} in each scanning line, with an overlapping of 30% between lines. The mean point density in the study area was 5.48 points·m^{-2}. Additionally, aerial images (60 Mpx/image) of 15 cm resolution (at ground level) were captured. These images had a minimum longitudinal overlap of 85% and transversal of 45%.

Two of the 21 flight lines were run transversely to increase the precision of the scan. 63% of the scan area (5000 ha) was forest.

Table 2. Czech LiDAR flight characteristics

Characteristics	Kostelec, Czech Republic
Date of *LiDAR* acquisition	April, 2020
LiDAR sensor	Leica *ALS70-CM*
Sensor type	Multi-return
Elevation range	1130 - 1400 m
Total coverage area	8000 ha
Number of passes	21 passes
Average pass width	1306.13 m
Vertical precision	0.07 – 0.10 m
Horizontal precision	0.14 – 0.17 m
Point density	20 points·m^{-2}
Reference system	*UTM-H*33 *ETRS*89

3.2.5 Point cloud processing

3.2.5.1 Extracting the laser data on each field plot

The geospatial information was in the form of a point cloud with .LAS format. To process it, the points corresponding to each field plot were extracted at first. For this, the software RStudio (version 1.1.453) was used. This program works with commands lines, which allow full control and customization for the analyzes. Furthermore, apart from executing commands specially designed for this program, running external software is also possible with RStudio. Due to the shape of the field plots (Spanish rectangular and Czech circular) the extraction of the point clouds was carried out in two different ways.

The Czech plots (CZ1, CZ2 and CZ3) were circular, which means that just the central coordinates and the diameter are necessary for clipping. This was done using the FUSION/LDV software (version 3.8) developed by the US Forest Service. It is a free software for processing laser data with .LAS format. “Clipdata” command allows clipping circular or rectangular plots and for this last one, both minimum and maximum X and Y coordinates are specified. However, the clips are oriented with geographic north. In the case of circular plots, orientation is not a problem, therefore it was the methodology chosen for these plots.

The Spanish plots (F1 and F2) had a rectangular shape (60x40 m^2) and their orientation in the field is parallel to the slope, that is, the longer side was placed parallel to the contour lines, while the shorter side was placed perpendicular to it (steepest slope). Therefore, the clip with FUSION/LDV was not possible. The "clip" command from RStudio ("lidR" package) uses a polygon in shapefile (.shp) format for clipping. This tool solved the problem of orientation. It is a much slower technique than the previous one, but it is a good alternative for irregular shape and orientation field plot.

The LAS density for each field plot can be consulted in Table 3:

Table 3. Point cloud density for each field plot both with and without ground points, respectively.

Plot	LAS density	LAS (no ground) density
	points·m^{-2}	*points·m^{-2}*
F1	11.84	8.92
F2	22.66	16.70
CZ1	9.89	6.89
CZ2	12.87	7.58
CZ3	12.81	8.57

3.2.5.2 LAS conversion into CHM

LiDAR vectorial data (point cloud) was transformed into raster with the CHM calculation. For this, the following chronological set of operations was caried out. First, the point clouds were normalized to eliminate the effect of the slope. After normalization, the point heights represent the real height above the ground. This process is usually carried out with the support of a digital elevation model (DEM). However, a high-precision DEM was not available for the Czech Republic study area. To homogenize the analysis process, an alternative and efficient solution was carried out by programming in RStudio. "Classify_ground" command ("lidR" package) was used, which classified the point clouds into ground and non-ground points using different algorithms. In this case, the "csf" function was chosen, which implements an algorithm for the segregation of ground points based on a cloth simulation filter, developed by (Zhang, et al., 2016). The calibration of some parameters such as slope and roughness of the terrain together with

some parameters of the cloth was necessary. After all, a classification of ground and non-ground points is obtained (Figure 11.A).

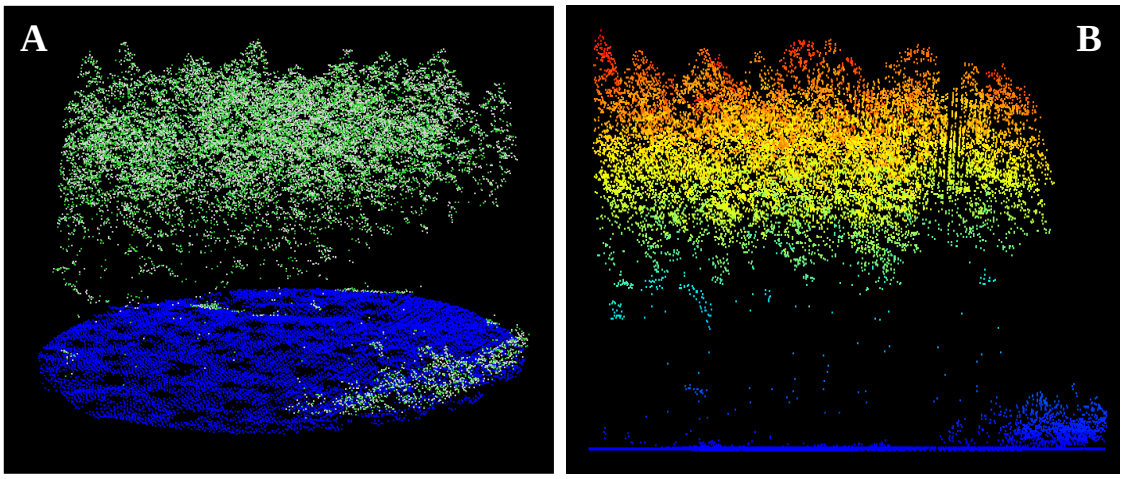


Figure 11. (A) CZ2 point cloud classified into ground and above-ground points; (B) CZ2 normalized point cloud.

After classifying the ground points, the normalization of the point cloud was performed using the "normalize_height" command (“lidR” package). Despite it allows to do the normalization using an external DEM, the automatic calculation of the ground surface was set for this study. For this, the “TIN” function was used, which implements a Delaunay spatial triangulation. Within each triangle a linear interpolation is performed between the ground points.

As a result, a point cloud file in which the ground points have no height was obtained (Figure 11.B), which is already prepared for the CHM extraction. To do this, the "grid_canopy" command from the "lidR" package was used in RStudio and the "dsmtin" algorithm was chosen for ease of calibration among the other two remaining algorithms. For this study, two CHM files were extracted for each field plot with different resolution (0.2 and 0.5 m). The chosen algorithm implements a Delaunay triangulation between the first LAS returns, so that within each triangle a linear interpolation is performed. The resulting CHM had a first smooth step because CHM heights slightly lower than the originals. Other algorithms, such as the "CanopyModel" command used by FUSION/LDV software assigns the height of the highest point to each cell.

3.2.6 CHM processing

3.2.6.1 Projection

Both Czech layers and the plot coordinates were in S-JTSK Krovak East North projection (a local coordinate system that is mainly used in the Czech Republic), while the point clouds and the CHMs were in UTM ETRS89 (Zone 33). Therefore, the CHMs were projected into UTM with the ArcMAP program (version 10.6). First step was to define the current coordinate system for each CHM with the "Define projection" tool, so that the layer has information about its current projection (Figure 12.A). Next, each CHM was projected to S-JTSK Krovak (NE) coordinate system, using the "Project raster" tool (Figure 12.B). Finally, the resulting layer was clipped with the shape of the field plot, since the point cloud clip was made with a diameter of 60 m, while the diameter of the plot was 35.7 m (1000 sq. m).

3.2.6.2 Smoothing the 0.2 m resolution CHM

In general, the CHMs contained some cells with anomalous data or no value, because within those specific cells there was not any LAS point, especially in the files with lower resolution. The lower the resolution of the raster file, the greater the probability of finding empty cells. Therefore, a smoothing filter was passed to the 0.2 m resolution CHMs prior to analysis, using the ArcMAP "Filter" tool. With this spatial function, the value of each cell is replaced by the mean value of the neighboring cells. This corrected anomalous values, although cells with correct values lost precision. The Czech plots had a lower density of points than the Spanish plots. For this reason, the CHMs generated for the Czech plots had more empty cells than for the Spanish plots.

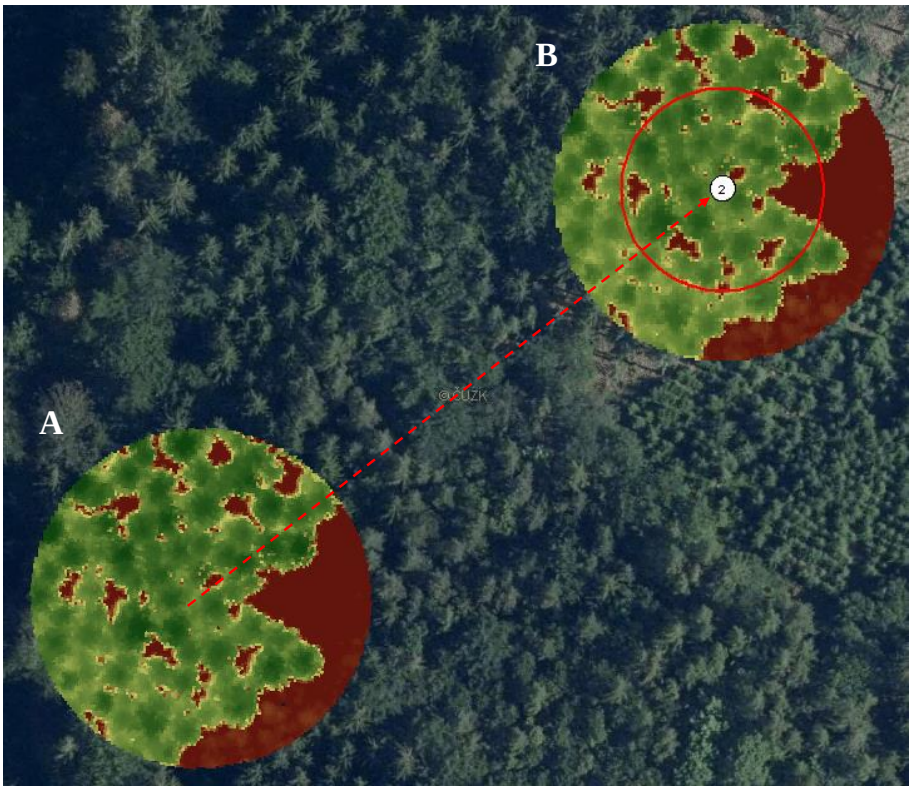


Figure 12. (A) CZ2 field plot in UTM ETRS89 (Zone 33); (B) CZ2 field plot in S-JTSK Krovak East North projection. Red circle is the 37.5 m in diameter field plot.

3.2.6.3 Ground and above-ground raster layers

Two raster layers were extracted from the CHM: ground and above-ground layers. For the 0.2 m resolution CHMs, the smoothed CHMs were used, since the original CHMs had many empty cells in general. The heigh distribution of each plot was evaluated to determine the limit height between the cells that were considered ground and those that were considered vegetation. The Spanish plots had a lot of regeneration, so the limit height was much lower than for the Czech plots (with hardly any regeneration). Ground cells were considered those with a height of less than 3.5 m for F1 and F2, less than 20 m for CZ1 and CZ3, and less than 24 m for CZ2 (Table 4). This process was done using conditionals with ArcMAP. For the calculation of the ground raster, those cells lower than the limit height were replaced with a value = 0 (ground cells), while those with a height equal to or greater than the limit height were replaced with a value = 1 (crown

cells). On the other hand, the vegetation cells (height ≥ limit height) were also extracted in an external layer (crown).

Table 4. Height limit to classify the CHM cells in ground and above-ground data.

Plot	Height limit (m)
F1	3.5
F2	3.5
CZ1	20
CZ2	24
CZ3	20

3.2.6.4 CHM smoothing (Kernel distances)

The tree crowns usually have very irregular shapes, with multiple branches that cause rises and falls in the profile of each crown. This is a problem when for detecting trees with basin segmentation since each crown can have several sub-basins associated. To solve this, different smoothing CHMs were performed, characterized by the evaluation distance (kernel distance). In particular, 4 types of distances were evaluated (0, 1, 3 and 7 radius cells). Each cell of the CHM was replaced by the mean value of its neighboring cells, considering these 4 radii. Kernel 0 means that each cell corresponds to itself (no smoothing). On the other hand, Kernels 1, 3, and 7 take all the cells that are in a radius of 1, 3 and 7 respectively to calculate the mean value that replaces each cell (Figure 13). The greater the smoothing, the greater the homogenization of the tree canopy. However, the accuracy in the total tree height also worsens.

