SEBASTIAN LAMPRECHT

vom Fachbereich VI

Raum- und Umweltwissenschaften

der Universität Trier

zur Verleihung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

Detection of Individual Tree Stems Using ALS and its Potential for Forest Research

Betreuender:

Univ.-Prof. Dr. Thomas Udelhoven

Berichterstattende:

Univ.-Prof. Dr. Thomas Udelhoven

Univ.-Prof. Dr. Markus Casper

Dr. Gilles Rock

Datum der wissenschaftlichen Aussprache:

11. Februar 2022

Trier, 2022

"What I see in Nature is a magnificent structure that we can comprehend only very imperfectly, and that must fill a thinking person with a feeling of 'humility.' This is a genuinely religious feeling that has nothing to do with mysticism."

Albert Einstein in Einstein, 1979

Acknowledgements

I would first like to thank my supervisor Univ.-Prof. Dr. Thomas Udelhoven, head of the Environmental Remote Sensing & Geoinformatics department at Trier University, for giving me the opportunity to write this PhD thesis and to work in the department in a great team. I am very thankful for his confidence in my work, sharing his scientific experience, providing constructive feedback and giving me the freedom to develop my thesis in the last years. Many thanks to Univ.-Prof. Dr. Markus Casper and Dr. Gilles Rock for agreeing to review this thesis. I would also like to thank the Open Access Publication Fund of Trier University which funded my open access articles almost completely.

Thanks to the whole Environmental Remote Sensing & Geoinformatics family of Trier University for their continuous support as well as their scientific and technical assistance. In this context I would like to highlight Dr. Rebecca Retzlaff, Dr. Gilles Rock and Dr. Henning Buddenbaum for sharing their scientific experience. A special thanks to Dr. Johannes Stoffels, for always being around, sharing his experience and being a sounding board. By always knowing where to find relevant data and by sharing his detailed knowledge on this data, he spared me weeks of research.

I would like to express my thanks to Dr. Andreas Hill, for having a couple of exciting discussions about forest inventories and tree trunk inclination, which I appreciate a lot. One of these discussions led to a joint paper, which would not have been doable without his experience. Writing such an interdisciplinary and inter-university paper with him was one of the most inspiring experiences during my Ph.D studies.

I would like to thank Martin Quintus for sharing his scientific expertise, for proof-reading this thesis and for giving me continuous mental support as well as Thomas Reusch for proof-reading this thesis, for providing welcome distractions and his ability to bring me back down to earth.

Thanks to all my family, who always showed interest in my work and motivated me to hold on in bad phases. Thank you Dorothee, without your love, continuous support and critical reviews I certainly would have given up my studies a long time ago. Last, but not least, I would like to thank Sophia, her zest for life brought me the final motivation I needed to finish this thesis.

Abstract

Forest inventories provide significant monitoring information on forest health, biodiversity, resilience against disturbance, as well as its biomass and timber harvesting potential. In recent years the importance of airborne laser scanning (ALS) and terrestrial laser scanning (TLS) to support modern forest inventories has increased. While the integration of ALS in plot-level surveys has reached operational level, the assessment on the individual tree level is still challenging. Individual tree crown detection and delineation (ITCD) using ALS can be seen as a mature discipline, whereas the detection of individual tree stems is a rarely addressed task. Although some detection methods have been developed, the informative value of the stem attributes—especially the inclination characteristics—is hardly known. In addition, a lack of tools for the processing and fusion of various forest-related data sources can be identified. In particular, ground data needs to be geo-referenced, trees of differing data sets need to be linked and terrestrial laser scans need to be aligned.

The given thesis addresses the identified research gaps in four peer-reviewed papers. A general focus is set on the suitability of ALS data for the detection and analysis of tree stems. The potential of the detected stems for enhancing forest surveys as well as providing novel parameters for tree growth behavior is investigated. To address these tasks and to exploit the full potential of ALS- and TLS-driven forest inventories, a Python package for the processing and analysis of point clouds with a focus on forest applications is developed.

Hence, two algorithms for detecting stems using ALS point clouds have been developed. Compared to ITCD, the detections are very reliable and the tree positions are accurate. In particular, the ALS detected stems have shown to be suited to study prevailing trunk inclination angles and orientations. In addition, the effect of topography, wind and soil properties have been investigated. A highly species-specific down-slope inclination of the stems and a leeward orientation of conifers have been observed. To geo-reference forest inventory plots, a novel post-processing strategy has been developed and successfully tested. Trained using synthetic forest stands, the algorithm matches ground truth trees with aerially detected trees, while it provides a matching probability for each link.

In conclusion, ALS detected stems are particularly suited to support the geo-referencing of forest inventory plots and to improve ITCD. They have great potential to provide further knowledge on tree growth behavior of various tree species under real-world conditions and might also be suited to identify landslides or to assess the risk of windthrow.

Contents

Appendix B

List of Figures

List of Tables

List of Abbreviations

2D	**Two d**imensional
3D	**Three d**imensional
ABA	**A**rea-**b**ased **a**pproach
ALS	**A**irborne **l**aser**s**canning
BWI	**B**undes**w**ald**i**nventur
CHM	**C**anopy **h**eight **m**odel
CEOS	**C**ommittee on **e**arth **o**bservation **s**atellites
CW	**C**ontinuous **w**ave
DAP	**D**igital **a**erial **p**hotogrammetry
DBH	**D**iameter at **b**reast **h**eight
DSM	**D**igital **s**urface **m**odel
DTM	**D**igital **t**errain **m**odel
GNSS	**G**lobal **n**avigation **s**atellite **s**ystem
HSR	**H**igh **s**patial **r**esolution
ICP	**I**terative **c**losest **p**oint
IMU	**I**nertial **m**easurement **u**nit
IFOV	**I**nstantaneous **f**ield **o**f **v**iew
ITA	**I**ndividual **t**ree **a**ppoach
ITCD	**I**ndividual **t**ree **c**rown detection and **d**elineation
LADAR	**La**ser **d**etection **a**nd **r**anging
LiDAR	**Li**ght **d**etection **a**nd **r**anging
MLS	**M**obile **l**aser **s**canning
NFI	**N**ational **f**orest **i**nventory
NICP	**N**ormal **i**terative **c**losest **p**oint
PLS	**P**ersonal **l**aser **s**canning
POS	**P**osition and **o**rientation **s**ystem
RADAR	**Ra**dio **d**etection **a**nd **r**anging
SAR	**S**ynthetic **a**perture **r**adar
SCIE	**S**cience **c**itation **i**ndex **e**xpanded
TLS	**T**errestrial **l**aser **s**canning
ULS	**U**nmanned **l**aser **s**canning
VHSR	**V**ery **h**igh **s**patial **r**esolution

List of Symbols

Category	Symbol	Unit	Description
General	λ	m	wavelength
	f	s^{-1}	frequency
	c	$m\,s^{-1}$	speed of light
	S/N		signal to noise ratio
	ρ		reflectivity of an object
LiDAR	β	mrad	beam divergence
	t_L	s	traveling time or delay of a light pulse
	Δt_L	s	time resolution of a pulsed LiDAR
	Φ	m	phase difference between transmitted and received signal
	$\Delta\Phi$	m	resolution of phase measurement
	R	m	distance between ranging unit and sampled object
	σ_R	m	ranging accuracy
ALS	a	m	spacing within an ALS scan line
	b	m	ALS scan line spacing
	D_t	m	diameter of the transmitting aperture
	D_r	m	diameter of the receiving aperture
	$IFOV$	mrad	instantaneous field of view

Chapter I

Introduction

"Forest resource information is gathered for planning and managing of various ecosystem services at various user-levels [...] and at various scales [...]" (Liang et al., 2016, p. 63). Forest inventories provide significant monitoring information on forest health, biodiversity, resilience against disturbance as well as its biomass and timber volume (Liang et al., 2016; White et al., 2016). But "[w]ith continuously increasing international agreements and commitments, the need for information has [...] grown drastically, and reporting requests have become more frequent and the content of the reports wider" (Tomppo et al., 2010, p. v). Thus, today's "information needs exceed the scope and design of many existing forest inventories" (White et al., 2016, p. 620) and push a search for new technologies and information layers. In this framework Gatziolis and Andersen (2008, p. iii) state that "[l]ight detection and ranging (L[i]DAR) is an emerging remote-sensing technology with promising potential to assist in mapping, monitoring, and assessment of forest resources". Although this statement refers to airborne laser scanning (ALS), it is not less true for terrestrial laser scanning (TLS).

Since the mid-2000s TLS has been used to monitor vegetation structures (Khorram et al., 2016b) and the Committee on Earth Observation Satellites (CEOS) highlights TLS as a key technology to assess above ground biomass (Duncanson et al., 2021; Disney et al., 2019). Vauhkonen et al. (2013, p. 1) note that ALS "[...] has emerged as one of the most promising remote sensing technologies to provide data for research and operational applications in a wide range of disciplines related to management of forest ecosystems". Although individual tree crown detection and delineation (ITCD) is increasingly used in modern forest inventories, Amiri et al. (2017) identify a lack of scholarly attention towards tree stem detection. However, TLS and ALS do not record forest biomass or timber volume directly, the attributes of interest can be predicted using structural features (e.g., tree height, stem diameter, or crown dimensions) and allometric functions (Khorram et al., 2016b). In consequence, remote sensing data needs to be fused with field or inventory data (ideally on the individual tree level) to achieve spatially complete, accurate, and cost-efficient estimations of forest parameters.

This thesis focuses on the suitability of ALS data for the detection and analysis of tree stems. The potential of the detected stems for providing novel parameters—in particular trunk inclination angles and directions—for research of tree growth behavior, as well as for enhancing forest surveys is investigated. To address these tasks and to exploit the full potential of ALS- and TLS-driven forest inventories, attention is also given to the development of tools for the processing and analysis of ALS- and TLS-derived point clouds with a focus on forest applications. Before the research questions are formulated in Section I.4 in detail, a general introduction will be provided. First, in Section I.1, the importance and design of forest inventories are described. Section I.2 gives background information on the principles of LiDAR, while Section I.3 provides insights on how TLS and ALS are used in modern forest surveys.

I.1 Forest Inventories

Forest inventories are designed to provide up-to-date information on the forest condition (namely biomass, biodiversity, health, mortality, and disturbances) and trends at a global and national scale, as well as the stem volume to plan timber harvesting at a regional scale (Liang et al., 2016; Tomppo et al., 2010; McRoberts and Tomppo, 2007). Precise and continuous forest inventory information is essential to manage ecosystem services, and is the basis for political and economical decision making on various scales and user-levels (Liang et al., 2016; White et al., 2016).

Sampling Design. As typical forest inventories rely on field samples (Liang et al., 2016; White et al., 2016; Tomppo et al., 2010; McRoberts and Tomppo, 2007), the amount and quality of the samples determine the accuracy of the information (Liang et al., 2016). In addition, the quantity and spatial distribution of the samples decide on how the information can be aggregated or interpolated (Liang et al., 2016). Tomppo et al. (2010) illustrate that the sample plots of national forest inventories (NFIs) are either distributed randomly (Italy), in a regular grid (Belgium, Japan), or in a multi-phase design, e.g., in a grid with clusters of plots (Austria, Germany) or in a more complex design (USA, Russia). The placement of the sample plots has implications on the (geo-)statistical analyses (Hill, Massey, and Mandallaz, 2021; Hill, Mandallaz, and Langshausen, 2018). Although the different inventories have varying sampling units (McRoberts and Tomppo, 2007), a typical sample plot covers a small (less than 0.1 ha) forested area and has a circular or square shape, while information is recorded for each tree individually (Liang et al., 2016; Tomppo et al., 2010). Some inventories, like the German

NFI (BMELV, 2011), use the angle count sampling technique (Bitterlich, 1984) to preferably select dominant trees, since it is statistically not required that all trees within a plot are sampled.

Tree attributes. In typical NFIs attributes like the tree height, stem length, stem diameter at breast height (DBH), diameter in upper stem region(s), crown dimension, tree species, tree age and canopy layer are gathered (Liang et al., 2016; BMELV, 2011; Tomppo et al., 2010; Hyyppä et al., 2008b). But also operational information like the accessibility for timber harvest or information on the forest biotope, habitat and biodiversity might be collected, while information on forest health can comprise the presence of fungi, pests, diseases, bark damages or deadwood (BMELV, 2011). To collect tree-by-tree samples, traditional tools are measuring tape, calipers and clinometers (Liang et al., 2016; BMELV, 2011). Liang et al. (2016) show that today terrestrial, personal (PLS) and mobile laser scanning (MLS) as well as image derived point clouds extend the pool of available data sources. They also clarify that, when selecting techniques to derive individual tree information, their suitability to measure relevant tree attributes, the accuracy of the measurements, the cost of equipment as well as the time consumption for data acquisition and post-processing need to be considered.

Trend to remote sensing data. Tomppo et al. (2010) see increasing demands for forest reports introduced by international agreements, while White et al. (2016) identify a general trend to more complex information needs for forest management, both exceeding the original purpose and capabilities of many traditional forest inventories. To comprehend the dimension of the problem, it needs to be noted that common European NFIs collect from 100 up to 400 attributes, which leads to a need to increase the efficiency and to reduce the costs of the sampling procedure (McRoberts and Tomppo, 2007). Thus, modern forest inventories—or forest surveys in general—are increasingly supplemented with remote-sensing data to ease forest mapping and to enable the spatially explicit estimation of forest attributes (Hill, Mandallaz, and Langshausen, 2018; Liang et al., 2016; White et al., 2016; McRoberts and Tomppo, 2007). Today, field data is often combined with the complementary remote-sensing technologies satellite-based synthetic aperture radar (SAR), ALS and TLS (Khorram et al., 2016b). If field measured tree positions are provided, a tree-by-tree linking to remote sensing data can be achieved, which opens the opportunity for a tree-based correlation of the attributes, rather than just estimating stand parameters (Breidenbach and Astrup, 2013). The given thesis will mostly focus on ALS data, but will also provide information on how to integrate TLS into forest inventories.

I.2   Principles of LiDAR

After Einstein (1916) postulated stimulated emission, the first working laser (*light a*mplified by *s*timulated *e*mission of *r*adiation) was described by Maiman (1960). A laser can produce a focused ray of optical light which is typically (depending on its type) highly energetic and particularly coherent both in space and time (Wehr and Lohr, 1999; Young, 1986) allowing for active and highly accurate remote ranging.

I.2.1   Laser Scanning

Wehr and Lohr (1999) point out that compared to microwave radar (*ra*dio *d*etection *a*nd *r*anging), lasers can emit very energetic pulses in high frequencies, which makes them particularly suited for precise range measurements. In addition, mobile devices can be easily equipped with lasers, since, as they note, small apertures are already sufficient to bundle their relatively short wavelengths. Since, as they state, the laser's ranging principle only relies on the high optical power and the collimation of the light beam, also other light sources (e.g., xenon or flash lamps) may be used. Thus, the two acronyms LADAR (*LA*ser *D*etection *A*nd *R*anging) and LiDAR (*Li*ght *D*etection *A*nd *R*anging) are common for optical ranging. In the given thesis the acronym LiDAR is preferred, while the acronym ALS is used for airborne LiDAR in particular.

A laser scanner consists of a ranging unit, an opto-mechanical scanner, as well as a control and processing unit (Wehr and Lohr, 1999). The range measurements are either based on the pulsed or the continuous wave (CW) ranging principle (Baltsavias, 1999; Wehr and Lohr, 1999). For both principles, a laser signal is emitted, whose response (backscattered by the scanned object) needs to be detected. The transmitting and the receiving apertures typically share the same beam path (Wehr and Lohr, 1999). The fundamental relations of both ranging principles are contrasted in Table I.1.

Pulsed LiDAR. As Baltsavias (1999) and Wehr and Lohr (1999) show, the pulsed LiDAR determines the distance to an object by measuring the time delay t_L between transmitting the laser beam and receiving the backscattered signal, which is also called echo. The range R corresponds to half of the time delay times the speed of light c. The ranging resolution is proportional to the minimum time resolution Δt_L at which the traveling time t_L can be measured. The ranging accuracy σ_R depends on the signal-to-noise ratio S/N and is proportional to the traveling time of the pulse.

Continuous wave LiDAR. A continuous wave LiDAR can be designed by applying (e.g., sinusoidal) phase modulations of wavelength λ to the signal (Liang et al., 2016; Wehr and Lohr, 1999). The phase difference Φ between the transmitted and the received signal is used to derive the range R (Liang et al., 2016; Nelson, 2013; Wehr and Lohr, 1999). But Wehr and Lohr (1999) show that, since phase differences greater than a full wavelength λ cannot be recognized, several frequencies (tones) with wavelengths ranging from λ_{short} to λ_{long} might have to be applied to reduce the limitation introduced by the maximum (unambiguous) range R_{max}. The range resolution σ_R is proportional to the minimum detectable phase difference $\Delta\Phi$. Keeping in mind $\lambda = \frac{c}{f}$, a higher range accuracy σ_R can be achieved with an increasing modulation frequency f (Wehr and Lohr, 1999).

TABLE I.1: Characteristics of pulsed and CW lasers (based on Baltsavias, 1999 and Wehr and Lohr, 1999).

Property	Pulsed Laser	CW Laser (sinusoidal modulated)
Traveling time or time delay	$t_L = 2\frac{R}{c}$	$t_L = \frac{\Phi}{2\pi} \cdot \frac{\lambda}{c}$
Range	$R = \frac{1}{2}\,c \cdot t_L$	$R = \frac{\lambda}{4\pi} \cdot \Phi$
Range resolution	$\Delta R = \frac{1}{2}\,c \cdot \Delta t_L$	$\Delta R = \frac{\lambda_{short}}{4\pi} \cdot \Delta\Phi$
Range accuracy	$\sigma_R \sim \frac{1}{4}\,c \cdot t_L \cdot \frac{1}{\sqrt{S/N}}$	$\sigma_R \sim \frac{\lambda_{short}}{4\pi} \cdot \frac{1}{\sqrt{S/N}}$
Maximum (unambiguous) range	$R_{max} = c \cdot \frac{1}{2}\,t_{L_{max}}$	$R_{max} = \frac{\lambda_{long}}{2}$

Wehr and Lohr (1999) point out that the signal-to-noise ratio of pulsed systems is typically superior to CW systems, because they can be easily equipped with a laser of high peak power. They conclude that the ranging of pulsed systems is typically more accurate than that of CW systems, if accuracies in the centimeter domain shall be achieved. But they also state that CW systems can achieve higher accuracies in the sub-centimeter domain, if higher modulation frequencies are applied.