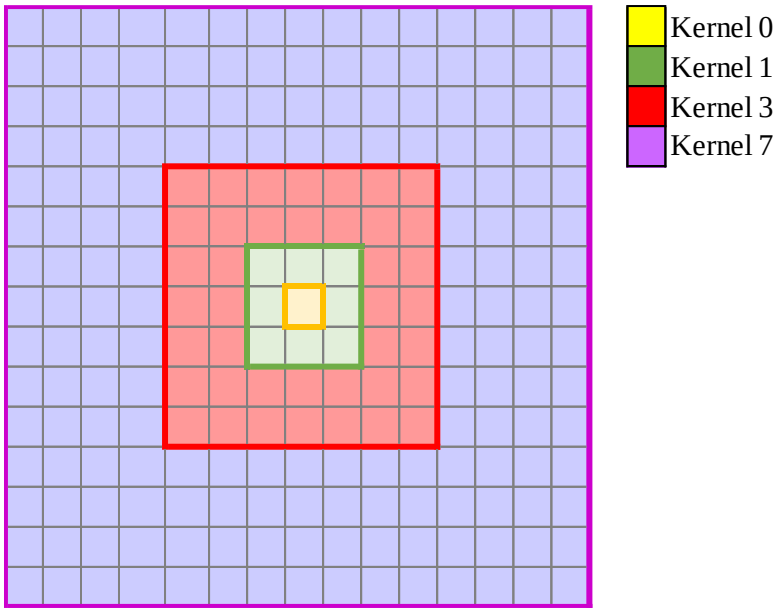


Figure 13. Representation of the cells that are taken for the smoothing considering each kernel situation.

3.2.7 Individual Tree Delineation

ITD were performed using three different methodologies. Two of them use the CHM raster layer to assess the position of the trees on each sample plot. The first one was an adaptation of the technique applied by Panagiotidis, Abdollahnejad, Surový, & Chiteculo (2017) and was developed with the ArcMAP 10.6 program. This is based on the river basin principle, so that by inverting the CHM (Inverted Watershed, IWS), each treetop can be understood as a watershed. The second one was based on different conditions established in the FUSION "TreeSeg" tool. Finally, an algorithm for vectorial tree extraction was evaluated in RStudio.

3.2.7.1 ArcMAP

Initially, the objective of this study was to apply the methodology developed by Panagiotidis, Abdollahnejad, Surový, & Chiteculo (2017), in which the location of each tree is detected by testing different kernel distances at first. Then, the tree crown was delineated by the watershed segmentation using previous trees locations as the pour points of the canopy basins. However, wrong delineations resulted due to the sub-basin effect. Tree crowns were delineated just partially in all of the kernel distance scenarios. Therefore, a modification of this methodology was proposed.

The alternative methodology (performed in this study) reversed the previous watershed segmentation process. First, all the tree basins of the CHMs were delineated considering all the smoothing cases (all crowns were detected). Then, the maximum height point for each crown corresponds to the height and position of the trees.

Panagiotidis, Abdollahnejad, Surový, & Chiteculo (2017)

1st	**Tree position**	→	2nd	**Crown delineation**
	Local maxima			Watershed segmentation

My alternative

1st	**Crown delineation**	→	2nd	**Tree position**
	Watershed segmentation			Local maxima

Therefore, 2 resolutions (0.5 and 0.2 m cell size) and 4 types of smoothing (0, 1, 3 and 7 kernel radius distance) were studied for each sampling plot. That is, 8 study cases. In total 40 case studies are required for the 5 sample plots. This process was semi-automated using the ArcMAP "Model Builder" tool, so that the CHM and some parameters such as the smoothing distance and the resulting layers were required to start the process.

First the CHM was smoothed considering the required kernel distance, using the "Focal statistics" tool. As mentioned in the "CHM smoothing" section, the value of each cell was replaced with the mean value of the cell values in the kernel neighborhood. Then, ground cells were corrected. If the smoothing is low, this correction is not necessary, but as the smoothing is stronger (Kernel 3 or 7) it must be corrected with the ground layer, because the trees would be greatly over dimensioned. This correction was made using the "Times" tool, which multiplied the ground layer and the smoothed CHM, replacing the ground cells with 0 value. In continuation, the ground cells were eliminated using a conditional (value > 0), while the remaining the cells (above-ground cells) were later inverted using the "Times" tool again, this time multiplying these cells by -1. At this moment, the trees appeared inverted, being the top of the three as the lowest points (pour

points). Once this is done, it was possible to evaluate the tree crowns as hydrographic basins, first with the calculation of the flow direction ("Flow direction" tool) and then with the calculation of the watershed using the "Basin" tool. Every basin raster layer was transformed into a polygon layer (.shp) using the "Raster to polygon" tool. Additionally, the surface of each polygon was calculated in the attributes table (crown size).

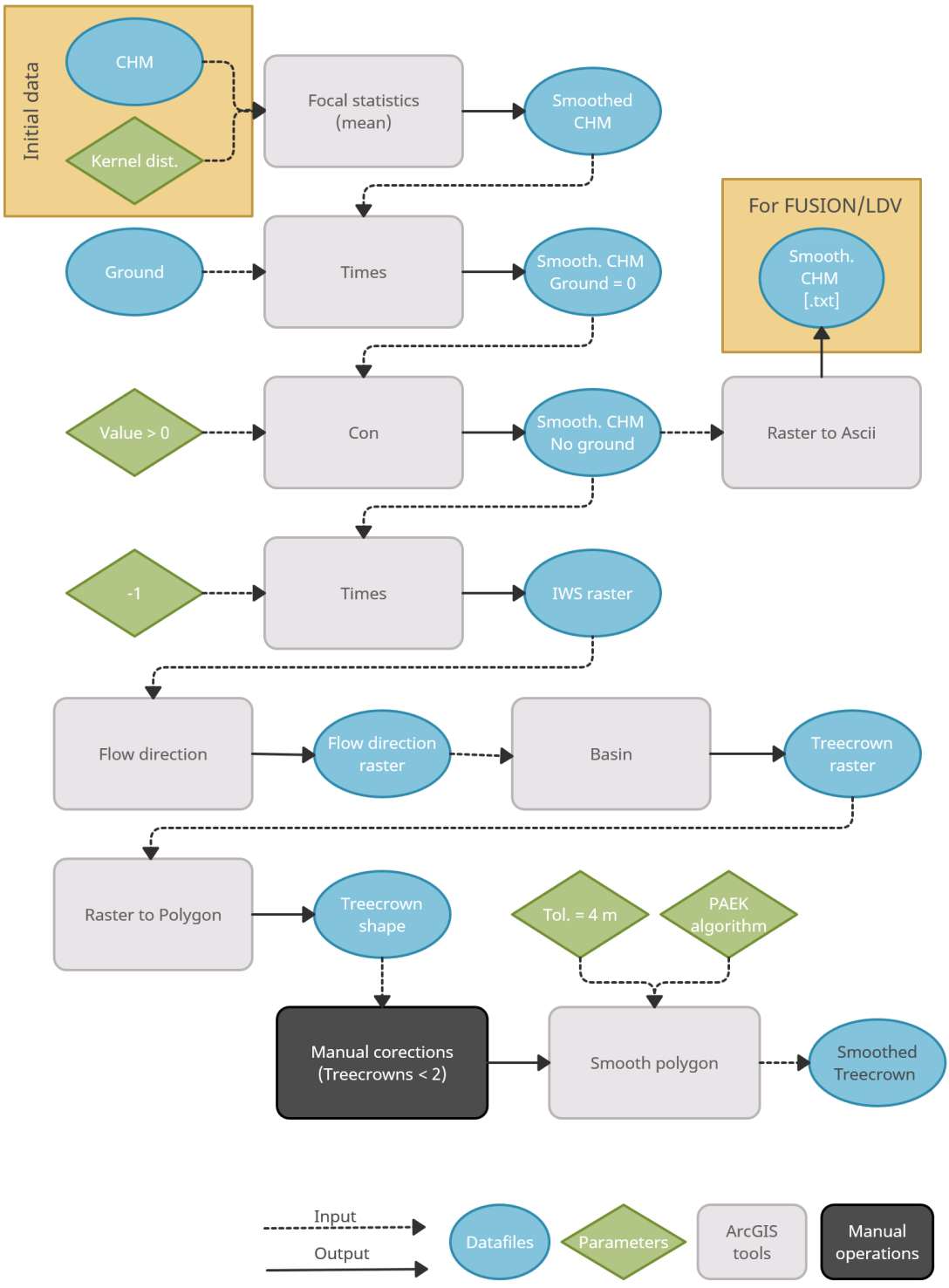
This process produced some errors, especially in CHMs with 0.2 m resolution and no smoothing (K0). Therefore, some corrections were made manually. First, the polygons that had less than 2 m^2 in surface were merged with the largest polygons with which they were in contact. This was done using the "Eliminate polygon" tool in ArcMAP, after selecting the polygons in the attributes table. Those isolated little polygons (mostly in the edge of the plot) were removed manually using the ArcMAP editor. Finally, all the polygons were smoothed for more natural shape using the "Smooth polygon" tool, which needs some calibration, as the application algorithm ("PAEK") and the tolerance (4 meters). This combination was used because it provided the best visual results.

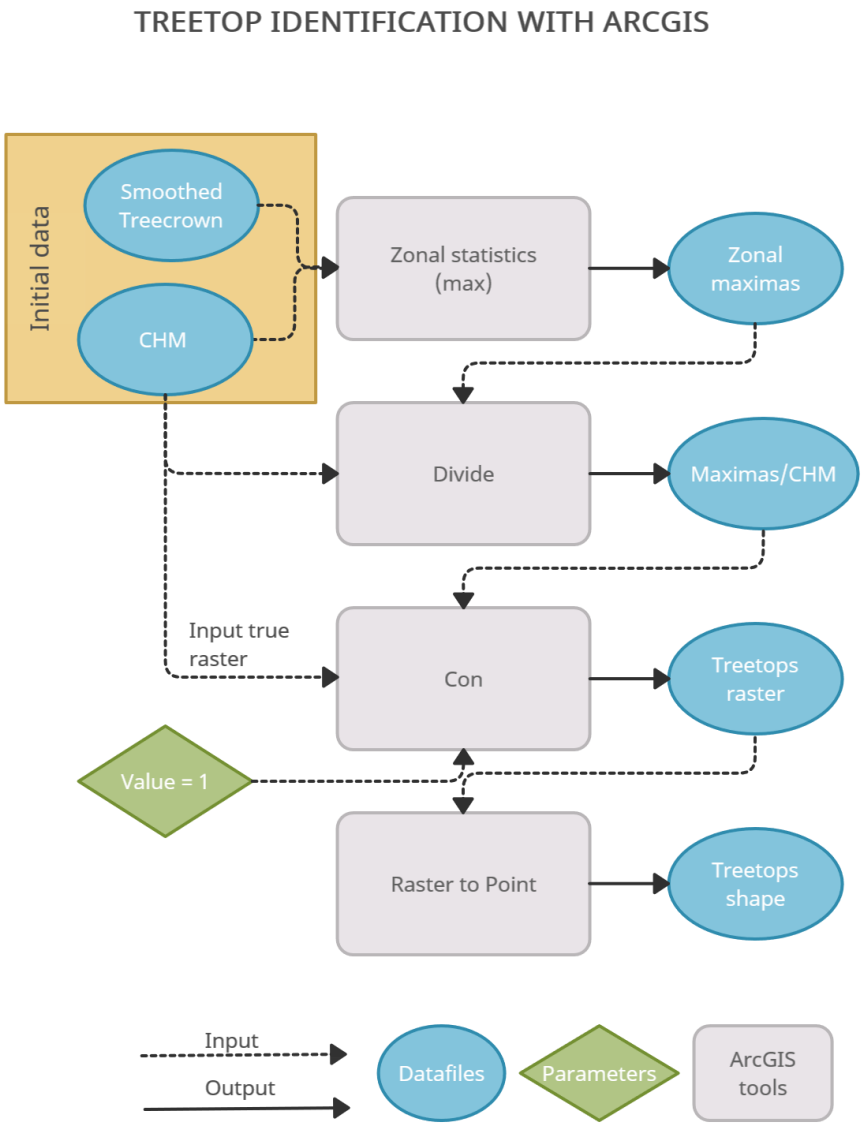
The highest point on each polygon was extracted. This process was also modeled with the "Model Builder" in ArcMAP because it was a mechanical process. The point shapefile was obtained applying the "Zonal statistics" tool using the original CHM (without smoothing) and the polygon layer as feature zone data. The result was a raster, in which the cells corresponding to each polygon had the same value (zonal maximum). By dividing this raster layer and the CHM, those cells with value = 1 corresponded to the maximum points, which were transformed into a point layer using the "Raster to point" tool. Additionally, the "Zonal statistics as table" tool was used to assign the maximum height to each polygon. This could also be done by a spatial join using the calculated point layer.