Interpretation of the backscattered signal. Roncat et al. (2013) point out that due to the short wavelengths of lasers typically used for terrestrial and airborne laser scanning, most surfaces can be treated as diffuse reflectors. As they illustrate, the effective backscatter cross-section σ of a target depends on the target reflectivity ρ, its area A and the scattering angle Ω as defined by Equation I.1 :

$$\sigma = \rho A \frac{4\pi}{\Omega} \tag{I.1}$$

To distinguish multiple targets and gain knowledge on their spectral characteristics, the received power P_r needs to be recorded continuously. As defined by Equation I.2, the received laser power P_r at time t is the superposition of k echos along the laser beam at the ranges R_i with $i \in \{1, 2, ..., k\}$

(Roncat et al., 2013). Next to the target ranges R_i and the effective backscatter cross-sections of the targets σ_i, the received power P_r depends on the transmitted laser power P_t, the laser's beam divergence β, the aperture diameter of the receiver D_r, as well as the atmospheric and system transmission factors η_a and η_s respectively (Roncat et al., 2013; Jelalian, 1992).

$$P_r(t) = P_t \, \eta_s \, \eta_a \, \frac{D_r^2}{4\pi\beta^2} \sum_{i=1}^{k} \frac{\sigma_i}{R_i^4} \left(t - 2\frac{R_i}{c} \right) \tag{I.2}$$

Based on these dependencies, individual or cluster echos can be identified and the energy amplitude, which is also known as intensity, can be estimated (Roncat et al., 2013; Gatziolis and Andersen, 2008; Wagner et al., 2006). Thus, CW and pulsed systems provide the opportunity to distinguish multiple echoes per modulation or per pulse respectively. These echoes are either discretized onboard by the hardware, or—if the full waveform is recorded—during post-processing (Gatziolis and Andersen, 2008).

I.2.2 Terrestrial LiDAR

TLS is capable of creating highly accurate three-dimensional (3D) point clouds of various objects and structures (Khorram et al., 2016b; Liang et al., 2016). To sample its surrounding area, the instrument typically rotates horizontally while the light beam is deflected vertically using a fast rotating mirror (Liang et al., 2016). Both the vertical and the horizontal angle steps are typically less than $0.01°$ resulting in high resolution point clouds (footprint diameter from $0.01\,\text{m}$ to $0.05\,\text{m}$; point spacing from $0.05\,\text{m}$ to $0.25\,\text{m}$), while the ranging precision is at the millimeter level (Beland et al., 2019; Liang et al., 2016). It needs to be noted that the scanning design leads to a distance depending decrease of the sampling density, while at the same time the footprint size increases (Beland et al., 2019; White et al., 2016). Sampling frequencies of up to one million points per second reduce the scanning time of modern TLS systems to a couple of minutes (Liang et al., 2016).

I.2.3 Airborne LiDAR

Airborne LiDAR can produce dense and geo-referenced 3D point clouds of large areas (Khorram et al., 2016a). The scanning and ranging unit of an ALS platform is complemented by an inertial measurement unit (IMU), a differential global navigation satellite system (GNSS) and a computer interface (Lohani and Ghosh, 2017; Khorram et al., 2016b; White et al., 2016; Gatziolis and Andersen, 2008). GNSS and IMU together form a position and orientation system (POS) (Wehr and Lohr, 1999),

which continuously records the geo-spatial pose of the ranging unit during measurement (Gatziolis and Andersen, 2008). Based on the trajectory measured by the POS and the scanning angles provided by the scanning unit, the desired 3D coordinate of each echo can be reconstructed using the timestamps of the measurement as reference (Lohani and Ghosh, 2017; Vauhkonen et al., 2013). The horizontal accuracy is typically in a range of $\pm 0.8\,\text{m}$ and the vertical accuracy in a range of $\pm 0.3\,\text{m}$, at 95 % confidence each (Quadros and Keysers, 2015).

Multi-wavelength ALS. In recent years, efforts have been made to design multi-wavelength systems (Lohani and Ghosh, 2017; Fernandez-Diaz et al., 2016; Kumar and Dasgupta, 2006; Tan and Narayanan, 2004). Although it would be optimal to have the full spectral information of each echo, this is practically not achievable (Ullrich, 2013). Thus, today for each channel a separate geo-referenced point cloud is created (Fernandez-Diaz et al., 2016). This is already an achievement compared to mapping the same target area with different scanners mounted simultaneously on the same aircraft, or even with different scanners at different flights (Pfennigbauer and Ullrich, 2011). Alternatively spectral information can also be introduced by complementing the airborne LiDAR platform by a multi-spectral camera (Lohani and Ghosh, 2017; Holmgren, Persson, and Söderman, 2008), if the spectral information is required for the surface only.

Sampling Patterns. The sampling characteristics of ALS are driven by the scanning pattern. Today two major patterns are common, the (stabilized) seesaw and parallel line patterns (Ullrich, 2013), but also elliptical patterns can be achieved (Gatziolis and Andersen, 2008; Wehr and Lohr, 1999). Assuming a flat terrain, an oscillating mirror creates (depending on the electronics) a triangular or sinusoidal seesaw pattern (Ullrich, 2013). Although manufacturers attempt to preserve point spacing at the ground, higher echo densities can be found near the turning points of the mirror (Ullrich, 2013; Gatziolis and Andersen, 2008). In contrast, continuously and smoothly rotating polygon mirrors can produce straight parallel scan lines, at the price of a certain amount of laser pulses not leaving the instrument (Ullrich, 2013). The Palmer scanner produces an almost elliptical pattern, which is generated by a mirror tilted to its rotation axis (Ullrich, 2013; Wehr and Lohr, 1999). Due to the movement of the plane, the Palmer scanner samples most objects in forward as well as in backward view (Wehr and Lohr, 1999).

Flight Planning. ALS campaigns cover the area of interest in parallel (overlapping) strips (Lohani and Ghosh, 2017). Gatziolis and Andersen (2008) highlight that when planning a campaign—next to the scanning pattern—the beam divergence, scanning frequency, scanning angle, and discretization settings need to be considered. They make clear that the footprint spacing, which is typically the parameter of interest, needs to be seen as the result of the scanning frequency and the aircraft's altitude and velocity. The beam divergence β of airborne lasers typically ranges from 0.3 mrad to 2 mrad (Wehr and Lohr, 1999). Gatziolis and Andersen (2008) note that small footprint lasers—with footprint diameters ranging from 0.1 m to 3 m (Beland et al., 2019)—are common, since their small beam divergence leads to high energy densities and consequently to high signal-to-noise ratios. They also state that low frequency scans have a higher ability to penetrate through loose structures—like vegetation—than high frequency scans, since higher energy densities increase the number of distinguishable echos. It needs to be noted that the complex interaction of the laser beam with dense vegetation can lead to multiple path effects (Hyyppä et al., 2008b), particularly for angles of incidence above 12° (Gatziolis and Andersen, 2008), or even to saturation effects for angles above 23° (Liang et al., 2016). Next to the these factors, the laser's wavelength needs to be considered when planning an ALS campaign. For a wide range of applications 1550 nm lasers are preferred, since these are eye-save and (compared to shorter wavelengths) are less influenced by background radiation as well as stable towards unfavorable atmospheric conditions (Quadros and Keysers, 2015; Pfennigbauer and Ullrich, 2011). However, for topographic and forest applications also lasers with wavelengths of 1040 nm to 1065 nm are common (Gatziolis and Andersen, 2008).

ALS Data Characteristics. Although the data characteristics of discrete return ALS depend on the sensor and software used (Vauhkonen et al., 2013), each echo is typically labeled with its return intensity, return number, number of returns, scan angle and GNNS time (Lohani and Ghosh, 2017; Gatziolis and Andersen, 2008). For full-waveform ALS, a temporal profile (waveform) of each pulse is recorded, providing the opportunity to adjust the echo discretization during post-processing (Roncat et al., 2013). After discrete point clouds are available, additional attributes (like the land cover class) can be added. Based on such data, aggregated information, like pulse density (pulses per square meter) or return density (echos per square meter) can be derived (Gatziolis and Andersen, 2008). In recent years, the *LAS* data format (ASPRS, 2013) has prevailed in storing LiDAR point clouds (Lohani and Ghosh, 2017).

I.3   TLS and ALS for Forest Surveys

Although Beland et al. (2019) identify five major LiDAR platform types, the given thesis will focus on TLS and in particular on ALS only. The reason for this decision is that even though the effect of forest structure on point cloud features is still not fully understood (White et al., 2016), ALS and TLS have evolved into operational tools for enhanced forest inventories.

I.3.1   TLS for Forest Surveys

The potential of TLS as an alternative tool to study tree characteristics has already been recognized by the forestry community in the early 1990s (Disney et al., 2019), while since the mid-2000s it is used operationally for vegetation monitoring (Khorram et al., 2016b). TLS is commonly used to provide structural parameters, like DBH and taper, which are also gathered by forest inventories (White et al., 2016). In forested areas multi-return TLS might be preferred over single-return systems, since the highly structured vegetation typically leads to a backscattering of each pulse by several targets (Liang et al., 2016).

TLS data of forest plots can be acquired in single scan, multi-scan or multiple single-scan mode (Liang et al., 2016). By just taking one scan in the center of the plot, the single scan approach spares time, unfortunately at the cost of hampered attribute estimation caused by occlusion effects (Liang et al., 2016; White et al., 2016). By scanning the trees inside and outside of the plot using the multi-scan or multiple single-scan approach, a larger coverage can be achieved, occlusion effects can be reduced and attributes can be estimated more accurately, although at the cost of an increased scanning time (Beland et al., 2019; Wilkes et al., 2017; Liang et al., 2016; White et al., 2016). Moreover, this approach requires the co-registration of either the scans—e.g., by identifying (artificial) reference targets—or of the trees (Wilkes et al., 2017; Liang et al., 2016). Although the multi-scan approach is expected to achieve the highest accuracy of the extracted attributes, the multiple single-scan approach might be preferred, since it is less prone to wind effects and provides a higher coverage in less acquisition and post-processing time (Liang et al., 2016).

I.3.1.1   Tree Reconstruction

In TLS-based surveys, tree attributes are estimated—either at individual tree level or at plot level—using 3D models of the trees (Liang et al., 2016). The main objective of this approach is to

provide a map of the trees and their attributes at the plot or stand level (Liang et al., 2016; White et al., 2016; Raumonen et al., 2015). Thus, it requires the identification and extraction of individual trees which are then reconstructed in 3D (Liang et al., 2016). Such 3D models are typically built step-wise (e.g., from root to twigs), but also graph-based segmentation methods are common (Li, Bu, and Wang, 2017; Liang et al., 2016). The trees are typically represented by a hierarchical graph (tree) of geometric primitives like circles (Pueschel et al., 2013), cylinders (Hackenberg et al., 2015; Calders et al., 2014; Raumonen et al., 2013), or spheres and truncated cones (Lamprecht, Stoffels, and Udelhoven, 2015). Based on the 3D tree models, key attributes, like tree height, DBH, stem volume, and crown size can be extracted (Beland et al., 2019; Disney et al., 2019; Liang et al., 2016). By aggregating the volume of the geometrical primitives (trunks, branches and leaves), the biomass can also be estimated (Disney et al., 2019; Khorram et al., 2016b). But even more sophisticated structural information, like the stem curve or taper, stem quality and stem inclination, as well as the branching characteristics can be estimated (Liang et al., 2016; White et al., 2016; Lamprecht, Stoffels, and Udelhoven, 2015; Razak et al., 2013). Attributes like the stem density, basal area or biomass can be transferred to the plot level by aggregating the individual tree information (Beland et al., 2019; Liang et al., 2016).

I.3.1.2 Integration into Forest Inventories

The latest achievements in the availability of software make an efficient use of TLS for forest applications increasingly feasible. But, due to technological, methodological, and operational challenges, a full integration of TLS is still difficult to achieve (White et al., 2016; Newnham et al., 2015). One aspect hampering TLS integration is its limited suitability to predict the mandatory attributes tree species, due to its limited spectral information, and tree height, due to occlusion of the tree tops (Liang et al., 2016; White et al., 2016). Thus, TLS rather needs to be seen as a complementary technology for forest surveys, since its strength is providing structural information at tree and plot level, which cannot be measured by traditional methods (White et al., 2016; Newnham et al., 2015). In particular, White et al. (2016) highlight the strength of TLS in calibrating allometric functions for stem volume or biomass estimation. They also see TLS as an intermediate technology which can be used to calibrate and validate models that use different remote sensing data (e.g., ALS) to estimate forest attributes at larger scales. A further critical aspect of its integration is that TLS—or similar technologies, like mobile (MLS) and personal laser scanning (PLS)—is not able to provide unbiased attributes, since typically not all trees can be detected (Liang et al., 2016; White et al., 2016). It needs to be noted that (next to the algorithm used) the tree detection rates depend on forest structure and scanning

setup, while in general multi-scan or multiple single-scan setups achieve better results than single scan setups (Wilkes et al., 2017; Liang et al., 2016; Pueschel et al., 2013). Despite the given conceptional challenges, White et al. (2016) are optimistic that TLS will eventually achieve the operational stage to estimate a range of tree attributes to support forest inventories.

I.3.2 ALS for Forest Surveys

LiDAR provides cost-efficient and accurate forest information not accessible using field surveys or passive optical remote sensing at large scales (Beland et al., 2019). In this context, ALS "[...] fills the gap between high-resolution below the canopy data provided by TLS [...] and coarser sattelite-based SAR measurements [...]" (Khorram et al., 2016b, p. 148). With their guide for LiDAR data acquisition and processing, Gatziolis and Andersen already emphasize in 2008 that ALS "[...] is an emerging remote sensing technology with promising potential to assisting mapping, monitoring, and assessment of forest resources" (Gatziolis and Andersen, 2008, p. 1). It has evolved into a standard of deriving digital surface models, and is frequently used to estimate biophysical attributes or to detect individual trees even at large scales (Fernandez-Diaz et al., 2016). The laser's characteristics allow its beam to penetrate the canopy layer and generate a vertical profile of the forest's structure, while an almost perfect spatial registration of the 3D data is achieved (Gatziolis and Andersen, 2008). Using regression models, large-footprint LiDAR can be used in combination with ground measurements to estimate forest parameters at stand level (McRoberts and Tomppo, 2007). In today's forest applications, typically small-footprint, discrete multi-return ALS—capable of recording four up to nine separate echoes—is used (Beland et al., 2019; Nelson, 2013; Vauhkonen et al., 2013). The small footprint ensures that the scanning data can be linked to forest stands, forest plots or even individual trees (Vauhkonen et al., 2013), while the relative spatial accuracy can be considered more relevant than the absolute accuracy (Gatziolis and Andersen, 2008).

I.3.2.1 Area-based vs. Individual Tree Approach

When using ALS in forest applications, the area-based approach (ABA) and the individual tree approach (ITA) are distinguished (Aubry-Kientz et al., 2019; Vauhkonen et al., 2013).

The ABA subdivides the ALS data into individual plots of typically at least $200 \, \text{m}^2$ (Vauhkonen et al., 2013). Within a plot, discrete metrics (e.g., mean or percentiles) of the vertical distribution

of the ALS echoes, penetration rates or other attributes (e.g., intensity) are calculated, which are expected to have statistical relationships to properties of the vegetation (White et al., 2016; Vauhkonen et al., 2013; Gatziolis and Andersen, 2008; Hyyppä et al., 2008b). Using these predictor variables, empirical models are trained to estimate attributes, like tree height, basal area, volume, biomass, carbon, or canopy closure (White et al., 2016; Vauhkonen et al., 2013; Gatziolis and Andersen, 2008).

With the ITA relevant attributes—e.g., DBH, tree height, basal area, crown dimension, or stem volume—can be estimated at the tree level (Zhen, Quackenbush, and Zhang, 2016; Koch et al., 2013). While in the past passive remote sensing technologies were common, today active sensors—particularly airborne LiDAR—are used for ITCD (Zhen, Quackenbush, and Zhang, 2016). The high density and accurate 3D information provided by LiDAR, which today is not achievable with other technologies (Koch et al., 2013), might be one reason for this trend. Vauhkonen et al. (2013, p. 8) describe the ITA as "a sequence of [...] tree detection, feature extraction, and estimation of tree attributes." They show that typically the ALS echo characteristics of the tree crowns are utilized to identify and delineate the trees. They also summarize that properties like tree position, tree height, tree crown dimensions, or intensity distributions can be extracted for each tree. After linking some ALS detected tree crowns with field measured trees, such metrics, but also species-specific and allometric properties, can be used to calibrate statistical models to finally predict the attributes of any detected tree (Breidenbach and Astrup, 2013; Vauhkonen et al., 2013). "Finally, the tree attributes can be aggregated to predict forest properties at the plot- or stand-level (White et al., 2016; Breidenbach and Astrup, 2013).

I.3.2.2   Methods for Tree Crown Delineation

In the past, various ITCD algorithms have been developed (Cabello-Leblic, 2015). Although common ITCD methods have their origin in the domain of high-resolution aerial imagery (White et al., 2016; Koch et al., 2013; Vauhkonen et al., 2013; Hyyppä et al., 2008a), modern algorithms take more advantage of the intrinsic characteristics of the active data (Zhen, Quackenbush, and Zhang, 2016). While traditionally ITCD methods have relied on rasterized data (e.g., of intensity or canopy height) (Koch et al., 2013), today methods exploiting the 3D information of voxelized data (Windrim and Bryson, 2020; Wang, Weinacker, and Koch, 2008) or point clouds (Williams et al., 2020; Aubry-Kientz et al., 2019; Ferraz et al., 2016; Dalponte and Coomes, 2016; Strîmbu and Strîmbu, 2015; Vega et al., 2014; Lindberg et al., 2013) become increasingly common (Zhen, Quackenbush, and Zhang,

2016). In general, ITCD is performed by a segmentation of the forest stand based on similar statistical features of the data (Koch et al., 2013). This is typically achieved by first identifying the positions of the trees, followed by a subsequent delineation of the tree crowns, or a combination of both at once (Cabello-Leblic, 2015; Koch et al., 2013). To detect and delineate the trees, (rasterized) canopy height models (CHMs), the point clouds, or both are used (Aubry-Kientz et al., 2019; Koch et al., 2013). Although the prime source for object identification using ALS data is geometry (Lohani and Ghosh, 2017), the segmentation can benefit from introducing a priori information, like tree species, tree top positions, or expected crown size and stand density (Williams et al., 2020; Breidenbach and Astrup, 2013; Koch et al., 2013; Heinzel and Koch, 2012).

Next to the ITCD method used, the forest structure (stand density and spatial pattern) is assumed to be an important factor influencing the performance of the tree identification (Kaartinen et al., 2012; Vauhkonen et al., 2011). Depending on the stand characteristics (stand density, spatial distribution, canopy structure) ITCD tends to identify dominant trees or groups of trees only, rather than identifying individual trees (Wang et al., 2016; White et al., 2016; Vauhkonen et al., 2013). ITCD methods also struggle to identify and delineate understory trees, due to transmission losses and occlusion caused by the upper canopy (Aubry-Kientz et al., 2019; Koch et al., 2013). In multi-layered forests this issue can partially be addressed by 3D segmentation rather than identifying the trees mostly based on the CHM (Aubry-Kientz et al., 2019). Wang et al. (2016) could show a significant improvement in the detection of suppressed trees by increasing the point density, although four returns per square meter have traditionally been perceived as sufficient for ITCD (Gatziolis and Andersen, 2008).

I.3.2.3 Integration into Forest Inventories

Since ALS has evolved into an operational, rapid and cost-efficient tool to obtain the vertical forest structure over large areas and entire states (Beland et al., 2019; Gatziolis and Andersen, 2008; McRoberts and Tomppo, 2007), efforts have been made for its integration into forest inventories (White et al., 2016). White et al. (2016) note that although the quality of ALS derived forest inventory information relies on the sensor characteristics and the parameters of the survey, the data is typically acquired for multiple stakeholders—who might have competing needs—in order to reduce costs. They also note that ITA usually requires a higher pulse density than ABA. This might be an issue when selecting the approach.

The ABA can be seen as the standard operational method, due to its precise prediction of several relevant structural attributes, its straightforward implementation, and its stability for varying pulse densities (White et al., 2016; Vauhkonen et al., 2013). At plot-level ALS has proven to meet or exceed the accuracy demands of forest inventories for most structural attributes (White et al., 2016).

However, in modern forest inventories, the ITA also plays a crucial role for providing timely, accurate, and complete forest information (Zhen, Quackenbush, and Zhang, 2016). Still, the ITA is not as operationally advanced as the ABA due to pending challenges in identifying trees and extracting the allometry (White et al., 2016). In particular, issues in identifying suppressed trees reduce the accuracy of attribute prediction in natural or semi-natural forests significantly (Vauhkonen et al., 2013). To overcome this issue and to reduce biases at the area-level, semi-ITCD and hierarchical methods have been developed (Koch et al., 2013; Kaartinen et al., 2012; Hyyppä et al., 2008b).

Since intensity metrics are not independent from tree size, foliage density, as well as leaf size or orientation (White et al., 2016; Korpela et al., 2010), discrete single wavelength ALS has shown not to be well suited for tree species classification. This can be seen as a major drawback of ALS, because the tree species is a crucial variable in forest inventories. For this reason, complex methods for the classification of tree species using discrete ALS data in combination with optical data and auxiliary information have been developed (White et al., 2016; Hyyppä et al., 2008b). Since impressive classification results can be achieved with full-waveform ALS (Heinzel and Koch, 2012) and detailed information on the understory can be gained (Koenig and Höfle, 2016; Quadros and Keysers, 2015), its use for forest inventories is more straightforward. Another way to improve classification capabilities is to exploit multi-wavelength ALS (Beland et al., 2019; Liang et al., 2016; Torabzadeh, Morsdorf, and Schaepman, 2014).

I.4   Research Questions

I.4.1   Gap Analysis

In the given thesis, four research gaps have been identified, which are explained in detail in the subsequent sections:

1. There is an imbalance between research on the identification of individual tree crowns and the identification of tree stems using ALS data. In particular, researchers and practitioners would benefit from need new algorithms for the detection of tree stems and further knowledge on the detection characteristics.

2. An opportunity to study the inclination characteristics of ALS-detected tree trunks is identified. It remains unclear whether the ALS-derived stem inclination might help to complete knowledge on tree growth behavior (e.g., in response to wind or incidence of light).

3. Several forest surveys require the fusion of various data sources. Thus, there is a need for methods simplifying the geo-referencing of inventory plots, the linking of individual trees, and the co-registration of TLS data.

4. A collection of tools for processing, analysis and fusion of point cloud (e.g., ALS or TLS) and image data with a focus on forest surveys might help to address fundamental research tasks.

I.4.1.1   Tree Stem Detection Using ALS

While today ALS-driven individual tree crown detection can be seen as a mature discipline, Amiri et al. (2017) note that the detection of individual tree stems has still gained a limited amount of scholarly attention. Even relevant reviews on the use of ALS for forestry applications (cf. Maltamo, Næsset, and Vauhkonen, 2014) or on trends in ITCD (cf. Zhen, Quackenbush, and Zhang, 2016) do not identify the detection of individual tree stems as a separate task. The non optimal sampling geometry of ALS for assessing tree trunks (White et al., 2016; Lovell et al., 2003), resulting in just a few of the pulses being reflected by the stems (Popescu and Hauglin, 2013), might be one reason for the little attention. To assess whether individual tree trunk detection using ALS data is reasonable, more detailed considerations on this issue can be found in Appendix A.

Although most trunks are not represented in the ALS data, a few methods for the detection of individual tree stems have been developed in the past. Two types of stem detection approaches are

common, which are heuristic methods and methods based on machine-learning. Typical heuristic methods (Chen et al., 2018; Shendryk et al., 2016; Polewski et al., 2015b; Lu et al., 2014; Reitberger et al., 2009) are designed for ALS data sets with point densities ranging from about 8 up to 40 points/m^2. Methods based on machine-learning (Windrim and Bryson, 2020; Amiri et al., 2017) require high density ALS or unmanned laser scanning (ULS) data with point densities ranging from about 300 up to 700 points/m^2. In principle, all methods use the vertical orientation of points associated with a stem to fit a 3D vector describing the position and orientation of the trunk.

Due to differing base data, the algorithms are hard to compare. Still, in general all approaches show low commission errors, while the detection rates can easily be outperformed by a detection of the tree crowns. The low detection rates are a consequence of the non-optimal representation of the trunks in the ALS data, due to occlusion by the canopy and the high demands on the sampling density (see Appendix A). Knowledge on the positional accuracy of ALS detected stem locations is incomplete, but they can be expected to be far more accurate than positions derived from the tree crowns. In this context, Valbuena (2013) notes that significant differences between tree top positions and tree locations at the ground might occur, if the stems are inclined.

In total, present literature addressing individual stem detection typically focuses on the development and testing of the algorithms rather than on the wide range of applications of the new methods. Thus, only small study sites are addressed, not allowing to generalize the findings or to exploit the novel features, like the stem inclination.

I.4.1.2   Causes of Tree Trunk Inclination

Although the major factors that influence the growth of trees—gravitropism, phototropism and continuous wind—are generally known, to date, the interactions between these factors and the species-specific differences are not well understood.

Gravitropism causes trees to grow contrarily to the gravitative field, while the roots follow the gravitative field (Chen, Rosen, and Masson, 1999). Phototropism causes higher plants to reorient their growth towards or against a light source (Christie and Murphy, 2013; Iino, 2001). Thus trees of the mid-latitudes tend to incline towards the equator to maximize light absorption (Johns et al., 2017). Next to this, phototropism in combination with gravitropism can cause trees to incline down-slope (Matsuzaki, Masumori, and Tange, 2006; Ishii and Higashi, 1997). Wind can cause trees to realign their foliage in the prevailing leeward direction to reduce wind drag and consequently reduce the

risk of stem or root damage (Gardiner, Berry, and Moulia, 2016; Telewski and Jaffe, 1986). Wind can also cause an adaptive growth of root systems expressed in an increased leeward allocation of structural root mass (Nicoll and Ray, 1996). In addition to these systematic effects, sudden events, like storms, landslides or snow break, can cause trees to realign their orientation and growth (Gardiner, Berry, and Moulia, 2016; Razak et al., 2013). All of the mentioned effects can be expected to have effects on the inclination of tree stems.

In summary, apart from small scale experiments, there is a lack of methods to study the inclination of tree stems on larger scales. But knowledge on the inclination and orientation of trees would be valuable information to assess the risk of windthrow and to identify or study landslides. More basically, to date, the species-specific differences with respect to gravitropism, phototropism and wind drag are not well understood. For addressing these knowledge gaps, transferable and cost-efficient methods to inspect the inclination characteristics of trees or stems on large scales are needed.

I.4.1.3 Data Fusion in Forest Applications

The "trend towards multisource inventories" (White et al., 2016, p. 620) requires the fusion of data of various (complementary) characteristics, spatial scales, and spatial coverage. The given thesis distinguishes between local scale (or plot level) and regional scale (or areal) data sources and applications, while the fusion of global scale (in particular space-borne) data is not further considered.

At the plot level, typically the position and attributes of the trees are either recorded manually, or by TLS or similar technologies. In consequence, common tasks for data fusion at the plot level are the alignment of multiple TLS data sets or the linking of manually measured trees with remotely detected trees. At the regional scale, the combination of LiDAR data with spectral imagery can enhance the accuracy and quality of the predicted attributes (Beland et al., 2019; Khorram et al., 2016a; Valbuena, 2013). But, since mismatches can result in a loss of all synergies among the sensors, an accurate geo-referencing of the data sets is required (Valbuena, 2013). Augmenting ground-sample data (e.g., TLS extracted trees) with areal data (e.g., ALS extracted trees) to transfer inventory information to large areas (White et al., 2016) can be seen as a major cross-scale task. Combining terrestrial data (like manual measurements or TLS) with regional data (like ALS or DAP) promises to gain insights on the interaction between the upper- and lower-canopy structure (White et al., 2016). Recent approaches take advantage of multiple LiDAR systems at different scales to estimate forest biomass

(Beland et al., 2019), or to achieve accurate height estimates by combining TLS derived tree positions with their corresponding ALS derived tree heights (Liang et al., 2016). In particular, when statistical models using areal or space-borne data need to be calibrated or validated, a fusion with ground measurements is required, regardless of the approach being used, ABA or ITA (Disney et al., 2019; Khorram et al., 2016b; White et al., 2016; Breidenbach and Astrup, 2013).

To address the previously described typical demands of data fusion in forest applications, the given thesis identifies three major tasks, which are presented in the subsequent paragraphs: 1) Linking of individual trees from various data sources; 2) Geo-referencing of forest survey data; 3) Marker-less alignment of point clouds (especially TLS scans).

Linking Individual Trees. Whenever a one-to-one relationship between trees of differing data sets is required (e.g., for calibrating statistical models, or to aggregate information) corresponding trees need to be identified. Valbuena (2013) identifies a need to use a systematic method to find such matching tree pairs. Although the usable information to comply this task strongly depends on the spatial scale and the data characteristics, the general principles are similar. To link corresponding trees, typically the spatial pattern of tree locations as well as tree attributes (e.g., tree height and DBH) are exploited (Aubry-Kientz et al., 2019; Eysn et al., 2015; Hauglin et al., 2014; Monnet and Mermin, 2014; Kaartinen et al., 2012; Dorigo et al., 2010; Olofsson, Lindberg, and Holmgren, 2008). In the traditional ITA it is presumed that one-to-one tree linking is possible (Breidenbach and Astrup, 2013). But, since the data sets are usually characterized by commission and omission errors as well as inaccurate tree locations and measurements, erroneous tree links might occur (Lindberg et al., 2012). Thus, linking of trees can be eased if reliable and accurate tree positions—which are achievable by stem detection—are provided. In summary, the problem for linking individual trees appears to be of similar nature regardless of the data source and spatial scale. Hence, generic and easily transferable methods are desirable, which utilize the spatial pattern of the trees and their attributes.

Geo-Referencing Inventory Data. The combination of spatial data from independent sources, like terrestrial inventory data with remote sensing data, requires a co-registration of the data sets to a common coordinate system (McRoberts, Andersen, and Næsset, 2013; Lindberg et al., 2012). In particular, an offset of the plot location can significantly reduce the accuracy of ALS derived biophysical parameters (Monnet and Mermin, 2014; White et al., 2013; Frazer et al., 2011; Gobakken and Næsset, 2009). Although nowadays GNSS is used for geo-referencing forest inventory plots, its accuracy is not optimal due to a poor signal under the canopy (McRoberts, Andersen, and Næsset,

2013; White et al., 2013; Valbuena et al., 2010; Andersen et al., 2009; Wing et al., 2008; Hoppus and Lister, 2007). A common approach to align terrestrial data with aerial data is to find matches between terrestrially and aerially detected trees (Valbuena, 2013). Although several (semi-)automated methods for co-registering forest inventory plots with ALS data have been developed (Monnet and Mermin, 2014; Dorigo et al., 2010; Olofsson, Lindberg, and Holmgren, 2008), the reliability of the results is hard to assess and the algorithms cannot easily be transferred to other sites due to the typically heuristic character of the algorithms. Thus, self-adapting strategies for geo-referencing inventory plots—which also provide an estimation of the reliability—might improve the integration of ALS data into forest inventories.

Aligning TLS Scans. As stated in Section I.3.1, scanning a sample plot using the multi-scan or multiple single-scan approaches requires a co-registration of either the TLS scans or the detected trees. Although both tasks can be seen as a rigid alignment of multi-dimensional point clouds, a focus will be set on the alignment of TLS scans. A typical protocol to align multiple scans is to place reference markers, which are detected during post-processing (Wilkes et al., 2017; Liang et al., 2016; Raumonen et al., 2015; Bienert, Pech, and Maas, 2011) or by identifying key points based on features of the object geometries (Zhang et al., 2021; Date et al., 2019; Bienert, Pech, and Maas, 2011). Beland et al. (2019) notice that next to the logistics of TLS, the alignment of the scans to a common reference system is a complex method in the field. To be able to identify markers, the understory might even have to be removed before scanning (Raumonen et al., 2015). To allow for an automated marker-less registration of scans, leading TLS manufacturers provide various software, e.g., *Cyclone* (by *Leica Geosystems*), *RiSCAN PRO* (by *RIEGEL*) or *SCENE* (by *FARO*). Methods like those proposed by Bienert, Pech, and Maas (2011), or more recently by Date et al. (2019) or Zhang et al. (2021), use variants of the iterative closest point (ICP) algorithm (Besl and McKay, 1992) to align point clouds using probably matching point pairs. Due to the complexity and similarity of forest objects, common methods using local features—typical in an industrial or urban context—tend to struggle finding suitable features for a coarse alignment (Zhang et al., 2021). Most recently, Ge and Zhu (2021) propose a target-based registration, since even the latest approaches cannot guarantee a successful registration in complex forests. In summary, there is still a need for alternative and flexible algorithms to enable a marker-less alignment of multiple terrestrial laser scans in complex forest environments.

I.4.1.4   Tools for Point Cloud Analyses for Forest Applications

Beland et al. (2019) observe that in recent years the research community has mainly focused on the identification of new applications of remote sensing technologies, rather than on developing broadly usable processing software. Hence, they conclude that software providing more advanced information than tree heights and tree distribution are slowly becoming available.

Commonly used software for processing (LiDAR) point clouds are the *PDAL* (PDAL contributors, 2018), *PCL* (Rusu and Cousins, 2011) and *Open3D* (Zhou, Park, and Koltun, 2018). Software like *LAStools* (Hug, Krzystek, and Fuchs, 2004), *FUSION* (McGaughey, 2016) and *LidR* (Roussel et al., 2020) are commonly used for processing ALS data (Beland et al., 2019). For TLS data Disney et al. (2019) and Beland et al. (2019) identify *Treeseg* (Burt, Disney, and Calders, 2019), *TreeQSM* (Åkerblom, 2020), *SimpleTree* (Hackenberg et al., 2015), *3D Forest* (Trochta et al., 2017), *Computree* (Othmani et al., 2011), *PyLidar* (Armston et al., 2015) and *Forestr* (Atkins et al., 2018) as commonly used software to extract and reconstruct individual trees, or to derive structural metrics.

Although various specialized software is available, Beland et al. (2019) encourage the development of cost-free, operational and well documented tools for processing raw LiDAR data. In line with their suggestion, a suitable set of tools for processing point cloud data is needed to address the previously described research gaps for tree trunk detection (Section I.4.1.1) and stem inclination analysis (Section I.4.1.2) as well as tree linking, geo-referencing and point cloud alignment (Section I.4.1.3). Hence, the following collection of tools and requirements has been devised:

- **Geospatial Operations**   Basic geo-spatial operations, like coordinate transformation and projection, need to be performed. Data structures for efficient point cloud, voxel and raster data processing (as well as data storage) are required. These should also be suited to perform efficient neighborhood analyses. In addition, a conversion and aggregation of different data structures (e.g., point to voxel) should be facilitated.

- **Point Cloud Processing**   For efficient point cloud processing, point filters—like local maxima filters for individual tree detection (Breidenbach and Astrup, 2013; Koch et al., 2013)—and clustering algorithms—e.g., k-means (Lloyd, 1982) or DBSCAN (Ester et al., 1996)—for tree crown, or stem delineation are required. These rely on efficient distance calculations and neighborhood analyses.

- **Model Fitting**   Functions to generate DTMs, digital surface models (DSMs) and CHMs are the

basis for most forest related applications (Gatziolis and Andersen, 2008; Hyyppä et al., 2008b), for example to detect (semi-)individual trees. In addition, functions to fit geometric primitives (like vectors, spheres or cylinders) need to be provided (e.g., to model trunks). These concepts also imply functions for multi-dimensional interpolation.

- **Data Alignment** To align ground truth data with remote data on an individual tree basis, individual trees need to be matched. In addition, several TLS scans need to be aligned. These tasks can be split into an alignment of n-dimensional point clouds (filtered point clouds or extracted tree positions), an assignment of point data (e.g., inventory tree and ALS detected tree positions), and a roto-translation of the data sets.

I.4.2 Objectives

In the given thesis the following four major objectives have been identified:

Objective I: Tree Trunk Detection using ALS – Based on the research gaps identified in Sections I.4.1.1 and I.4.1.2, the first objective is to develop an algorithm for the efficient identification of tree trunks using ALS data that does not need to identify tree crowns beforehand. Next to the opportunity to apply the algorithm to large scale study areas (to be able to potentially support forest surveys) the general detection characteristics shall be inspected. In particular, the positional accuracy shall be studied in detail.

Objective II: Causes of Tree Trunk Inclination – ALS detected tree stems shall be inspected for their suitability to provide valid information on the tree stem inclination. If this can be verified, the observations shall be compared with today's knowledge on tree growth. In this context, a particular objective is to examine whether the ALS detected stems can confirm the postulations on the effect of the terrain, wind and differences between the tree species found in literature.