To sum up, two shapefiles were obtained for each sample plot in every scenario (kernel-resolution):

- Point shapefile corresponding to the location and height of the trees.
- Polygon shapefile corresponding to the shape of the tree crowns.

TREECROWN DELINEATION WITH ARCGIS





3.2.7.2 FUSION

The "TreeSeg" tool of FUSION software was used as the second methodology for the detection and delineation of individual trees from the CHM layers. This tool process twice the CHM for tree crown delimitation. First, it develops the watershed segmentation methodology by inverting the CHM layer, which produced similar results than the segmentation performed with ArcMAP. Then, it can implement a second algorithm. The highpoint of each basin is used as the central point for 18 evenly-spaced radial profiles

that are evaluated. Three criteria were used to select the point of each profile that represents the edge of the crown:

1. The selected point is a local minima (Figure 14.A)
2. The selected point and its neighbors have a height less than 2/3 of the treetop point height (Figure 14.B)
3. The change in height before the point is more than 4 times the change in height after the point (Figure 14.C)

This set of rules produces basin/crown polygons that seem to represent objects visible on the canopy surface. The resulting polygons can overlap and they will not usually fully encompass the basin (McGaughey, 2018).

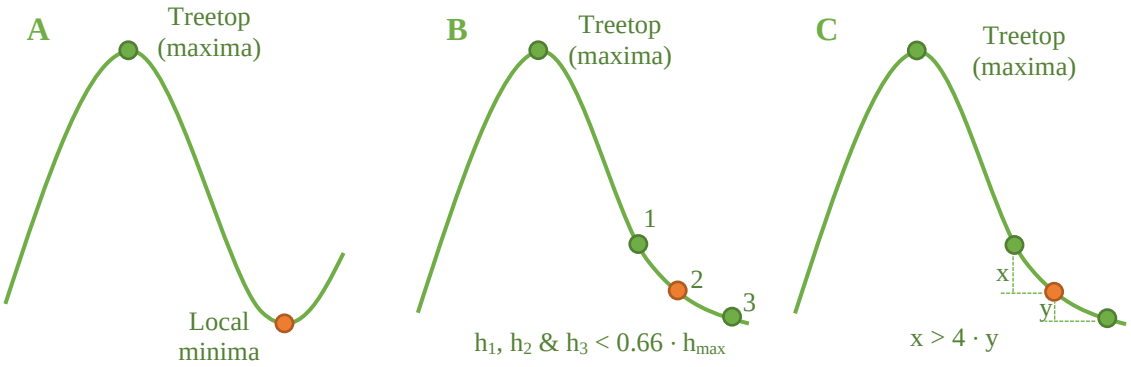


Figure 14. Criteria for the tree crown edge point selection on each radius profile using "TreeSeg" tool in FUSION/LDV.

The RStudio software was used to execute "TreeSeg" tool, by controlling the Windows console in the background through the command line. These commands required some parameters such as the directory of the CHM (previously smoothed in ArcGIS) and the directory of the result layers. The height threshold was also detailed, which represents the minimum height for object segmentation. For this parameter, the heights of the Table 4 were used.

3.2.7.3 RStudio

Apart from the tree delineation applying the watershed segmentation over the CHMs, the detection of individual trees has been also evaluated by processing the point cloud directly. For this, the "segment_trees" tool from the RStudio "lidR" package has been used, which perform the tree segmentation based on different algorithms and the point cloud. The algorithm developed by Li, Guo, Jakubowski, & Kelly (2012) was used, which is based on the distance between trees to classify the points. Even if the trees are overlapped in close canopies, there is always some distance between them at higher levels. This algorithm runs the classification from the highest point and continues including points by evaluating the relative distance between nearby trees (Figure 15).

The point cloud files were processed in the same way as to obtain the CHM (classification of the ground and height normalization). Additionally, those classified as ground points were eliminated to improve the tree crown delimitation. All the tree points were projected horizontally and the crown perimeter was delineated by calculating the convex hull. Therefore, if there are ground points classified as part of the tree, these could oversize the tree canopy. To eliminate the ground points, the point cloud was filtered, eliminating those with value 2 (ground points).

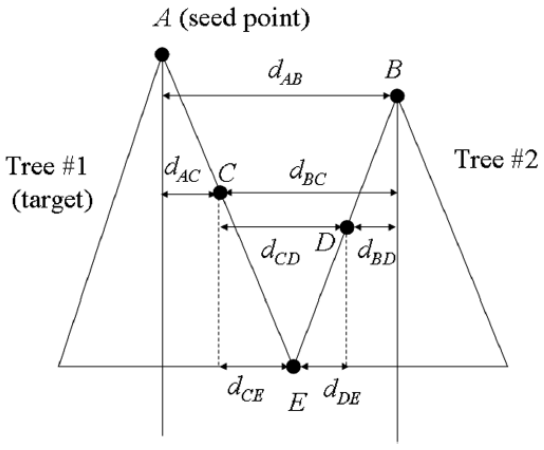


Figure 15. Individual tree segmentation principle based on the relative distance evaluation (Li, Guo, Jakubowski, & Kelly, 2012)

As a result, two shapefiles were obtained for each sampling plot:

- Point shapefile corresponding to the location and height of the trees.
- Polygon shapefile corresponding to the shape of the tree crowns.

3.2.8 Statistics

3.2.8.1 Height distribution

On each sample plot the height and the DBH of all the trees were known, but not their location. Therefore, the analysis could not be validated by comparing real and estimated trees directly (tree by tree). As an alternative, the real and estimated height distributions were compared on each scenario. In this way, the height data were classified in 2 meters ranges. The Spanish trees were smaller in height than the Czech trees. Even so, ranges of heights were distinguished from 0 to 52 meters every two meters (0 - 2; 2 - 4, ..., 50 - 52).

3.2.8.2 Low and high vegetation

Previous to the statistical analyzes, the information was divided into low and high vegetation, considering the height distributions in every sample plot (Figure 9). These height ranges can be consulted in the next table (Table 5). There was great representation of trees on each classification for Spanish plots, because regeneration is very high and the forest is uneven-aged. However, the Czech plots have very little representation of low vegetation, with no small trees in CZ2.

Table 5. Height range classification for low and high vegetation.

Plot	Low vegetation	High vegetation
F1	0 – 16	16 – 52
F2	0 – 16	16 – 52
CZ1	0 – 26	26 – 52
CZ2	0 – 26	26 – 52
CZ3	0 – 26	26 – 52

3.2.8.3 RMSE & nRMSE

Root Mean Squared Error (RMSE) was used to evaluate the goodness of the estimations on each scenario. It is an error parameter commonly used in many scientific and statistical investigations, as it is considered an excellent error metric for numerical predictions. However, this error is very sensitive to scale, so it is only recommended for comparing two variables of the same nature. It is defined as the square root of the mean value of all squared errors (Formula 1).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - O_i)^2} \quad [1]$$

where O_i are the observations, S_i predicted values of a variable, and n the number of observations available for analysis (Neill & Hashemi, 2018)

The RMSE is an absolute error and its magnitude is difficult to interpret. For this, the normalized RMSE (nRMSE) or relative RMSE (Formula 2), which is easier for understanding.

$$nRMSE = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n} {O_i}^2} \quad [2]$$

The estimations presented some mismatches with the observed values, in form of trees detected in height ranges where no tree had been observed in the field. This caused the nRMSE to tend to infinity when dividing by zero many times. To solve this, just the evaluation of RMSE in the height ranges with real observations, neglecting those estimations in ranges without observed trees. This measure did not explain the error completely. Therefore, the similarity in the height distributions (observed and estimated) using the Kolmogorov-Smirnov test.

3.2.8.4 Kolmogorov-Smirnov test

The Kolmogorov–Smirnov test (KS test) compares if two datasets differ significantly by evaluating the maximum distance between the experimental cumulative distribution function and the theoretical cumulative distribution function. The maximum discrepancy is defined by the next equation (Formula 3):

$$D = \max |F_1 - F_2| \quad [3]$$

where F_1 is the experimental distribution and F_2 = theoretical distribution.

The graphical representation that describes the K–S test is demonstrated in Figure 16, which demonstrates the maximum discrepancy D.

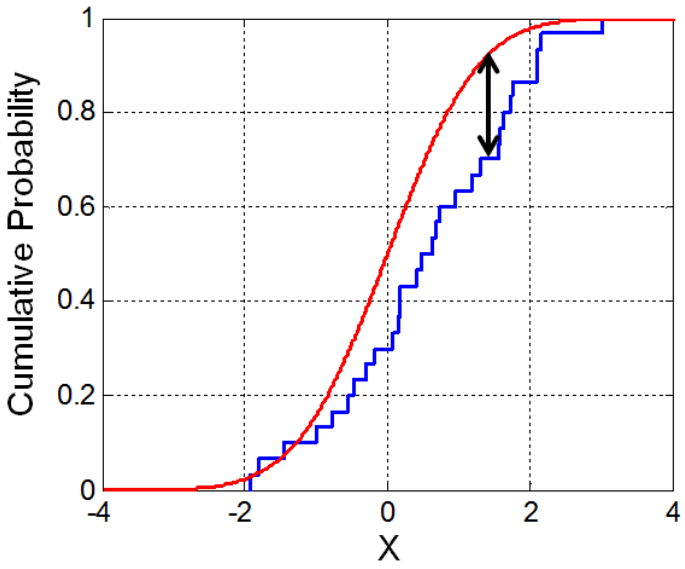


Figure 16. KS-test illustration. Black line represents the maximum discrepancy between both cumulative functions (empirical and estimation).

KS-test p-value was considered to evaluate the similarity of both distributions. The P-value is a probability value and the closer to 1 means that the distance D is smaller

and therefore that there is more similarity between the observed values and the predicted values, in this case.

3.2.8.5 Analysis of variance

The analysis of variance (ANOVA) test was performed to evaluate the effect of the different parameters (kernel distance for smoothing, CHM resolution and methodology) on the estimation. A confidence level of 95% was considered for signification.

3.2.8.6 Correlation analysis

DBH, total height and crown diameter were correlated using the Spanish field data of 2002 because tree crown diameters were not measured in the 2015 Spanish and the 2020 Czech field inventories. Single Pearson correlation was calculated for DBH with the other two remaining tree variables. In addition, the best regression model was searched.

3.2.8.7 Software for calculations

Data processing, the heigh distributions, the error calculation, and the KS-tests were performed with the RStudio program, because the ease of programming the calculations. This program has the advantage that, once the execution code has been programmed, the results are immediate. Subsequently, the ANOVA tests and correlation analysis were carried out in Statgraphics Centurion 18 (v. 18.1.13).

4 RESULTS

4.1 Low vegetation

The classification of vegetation in different layer heights (low and high) allowed to evaluate the precision of the detection both in dominant and suppressed trees. The use of different types of smoothing has produced different results in high vegetation detection. In general, individual trees detection was better as the smoothing was stronger. Raw CHMs (no smoothing) have produced very high errors (nRMSE = 2) in dominant trees detection (Figure 17.A), which decreased as smoothing was higher until reaching values around 80%. These errors correspond to mean values since they comprise all the sampling plots. The goodness in estimating suppressed trees (low vegetation) was independent of the CHM smoothing and the applied methodology, since the general mean error produced was set at 80% in all the study cases (Figure 17.A). In order to evaluate the suppressed trees, the KS p-values were also compared in the different sampling plots, since the Czech plots had hardly any low vegetation (Table 6) and they influenced the mean results.

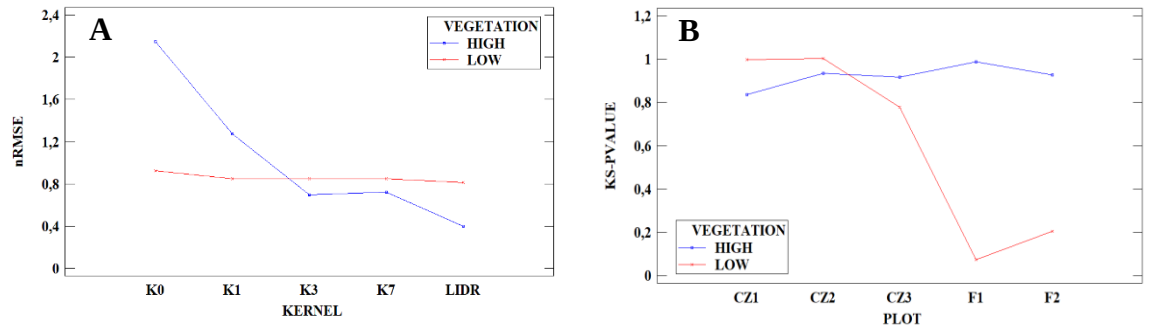


Figure 17. Estimation of low and high vegetation results. (A) Evolution of nRMSE in different cases of smoothing. (B) KS p-value on each plot. Decimal numbers with commas.