Objective III: Methods for Data Fusion – As described in Section I.4.1.3, three main tasks of data fusion for forest applications have been identified. It needs to be noted that linking individual tree positions, geo-referencing forest survey plots based on tree positions, and aligning point clouds are algorithmically related problems. Understanding trees with their positions and attributes as multi-dimensional point clouds, the problem of finding corresponding tree pairs is similar to finding matching point pairs. In addition, an alignment of multi-dimensional point clouds or geo-referencing inventory plots based on tree positions can be achieved easily if matching point pairs are known. In

consequence, a third objective is to address the previously mentioned tasks of data fusion. In particular, synergies between the tasks shall be exploited to achieve efficient, generic and easily transferable methods. Most importantly, a post-processing strategy to geo-reference forest inventory plots and to match terrestrially measured trees with aerially detected trees shall be developed and tested.

Objective IV: Tools for Point Cloud Processing in Forestry – To achieve the main objectives of the given thesis an additional objective is to develop an application-oriented software package providing tools for processing discrete point LiDAR and raster data with a focus on typical tasks occurring in forest research. This package shall implement the requirements identified in Section I.4.1.4 in a generic manner. In consequence, suitable data structures need to be developed, existing tools need to be integrated and existing or novel algorithms for data processing need to be implemented.

I.5   Overview and Structure

The objectives of this thesis were addressed in four peer-reviewed standalone papers. Three papers—Lamprecht et al. (2015), Lamprecht et al. (2017a) and Lamprecht, Stoffels, and Udelhoven (2020)—have been published in journals listed in the Science Citation Index Expanded (SCIE). One paper—Lamprecht (2019a)—has been published in a journal focusing on a peer-reviewed publication of well documented and tested research software. As a cumulative dissertation, each of the papers forms a chapter of the thesis. In the following, the study setting of each paper is briefly summarized. A discussion of the main findings and their scientific relevance can be found in Chapter VI.

Chapter II: aTrunk—An ALS-Based Trunk Detection Algorithm – In this study, a point-based method for individual tree trunk detection is developed and tested. Next to the algorithm design, the focus of the study is set on an accuracy assessment, as well as investigating the positional accuracy of the detected stems (Objective I). The algorithms are applied to a study site of about 0.5 ha dominated by Norway Spruce and European Beech and compared to a traditional ITCD approach. To gain highly accurate ground truth for the accuracy assessment, the reference tree positions are measured using eight terrestrial laser scans. To align these scans, a first version of a marker-less point cloud alignment algorithm is developed (Objective III).

Chapter III: A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data – This study focuses on geo-referencing inventory plots of an NFI by aligning trees measured on the ground with remotely detected trees

based on methods of machine learning. A random forest classifier is used to estimate the matching probability of each terrestrial-reference and aerially-detected tree pair, which offers the opportunity to assess the reliability of the matches. The method is trained and validated with synthetic forest stands. In summary, the study addresses fundamental tasks of data fusion (Objective III). It needs to be noted that due to an incorrect equation in the work of Lamprecht et al. (2017a), the chapter is complemented by the erratum Lamprecht et al. (2017b).

Chapter IV: Pyoints: A Python Package for Point Cloud, Voxel and Raster Processing – In this chapter the *Pyoints* package, which provides algorithms and tools for point cloud analysis and data fusion with a focus on forestry, is presented. Thus, it implements Objective IV and provides fundamental tools to address Objective III. All program code is packaged as an open source *Python* repository (Lamprecht, 2019b; Lamprecht, 2019c). To develop the concept, next to the requirements defined in Sections I.4.1.4 and I.4.1.3, best practices identified in Chapters II and III are taken into account. The algorithms presented in Chapter V exploit the data structures and algorithms of *Pyoints*. Since the journal focuses on a well documented program code and reasonable tests, rather than on the concept and the originality of the software, the publication provides a general overview of the package only. Thus, a brief summary on implementation details can be found in Section VI.1.4.

Chapter V: ALS as Tool to Study Preferred Stem Inclination Directions – This study sets a focus on a cause analysis for individual tree trunk inclination as demanded in Objective II. Next to empirically verifying the occurrence of preferred stem inclination angles, linear regression models are fitted to verify known causes for trunk inclination. Beside the cause analysis, a newly developed stem detection algorithm is presented which uses simplified heuristics with just one core parameter to overcome issues with the parametrization of the original approach presented in Chapter II. Thus, the paper also addresses Objective I.

Chapter II

aTrunk—An ALS-Based Trunk Detection Algorithm

S. Lamprecht, J. Stoffels, S. Dotzler, E. Haß, and T. Udelhoven

Article

aTrunk—An ALS-Based Trunk Detection Algorithm

Sebastian Lamprecht *, Johannes Stoffels, Sandra Dotzler, Erik Haß, Thomas Udelhoven

Remote Sensing & Geoinformatics Department, Trier University, Behringstraße, Trier 54286, Germany; E-Mails: stoffels@uni-trier.de (J.S.); dotzler@uni-trier.de (S.D.); hass@uni-trier.de (E.H.); udelhoven@uni-trier.de (T.U.)

* Author to whom correspondence should be addressed; E-Mail: lamprecht@uni-trier.de; Tel.: +49-651-2014612.

Abstract: This paper presents a rapid multi-return ALS-based (Airborne Laser Scanning) tree trunk detection approach. The multi-core *Divide & Conquer* algorithm uses a CBH (Crown Base Height) estimation and 3D-clustering approach to isolate points associated with single trunks. For each trunk, a principal-component-based linear model is fitted, while a deterministic modification of *LO-RANSAC* is used to identify an optimal model. The algorithm returns a vector-based model for each identified trunk while parameters like the ground position, zenith orientation, azimuth orientation and length of the trunk are provided. The algorithm performed well for a study area of 109 trees (about $2/3$ Norway Spruce and $1/3$ European Beech), with a point density of 7.6 points per m^2, while a detection rate of about 75% and an overall accuracy of 84% were reached. Compared to crown-based tree detection methods, the *aTrunk* approach has the advantages of a high reliability (5% commission error) and its high tree positioning accuracy (0.59 m average difference and 0.78 m RMSE). The usage of overlapping segments with parametrizable size allows a seamless detection of the tree trunks.

Keywords: airborne LiDAR; stem detection; tree recognition; trunk orientation; clustering; forest; 3D

1. Introduction

1.1. Relevance

For a quantification of environmentally sustainable forest management an exact knowledge of various parameters is needed, e.g., the available number and distribution of trees as well as their species, the timber volume or the LAI (Leaf Area Index) [1]. Remote sensing based single tree characterisation procedures have been proven suitable to enhance forest inventories regarding, for example, ecology, wildlife or biodiversity (e.g., [2,3]). Moreover, growth simulations require single tree models (with features such as height, crown size, tree species, spatial distribution) as input variables [4]. In addition, single tree models have the potential to characterise features of biodiversity (for example, the identification of habitat trees [5]), the tree species [6–8] or the spatial distribution of trees.

1.2. State of the Art

Various approaches have been developed focusing on single tree detection using remote sensing data, while initially spectroscopy images and today mostly ALS (Airborne Laser Scanning) have been used as input data [9]. ALS data has been proven suitable to characterise forest stands on a single tree level ([9–11]). An airborne laser scanner is an active remote sensing system, which provides 3D (three-dimensional) information. Full-waveform and discrete-return ALS systems are distinguised, which are processed in the form of point clouds with additional intensity information [12,13]. Single tree detection relies on pattern recognition methods, which take advantage of the tree phenotype and, if ALS is used, optionally the neighbourhood data of the laser returns [3]. Jakubowski *et al.* [9] mention a general trend towards the usage of ALS alone and a more complex analysis with an increased accuracy of results. For a 2D (two-dimensional) identification and delineation, image-based approaches have been developed, for example watershed or valley-following segmentation techniques [14,15], the *marked point process* technique [16] or object-based image analysis [9]. An adequate identification of suppressed trees relies on the analysis of full-waveform or laser-point data ([11,12,17]). A 3D delineation is possible using voxel-based or vector-based approaches, for example multi-layer imaging techniques [18], variants of *k-means* clustering [19–21], adaptive 3D clustering [22], multi-scale clustering [23] or graph-based methods [2,12]. Generally the 3D approaches result in improved detection rates in vertically inhomogeneous forest stands and more complex and realistic crown shapes [9] compared to raster-based approaches.

1.3. Related Work

The mentioned 3D tree delineation methods are focusing on a crown segmentation which mostly relies on a detection of tree tops based on local maxima (*cf.* [24]). Alternative tree identification approaches focusing on a tree trunk detection are rare.

Reitberger *et al.* [12] present a single tree detection approach, which includes an identification and modelling of tree trunks (ALS point density of 10–25 points per m^2). This method distinguishes trees using a *normalized cut segmentation*, which is capable of using already known tree positions to increase

the accuracy. The initial tree positions are derived by a watershed segmentation method related to Vincent and Soille [25]. Within the segments the CBH (Crown Base Height) is estimated by analysing the vertical histogram. The trunk points are separated within a segment between the CBH and a ground cover threshold. The trunks are identified by a 2D hierarchical clustering, and subsequently modelled by a *RANSAC*-based (*cf.* [26]) line fitting. Reitberger *et al.* [12] defines several rules to the trunk models—minimum number of tree points, minimum length, maximum angle—which shall keep the commission errors low.

The method of Abd Rahman *et al.* [27] uses high density ALS data (70 points per m^2) to identify single trees and model their trunks. Crown segments are identified using a watershed segmentation approach, while the vertical histogram of the points within a segment is evaluated to estimate the crown base height (reference height). The trunk of a segment is identified using a three-dimensional top-down trunk-growing algorithm. The 3D distance, the vertical histogram as well as the trunk diameter are evaluated to decide which points are assigned to the trunk.

The method of Lu *et al.* [28] takes advantage of the ALS intensity values to isolate points associated with trunks from leaf-off deciduous trees (10 points per m^2). The extracted points are used for a three-dimensional bottom-up trunk-growing process, while the assignment is controlled by the least 3D distance and a horizontal distance threshold. To avoid false positives, the trunk length and the maximum height of the lowest trunk point are controlled by threshold values. Furthermore the identified trunks are used for a 3D crown delineation.

Crown-based tree detection approaches usually reach an RMSE in positioning of $0.7\,$m up to above $3\,$m [11,24,29], which relies—besides the detection technique—on the tree height class (*cf.* [11]). In combination with the derived trunk positions Reitberger *et al.* [12] notices an increased accuracy in positioning by up to $25\,$%.

1.4. Objective

The *aTrunk* approach shall focus on a rapid ALS-based trunk detection without a dependency on a previous crown identification and shall work seamlessly on the entire observation area. The results shall be suitable to support large scale operational forest surveys, which makes a limited computation effort and data reduction crucial (caused by cost limitations [3]). The unique linear representation of a trunk promises a high reliability of the models, while potentially suppressed trees or dead trunks could be detected too. Applications of the approach could be to use the models as initial positions for crown segmentation approaches (e.g., [12,16]) or to confirm alternatively derived tree positions. An analysis of the trunk characteristics, regarding trunk length, trunk orientation (zenith, azimuth), point distribution in growth direction or spatial tree distribution could give additional information for characterising forest stands and biodiversity.

2. Materials and Methods

2.1. Study Area

The study area—illustrated in Figure 1—is located near the city of Hermeskeil in Rhineland-Palatinate, Germany, at the coordinates $49°48'49''$N, $7°10'3''$E, with an expanse of about $100\,\mathrm{m} \times 80\,\mathrm{m}$ and an area of about $0.5\,\mathrm{ha}$. The examined open forest compartment with tree heights of up to $35\,\mathrm{m}$ was dominated by Norway Spruce (59 trees recorded), but edged by European Beech (16 trees) at the western part of the study area. A measurement campaign on Tuesday, 19 August 2014 derived different types of validation data. This study site was selected, because the forest compartment was to be cleared by a harvester. With this information an analysis of the wood volume can be performed in a further study.
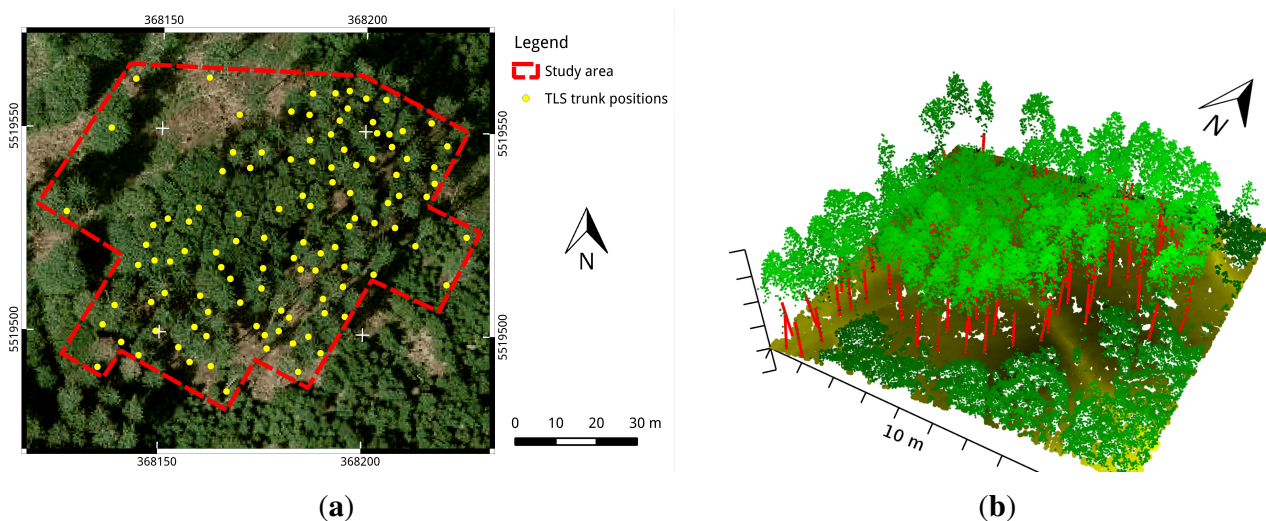
Figure 1. Study area. (**a**) Study area near city of Hermeskeil, Germany with trunk positions predicted by a TLS-based (Terrestrial Laser Scan) trunk detection approach and background WMS-Service [30]. Coordinates are specified in the EPSG:25832 system; (**b**) 3D ALS view of the study area. Yellow-coloured points correspond to ground and green points to vegetation. The brightness values correspond to the elevation. Red lines mark the finally modelled trunks.

2.2. ALS Data

The multi-return ALS data used in this study was collected by the state forest service of Rhineland-Palatinate for the whole federal state, which specifies an absolute horizontal accuracy of $0.30\,\mathrm{m}$ and a vertical accuracy for open terrains of $0.15\,\mathrm{m}$ while a point density of at least 4 points per m^2 is guaranteed. The data was available in form of *ASCII*-files with a spatial expanse of $1\,\mathrm{km} \times 1\,\mathrm{km}$ for each dataset. For the evaluation of the *aTrunk* approach, a subset of the study area of about $125\,\mathrm{m} \times 116\,\mathrm{m}$ (with a extent of [368,112.5, 5,519,462.5, 368,237.5, 5,519,578.0] in EPSG:25832 system) and an overall point density of about 7.6 points per m^2 was chosen. Figure 1b illustrates the

3D structure of the study area. In addition, a full $1\,\text{km}^2$ dataset with the extent [368,000, 5,519,000, 369,000, 5,520,000] and an average point density of 7.7 points per m^2 was analysed.

2.3. Validation Data

The validation data collection aimed especially at the position of the trunks because this information can be easily used to estimate the detection rate of the algorithm and the accuracy in modelling the trunk positions. Moreover, ground measurements of the diameter at breast height (DBH) were taken for each trunk, which resulted in an average DBH of $0.255\,\text{m}$ with a standard deviation of $0.049\,\text{m}$.

For a measurement of the trunk positions, a differential GNSS (Global Navigation Satellite System) of the type *Topcon HiPer V* (*cf.* Topcon Corporation [31]) was used. Because of a poor signal quality, the total accuracies of the measurements were quite low, with location differences clearly above $0.5\,\text{m}$.

To gain accurate information about the topology of the trunks, eight terrestrial laser scans (TLS) were taken with a *Laser Scanner Photon 120* of the manufacturer FARO®, which measures at a wavelength of $758\,\text{nm}$ and reaches a ranging accuracy of $\pm 2\,\text{mm}$ at a distance of $25\,\text{m}$ (*cf.* FARO Europe GmbH [32]). Each of the all-around scans had a scan size of 8044×3446 pixels with the scanner-specific parameters: $1/5$ resolution and $3\times$ quality. The positions of the scans were chosen in such a manner that the study area was mapped completely. The alignment of the single scans to each other was prepared by placing reference spheres which can be identified in the post-processing software.

These TLS datasets were used to estimate the trunk position and diameter by applying a slicing approach (similar to [33]). The slices were $0.05\,\text{m}$ thick, beginning at a level of $1.3\,\text{m}$ up to the crowning height with a vertical distance of $1\,\text{m}$. To compare these models to the *aTrunk* positions, the positions at ground level were estimated by fitting a linear model to the slice centres of each trunk.

2.4. Methods

2.4.1. Preprocessing

The *aTrunk* approach relies on a height-normalised ALS point cloud using a digital terrain model. The delivered ALS data of the study area had already been classified into ground points and non-ground points by the provider. A *Delaunay*-triangulation of the ground points was used to calculate the height above ground for the non-ground points using *CloudCompare* [34].

2.4.2. Assumptions on Trunk Representation

ALS point clouds of different forested areas were visually inspected to define rules for a knowledge-based trunk modelling. The 3D view of the vegetation layer (green highlighted dots in Figure 1b) shows the complex crown structure and allows to identify single trunks because of a linear structure of the corresponding points. The following characteristics of a trunk mapped by ALS are suggested, which form the basis of the trunk model concept outlined in Sections 2.5.3 and 2.5.4:

The LiDAR points associated with a trunk are:

- widely spatially separable from the crown portion and the ground covering vegetation.

- moderately surrounded by points associated with branches, foliage or other objects.
- arranged in a straight line, which is oriented along the growth direction of the trunk. The maximum deviation from this line depends on the length of the trunk, for example, caused by irregular growth or branching.
- largely uniformly distributed in growth direction of the trunk, which is substantiated in the spatial resolution of the ALS data.

2.5. Trunk Detection Algorithm

Figure 2 illustrates the major steps of the presented *aTrunk* approach. The current implementation was realised in *Python* [35]. As base data, the algorithm uses a height-normalised ALS point cloud in the *ASCII*-format.

(a)   (b)   (c)

(d)   (e)   (f)
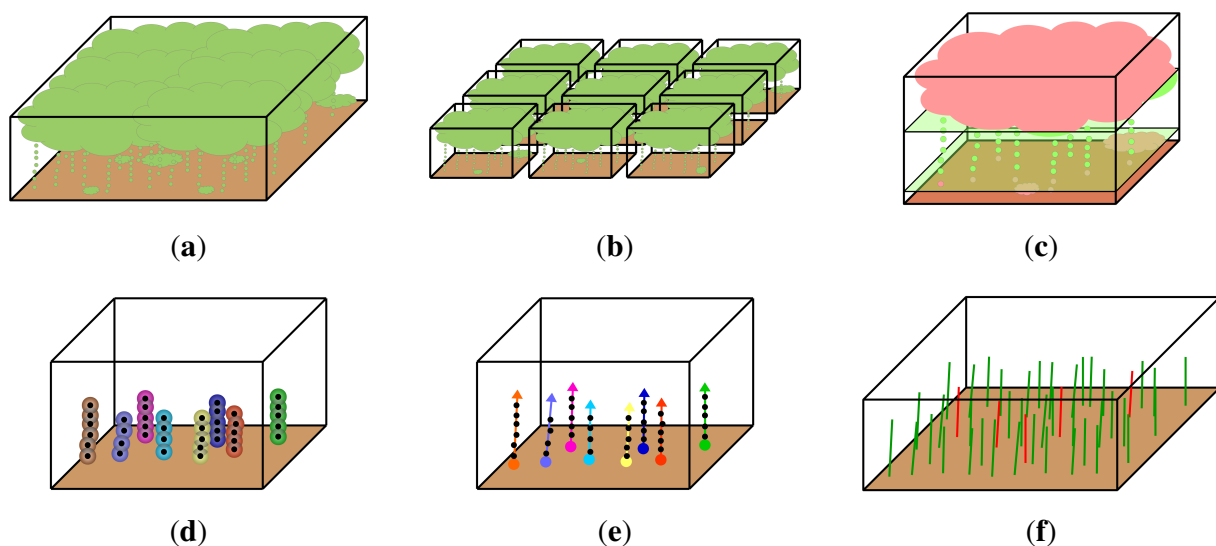
Figure 2. Major steps of the *aTrunk* approach. (**a**) Organisation of the height-normalised points (green) above ground (brown) as a sample S; (**b**) *Divide & Conquer* of the samples to get multiple subsamples; (**c**) Separation of the trunk section for each sample, by using two hyperplanes; (**d**) Trunk identification using a pairwise-distance-based clustering approach; (**e**) Fitting of a 3D principal-component-based linear model to each cluster; (**f**) Merging of the samples, elimination of duplicated trunk models (red lines) and data storage.

2.5.1. Divide & Conquer

The large amount of points makes a direct analysis of the point cloud almost unmanageable. For this reason the idea of introducing a splitting step is used to divide the point cloud into subsets that are easier to handle, which is sketched in Figure 2b. In addition, this procedure facilitates a local estimation of the CBH, which allows a separation of points associated with trunks from other points (see Section 2.5.2).

The input point cloud is stored as an object called sample S, which allows a splitting into two subsamples, while a split is always done in an xy-direction. The size of a sample needs to be selected in such a way that it is small enough to estimate the CBH accurately, but large enough to include at least

one tree. Each sample is separated into two new samples until the extent of the sample falls below a predefined threshold ($maxSampleSize$). This successive splitting allows a multi-core implementation of the algorithm in which it is possible to add a new CPU core to the program after each split.

Nevertheless, the splitting of the point cloud results in some disadvantages, which have to be dealt with. It is possible that points associated with a single trunk are separated from each other at random. This would result in an omission of trees. Therefore an overlapping area along the cutting edge of width *overlap* is suggested. The implementation of an overlap inevitably results in an additional analysis effort and the possibility of a multiple detection of a single trunk. These repeatedly detected trunks are merged, as described in Section 2.5.5.

2.5.2. Separation of the Trunk Section

The separation of the trunk section is based on the assumption that the potential trunk points P_{TS} lie between points which are associated with ground-covering vegetation and tree crowns (illustrated in Figures 2c and 3b). A first threshold plane at the user-specified height Z_{GCV} detaches the low-growing vegetation. For a sample with a maximum elevation z_{max} it is assumed that the crown base height Z_{CBH} is located in the range $R_{CBH} := [\rho_{min} \cdot z_{max}, \rho_{max} \cdot z_{max}]$. The parameters ρ_{min} and ρ_{max} correspond to the assumed minimum and maximum relative CBH $\rho_{CBH} := \frac{Z_{CBH}}{z_{max}}$ occurring in the study area.
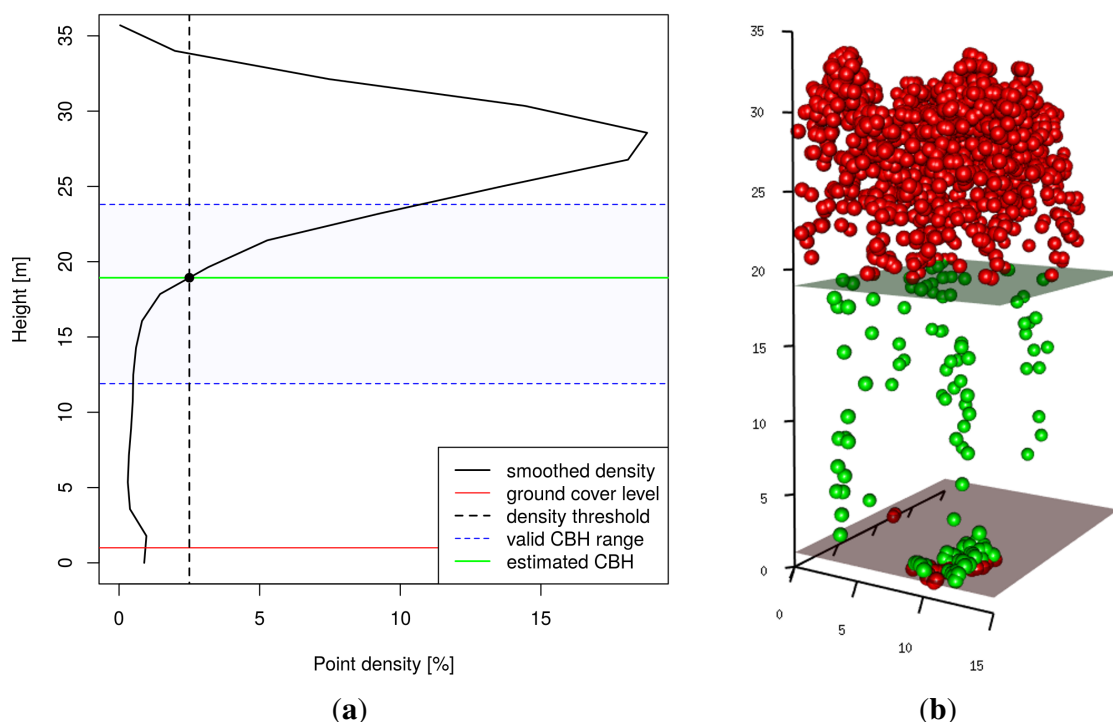

Figure 3. Crown Base Height (CBH) estimation and trunk section separation within a sample. (**a**) The CBH is estimated by identifying the intersection point between the smoothed vertical histogram and a threshold line. Here $N_{layers} = 20$ horizontal layers were used; (**b**) Separation of potential trunk points (green dots) from crown points or low-growing vegetation (red dots).

The CBH is estimated for each sample by analysing the vertical histogram (similarly to [12]). To become independent from the tree height, the points are divided into N_{layers} horizontal layers between Z_{GCV} and z_{max}. The histogram is normalised by dividing the number of points of each layer by the total number of points, and filtered by applying a moving average filter of width 3 (Figure 3a). The estimated CBH corresponds to the highest intersection point where the histogram exceeds the threshold value th. To reduce the dependency of th from the number of layers, it is defined by $th := \frac{thCBH}{N_{layers}}$, while the parameter $thCBH$ is set by the user. If the estimated CBH is located outside the range R_{CBH}—for example caused by a vertically uniform distribution of the points which occurs sometimes for dead trees—it is guessed by $\rho_d \cdot z_{max}$, while the proportion ρ_d is estimated by the user for the study area.

2.5.3. Clustering

The clustering step is used to identify points associated with trunks in the point set P_{TS}. It is assumed that several points which are spatially close together will form a trunk. Isolated points without or with just a few neighbours are assumed to be noise or sparse vegetation. The major problem is to identify point groups without knowing the number of clusters a priori (in contrast to, for example, a *k-Means*-based clustering approach). The proposed cluster definition—as a multi-dimensional extension of the *DBSCAN* [36] approach—takes advantage of the pair-wise spatial neighbourhood of points and the point density to solve this problem.

Cluster Definition: A cluster $C_{p_0,\delta,cMinPts} \subseteq P$ of a point cloud $P \subseteq \mathbb{R}^N$ with dimension N is developed around the initialisation point p_0 using the threshold distance $\delta \in \mathbb{R}^+$ (see Equation (1)). To limit the influence of outliers, each point of the cluster has to have at least $cMinPts$ neighbours with distance δ.

$$C_{p_0,\delta,cMinPts} := \{ p \in P : |\{ p_c \in C_{p_0,\delta,cMinPts} \cup \{ p_0 \} : ||p - p_c||_2 \leq \delta \}| \geq cMinPts \} \quad (1)$$

For the purpose of this paper, this cluster definition allows to identify points associated with a single trunk P_{Ct} out of P_{TS} with $N = 3$ (illustrated in Figure 2d). Due to the high vertical variability of the ALS points, the z-coordinates are scaled by the parameter $zBufferScale$ (similar to [23]).

2.5.4. Trunk Model

To get the desired orientation information of the trunks, a linear model is fitted to the points associated with the trunk P_{Ct}. To receive a 3D vector-based regression model, a principal component analysis (PCA) is used.

Basic PCA-Model: A PCA of a point cloud $P \subseteq \mathbb{R}^N$ with $N \in \mathbb{N}$ dimensions provides N pair-wise orthogonal vectors $\overrightarrow{PC_i}$ (with $i \in 1 \dots N$) which are called principal component (PC) vectors. The PCs define a projected coordinate system, in which the PC scores ($\text{scores}(P) \subseteq \mathbb{R}^N$) of the points P correspond to the projected coordinates of these points in the new coordinate system (*cf.* Wold *et al.* [37]).

The first principal component $t_{\overrightarrow{PC_1}}$ of the point cloud P_{Ct} is oriented in the direction of the highest variance. For a nearly perfect linear alignment of the points associated with the trunk, the $t_{\overrightarrow{PC_1}}$ vector should be oriented in the direction of the trunk. As a data preprocessing, a mean centring of the P_{Ct} points is applied. So the trunk model corresponds to the $t_{\overrightarrow{PC_1}}$ vector which is translated to the original

centre point of the defining P_{Ct} points. The model residuals correspond to the PC scores of the second and third principal component, so these are calculated by Equation (2) (*cf.* p. 9985). Figure 4 illustrates the expected principal component model for a trunk.
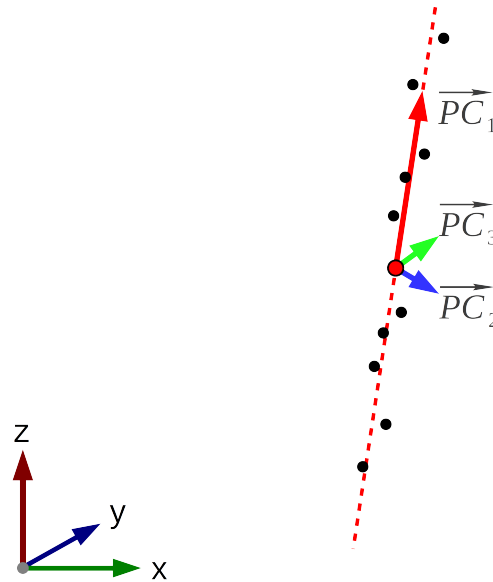

Figure 4. Three-dimensional one component PC trunk modelling concept with P_{Ct} points (black dots) and centre point (red dot).

Unfortunately, the assumption of perfectly linearly arranged points with no kind of outliers cannot be maintained. Therefore an attempt was made to find one PC model which fits most of the points by excluding outliers.

Idea of Best Model Selection: One way to solve the problem of fitting a regression model to highly outlier-affected data is the *RANSAC* approach by Fischler and Bolles [26], which was also used by Reitberger *et al.* [12]. The *RANSAC* approach is based on the assumption that outlying points can be identified by fitting multiple models which each rely on the minimum number of necessary points.

A basic assumption is that the points are randomly chosen and all models are independent from each other. In the case of a PC model, exactly two points are needed for each *RANSAC* sample. After this initialisation, all those points which have a residuum to the model below a specific threshold are assumed to support the model, while the others are assumed to be outliers. The model with the smallest proportion of outliers is accepted as the best model.

In this study the idea of the *LO-RANSAC* approach [38] was used, which optimises the outlier identification by running an additional model fitting for each *RANSAC* sample, using the model-supporting points. This technique relies on $k \approx \frac{\log(\eta)}{\log(1-\epsilon^m)}$ (*cf.* Chum *et al.* [38], p. 2 & 3) random samples to find—with a probability of $P(O)$ % a maximum proportion of O % outliers, with $\epsilon := P(O) \div 100 + 1$ and $\eta := O \div 100 + 1$—a set of at least m inliers. This equation clarifies that the number of points associated with a trunk is insufficient in order to find enough independent points for the common *LO-RANSAC* approach. This is caused by the low number of points associated with a trunk (usually clearly below twenty) and an extremely high proportion of outliers (often caused by ramifications in the upper trunk section or by low-growing vegetation). To derive a deterministic

model, every point pair combination of P_{TS} is defined as an initial pair, whether or not the models are independent.

Model Quality: The quality analysis of a trunk model is divided into two steps. In a first step, the validity of the model is tested, where already one fail results in rejecting the model. In a second step, the model with most model-supporting points is assumed to be the optimum. In case of the same point number, the model with the smallest mean squared error (MSE) is selected. For the validity check, some properties of each model are evaluated according to the modelling parameters summarised in Table 1 (see p. 9988).

A model is assumed to be valid if ...

- it contains enough points to ensure an accurate adaptation but unlikely false detections.
  $\Rightarrow |t_P| \geq n_{min}$
- it contains only some points, because it is assumed that a high number of neighboured points is probably caused by leaves or branches. The value ρ corresponds to the point density of the sample.
  $\Rightarrow |t_P| \leq f_{max} \cdot \rho$
- the range of z is large enough to contain a trunk.
  $\Rightarrow \text{range}\,(t_{Pz}) \geq minZRange$
- the ratio between the z-range (height) and xy-range (width) is comprehensible.
  $\Rightarrow \frac{\text{range}(t_{Pz})}{\max\left(\text{range}(t_{Px}),\text{range}(t_{Py})\right)} \geq hwRel$
- the zenith angle of the trunk is reasonable.
  $\Rightarrow t_\zeta \leq maxZenith \in [0,90[$
- the model has a favourable ratio between model-supporting points and outliers.
  $\Rightarrow \frac{|t_P|+|t_O|}{|t_O|} \leq relOutliers \in [0,1[$
- the points associated with the trunk are largely uniformly distributed in $t_{\overrightarrow{PC_1}}$ direction.
  $\Rightarrow t_{\chi^2} < uniformProb \in [0,1]$

Due to the assumption of residuals increasing with the length of the trunk, an adaptable threshold is needed to identify outliers. Therefore a length-dependent quality criterion—called $MEPL$ (Maximum Error Per Length)—is proposed, which shall privilege large trunks in residual weighting. Equation (9) (p. 9987) illustrates the calculation of this criterion, while the still unknown length of the trunk is approximated by the z-values of the points.

Figure 5 illustrates the modified *LO-RANSAC* concept for the PC-based trunk modelling approach, while Figure 2e sketches the modelling results within a sample.

2.5.5. Merge Duplicated Trunks

The split of the samples—done in step 2.5.1—with an intended overlap area can result in a multiple detection of single trunks in the overlapping separation section. This undesired effect is compensated for by an identification and subsequent merging of duplicated trunks whose centres are close together. As it cannot be ruled out that the models differ from each other (caused by a different point basis), all raw points are joined, while duplicates are eliminated. After the merging, a new trunk model is fitted

just as explained in Section 2.5.4. All the derived trunk models are combined again in a result dataset as sketched in Figure 2f.
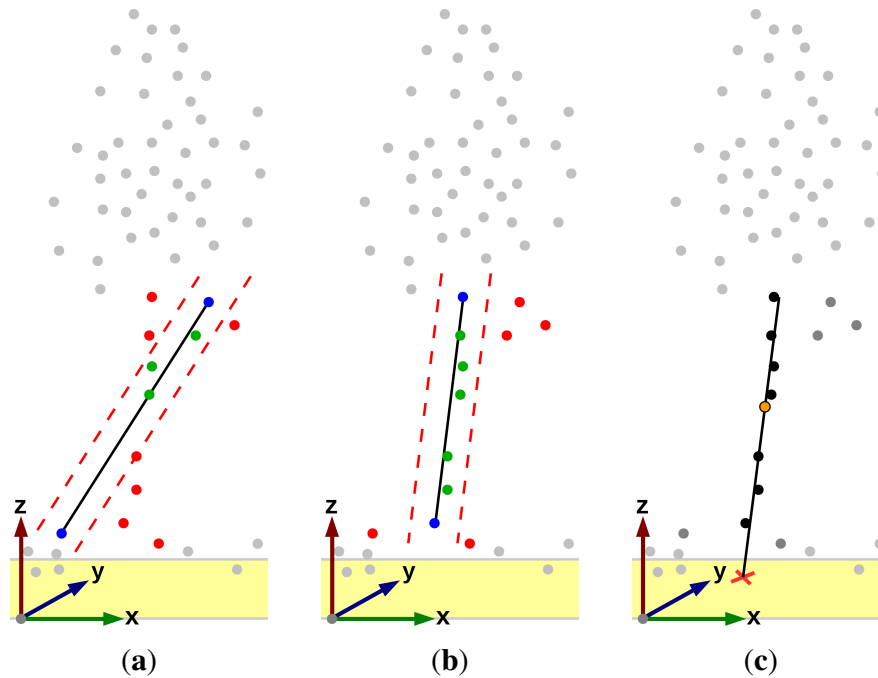

Figure 5. LO-RANSAC-based line fitting approach. Two possible raw trunk models (Figure 5a,b) using two randomly chosen points (blue dots), inlier threshold (red dashed lines), identified inliers (green dots) and outliers (red dots). Grey dots correspond to excluded points. Final model (Figure 5c) using inlier points t_P (black dots), centre point t_c (orange dot) and trunk position t_p (red cross). (**a**) Low quality model in terms of RANSAC; (**b**) High quality model in terms of RANSAC; (**c**) Final trunk model.