Figure 17.B shows the KS p-values as a function of the sampling plot and the vegetation classes. The p-values of the high vegetation were very similar in all of the plots (> 0.8). This means that the estimating height distribution closely resembles the measured height distribution in the field. Values greater than 0.05 (95% significance) allow

accepting the null hypothesis of the KS test (similarity between distributions). However, the closer this value is to 1, the lower the D statistic and therefore the more similar both distributions are.

Clearer differences were obtained in low vegetation detecting. First, the Czech plots had KS p-values greater than 0.7, being equal to 1 for CZ1 and CZ2, since there are hardly any small trees in these plots (0 and 1 respectively). On the other hand, in CZ3 six small trees were noticed, which represents 14% of all of the measured trees. The KS p-value in this plot was lower than those in the other Czech plots, because suppressed trees were not detected in any of the study cases. The Spanish plots had a higher percentage of low vegetation. F1 had 125 small trees, representing a percentage of 72%, while F2 had 54 small trees, 42% of the total. In general, including all the methods and parameters studied, the low detected vegetation in the Spanish plots were few compared to the observed field trees (16% in F1 and 24% in F2). Therefore, the KS p- were much lower than those of the Czech plots (Figure 17.B).

Table 6. Observed and predicted low and high vegetation on each sampling plot. An average estimated value was calculated for each plot and type of vegetation considering all the study cases (scenarios).

	LOW VEGETATION		HIGH VEGETATION	
	Field	Esimation	Field	Estimation
F1	125	21	49	57
F2	54	13	74	71
CZ1	0	0	35	51
CZ2	1	0	33	35
CZ3	6	0	36	43

4.2 ArcMAP vs FUSION

The CHM segmentation methods were analyzed in the following graphs (Figure 18). Both methodologies were simple to apply and depended on the raster layer quality. However, they used different methodologies. ArcMAP developed the watershed segmentation, while FUSION established different criteria for tree crowns delineation. The errors (nRMSE) produced by watershed segmentation in ArcMAP were lower than

those produced with the methodology developed in FUSION. The difference between both analyzes was statistically significant (ANOVA p-value = 0.0298 < 0.05).

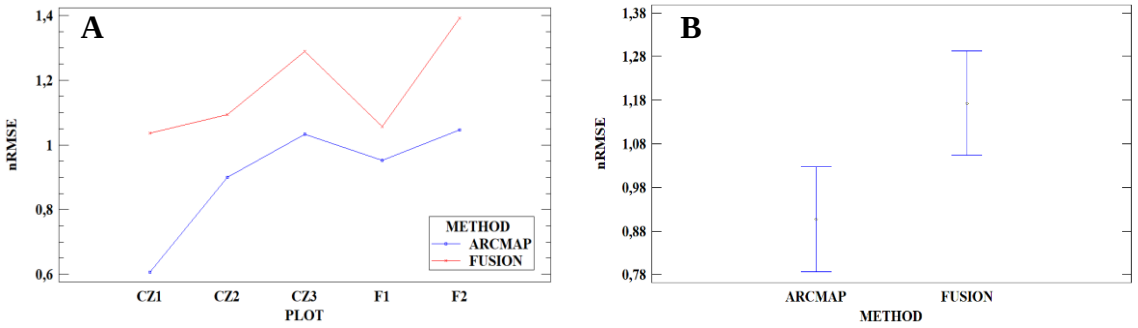
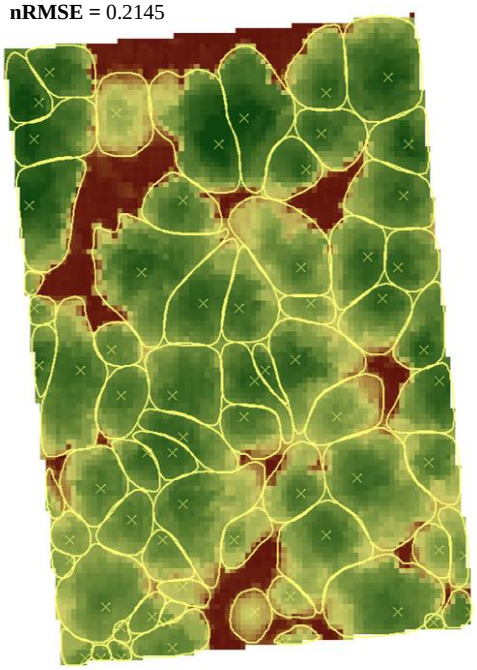


Figure 18. Effect of ITD in both CHM segmentation. (A) nRMSE for ArcMAP and FUSION methods on each sampling plot. (B) General nRMSE on each CHM methodology. Decimal numbers with commas.

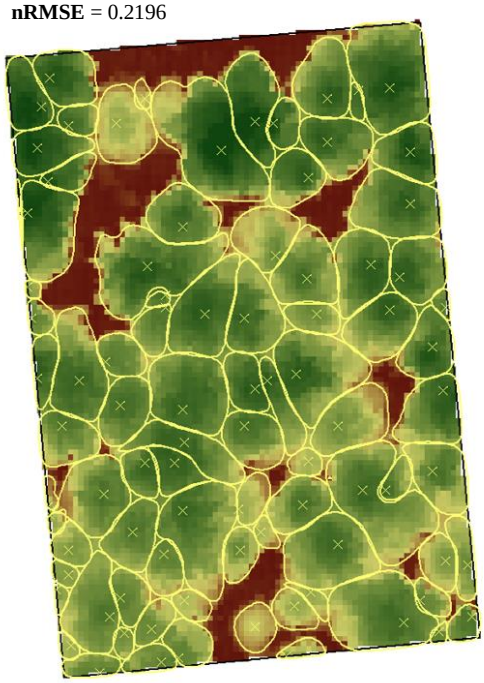
Figure 18.A shows the errors produced according to the CHM segmenting method and the sampling plot. CZ1 had the best results with the ArcMAP method, although its results are largely influenced by low vegetation, same as CZ2 and CZ3 plots. The influence of this type of vegetation in Figure 18.A was shown in Figure 17.B, where the KS p-values were clearly higher in the Spanish plots compared to the Czech ones. In general, an improvement of almost 20% using ArcMAP instead of FUSION was obtained (Figure 18.B).

These results refer to the trees detection, but they do not consider their crown delineation. Apart from producing the best results in tree detection analyses (comparing height distributions), ArcMAP also provided a more realistic crown delineation. Visually, more natural shapes can be seen in the layers obtained with ArcMAP watershed segmentation. In the other hand, FUSION delineated the tree crowns with an irregular and overlapping perimeter (Figure 19).

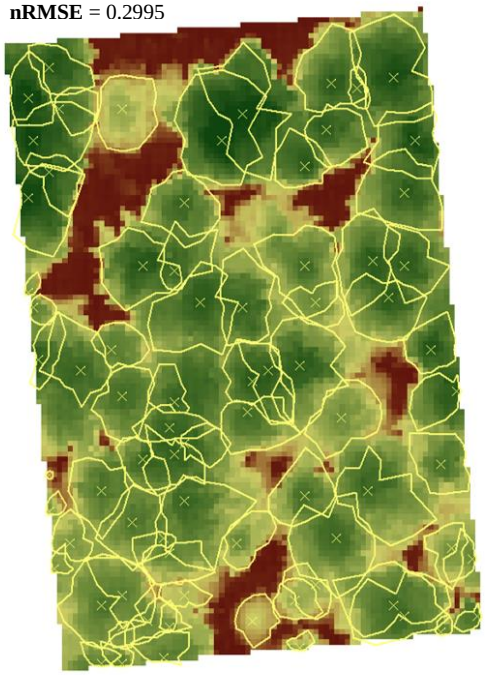
Method: ArcMAP
CHM resolution: 0.5 m
Kernel: K1
nRMSE = 0.2145



Method: ArcMAP
CHM resolution: 0.2 m
Kernel: K3
nRMSE = 0.2196



Method: FUSION
CHM resolution: 0.5 m
Kernel: K1
nRMSE = 0.2995



Method: FUSION
CHM resolution: 0.2 m
Kernel: K1
nRMSE = 0.3070

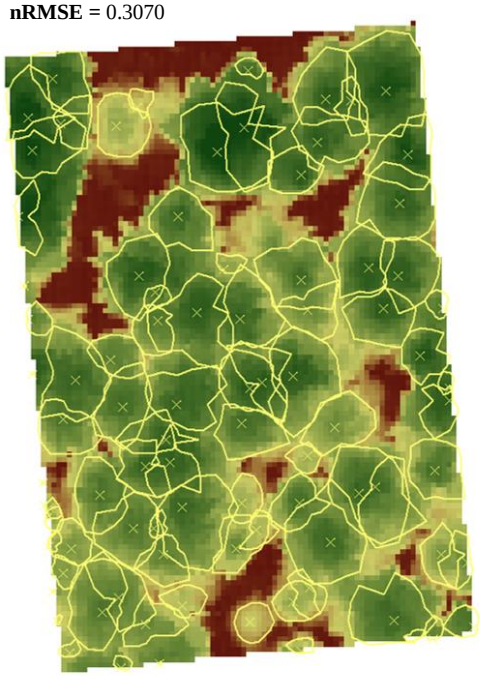


Figure 19. ITD in F2 plot using ArcMAP and FUSION.
The best kernel case has been selected for each CHM resolution.

4.3 Kernel – Resolution

The ANOVA analyzes confirmed that the CHM resolution (ANOVA p-value = 0) and the smoothing distance (kernel) had a significant influence on ITD. Figure 20.A shows the improvement in the estimation as the smoothing of the CHM was stronger. Although there were particular differences between the different types of smoothing, the differences between K1, K3 and K7 were not statistically significant in general terms due to the dispersion of the results. However, raw layers produced significantly worse results than the smoothed CHMs (Figure 20.A). Figure 20.B shows that the CHMs with 0.5 m resolution produced significantly better estimates (nRMSE = 0.83) than the CHMs with 0.2 m resolution (nRMSE = 1.24). Therefore, a general improvement of 41% in the estimation was achieved by expanding the raster resolution.

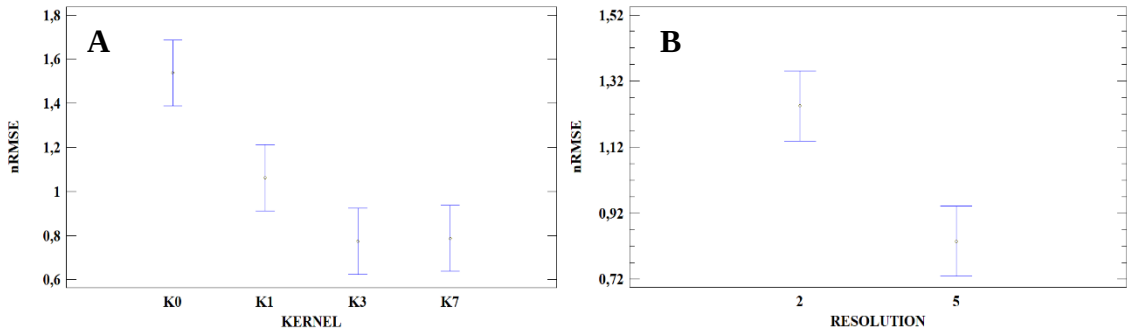


Figure 20. Effect of kernel and CHM resolution on the ITD. Decimal numbers with commas.

No significant differences were found between the sampling plots (ANOVA p-value = 0.15), nor in the plot-smoothing (ANOVA p-value = 0.95) and plot-resolution (ANOVA p-value = 0.85) interactions. However, the interaction between smoothing distance and raster resolution did produce significant differences (Figure 21). The best estimates with the 0.5 m resolution CHMs were obtained in K1 and K3 smoothing layers (nRMSE = 0.71), slightly worse with the K7 smoothing (nRMSE = 0.81) and significantly worse with the K0 (nRMSE = 1.11). On the other hand, the weakest smoothing produced the worst results in the 0.2 m resolution CHMs, being with K7 the most appropriate smoothing case (nRMSE = 0.76).

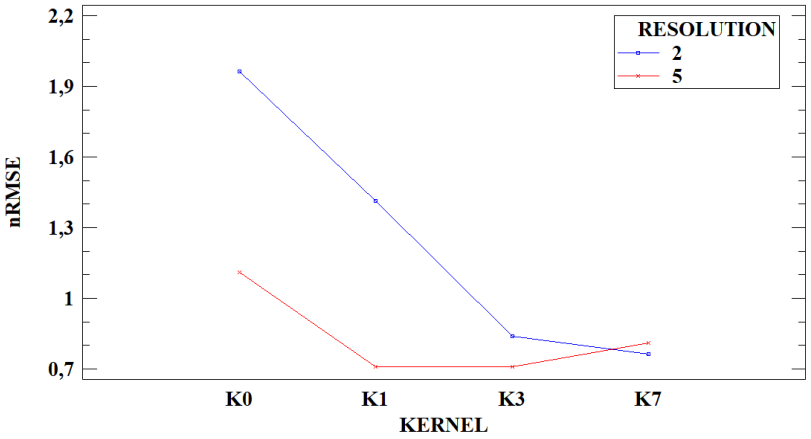


Figure 21. Effect of Kernel-CHM interaction on the ITD. Decimal numbers with commas.

4.4 LIDR algorithm

The algorithm developed by Li, Guo, Jakubowski, & Kelly (2012) produced a direct classification of the trees in the point cloud directly, as can be seen in Figure 22.

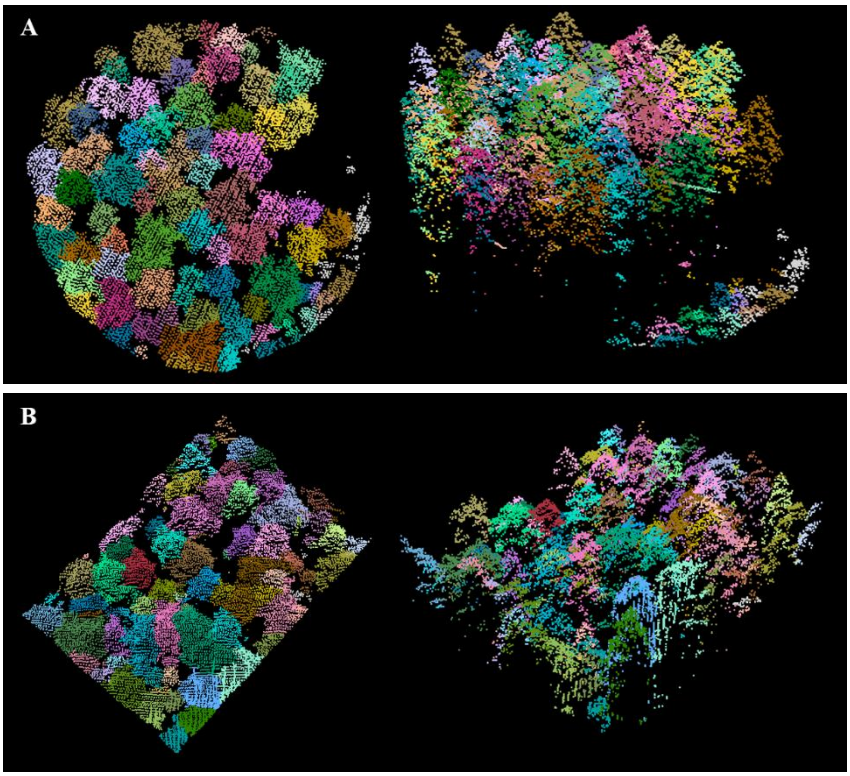


Figure 22. Individual tree classification in the point cloud directly using the Li, Guo, Jakubowski & Kelly (2012) algorithm. (A) CZ2; (B) F1

This vectorial classification was less laborious than the tree segmentation from the CHM, since it worked directly on the point cloud. In addition, the errors in estimation (nRMSE) using this methodology have been very similar to those obtained with the watershed segmentation in ArcMAP or with the FUSION criteria (Table 7).

For the highest resolution CHMs (0.5 m), the best estimations have been obtained with K3 smoothing for F1 (nRMSE = 0.17), CZ1 (nRMSE = 0.59), CZ2 (nRMSE = 0.45) and CZ3 (nRMSE = 0.40) using the watershed segmentation in ArcMAP, which supports the results shown in Figure 21. For the same combination of parameters (0.5 m raster's resolution and ArcMAP methodology), the best estimation was achieved with K1 smoothing in plot F2 (Table 7). FUSION produced heterogeneous results. In contrast, with CHMs of 0.2 m resolution, the pattern is common for both methodologies. The best results were obtained with the highest smoothing (K7) for the F1, CZ1, CZ2 and CZ3 plots, while the trees of F2 plot were detected better with the weakest smoothing (K3). The results are always better in ArcMAP than in FUSION, which supports the graph in Figure 18.

The best results have been obtained with ArcMAP using the 0.5 m resolution CHM with 0.5 m, with the lowest error in F1 (nRME = 0.17) and F2 (nRME = 0.21), while higher in the Czech plots. Similar results were obtained using the LIDR algorithm in RStudio. For the Spanish plots, the differences between ArcMAP and LIDR were less than 10% (0.07 for F1 and 0.09 for F2), while for the Czech plots these differences were less than 5%. Even for the CZ2 plot, the LIDR results was better than with the watershed segmentation (Table 7).

Table 7. nRMSE(%) for high vegetation estimation in every study case (method, CHM resolution and smoothing), comparing with vectorial analysis (LIDR).

		0,5				0,2			
		K0	K1	K3	K7	K0	K1	K3	K7
F1	ARCMAP	91,04	23,71	**17,07**	57,00	292,34	214,04	53,02	**33,84**
	FUSION	240,39	**34,15**	43,29	88,84	316,79	163,07	54,76	**34,75**
	LIDR	24,15							
F2	ARCMAP	77,56	**21,45**	35,80	77,06	281,92	183,58	**21,96**	38,89
	FUSION	210,71	**29,05**	55,29	82,75	426,15	207,11	**30,70**	66,45
	LIDR	30,52							
CZ1	ARCMAP	108,42	80,41	**59,38**	61,01	206,19	189,00	159,39	**105,68**
	FUSION	190,55	143,20	123,88	**104,43**	466,69	317,13	167,97	**144,79**
	LIDR	63,12							
CZ2	ARCMAP	68,77	61,36	**44,95**	52,92	149,38	132,09	82,43	**47,91**
	FUSION	152,57	**61,73**	66,74	95,11	251,26	186,06	76,06	**60,98**
	LIDR	41,22							
CZ3	ARCMAP	95,26	45,13	**40,25**	64,19	162,73	156,05	104,31	**60,48**
	FUSION	149,69	75,77	**67,36**	91,03	356,16	224,33	92,30	**80,51**
	LIDR	41,39							

LIDR delineation was also good visually, both the three-dimensional point cloud classification (Figure 22) and the treetops delimitation in the shapefiles (Figure 23). Although there are some overlapping, the polygon shapes were more natural than those obtained by FUSION (Figure 19).

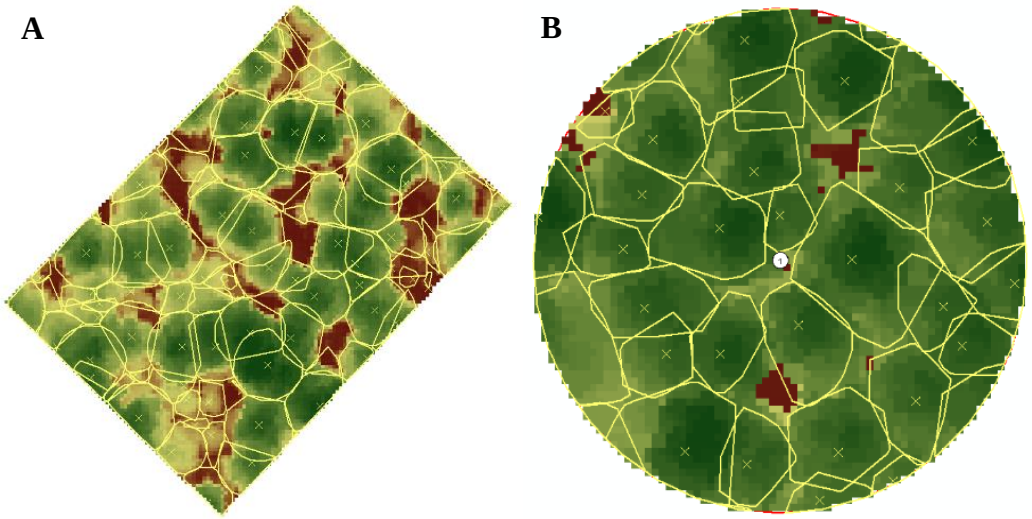


Figure 23. ITD using the Li, Guo, Jakubowski & Kelly (2012) algorithm in RStudio. (A) F1; (B) CZ1.

4.5 Regression analysis

Once evaluated the single tree detection, the relation between different biophysical tree parameters was performed to develop a good regression model. This will allow to know if obtaining the diameter distribution through this methodology is viable.

The Pearson correlation matrix has been calculated for the DBH, height and tree crown area variables (Figure 24). These correlation coefficients range from -1 to +1, and they measure the strength of the linear relationship between the variables. The closer to 1 in absolute value, the greater the correlation between the variables.

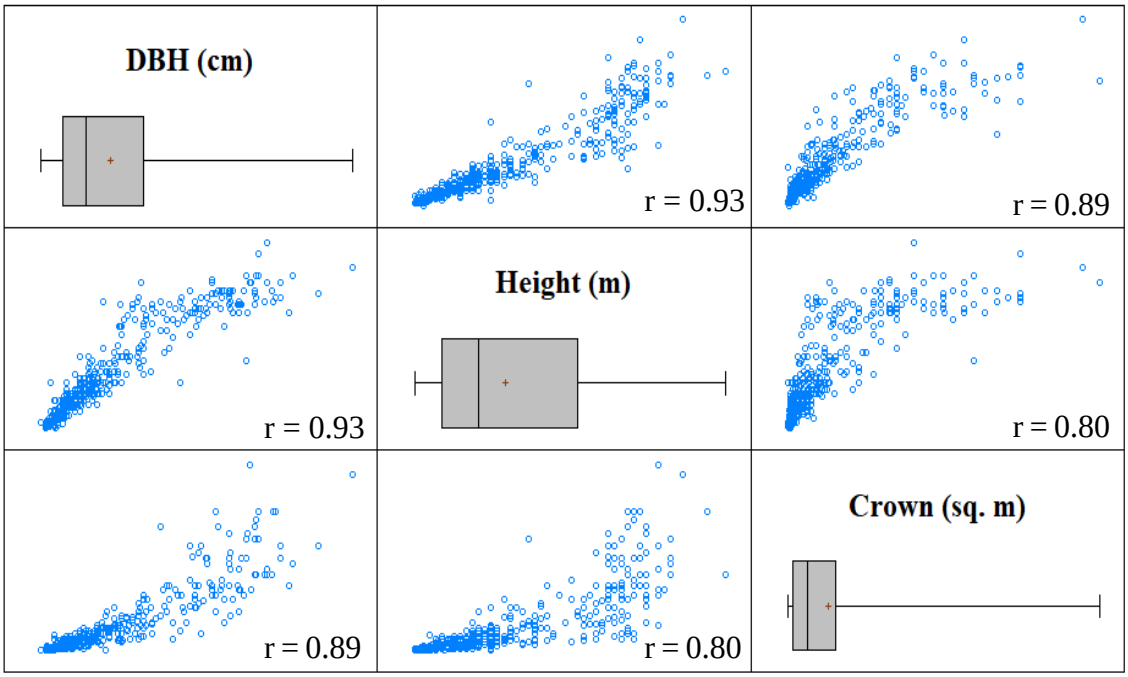


Figure 24. Pearson correlation matrix for DBH, height and tree crown area variables. r = Pearson correlation value.

The most correlated variable with DBH has been the tree height (r = 0.93), although the tree crown size had a very similar correlation (r = 0.89). Both tree height and crown size variables were also highly correlated with each other (r = 0.80), although slightly lower than the correlations with DBH.

Table 8. Statistical analysis for each variable.

	DBH (cm)	**Height (m)**	**Crown (sq. m)**
Count	427	427	427
Mean	16.33	8.68	12.21
Standard deviation	15.28	6.23	16.30
Coefficient of variation	93.56%	71.76%	133.50%
Minimum	0	1.3	0
Maximum	73	26.5	95.03
Range	73	25.2	95.03
Standardized bias	10.11	6.29	17.69
Standardized kurtosis	1.94	-3.30	19.22

Standardized bias and standardized kurtosis are parameters of special interest to determine if a variable has a normal distribution (gaussian distribution). Values of these statistics higher than 2 in absolute value indicate significant deviations from normality, which would tend to invalidate many of the statistical procedures that are commonly applied to these data, such as regression models. In this case, all the variables showed values of standardized bias and the standardized kurtosis out of this range. Therefore they present normality deviations (Table 8).

Different transformations of the variables have been tested in order to achieve a normal distribution. The logarithmic, the squared root and the inverse transformation were the most common ones. Several combinations were tested, being the logarithmic transformation (LOG = LN) of all the variables which provided best result (Figure 26, Table 9).