2.6. Trunk Model Properties

2.6.1. Principal Component Model

The PC model is based on a set of points associated with the trunk t_P, whose selection was explained in 2.5.4. In addition, the points assumed to be outliers t_O, which surround the trunk, are provided by the model. The three principal components $t_{\overrightarrow{PC_1}}$, $t_{\overrightarrow{PC_2}}$ and $t_{\overrightarrow{PC_3}}$ depend on the inliers t_P only. The $t_{\overrightarrow{PC_1}}$ corresponds to a linear regression model of the trunk, while the residuals $\epsilon(t_P)$ of this regression vector are defined by the PC scores of the second and third component (see Equation (2)).

$$\epsilon(t_P) = \sqrt{\text{scores}(t_P)_2{}^2 + \text{scores}(t_P)_3{}^2} \tag{2}$$

2.6.2. Trunk Orientation

The zenith angle $t_\zeta \in [0, 90[$ (Equation (3)) describes the angle deviation of $t_{\overrightarrow{PC_1}}$ (the trunk) to a vertical line.

$$t_\zeta = \frac{\cos^{-1}\left(t_{\overrightarrow{PC_1}_3}\right)}{\sqrt{t_{\overrightarrow{PC_1}_1}{}^2 + t_{\overrightarrow{PC_1}_2}{}^2 + t_{\overrightarrow{PC_1}_3}{}^2}} \tag{3}$$

The azimuth angle $t_\alpha \in [0, 360[$ (Equation (4)) describes the deviation of $t_{\overrightarrow{PC_1}}$ from the northern direction which indicates the leaning direction of the trunk.

$$t_\alpha = \tan^{-1}\left(\frac{t_{\overrightarrow{PC_1}_1}}{t_{\overrightarrow{PC_1}_2}}\right) \tag{4}$$

2.6.3. Position

The centre point t_c (Equation (5)) of the model corresponds to the spatial centre of the points t_P associated with the trunk which were used to fit the model.

$$t_c = \left(\overline{t_{Px}}, \overline{t_{Py}}, \overline{t_{Pz}}\right) \tag{5}$$

An advantage of the principal component line fitting approach is that the coordinates of every point of the trunk can be calculated as a linear combination of the first principal component $t_{\overrightarrow{PC_1}}$ and the centre point t_c. Therefore a point p_k on this regression line can be calculated by Equation (6), in which the scalar k corresponds to the desired height along the growth direction of the trunk relative to the centre point c_t:

$$p_k = k \cdot t_{\overrightarrow{PC_1}} + t_c \tag{6}$$

This feature is used to estimate the position of the trunk t_p, while the z-component of the t_p coordinate should be zero. Therefore the parameter k is estimated by Equation (7), in which t_ζ corresponds to the zenith angle of the trunk and h to the height of the centre point t_c.

$$k = \frac{h}{\cos(t_\zeta)} \tag{7}$$

The trunk top position t_{top} corresponds to the modelled centre of the trunk at the assumed CBH height. It can be calculated corresponding to the ground position by using Equation (6), while the parameter k is set to the assumed height of the trunk (see Section 2.6.4).

2.6.4. Trunk Height

The height of a trunk t_h is implicitly given by the expected crown base height Z_{CBH} (see Section 2.5.2). The length of a trunk t_l is calculated by Equation (7), in which the variable h is set to Z_{CBH}.

2.6.5. Quality Criteria

The quality of a trunk model is evaluated by different criteria. Apart from the classical MSE, the MEPL (mentioned in 2.5.4) or the uniform distribution criterion (t_{χ^2}) of the points associated with the

trunk can be used. The t_{MSE} is based on the residuals $\epsilon(t_P)$ of the regression line (Equation (8)), while t_{MEPL} is defined by Equation (9). The uniform distribution criterion of the inliers is calculated by applying a *Chi-Square Test* on the pair-wise distance according to their sorted $t_{\overrightarrow{PC_1}}$ scores.

$$t_{MSE} = \overline{\epsilon(t_P)^2} \tag{8}$$

$$t_{MEPL} := \frac{\max\left(\epsilon(t_P)\right)}{\max(t_{P_z}) - \min(t_{P_z})} \tag{9}$$

2.7. Methods of Evaluation

To carry out an evaluation of the *aTrunk* approach, it was applied to the ALS point cloud of the study area. The measured GNSS positions and derived TLS trunk positions served as reference datasets. To minimise systematic errors both the GNSS and the TLS positions were adjusted to the detected positions using an affine point set registration. A detected trunk position was assumed to correspond to its reference position when its distance was below $4\,\mathrm{m}$. In addition, the selection was limited to the extent of the reference datasets. It should be noted that both reference datasets did not include all trunks because some trees were not measured or not detected by the slicing approach.

A local maxima-based *watershed* segmentation—similar to the approach used by Hyyppä *et al.* [3])—served as benchmark, which identifies the trees using assumed water basins and dams. A raster model with $0.5\,\mathrm{m} \times 0.5\,\mathrm{m}$ pixels, using the maximum point elevation for each pixel, was generated and smoothed by a $5\,\mathrm{m} \times 5\,\mathrm{m}$ *Gaussian*-filter. The trees were identified by the watershed-segmentation-tool of *SAGA* (*cf.* [39]). For each segment the coordinates of the pixel with maximum value were considered as a tree position when the value was above $15\,\mathrm{m}$.

To get information about the potential accuracy improvement of the *aTrunk* approach to the *watershed* approach, the *matching* positions (distance below $4\,\mathrm{m}$) were identified. A *combined* dataset was generated using both the *aTrunk* and *watershed* positions. In case of corresponding positions the *aTrunk* positions were preferred. An accuracy assessment of the *aTrunk*, *watershed*, *matching* and *combined* positions was performed using the TLS positions as reference. Due to the large location residuals, the GNSS reference dataset was rejected for further evaluation.

A sensitivity analysis of the the *aTrunk* approach was performed in two steps. Firstly the parameters were optimised by applying the algorithm repeatedly to the study area with a large number of randomly selected parameter sets. The TLS positions served as reference to quantify the accuracy, while the positions were assigned automatically (with a maximum distance of $3\,\mathrm{m}$). A parameter combination (illustrated in Table 1) with a high detection rate maintained for the study area. For the second step of the sensitivity analysis, the chosen set served as initial values, while each parameter was varied in its reasonable range of values.

Table 1. Model Parameters

Parameter Name	Values' Range	Unit	Description	Value in This Study	Reference Section
$minPoints$ (n_{min})	$\mathbb{N}_{>1}$	–	Minimum number of points assumed to form a trunk	4	2.5.4
$maxPointsF$ (f_{max})	$\mathbb{R}^+$	–	Adaptive maximum number of points forming a trunk	5.0	2.5.4
$overlap$	$\mathbb{R}_0^+$	m	Width of the overlapping area	5	2.5.1
$maxSampleSize$	$\mathbb{R}^+$	m	Maximum xy-size of a sample before trunk identification	5	2.5.1
$hwRel$	$\mathbb{R}^+$	–	Minimum ratio between z- and xy-range of a trunk	3.0/1.0	2.5.4
$minZRange$	$\mathbb{R}_0^+$	m	Minimum height of a trunk	3.0	2.5.4
$groundCoverLevel$ (Z_{GCV})	$\mathbb{R}$	m	Maximum height of ground-covering vegetation	1.0	2.5.2
$minCBH$ (ρ_{min})	$[0,1]$	–	Assumed minimum relative crown base height	0.35	2.5.2
$maxCBH$ (ρ_{max})	$[\rho_{min},1]$	–	Assumed maximum relative crown base height	0.65	2.5.2
$defaultCBH$ (ρ_d)	$[\rho_{min},\rho_{max}]$	–	Default relative crown base height	0.45	2.5.2
$thCBH$	$\mathbb{R}^+$	–	Threshold for crown base height estimation	0.3	2.5.2
$delta$ (δ)	$\mathbb{R}^+$	m	Maximum distance of clustering algorithm.	1.5	2.5.3
$cMinPts$	$\mathbb{N}$	-	Minimum neighbours of a point in a cluster	2	2.5.3
$zBufferScale$	$\mathbb{R}_0^+$	–	Scale factor of z-axis for 3D clustering	0.1	2.5.3
$MEPL$	$\mathbb{R}_0^+$	m	Expected maximum error per length of trunk	0.07	2.5.4
$maxZenith$	$[0,90[$	°	Maximum assumed zenith angle of a trunk	10	2.5.4
$relOutliers$	$[0,1[$	–	Expected maximum ratio of t_P and t_O *vs.* t_O	0.7	2.5.4
$uniformProb$ (χ^2)	$[0,1]$	–	Assumed minimum unique distribution of the z-values	0.001	2.6.5
$mergeBuffer$	$\mathbb{R}_0^+$	m	Assumed minimum distance between two trunks	1.8	2.5.5

3. Results and Discussion

3.1. Sensitivity Analysis

For the evaluation of the results of the sensitivity analysis, the parameters were grouped by their qualitative effect on the results (see Table 2). Selected parameters are illustrated in Figures 6–8 while other parameters can be found in the supplement.
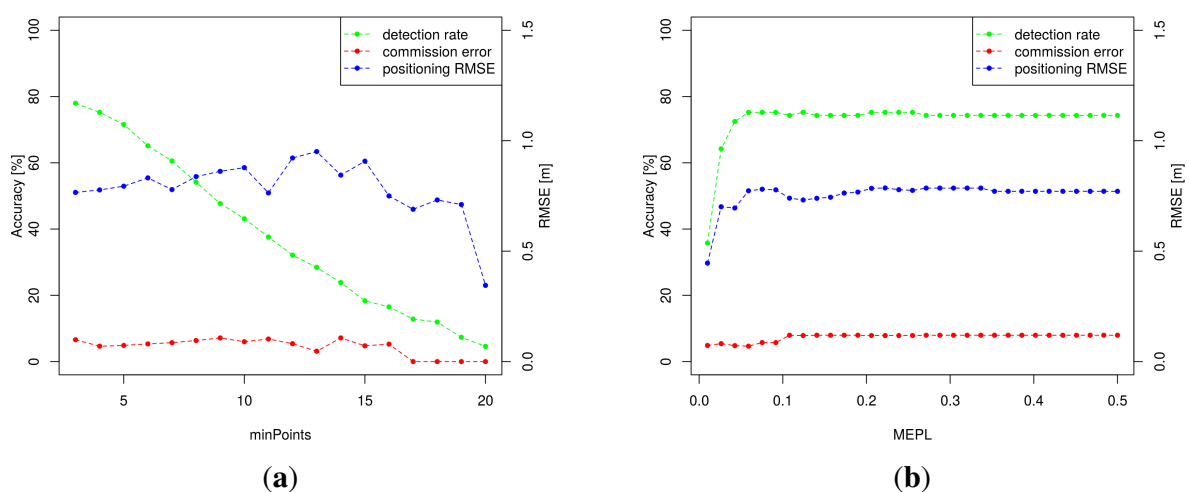

Figure 6. Sensitivity analysis of selected parameters controlling the model quality (Group 2). (**a**) Minimum number of points n_{min}; (**b**) Quality criterion $MEPL$.
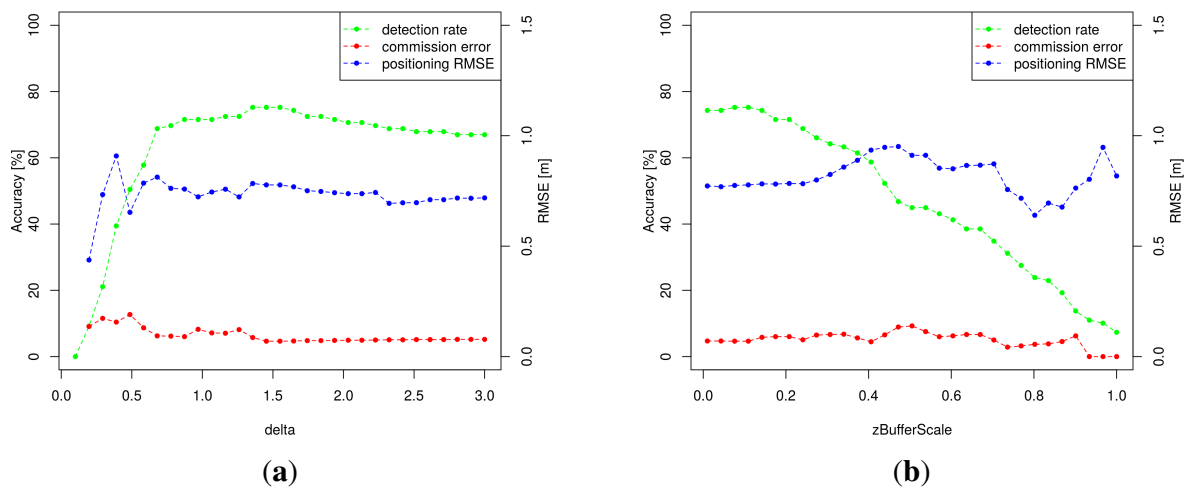
Figure 7. Sensitivity analysis of selected parameters controlling the clustering (Group 4). (**a**) Cluster parameter δ; (**b**) Cluster parameter $zBufferScale$.
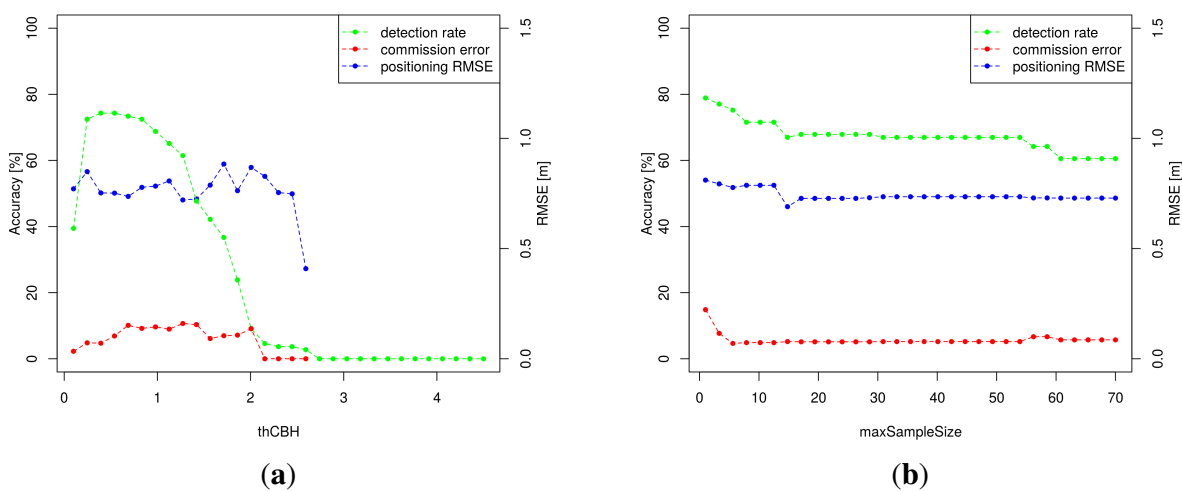

Figure 8. Sensitivity analysis of selected parameters controlling the CBH estimation (Group 5). (**a**) Crown base height threshold $thCBH$, with $\rho_d = 0$; (**b**) Sample size parameter $maxSampleSize$.

Table 2. Parameter Groups

Group	Parameters	Expected Effect on Results
1	$overlap, f_{max}, hwRel, relOutliers$	Control the computation effort
2	$n_{min}, MEPL, \chi^2, maxZenith$	Control the trunk model accuracy
3	$mergeBuffer, \rho_{min}, \rho_{max}, \rho_d, Z_{GCV}, minZRange$	Rely on stand structure
4	$\delta, zBufferScale, cMinPts$	Control the clustering
5	$maxSampleSize, thCBH$	Control the CBH estimation

Group 1: *Group 1* controls the computation effort and has—in a reasonable range of values—a low influence on the accuracy. The parameter *overlap* should be larger than the assumed average tree distance to ensure identifying all points of a trunk.

Group 2: *Group 2* defines the demands of the user to the trunk model accuracy while stricter demands result in reduced commission errors or improved positioning accuracies but also in reduced detection rates. Figure 6a illustrates that the detection rate decreases the higher n_{min} is chosen. Although the effect of n_{min} on the commission errors is low for the study area, a positive effect is expected for more dense point clouds and larger n_{min} values. The parameter $MEPL$ (Figure 6b) leads to reduced commission errors for values below 0.1 m while the detection rate decreases clearly for values below 0.05 m caused by an overestimation of the number of outliers. The parameter $maxZenith$ ensures finding almost vertical trunk models which results in a slightly optimised RMSE. A value above $5°$ is recommended to reduce omissions. The parameter χ^2, which ensures a certain uniform distribution of the point in growing direction, has shown to be unsuitable to reduce the commission errors.

Group 3: The values of *Group 3* conform to forest characteristics, which are not known exactly a priori. Nevertheless the parameter $mergeBuffer$ is widely stable for about the assumed minimum distance between two trunks. The parameters ρ_{min}, ρ_{max} and ρ_d—which support the automated CBH estimation—are estimated by expert knowledge. Assuming a correct automated CBH estimation, these parameters should have no influence on the results. The parameter Z_{GCV} reduces the detection rate for higher values, because the number of available points in the lower trunk section is decreased. For the study area the minimum trunk height $minZRange$ has a low influence on the commission error, while values above 5 m reduce the detection rate. Nevertheless it is assumed that for vertically inhomogeneous forest stands stricter demands on $minZRange$ will minimise the commission errors.

Group 4: The parameters of *Group 4* are necessary due to the clustering concept and are affected by the point density, vertical point distribution and noise. It is recommended to optimise the parameters δ (Figure 7a), $zBufferScale$ (Figure 7b) and $cMinPts$ for each dataset.

Group 5: *Group 5* controls the CBH estimation accuracy. The parameter $maxSampleSize$ (Figure 8b) ensures a local estimation of the CBH, which results in an increased detection rate for small values, but also in increased commission errors. These incorrect detections are caused by trees, which are divided into small point groups. Such points are identified as a trunk, because the low horizontal extent results in a vertical arrangement similar to a trunk. A $maxSampleSize$ value of about the average crown diameter is proposed. The threshold value $thCBH$ has a great impact on the detection rate (Figure 8a), because for values chosen too large the trunk models reach into the crown and for small values the available number of points is reduced, which both result in inaccurate trunk models and omissions. For the analysis of $thCBH$ the parameter ρ_d was set to zero, because the threshold range defined by ρ_{min} and ρ_{max} as well as ρ_d conceal wrong CBH estimations.