Despite of the transformations, the standardized bias and kurtosis were still higher than the established values for consider the normal distribution of the variables. Furthermore, the height and crown size variables continued being highly correlated with each other. Therefore they should not act as explanatory variables at the same time since they can produce an overfitting of the regression model.

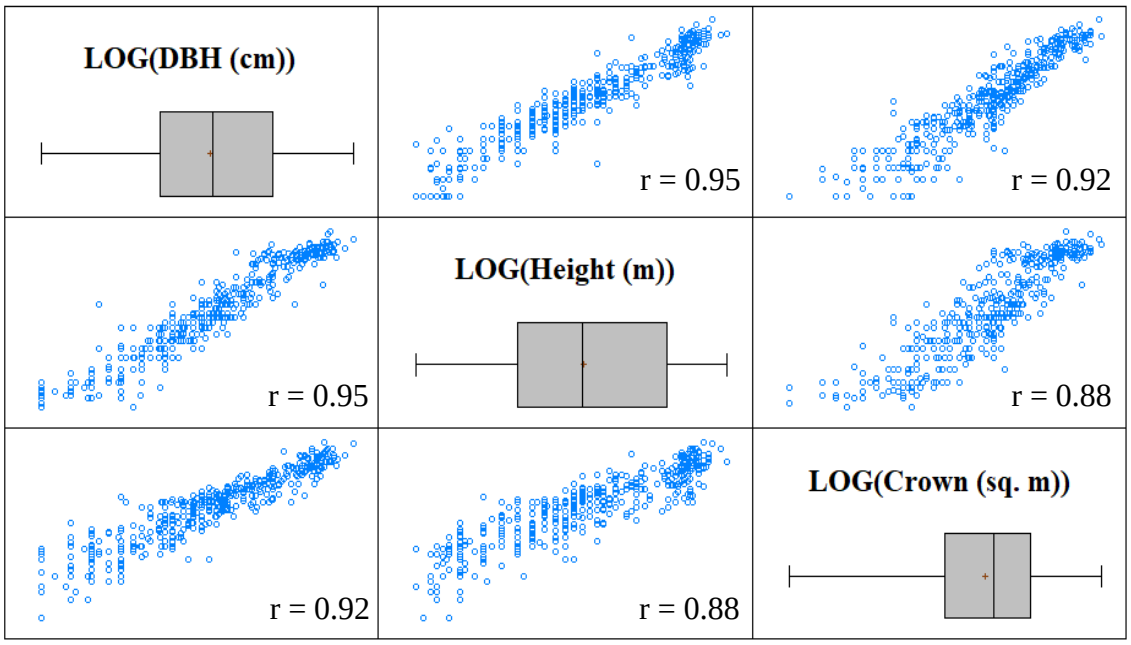


Figure 25. Pearson correlation matrix for logarithmical DBH, height and tree crown area variables. r = Pearson correlation value.

Table 9. Statistical analysis for each logarithmical variable.

	LOG(DBH)	**LOG(Height)**	**LOG(Crown)**
Count	424	424	424
Mean	2.32	1.88	1.56
Standard deviation	1.06	0.79	1.59
Coefficient of variation	45.58%	41.72%	101.46%
Minimum	0	0.26	-3.51
Maximum	4.29	3.30	4.55
Range	4.29	3.01	8.06
Standardized bias	-2.11	-1.19	-4.02
Standardized kurtosis	-3.06	-4.37	-0.75

The tree height was the explanatory variable most correlated with the DBH, so it would be the first option to test a regression model. However, both options have been evaluated.

4.5.1 DBH-Height regression model

The equation for the best regression model is detailed below (Formula 4), with the logarithmic transformation of both the response (DBH) and the explanatory (Height) variables:

$$\ln(DBH\ (cm)) = 1.28283\ \cdot \ln(Height\ (m)) - 0.0994809 \qquad [4]$$

Table 10 shows the results of the regression analysis performed. The correlation between variables has already been studied in Figure 25 (r = 0.95). The model explained 91% of the DBH variability (adjusted R-squared = 0.908). On the other hand, the Durbin-Watson (DW) statistic had a p-value lower than 0.05, which indicates a possible residuals correlation with a 95% confidence level.

Table 10. Results for DBH-Height regression model.

Pearson correlation coef.	0.953
R-squared	0.909
Adjusted R-squared	0.908
Residual standard deviation, RSD	0.320
Mean absolute error	0.242
Durbin-Watson statistic	1.631 (P=0.0001)

Residuals correlation was verified visually with the scatter plot of studentized residuals as a function of row number (Figure 26). A slight sinus pattern could be appreciated.

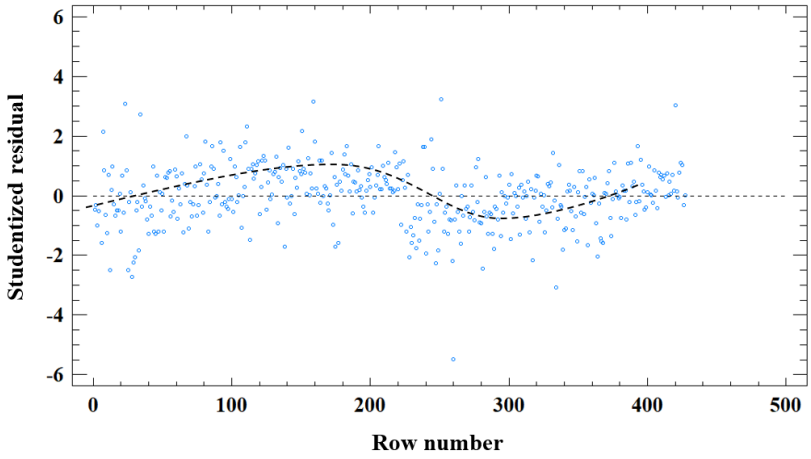


Figure 26. Studentized residuals of DBH-Height model dispersion

The predicted – observed graph (Figure 27) showed a clear linear trend, with some homogeneous dispersion of the points.

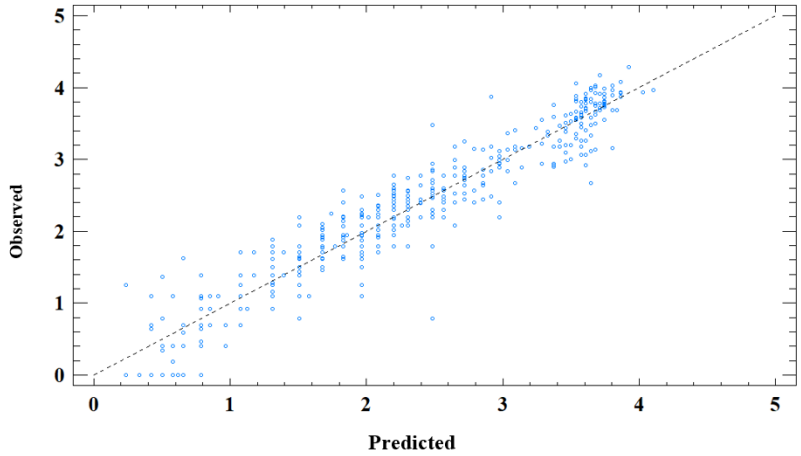


Figure 27. Observed vs Predicted logarithmical DBH values, using DBH-Height model.

4.5.2 DBH-Crown regression model

The equation for the best regression model is detailed below (Formula 5), with the logarithmic transformation of both the response (DBH) and the explanatory (Crown) variables:

$$\ln(DBH\ (cm)) = 0.615728 \cdot \ln\big(Crown\ (sq.m)\big) + 5.42609 \qquad [5]$$

Table 11 shows the results of the regression analysis performed. The correlation between variables has already been studied in Figure 25 (r = 0.92). The model explained 86% of the DBH variability (adjusted R-squared = 0.855). On the other hand, the DW statistic had a p-value lower than 0.05, which indicates a possible residuals correlation with a 95% confidence level.

Table 11. Results for DBH-Crown regression model

Pearson correlation coef.	0.925
R-squared	0.855
Adjusted R-squared	0.855
Residual standard deviation, RSD	0.402
Mean absolute error	0.306
Durbin-Watson statistic	1.551 (P=0.0000)

Residuals correlation was verified visually with the scatter plot of studentized residuals as a function of row number (Figure 28). A slight sinus pattern could be appreciated. The residuals dispersion seems more homogeneous than in the previous model, although there is more dispersion in the initial rows, while decreasing at the ending.

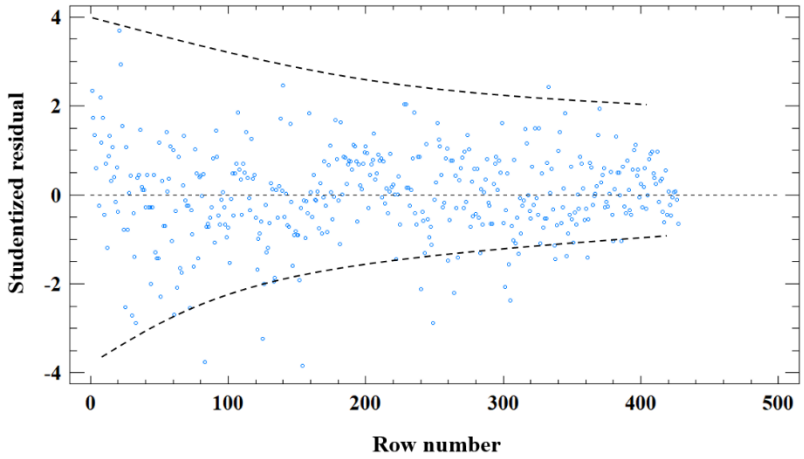


Figure 28. Studentized residuals of DBH-Crown model dispersion.

The predicted – observed graph (Figure 29) showed a clear linear trend. Compared with the results of the previous model (Figure 28), the point dispersion is wider and more irregular in this case.

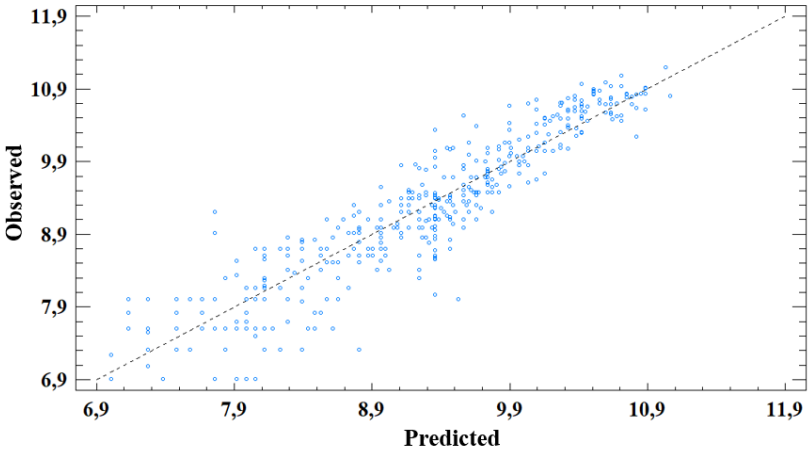


Figure 29. Observed vs Predicted logarithmical DBH values, using DBH-Crown model.

4.5.3 Best regression model

Despite the possible correlation between the residuals, both models could be considered as valid. These show spatial residuals patterns, although they appear slight in both cases. Objectively, the DBH-Height model showed a better fit (adj. R-squared = 0.91) and a lower residual dispersion (RSD = 0.32). Therefore, this model was better.

To make it practical, the logarithmic transformation of its variables has been returned. The logarithmic transformation introduces a bias (underestimation) in the calculations. To minimize this bias, the final result must be multiplied by a correction factor (CF), calculated with the RSD. CF was calculated according to the following expression (Formula 6):

$$CF = e^{\frac{RSD^2}{2}} \quad [6]$$

$$CF = e^{\frac{0.32^2}{2}} = 1.053 \quad [7]$$

The model expression is detailed below. Formula 8 is the general expression of the transformed model and formula 9 is the expression without transformation:

$$\ln(y) = a \cdot \ln(x) - b \quad [8]$$

$$y = x^a \cdot \frac{CF}{e^b} \quad [9]$$

Considering formula 9, the model has the following expression (Formula 10):

$$\boldsymbol{DBH\ (cm) = Height\ (m)^{1.28283} \cdot 0.9533} \quad [10]$$

This model is only valid for the study area (Cercedilla, Spain), where the species was Scots pine (Pinus sylvestris L.), and for the diameter (0 - 73 cm) and height (0 - 26.5 m) studied ranges.

5 DISCUSSION

5.1 Problems in detecting suppressed vegetation

ITD using LiDAR data has been used successfully in numerous scientific studies (for example in Jakubowski, Guo, & Kelly, 2013 and Panagiotidis, Abdollahnejad, Surový, & Chiteculo, 2017). However, the best results usually occur in homogeneous forests, with well-defined and separated trees, such as forest plantations (Reitberger, Heurich, Krzystek, & Stilla, 2007). In this study, ITD has been tested in two different types of stands. First, the Spanish plots were located in a completely irregular forest with a high regeneration capacity (low vegetation). On the other hand, the plots in the Czech Republic were located in a managed forest with fewer suppressed trees.

It has been found that the trees that are covered by the crowns of the dominant trees have not been detected in this study with any of the techniques used, in accordance with previous studies (Næsset, et al., 2004). ITD using CHM is ineffective in detecting small trees that are dominated by the crowns of upper trees. The reason is found in the definition of CHM. Each cell of the raster layer is influenced by the highest LiDAR points, neglecting those with the lowest heights. In addition, the point density for the suppressed trees is lower than in the upper layers since they intercept most of the radiation emitted by the LiDAR sensors.

Other studies have also had worse results in detecting low vegetation (Reitberger, Heurich, Krzystek, & Stilla, 2007; Jakubowski, Guo, & Kelly, 2013). The first study was located in the Bavarian Forest National Park and used full waveform LiDAR data with a density of 25 points·m^{-2}. The percentage of trees detected was much higher in the upper layer (56% - 94%) than in the intermediate and low layer, where only a few trees could be detected, since they were covered by the main vegetation. Similarly, the second study detected 100% of the dominant trees, but the estimation accuracy decreased according to the vertical distribution. In this way, a range from 8% to 19% of the suppressed trees were detected.

The detection of low vegetation has been very poor in this study. However, some improvement could be seen when using point cloud trees classification instead of CHM segmentation. The advantage of working on the point cloud is that all points are

preserved, including the points of small trees (no matter how few). The limitation is the quality of the point classification since the shape of the trees is not as well defined in these layers as in the upper ones due to their lower density of points. Table 12 shows lower nRMSE (%) using the LIDR algorithm in comparison with the CHM segmentation, although the differences are very small. In the Czech plots, the dominated trees were not detected.

Table 12. nRMSE(%) for low vegetation estimation on each study case (method, CHM resolution and smoothing), in comparison with point cloud classification (LIDR).

		0.5				0.2			
		K0	K1	K3	K7	K0	K1	K3	K7
F1	ARCMAP	**89.37**	93.11	100.78	101.09	**82.71**	92.48	99.41	92.02
	FUSION	**83.60**	83.92	89.97	97.96	**82.38**	76.32	90.72	97.06
	LIDR	**82.60**							
F2	ARCMAP	**112.90**	115.74	117.79	115.53	**119.31**	111.08	114.07	109.56
	FUSION	129.81	116.15	**112.05**	115.74	291.80	148.61	111.40	**111.19**
	LIDR	**109.23**							
CZ1	ARCMAP								
	FUSION								
	LIDR								
CZ2	ARCMAP	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	FUSION	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	LIDR	100.00							
CZ3	ARCMAP	115.47	115.47	115.47	115.47	115.47	115.47	115.47	115.47
	FUSION	115.47	115.47	115.47	115.47	115.47	115.47	115.47	115.47
	LIDR	115.47							

5.2 FUSION methodology imprecisions

The smoothed CHM layers were the same for the watershed segmentation in ArcMAP and tree delineation in FUSION. However, the differences between these methodologies were statistically significant (ANOVA p-value = 0.0298; Figure 18). Watershed segmentation in ArcMAP improved FUSION estimation by 20% on average. Possibly, the quality of the CHMs had an influence on these differences. Their extraction using the resolutions of 0.5 m and 0.2 m produced many empty cells, which were corrected with a mean smoothing. The quality of these CHMs was not then the highest. Therefore, the criteria used by FUSION were ineffective. In addition, the subsequent delimitation of the tree canopy has been ineffective in some study cases, especially in the layers with less resolution and under lower LiDAR density conditions (Appendix 2).

"TreeSeg" is very sensitive to the resolution and smoothing used in the production of CHM, working better in open stand conditions (McGaughey, 2018).

5.3 Effect of Kernel – Resolution combination

The Kernel - Resolution interaction has produced significant differences in the estimates. In general, the higher the resolution of the CHM, the smaller the radius (in number of cells) of the smoothing area (kernel). In physical distance, the radius is similar. In a 0.5 m resolution CHM, a radius of 3 cells means a radius of 1.5 meters. On the other hand, in a 0.2 m resolution CHM, a radius of 7 cells means 1.4 meters of radius. However, the results of the combination 0.5-K3 (resolution - kernel) were not the same as the combination 0.2-K7, although in some cases they were similar.

F2 plot produced similar nRMSE in both combinations (Table 7). One possible explanation is that the density of the point cloud in this plot is the highest (22.66 points·m^{-2}; Table 3). The higher the density of points, the lower resolution CHMs are better since there will be a greater probability of finding points belonging to the canopy on each cell. If the point density is not sufficient, many cells in the CHM will be empty and when performing a mean smoothing, the cells are underestimated. This causes an overestimation of the trees due to the subcanopy effect, that is, several trees are detected where there is only one.

Point cloud density also influences the search distance to detect trees. The higher the density of points, the shorter the search distance necessary to detect the trees (Picos, Bastos, Míguez, Alonso, & Armesto, 2020). This can be applied to this case study considering the smoothing kernel distance. F2 plot, with the highest point cloud density, produced better results with lower kernels than the other plots. In general, with a resolution of 0.5 m, the K3 smoothing produced the best results (Table 7), but in F2, the best smoothing was K1. Similarly, K7 smoothing was the best option in the 0.2 m resolution CHMs, while F2 had better results with a smaller smoothing (K3, Table 7).

Another parameter that influenced the detection of trees is the spacing between them (stand density). In open stands is easier to isolate the trees individually (Picos, Bastos, Míguez, Alonso, & Armesto, 2020). In this study, the effect of stand density is appreciated in the F1 plot, where the detection of the trees (high vegetation) was the most

accurate (nRMSE = 0.17). The stand density in this plot was the highest if low vegetation is considered (Figure 9), but the dominant trees (height > 16 m; Table 5) were only 49, which means a high vegetation density of 204.16 trees·ha^{-1}. In addition, the trees were well defined in the CHM, which resulted an accurate ITD. Considering the same criterion, F2 had a stem density of 308.33 trees·ha^{-1}, less than the density of the Czech plots (n > 340 trees·ha, Figure 9). This, coupled with a higher LiDAR point density, has resulted in the second-best estimation (nRMSE = 0.21).

On the other hand, the Czech plots produced the worst estimations despite having the highest and most defined trees. The stem density did not have influence on the estimations, since all of them had the same (approximately 35 dominant trees, Figure 9). However, there was a slight influence of the point cloud density. CZ3 plot had the lowest density (9.89 points·m^{-2}) while CZ1 and CZ2 had higher values (around 12 points·m^{-2}). This resulted in nRMSE = 0.59 in plot CZ3, being lower in the other ones (nRMSE = 0.45 in CZ2 and 0.40 in CZ3). Other possible sources of error are discussed in section 4.1.5).

5.4 Point cloud classification

The advantage of point cloud trees classification is that the processing and extraction of the CHM are eliminated, in addition to its subsequent corrections. Both methodologies are influenced by the point cloud, since the extraction of the CHM is better in higher point cloud densities, as is the classification of the point cloud. The limitation of the watershed segmentation is mainly found in the suppressed trees (low vegetation). Only a few individuals can be detected with this methodology (section 4.1.1), while detecting dominated trees is possible using with the point cloud classification algorithms, especially in higher point cloud densities (Li, Guo, Jakubowski, & Kelly, 2012). In this study, the LIDR algorithm produced better results than the CHM segmentation (Table 12), despite its calibration was very simple (the predefined parameters were used).

On the other hand, the high vegetation segmentation was similar both with LIDR algorithm and raster analysis. As mentioned, the algorithm was not carefully calibrated for each plot, while the watershed segmentation was. Therefore, the potential of the point cloud classification is very high, since they take advantage of the three-dimensional structure of these layers (Li, Guo, Jakubowski, & Kelly, 2012) instead of transforming

them to a two-dimensional layer, whose transformation can produce some interpolation errors (Guo, Kelly, Gong, & Liu, 2007).

5.5 Possible sources of error

The following aspects could have significant influence on the ITD using ALS data:

1. **CHM quality:** The CHMs had low resolutions (0.2 m and 0.5 m) and the point clout density ranged from 9.89 to 22.66 points·m^{-2}. These densities should be enough to do a good ITD from ALS data since it is recommended to use a density greater than 4 - 5 points·m^{-2} (Wulder, Bater, Coops, Hilker, & White, 2008; Reutebuch & Andersen, 2005). However, many CHM cells had empty data, especially in the Czech ones. Point cloud density was similar in Czech plots than in the Spanish F1. Perhaps the dispersion of the points in the first ones was lower. In this situation, performing a mean smoothing does not completely solve the problem, even reducing the precision in those cells where the values were adequate.
2. **Location of the plot coordinates:** The coordinates of the Czech plots were not precisely located. The spatial correction was performed visually, which induces considerable errors. The Spanish parcels correspond to two permanent plots, whose spatial location was carefully worked out. Therefore, nRMSE were lower. In addition, the transformation from the Czech projection (S-JTSK Krovak East North) to the UTM ETRS89 (Zone 33) could cause an additional source of error when extracting the plots from the point cloud.
3. **Temporal gap between the LiDAR and the dasometric inventory:** The dasometric inventory in Spain was performed in 2015, 4 years after the LiDAR flight. During that time, the stand has changed, which also influence the estimates.
4. **Field work:** The Spanish plots were located on steep slopes, while the trees in the Czech Republic have heights up to 42 meters. All of this could affect the accuracy of the measurement. The first because the field work on steep slopes is difficult to perform, while the second because of the imprecision when taking the heigh measurements (lack of vision over the treetop). In general, the height of trees is overestimated in classical inventories, while LiDAR tends to underestimate it

since it is difficult for the LiDAR point to coincide with the exact point of the treetop. In this study, height distributions were compared in 2 meters wide ranges. It is easy to make a mistake in such small ranges. For example, a tree measuring 30.1 m and its LiDAR counterpart 29.9 m would be in different ranges, although an error of 20 cm would be a 0.66% relative error (almost negligible).

5.6 DBH relation with the height and the crown size

The correlations between DBH, height and crown area were high, like those obtained in other studies (Valbuena Rabadan, Santamaria Pena, & Sanz Adan, 2016). In this study, the correlations are so high (r = 0.95 for DBH-Height and r = 0.92 for DBH-Crown size) because only trees from two sample plots (427 trees) have been considered. However, a high dispersion was observed in the largest trees. For these, the variability in DBH increases as the height is higher, being worse their relation (Arias-Rodil, et al., 2018).

Figure 24 shows a greater dispersion in the taller trees, while in the smaller ones the dispersion decreases. This is also seen in other studies (Gonzalez-Benecke, et al., 2014). In this study, the dispersion was even greater than in this study since they analyzed 127 plots. When the trees reach their maximum height, the diameter continues growing, which explains this variability. For the regression analyzes, the logarithmic transformation of all the variables was considered, so that a potential model was obtained, in accordance with other studies (Gonzalez-Benecke, et al., 2014). This means that as the diameter increases, the height tends to stabilize, for the same reason explained above. Both the model and the high Pearson correlations show that there are good relationships between the biophysical parameters of the tree. However, this model needs to be improved and more sample plots are needed to support this strong relationship.

In addition, more study is necessary to extract the suppressed vegetation with ALS data. This is the main limitation to calculate the diameter distribution with this methodology.

5.7 Suggestions and proposals for improvement

ITD is a technique with great potential, since it allows obtaining high detailed information about the stands. ABA methods are also useful, but the level of detail obtained with ITD is higher. In contrast, the economic investment, and the errors (especially for suppressed vegetation) are the main disadvantages of this methodology.