3.2. Evaluation

Table 3 illustrates the accuracy assessment using the TLS-derived trunk positions as reference. The assignment of the detected positions to the reference positions is visualised in Figure 9. The detection rate, precision and overall accuracy is calculated according to Yu *et al.* [13].

Table 3. Accuracy assessment of the different approaches.

Approach	Detection Rate	Precision	Overall Accuracy	Position Error	
				Average	RMSE
watershed	91%	85%	88%	1.04 m	1.25 m
aTrunk	75%	95%	84%	0.59 m	0.78 m
matching	69%	96%	80%	0.64 m	0.82 m
combined	98%	86%	92%	0.67 m	0.85 m
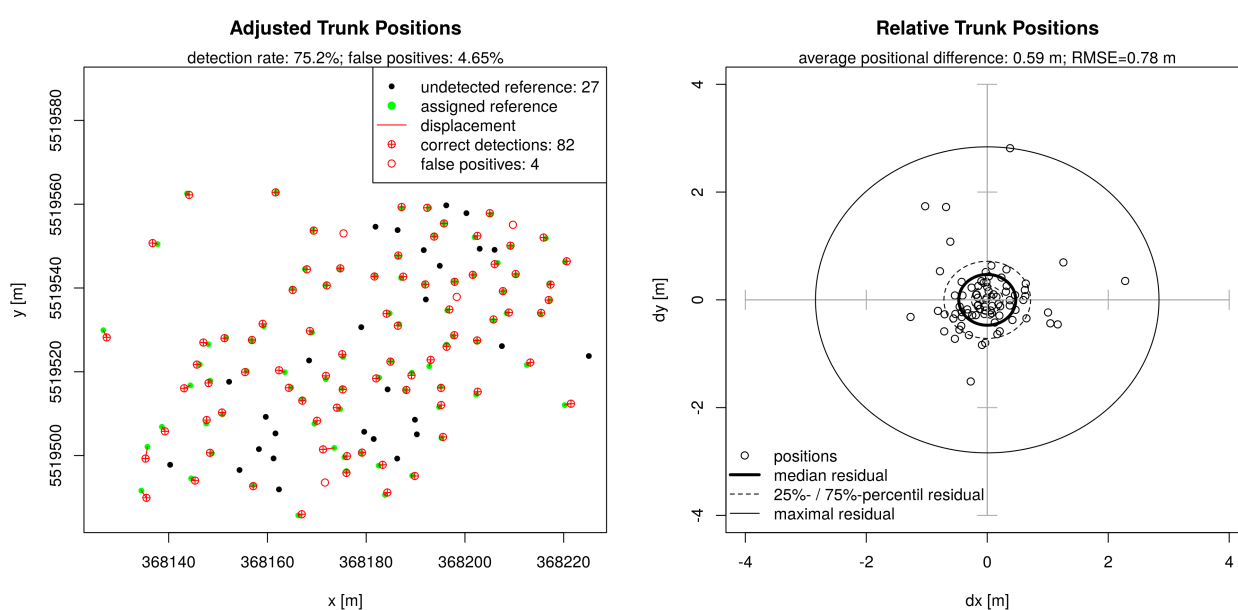
Figure 9. Evaluation of the detected *aTrunk* positions using the TLS positions as reference. The left image illustrates the assignment of the positions using absolute coordinates. The right image illustrates the relative positional differences between the reference positions and the corresponding detected positions.

The detection rate of the *aTrunk* approach is clearly outnumbered by the detection rate of the *watershed* approach. This is mainly due to the effect that not all trunks are recorded by the ALS data. In addition the homogeneous stand structure benefits the crown identification. The precision of the *aTrunk* approach is clearly better compared to the *watershed* results, because the point structure of trunks is less ambiguous compared to crown shapes which are partially merged. Nevertheless, other crown segmentation approaches could reach better results for this dataset.

The actual benefit of a trunk detection compared to a crown detection is the high accuracy in positioning, which significantly increases the positioning accuracy of the *combined* approach compared to the *watershed* approach. The improvement of the mean positioning error of about 32% supports the results of Reitberger *et al.* [12] which mention an improvement of up to 25%. The remaining residuals are mainly caused by branches and the trunk diameter. In addition uncertainties of the ALS positions (horizontal accuracy of about 5 to 15 cm assumed by May and Toth [40]) and of the TLS reference

positions are remaining. Nevertheless the inaccurate GNSS measurements imply that the trunk detection can be superior to ground-based GNSS positioning.

As the trunk detection is independent from the crown detection, the *matching* positions can be used as particularly reliable tree positions. The remaining commission error is partially caused by incomplete reference positions because of covered trees in the TLS data.

3.3. Modelling Results

Figures 10 and 11 illustrate the modelling results of the study area and the larger $1\,\text{km}^2$ sample using the same parametrisation. The homogeneously tall trees of the study area result in long detected trunks, with a mean value of $16.5\,\text{m}$ and a standard deviation (SD) of $2.5\,\text{m}$. In comparison, the $1\,\text{km}^2$ sample consists of different areas with varying tree heights. This results in a high standard deviation of about $2.7\,\text{m}$. The MSE distribution of the points associated with the trunk is right-skewed for both datasets. A median of $0.15\,\text{m}$ is reached for the study area and of about $0.06\,\text{m}$ for the $1\,\text{km}^2$ sample.
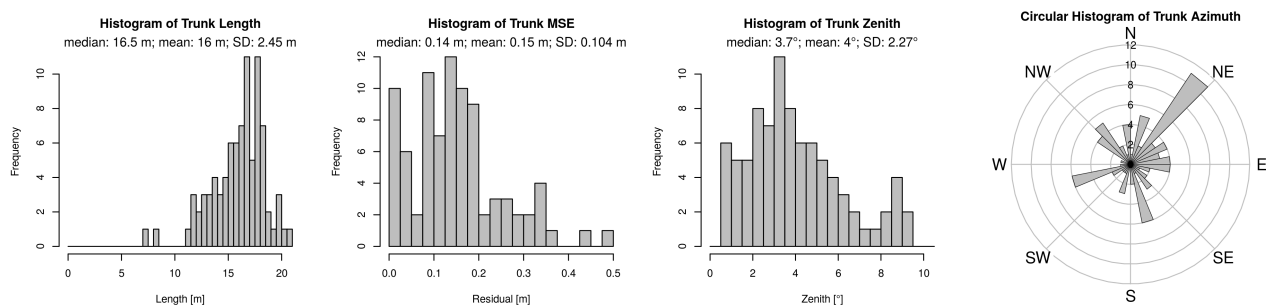
Figure 10. Modelling results of the study area (86 detected trunks).
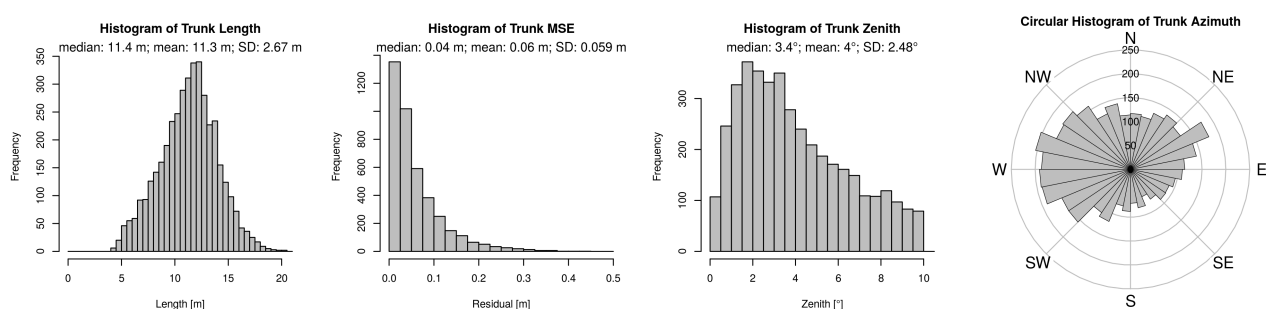
Figure 11. Modelling results of the $1\,\text{km}^2$ sample (4078 detected trunks).

The *zenith* is right-skewed distributed with a median of about $3.5°$, while most of the trunks clearly differ from a vertical orientation. The circular histogram of the trunk azimuth implies a preferred east-west or west-east orientation of the trunks. Under the assumption of low systematic measurement errors, this effect could be caused by the stand characteristics (e.g., soil conditions, topography or dominant wind direction).

3.4. Model Performance

The objective of this paper was the development of a rapid algorithm which is capable of detecting single tree trunks independent from a crown identification. It has been shown that the number of false detections of the *aTrunk* approach is low. On a *Dell Precision T1500* with an *Intel Core i5-750* processor and 8 GB RAM, the algorithm reached a computation time of about 13 s (0.2 s data loading and 12.7 s analysis) for the study area and about 7 min (13 s data loading and 461 s analysis) for the larger $1\,\text{km}^2$ sample.

A sensitivity analysis of the effects of data quality (e.g., point density, recording angle), stand structure (e.g., tree species, suppressed trees, inhomogeneous vertical structure, tree number, canopy density), seasonal conditions and terrain (e.g., morphology, steep or varying terrain) has to be performed in a further study including an evaluation of the orientation information, to provide a tool which is capable of supporting operational forest surveys. The additional information about the trunk orientation could help to improve the characterisation of forest stands.

3.5. Model Concept

The presented approach is related to the trunk detection module of Reitberger *et al.* [12], so the overall concepts—ground covering vegetation cut-off, CBH estimation, clustering and line fitting—were adopted. The *RANSAC* concept has been replaced by a deterministic modification of *LO-RANSAC* using a length-dependent quality criterion, and the 2D hierarchical clustering has been replaced by a 3D variation of *DBSCAN*. Especially, the non-overlapping crown segmentation has been exchanged with partially overlapping rectangular segments allowing a seamless tree trunk detection and a multi-core computation.

Current drawbacks are caused by the simplified CBH estimation, insufficient rules to decide if a point group represents a trunk, the simplified removal of ground covering vegetation, and not explicitly considering the characteristics of dominated trees. The CBH estimation could be optimised by more advanced methods (e.g., [41,42]). The sampling technique with overlapping areas leads to an additional computation effort. An alternative method which joins the points associated with trunks to a single dataset may reduce the computation effort under the assumption that the clustering can be done efficiently. Visual inspections had shown that trees which slightly cross the border of a segment can result in incorrect detections, because the isolated points may be aligned in a line. The rules which are used to decide if a point group represents a trunk need to be evaluated and refined to get optimal results.

Independently from the performed sensitivity analysis, the modelling concept implies that the approach depends on a high point density in the trunk section and clearly separable trunks, caused by the assumption of a linear arrangement of points associated with the trunks. So the approach is limited to forest stands with spatially distinguishable trunks where the canopy is permeable for the ALS beams. In addition low-branched trunks (like beech compared to spruce) and leaf off conditions are preferred, while a higher point density could optimise the trunk detection accuracy.

4. Conclusions

This paper presented a trunk detection approach which is independent from a crown identification and works seamlessly on the entire observation area avoiding non-overlapping segments. It derives reliable models (below 5% commission error), which are accurate in positioning ($0.78\,\text{m}$ RMSE). The models provide—next to the position—information about the trunk length, trunk orientation and point distribution (e.g., in growing direction). With a general focus on the computation effort, the algorithm is designed to be a potential tool for supporting operational forest surveys.

It has also been confirmed that the combination of the trunk positions with a crown-based tree detection is suitable to significantly improve the accuracy in tree positioning (32% for the study site). The independent detections result in a mutual confirmation of the tree positions which leads to decreased commission errors. In addition a preferred leaning direction of trunks was observed, which should be evaluated in further studies.

The algorithm has to be further evaluated regarding data quality, stand structure, seasonal conditions and terrain. It should also be evaluated whether the trunk detection is suitable to reduce the dependency of the tree size on the accuracy of detected tree locations which results in more accurate positions for taller trees [11].

Author Contributions

Sebastian Lamprecht as the main author developed and implemented the approach, led the data acquisition, analysed the data and is responsible for the content of the manuscript. Sandra Dotzler, Erik Haß and Johannes Stoffels took part in the validation data acquisition. Thomas Udelhoven and Johannes Stoffels initiated the study and cross-checked the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Maes, W.H.; Fontaine, M.; Rongé, K.; Hermy, M.; Muys, B. A quantitative indicator framework for stand level evaluation and monitoring of environmentally sustainable forest management. *Ecol. Indic.* **2011**, *11*, 468–479.
2. Strîmbu, V.F.; Strîmbu, B.M. A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 30–43.

3. Hyyppä, J.; Yu, X.; Hyyppä, H.; Vastaranta, M.; Holopainen, M.; Kukko, A.; Kaartinen, H.; Jaakkola, A.; Vaaja, M.; Koskinen, J.; *et al*. Advances in forest inventory using airborne laser scanning. *Remote Sens.* **2012**, *4*, 1190–1207.

4. Lafond, V.; Lagarrigues, G.; Cordonnier, T.; Courbaud, B. Uneven-aged management options to promote forest resilience for climate change adaptation: effects of group selection and harvesting intensity. *Ann. For. Sci.* **2013**, *71*, 173–186.

5. Kania, A.; Lindberg, E.; Schroiff, A.; Mücke, W.; Holmgren, J.; Pfeifer, N. Individual tree detection as input information for Natura 2000 habitat quality mapping. In Proceedings of the Remote Sensing and GIS for Monitoring Habitat Quality—RSGIS4HQ, Vienna, Austria, 24–25 September 2014.

6. Holmgren, J.; Persson, A.; Söderman, U. Species identification of individual treesHolmgren by combining high resolution LiDAR data with multi-spectral images. *Int. J. Remote Sens.* **2008**, *29*, 1537–1552.

7. Ørka, H.O.; N\a esset, E.; Bollandsås, O.M. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sens. Environ.* **2009**, *113*, 1163–1174.

8. Yu, X.; Litkey, P.; Hyyppä, J.; Holopainen, M.; Vastaranta, M. Assessment of low density full-waveform airborne laser scanning for individual tree detection and tree species classification. *Forests* **2014**, *5*, 1011–1031.

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

10. Leiterer, R.; Mücke, W.; Hollaus, M.; Pfeifer, N.; Schaepman, M.E. Operational forest structure monitoring using airborne laser scanning. *Photogramm. Fernerkund. Geoinf.* **2013**, *2013*, 173–184.

11. Kaartinen, H.; Hyyppä, J.; Yu, X.; Vastaranta, M.; Hyyppä, H.; Kukko, A.; Holopainen, M.; Heipke, C.; Hirschmugl, M.; Morsdorf, F.; *et al*. An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sens.* **2012**, *4*, 950–974.

12. Reitberger, J.; Schnörr, C.; Krzystek, P.; Stilla, U. 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 561–574.

13. Yu, X.; Litkey, P.; Hyyppä, J.; Holopainen, M.; Vastaranta, M. Assessment of low density full-waveform airborne laser scanning for individual tree detection and tree species classification. *Forests* **2014**, *5*, 1011–1031.

14. Chen, Q.; Baldocchi, D.; Gong, P.; Kelly, M. Isolating individual trees in a savanna woodland using small footprint LiDAR data. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 923–932.

15. Koch, B.; Heyder, U.; Weinacker, H. Detection of individual tree crowns in airborne LiDAR data. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 357–363.

16. Zhou, J.; Proisy, C.; Descombes, X.; Hedhli, I.; Barbier, N.; Zerubia, J.; Gastellu-Etchegorry, J.P.; Couteron, P. Tree Crown Detection in High Resolution Optical and LiDAR Images of Tropical Forest. Available online: http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.393.1014& rep=rep1&type=pdf (accessed on 16 April 2015).

17. Duncanson, L.I.; Cook, B.D.; Hurtt, G.C.; Dubayah, R.O. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. *Remote Sens. Environ.* **2014**, *154*, 378–386.

18. Wang, Y.; Weinacker, H.; Koch, B. A LiDAR point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest. *Sensors* **2008**, *8*, 3938–3951.

19. Morsdorf, F.; Meier, E.; Kötz, B.; Itten, K.I.; Dobbertin, M.; Allgöwer, B. LiDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sens. Environ.* **2004**, *92*, 353–362.

20. Gupta, S.; Koch, B.; Weinacker, H. Tree Species Detection Using Full Waveform LiDAR Data in a Complex Forest. Available online: http://www.isprs.org/proceedings/xxxviii/part7 /b/pdf/249_XXXVIII-part7B.pdf (accessed on 17 April 2015).

21. Lindberg, E.; Holmgren, J.; Olofsson, K.; Wallerman, J.; Olsson, H. Estimation of tree lists from airborne laser scanning using tree model clustering and k-MSN imputation. *Remote Sens.* **2013**, *5*, 1932–1955.

22. Lee, H.; Slatton, K.C.; Roth, B.E.; JR, W.P.C. Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. *Int. J. Remote Sens.* **2010**, *31*, 117–139.

23. Leiterer, R.; Morsdorf, F.; Torabzadeh, H.; Schaepman, M.; Mucke, W.; Pfeifer, N.; Hollaus, M. A voxel-based approach for canopy structure characterization using full-waveform airborne laser scanning. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 3399–3402.

24. Eysn, L.; Hollaus, M.; Lindberg, E.; Berger, F.; Monnet, J.M.; Dalponte, M.; Kobal, M.; Pellegrini, M.; Lingua, E.; Mongus, D.; *et al*. A benchmark of LiDAR-based single tree detection methods using heterogeneous forest data from the Alpine space. *Forests* **2015**, *6*, 1721–1747.

25. Vincent, L.; Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 583–598.

26. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395.

27. Abd Rahman, M.; Gorte, B.; Bucksch, A. A new method for individual tree delineation and undergrowth removal from high resolution airborne lidar. In Proceedings of the ISPRS Workshop Laserscanning 2009, Paris, France, 1–2 September 2009.

28. Lu, X.; Guo, Q.; Li, W.; Flanagan, J. A bottom-up approach to segment individual deciduous trees using leaf-off LiDAR point cloud data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *94*, 1–12.

29. Edson, C.; Wing, M.G. Airborne light detection and ranging (LiDAR) for individual tree stem location, height, and biomass measurements. *Remote Sens.* **2011**, *3*, 2494–2528.

30. Landesamt für Vermessung und Geobasisinformation Rheinland-Pfalz (LVermGeo). Luftbild RP Basisdienst. Available online: http://www.geoportal.rlp.de/mapbender/php/wms.php?layer_id= 30692&PHPSESSID=02dbaf5a20e411b1c46de1f8ef2a9cdd&REQUEST=GetCapabilities&VERS ION=1.1.1&SERVICE=WMS (accessed on 25 September 2014).