Considering the results of this study and the performed methodology, some recommendations are detailed below to improve the estimations in future research:

1. **CHM extraction:** ITD using CHM segmentation needs accurate raster layers. For this, some details must be considered. First is evaluate the appropriate resolution, which will be greatly influenced by the point cloud density. The higher point density, the lower the resolution that can be used. It will be a more detailed CHM and better results can be achieved. In addition, the appropriate algorithm must be used. The one applied in this study has great deficiencies, since first returns points were interpolated on each raster cell. Perhaps, another criterion such as highest point of returns would be more appropriate.
2. **CHM correction:** Correcting the CHMs with a mean smoothing is not the best solution in cases where there are many empty cells. Another type of filters such as those developed in other studies (Jakubowski, Guo, & Kelly, 2013) are more complex to use but produce better results. In this study, the CHM could have been improved if a filter of close maximum values had been applied before the mean smoothing, so that, the highest value of the neighborhood is assigned to each cell. It would be similar to increasing the resolution. Therefore, CHMs with 0.5 m resolution produced better results than CHMs with 0.2 resolution.
3. **Other classification methods:** Working with the LiDAR point cloud directly has many advantages. The main one is that the three-dimensional structure of the point cloud is used and intermediate processes are reduced (Li, Guo, Jakubowski, & Kelly, 2012). The LIDR results were similar to those obtained in the best CHM segmentation cases. Furthermore, the suppressed vegetation can be estimated better than using CHMs.
4. **Expand the plot radius:** One aspect to improve and that has not been considered in this study is to expand the radius of the sample plot. It may be trees measured

in field, but their crown is not detected with LiDAR because it falls out of the plot or that its maximum height is outside the point cloud clip.

5. **Precise location of the plot coordinates:** It is very important to locate the exact coordinates of the sampling plots. Otherwise, the estimation errors could be very high since the LiDAR information do not correspond to the field measurements.

6 CONCLUSIONS

ITD extracts detailed information on the forest with the individual segmentation of its trees and the subsequent calculation of their biophysical parameters. However, it has some limitations, both technical and economic. Applying this methodology in uneven-aged stands or forests with a high stem density is difficult, especially because the suppressed trees (under the canopy) detection.

CHM segmentation requires high quality raster layers, which is related with a high LiDAR point density. The CHM methods studied were ineffective in detecting suppressed trees, although good results were obtained in detecting the dominant ones, especially with ArcMAP watershed segmentation. The quality of the CHM and the imprecision of the plots location did not allow to obtain good results with the CHM of 0.2 m resolution nor in the plots of the Czech Republic. Point cloud density was a limiting factor, so that the optimal resolution of the CHM layers was 0.5 meters, which required K1 or K3 smoothing. The detection of the trees directly from the LiDAR point cloud offered good results, similar to those obtained with the CHM segmentation, despite not having worked on their calibration. In addition, the detection of the suppressed trees was better in vectoral instead of raster classification.

It is advisable to delve into the ITD using different point cloud classification algorithms since it is a more dynamic technique that takes advantage of the three-dimensional structure of the layer. On the other hand, the segmentation of the CHM is very intuitive, but requires a detailed study for its optimal extraction and subsequent correction. Both methodologies are valid for detecting and extracting individual trees (at least the dominant vegetation which together with the high relationships between their biophysical parameters allows obtaining very detailed information on the stand using remote sensing information. However, diameter distribution needs also the suppressed vegetation information, which is the main challenge for future studies in order to calculate the diameter distribution with ALS data.

7 REFERENCES

Andersen, H. E., McGaughey, R. J., & Reutebuch, S. E. (2005). Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment, 94*(4), 441-449.

Andersen, H. E., Reutebuch, S. E., M. R., d'Oliveira, M. V., & Keller, M. (2014). Monitoring selective logging in western Amazonia with repeat lidar flights. *Remote Sensing of Environment, 151*, 157-165.

Arias-Rodil, M., Diéguez-Aranda, U., Álvarez-González, J. G., Pérez-Cruzado, C., Castedo-Dorado, F., & González-Ferreiro, E. (2018). Modeling diameter distributions in radiata pine plantations in Spain with existing countrywide LiDAR data. *Annals of forest science, 75*(2), 1-12.

Baltsavias, E. P. (1999). Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing, 54*(2-3), 199-214.

Brede, B., Lau, A., Bartholomeus, H., & Kooistra, L. (2017). Comparing RIEGL RiCOPTER UAV LiDAR Derived Canopy Height and DBH with Terrestrial LiDAR. *Sensors, 17*(10), 2371.

Breidenbach, J., Gläser, C., & Schmidt, M. (2008). Estimation of diameter distributions by means of airborne laser scanner data. *Canadian Journal of Forest Research, 38*(6), 1611-1620.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint LIDAR data. *Photogramm. Eng. Remote Sensing, 72*, 923–932.

Daly, S. F., Vuyovich, C., & Finnegan, D. (2011). *Situk River hydrology following closure of Russell Fiord by Hubbard Glacier.* Washington, DC: US Army Corps of Engineers.

Del Río, M., Montes, F., Cañellas, I., & Montero, G. (2003). Del Río, M., Montes, F., Cañellas, I., & Montero, G. *Revisión: Índices de diversidad estructural en masas forestales, 12*(1), 159-176.

FAO and UNEP. (2020). *The State of the World's Forests (SOFO).* Rome.

Führer, E. (2000). Forest functions, ecosystem stability and management. *Forest Ecology and management, 132*(1), 29-38.

Gatziolis, D., & Andersen, H. E. (2008). *A guide to LIDAR data acquisition and processing for the forests of the Pacific Northwest.* Portland: USDA, Forest Service.

Gobakken, T., & Næsset, E. (2004). Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research, 19*(6), 529-542.

Gonzalez-Benecke, C. A., Gezan, S. A., Samuelson, L. J., Cropper, W. P., Leduc, D. J., & Martin, T. A. (2014). Estimating Pinus palustris tree diameter and stem volume from tree height, crown area and stand-level parameters. *Journal of Forestry Research, 25*(1), 43-52.

Goodwin, N. R., Coops, N. C., & Culvenor, D. S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment, 103*(2), 140-152.

Guo, Q., Kelly, M., Gong, P., & Liu, D. (2007). An object-based classification approach in mapping tree mortality using high spatial resolution imagery. *GIScience & Remote Sensing, 44*(1), 24-47.

Hinsley, S. A., Hill, R. A., Bellamy, P. E., & Balzter, H. (2006). The Application of Lidar in Woodland Bird Ecology. *Photogrammetric Engineering & Remote Sensing, 72*(12), 1399-1406.

Hopkinson, C., Chasmer, L., & Hall, R. J. (2008). The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets. *Remote Sensing of Environment, 112*(3), 1168-1180.

Hyyppä, J., & Inkinen, M. (1999). Detecting and estimating attributes for single trees using laser scanner. *Photogramm. J. Finl, 16*, 27-42.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., & Yu X., M. M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing, 29*(5), 1339-1366.

Jaakkola, A., Hyyppä, J., Kukko, A., Yu, X., Kaartinen, H., Lehtomäki, M., & Lin, Y. (2010). A low-cost multi-sensoral mobile mapping system and its feasibility for

tree measurements. *ISPRS journal of Photogrammetry and Remote Sensing, 65*(6), 514-522.

Jakubowski, M. K., Guo, Q., & Kelly, M. (2013). Tradeoffs between lidar pulse density and forest measurement accuracy. *Remote Sensing of Environment, 130*, 245-253.

Lee, H., Slatton, K., Roth, B., & Cropper, W. (2010). Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. *International Journal of Remote Sensing, 31*(1), 117–139.

Lefsky, M. A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar remote sensing for ecosystem studies. *BioScience, 52*(1), 19-30.

LHP. (2011). *Lesní hospodářský plán ŠLP Kostelec nad Černými lesy. Kostelec nad Černými lesy: Lesprojekt Stará Boleslav, s.r.o.*

Li, W., Guo, Q., Jakubowski, M. K., & Kelly, M. (2012). A new method for segmenting individual trees from the lidar point cloud. *Photogrammetric Engineering & Remote Sensing, 78*(1), 75-84.

Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J., & Vehmas, M. (2004). Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote sensing of environment, 90*(3), 319-330.

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A., & Kangas, J. (2006). Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Canadian Journal of Forest Research, 36*(2), 426-436.

Maltamo, M., Suvanto, A., & Packalén, P. (2007). Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. *Forest Ecology and Management, 247*(1-3), 26-34.

McGaughey, R. J. (2018). *FUSION/LDV: Software for LIDAR data analysis and visualization.* Seattle: USDA, Forest Service.

McRoberts, R. E., & Tomppo, E. O. (2007). Remote sensing support for national forest inventories. *Remote sensing of environment, 110*(4), 412-419.

Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L., & Hendrix, C. J. (2000). Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing, 66*(11), 1367-1372.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote sensing of environment*, *80*(1), 88-99.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., . . . Söderman, U. (2004). Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, *19*(6), 482-499.

Neill, S. P., & Hashemi, M. R. (2018). *Fundamentals of ocean renewable energy: generating electricity from the sea*. Academic Press.

Panagiotidis, D., Abdollahnejad, A., Surový, P., & Chiteculo, V. (2017). Determining tree height and crown diameter from high-resolution UAV imagery. *International journal of remote sensing*, *38*(8-10), 2392-2410.

Pascual, C., García-Abril, A., García-Montero, L. G., & Martín-Fernández, S. C. (2008). Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous Pinus sylvestris stands. *Forest Ecology and Management*, *255*(11), 3677-3685.

Persson, A., Holmgren, J., & Soderman, U. (2002). Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, *68*(9), 925-932.

Picos, J., Bastos, G., Míguez, D., Alonso, L., & Armesto, J. (2020). Individual tree detection in a eucalyptus plantation using unmanned aerial vehicle (UAV)-LiDAR. *Remote Sensing*, *12*(5), 885.

Reitberger, J., Heurich, M., Krzystek, P., & Stilla, U. (2007). Single tree detection in forest areas with high-density LiDAR data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*(6), 139-144.

Reutebuch, S. E., & Andersen, H. E. (2005). Light Detection and Ranging (LIDAR): An Emerging Tool for Multiple Resource Inventory. *Journal of Forestry*, *103*(6), 286-292.

Rodrigues de Souza, C., de Azevedo, C., Brum Rossi, L., dos Santos, J., & Higuchi, N. (2014). Projection of diametric distribution and carbon stock of a managed forest in Manaus/AM. *Floresta*, *44*(3), 525-534.

Smith, D. M., Larson, B. C., Kelty, M. J., & Ashton, P. M. (1997). *The practice of silviculture: applied forest ecology* (9 ed.). John Wiley and Sons, Inc.

Spies, T. A., & Franklin, J. F. (1991). The structure of natural young, mature, and old-growth Douglas-fir forests in Oregon and Washington. *Wildlife and vegetation of unmanaged Douglas-fir forests*, 91-109.

Tordesillas Torres, A. (2014). Estimación y evolución temporal de variables de variables forestales con tecnología LiDAR en el valle de la Fuenfría (Cercedilla, Madrid). (Bachelor Thesis, E.T.S.I Montes, Forestal y del Medio Natural - UPM), Madrid, Spain.

Treitz, P., Lim, K., Woods, M., Pitt, D., Nesbitt, D., & Etheridge, D. (2012). LiDAR sampling density for forest resource inventories in Ontario, Canada. *Remote Sensing, 4*(4), 830-848.

Valbuena Rabadan, M. A., Santamaria Pena, J., & Sanz Adan, F. (2016). Estimation of diameter and height of individual trees for Pinus sylvestris L. based on the individualising of crowns using airborne LiDAR and the National Forest Inventory data. *Forest Systems, 25*(1).

Vauhkonen, J., Korpela, I., Maltamo, M., & Tokola, T. (2010). Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment, 114*(6), 1263-1276.

Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Computer Architecture Letters, 13*(6), 583-598.

Wallace, L., Musk, R., & Lucieer, A. (2014). An assessment of the repeatability of automatic forest inventory metrics derived from UAV-borne laser scanning data. *IEEE Transactions on Geoscience and Remote Sensing, 52*(11), 7160-7169.

Wasser, L. A. (2020, October 07). *The Basics of LiDAR - Light Detection and Ranging - Remote Sensing*. Retrieved March 31, 2021, from NEON Science: https://www.neonscience.org/lidar-basics

Wei, L., Yang, B., Jiang, J., Cao, G., & Wu, M. (2017). Vegetation filtering algorithm for UAV-borne lidar point clouds: a case study in the middle-lower Yangtze River riparian zone. *International journal of remote sensing, 38*(8-10), 2991-3002.

World Commission on Environment and Development. (1987). *Our common future.* New York: Oxford.

Wulder, M., Bater, C., Coops, N., Hilker, T., & White, J. (2008). The role of LiDAR in sustainable forest management. *Forestry Chronicle, 84*(6), 807-826.

WWF. (2020). *WWF*. Retrieved from Importance of forest: https://wwf.panda.org/discover/our_focus/forests_practice/importance_forests/

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 Sensing, 8*(6), 501.

Zhao, K., & Popescu, S. (2007). Hierarchical watershed segmentation of canopy height model for multi-scale forest inventory. *Proceedings of the ISPRS working group'*, 436-442.

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment, 113*(1), 182-196.

Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C., & Gerard, P. D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote sensing of Environment, 87*(2-3), 171-182.