31. Topcon Corporation. HiPer V—Dual-Frequency GNSS Receiver. Available online: http://www.topconpositioning.com/sites/default/files/HiPer_V_Broch_7010_2121_RevB_TF_sm. pdf (accessed on 8 Dezember 2014).

32. FARO Europe GmbH. FARO® Laser Scanner Photon 120/20. Available online: http://www. faroeurope.com/portal/htdocs/download.php?id=1794&type=DOC&SiteCatalyst=true (accessed on 13 September 2014).

33. Bienert, A.; Maas, H.G.; Scheller, S. Analysis of the information content of terrestrial laserscanner point clouds for the automatic determination of forest inventory parameters. In Proceedings of the Workshop on 3D Remote Sensing in Forestry, Vienna, Austria, 14–15 February 2006.

34. EDF R&D. CloudCompare – 3D Point Cloud and Mesh Processing Software Open Source Project. Available online: http://www.cloudcompare.org/ (accessed on 14 November 2014).

35. Python Software Foundation. python. Available online: https://www.python.org (accessed on 9 February 2015).

36. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Available online: https://www.aaai.org/ Papers/KDD/1996/KDD96-037.pdf (accessed on 17 April 2015).

37. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.

38. Chum, O.; Matas, J.; Kittler, J. Locally optimized RANSAC. In *Pattern Recognition*; Michaelis, B., Krell, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 236–243.

39. SAGA User Group Association. System for Automated Geoscientific Analyses. Available online: http://www.saga-gis.org/en/index.html (accessed on 28 May 2015).

40. May, N.C.; Toth, C.K. Point positioning accuracy of airborne LiDAR systems: A rigorous analysis. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2007**, *36*, 107–111.

41. Popescu, S.C.; Zhao, K. A voxel-based LiDAR method for estimating crown base height for deciduous and pine trees. *Remote Sens. Environ.* **2008**, *112*, 767–781.

42. Vauhkonen, J. Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. *Int. J. Remote Sens.* **2010**, *31*, 1213–1226.

Chapter III

A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data

S. Lamprecht, A. Hill, J. Stoffels, and T. Udelhoven

remote sensing

MDPI

A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data

Sebastian Lamprecht [1,*], Andreas Hill [2], Johannes Stoffels [1] and Thomas Udelhoven [1]

[1] Remote Sensing & Geoinformatics Department, Trier University, 54286 Trier, Germany; stoffels@uni-trier.de (J.S.); udelhoven@uni-trier.de (T.U.)

[2] Department of Environmental Systems Science, ETH Zurich, 8092 Zurich, Switzerland; andreas.hill@usys.ethz.ch

* Correspondence: lamprecht@uni-trier.de; Tel.: +49-651-201-4612

Abstract: Determining the exact position of a forest inventory plot—and hence the position of the sampled trees—is often hampered by a poor Global Navigation Satellite System (GNSS) signal quality beneath the forest canopy. Inaccurate geo-references hamper the performance of models that aim to retrieve useful information from spatially high remote sensing data (e.g., species classification or timber volume estimation). This restriction is even more severe on the level of individual trees. The objective of this study was to develop a post-processing strategy to improve the positional accuracy of GNSS-measured sample-plot centers and to develop a method to automatically match trees within a terrestrial sample plot to aerial detected trees. We propose a new method which uses a random forest classifier to estimate the matching probability of each terrestrial-reference and aerial detected tree pair, which gives the opportunity to assess the reliability of the results. We investigated 133 sample plots of the Third German National Forest Inventory (BWI, 2011–2012) within the German federal state of Rhineland-Palatinate. For training and objective validation, synthetic forest stands have been modeled using the *Waldplaner* 2.0 software. Our method has achieved an overall accuracy of 82.7% for co-registration and 89.1% for tree matching. With our method, 60% of the investigated plots could be successfully relocated. The probabilities provided by the algorithm are an objective indicator of the reliability of a specific result which could be incorporated into quantitative models to increase the performance of forest attribute estimations.

Keywords: co-registration; individual tree detection; tree matching; point set registration; machine-learning; forest inventory

1. Introduction

Modeling and characterizing forest stands on small scales using high-resolution remote sensing data requires spatially explicit linking of inventory information and remote sensing data [1–3]. The exact sampling positions of field inventory plots are often determined using non-differential [4] or differential Global Navigation Satellite System (GNSS), which gives rise to location errors of up to several meters. In comparison, airborne laser scanning (ALS) provides absolute horizontal accuracies of about 25 cm (at 400 m flight altitude above ground) [5].

The effect of positional displacements between terrestrial reference data and ALS data has been investigated by Gobakken and Næsset [6] and Frazer et al. [7] at the sample plot level. They found that with increasing positional displacements the performance of biophysical models decreases in relation to the variable and stand characteristics. This problem exacerbates on individual tree level, since survey trees might get incorrectly linked to aerial detected trees (e.g., using ALS), which results in erroneous output data for further analyses.

Individual tree information has a specific added value for quantitative modeling compared to plot-based methods insofar as it includes features which address structural tree characteristics [8,9]. A general trend towards the usage of ALS data is recognizable [2]. Densities of about five to ten pulses per m^2 are sufficient for tree detection [10,11]. It has been shown that most methods perform well in the upper canopy layer, but the identification of understory trees is challenging, because the laser beam is not able to completely penetrate each canopy layer [3,11]. Tall or isolated trees can be detected most reliably with detection rates between 55% and 100%, while detection rates for trees in a group of similar trees decline to 30% to 80% [11]. Usually, only approximately 40% of the trees next to or below taller protruding trees are detectable [11]. Additionally, varying forest conditions result in a bias in the detection of large trees [12].

1.1. Positional Accuracy of Forest Inventory Data

Numerous research groups have analyzed the performance of different GNSS receivers considering various forest types and conditions [13–16]. They concordantly found that the absolute planar accuracy of field survey plot centers strongly depends on the GNSS receiver used, the GNSS signal quality, measurement duration, and whether or not a differential correction has been applied. The signal quality relies on the satellite geometry during recording, open sky conditions, and forest stand density [15]. With increasing stand density, the positional accuracy significantly decreases [14,15].

Wing et al. [14] investigated different mapping-grade GPS receivers and found average absolute horizontal GNSS errors under closed canopy ranging from 1.8 to 2.6 m. The average standard deviation of the horizontal errors was 1.5 m. Using differential GNSS, the average errors decreased to 1.2–2.0 m. In comparison, Valbuena et al. [16] observed average errors ranging from 0.90 to 2.14 m, and from 0.75 to 1.0 m after 10 or 20 min measurement time, respectively. Based on these findings and related previous studies on co-registration [4,17], GNSS errors of up to 20 m can be expected.

In forest inventories (e.g., in the German National Forest Inventory, BWI), individual tree positions are typically determined by measuring their distance to the plot center (e.g., using measuring tape or an ultrasonic distance meter and laser meter), and the corresponding azimuth angle using a compass. Thus, the accuracy of a tree position depends on the experience and care of the field staff, the accuracy of the measuring tools, and the distance of the tree to the plot center. The relative horizontal accuracy-level of tree positions can be assumed to be in a range between 0.3 m and 1.0 m [3,4,12].

Tree heights are typically determined using distance and zenith angle measurements that result in similar error sources. Luoma et al. [18] investigated the precision of conventional field measurements in boreal forests, and observed a standard deviation of height measurements of 2.9%.

1.2. Co-Registration and Individual Tree Matching

Typically, a manual co-registration is performed by identifying pairs of corresponding survey trees and aerial detected trees using canopy height models (CHMs) or high spatial resolution aerial imagery [3,4,19]. Research groups have developed automatic or semi-automatic methods to overcome this time consuming and complex procedure:

1.2.1. Automated Co-Registration of Field Survey Plots

Olofsson et al. [20] propose a tree linking and co-registration method based on cross-correlating the Gaussian surface images of terrestrial survey trees and aerial-detected trees for each pixel. To generate these images, the method uses the tree height or crown diameter for the survey trees and the CHM

tree height for the aerially-detected trees. Within an inventory plot, all trees within a 10 m radius and with a diameter at breast height (DBH) larger than 5 cm were recorded. Tested with simulated forest stands, the method performed well (above 75% correctly linked trees) if the detection method was characterized by less than 20% commission and omission errors and a random position error with standard deviation below 1 m.

The method of Dorigo et al. [4] minimizes the weighted difference between the terrestrial individual tree heights and the CHM for each pixel within a given search window. The weighting factor represents the visibility of a tree in the CHM. A heuristic quality assessment is applied to label a proposed plot center position as "certain" or "uncertain". The method achieved robust results for inventory plots recorded by angle count sampling. Of the co-registered plot center positions, 68–74% were characterized by a distance of less than 5 m compared to the locations (a priori) identified by a human interpreter.

Like Olofsson et al. [20], Monnet and Mermin [17] use cross-correlation for the co-registration of fixed radius inventory plots and ALS data. The method uses the DBH or tree height of the survey trees to generate a smoothed reference image. The cross-correlation is calculated between the reference image and the CHM for each pixel. If at least five inventory trees were used, more than 91% of the co-registered plot center positions were characterized by a distance of less than 2 m compared with the locations (a posteriori) identified by a human expert. They also found that the method achieved no better results for plots with "[...] already accurate GNSS positions, low-density ALS data and complex, mixed forests [...]" ([17], p. 2324).

Existing co-registration methods are especially challenged when the number of reference trees is low and the stand characteristics are complex for tree matching (high stand density, occurrence of similar tree patterns, evenly-aged stands, high proportion of deciduous trees). Heuristic methods have the disadvantages that the assumptions are defined based on the intuition of the expert, the suitability of these heuristics is hard to evaluate, and finally, the heuristics might compete. In addition, the algorithms provide no higher-order information (beyond quality flagging) about the reliability of the results.

1.2.2. Automated Individual Tree Matching

Available automated algorithms for matching survey trees and aerial tree detections are very similar, and are all based on rules defined by an expert user. For each terrestrial reference tree, neighbored detections (candidates) are selected within a predefined [11] or adapting [3,10,12] radius. Filters are applied (e.g., based on height difference: [10,11]) to exclude implausible candidates. The closest detected tree [12,21], the detected tree with lowest 3D distance [10], or the lowest value of a combined criterion (e.g., 2D/3D distance or height difference: [3,11,20]) is then assigned to the reference tree.

1.2.3. Related Methods for Co-Registration and Tree Matching

Alternative methods which might be suitable for co-registration and tree matching can be found in the fields of image pair registration, object recognition, or point set registration. These disciplines deal with finding a transformation function between two 2D or 3D point sets.

In the case of non-distorted point sets, the usage of the random sample consensus (RANSAC) [22] technique (to find corresponding point pairs) and the application of Helmert transformations is sufficient [23,24]. In the case of roughly preregistered point sets, the iterative closest point method is expedient [25].

For both distorted and non-distorted point sets, algorithms which identify matching point pairs by iteratively applying affine transformations while minimizing some kind of an objective function are popular. Examples comprise the Softassign algorithm [26], the coherent point drift algorithm [27], or more recently, a rigid registration algorithm inspired by gravitational force fields [28]. A further solution is to describe each point by some kind of feature which makes it recognizable in the second

point set. These features are defined by image statistics [29] or surface shape [30] in the proximity of the point. Relevant representatives use similarity measures [30] or machine learning [31] to find corresponding points based on the features.

1.2.4. Research Gap

Existing methods for co-registration are challenged by complex and variable stands (unevenly or evenly aged, mixed stands, or purely coniferous/deciduous)—especially if the number of survey trees is low. The limited number of survey plots particularly impedes the parametrization of the algorithms. To overcome these limitations, a procedure is needed which uses synthetic (modeled) training data to achieve optimal parameters and to reduce the need for additional validation data.

Since mismatched surveyed and detected tree pairs reduce the performance of empirical models (e.g., individual tree-based), a higher-order valuation of reliability (e.g., an a posteriori matching probability) is needed. This information could be suitable to enhance the performance these models by increasing the amount of usable reference data. This could be achieved by weighting the importance of a training dataset in relation to its reliability.

1.3. Objective

The objective of this study is to develop an automatic method for the co-registration of forest survey plots and ALS data which also provides information about matching trees and the reliability of a specific result. To provide an objective accuracy assessment, a process chain needs to be implemented that takes the inventory and tree detection characteristics into account.

In a further step, the method will be applied to data from the Third German National Forest Inventory Bundeswaldinventur (BWI), which has shown to be an unreliable reference data set at the individual tree level because of gross GNSS errors and a small number of trees per sample plot. This application example was chosen as a preparatory step for a subsequent study focusing on forest characterization on the individual tree level.

2. Data and Materials

2.1. Inventory Data

To develop and evaluate our co-registration method, we used inventory plots from the latest BWI (2011–2012) within the area of Hunsrück-Hochwald National Park located in Rhineland-Palatinate, Germany (Figure 1). The plot centers are arranged in a regular 2 km × 2 km grid. The plot centers were measured with differential GPS and projected into the ETRS89/UTM coordinate system.

In the BWI, a tree is recorded as a sample tree according to the angle count sampling technique [32]. As the BWI uses an angle count factor of 4, a tree is hence recorded if its distance to the plot center is less than 25 times its DBH. A further selection criterion for a tree to be recorded is a DBH of at least 7 cm. For each sampled tree, its species and DBH are recorded, whereas the tree height is measured only for a subset of the sample. The remaining tree heights are estimated using regression models based on the DBHs. Finally stand-specific taper functions [33,34] are applied. The tree location is determined by measuring the distance to the plot center and the corresponding azimuth angle.

As a result of this sampling technique, each plot is characterized by an individual maximum radius (determined by the most distant tree) as well as the maximum limiting circle (maximum distance where the strongest tree would still have been selected). The angle count sampling technique realizes the sample inclusion probability of a tree being proportional to its diameter. Therefore, this technique prefers in particular the selection of diameter strong (usually tall) trees.
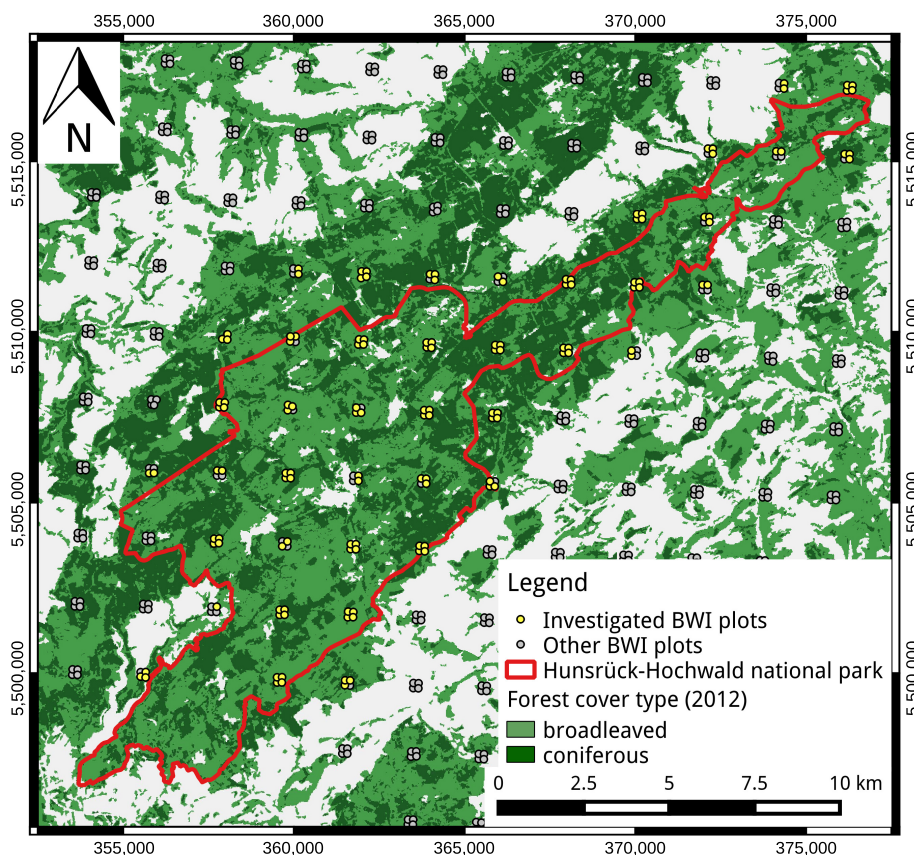
Figure 1. Study area and German National Forest Inventory (*Bundeswaldinventur*, BWI) sampling design. Background: Web Map Service (WMS) of forest types provided by Copernicus [35].

We focused on plots with at least two recorded trees of at least 4 m height, which resulted in 133 plots (65 plots surveyed 2011 and 76 surveyed 2012). These plots comprise 1015 trees in total, consisting of 43.3% Norway spruce (*Picea abies*), 37.6% European beech (*Fagus silvatica*), 4.8% sessile oak (*Quercus petraea*), 4.4% Douglas fir (*Pseudotsuga menziesii*), 4.1% European larch (*Larix decidua*), 3.3% European white birch (*Betula pendula*), and 2.4% others. Detailed information on the investigated plots is summarized in Table 1. As an indicator for the GNSS satellite constellation, the horizontal dilution of precision (HDOP, c.f. [36]) is specified.

Table 1. Basic statistics of the 133 investigated BWI plots. DBH: diameter at breast height; HDOP: horizontal dilution of precision.

Attribute	Minimum	1th Quartile	Median	Mean	3th Quartile	Maximum
DBH (cm)	7.0	23.8	36.2	36.73	47.9	112.5
Tree height (m)	5.2	19.8	24.8	24.65	29.9	50.6
Stems per hectare	48	202	412	742.4	839	6014
Number of recorded trees per plot	2	5	7	7.6	10	16
Maximum radius (m)	0.3	3.0	5.25	5.9	8.1	20.8
Limiting circle radius (m)	2.6	8.9	12.6	12.4	15.7	28.1
HDOP	0.8	1.1	1.2	1.32	1.5	3.2

2.2. ALS Data

The ALS data acquisition was accomplished from 24 March to 7 April 2015 using a *Riegl Q560* [37] under leaf off conditions. The mean flying height above ground was 600 m, and resulted in a footprint diameter of about 0.3 m and an average pulse density of 11.2 pulses per m^2. The ALS datasets were

provided by the state forest service of Rhineland-Palatinate in the form of pre-classified (ground vs. other classes) LAS-files. For each of the investigated BWI plots, we selected a circular subset with a radius of 38 m around the plot center.

3. Methods

3.1. Definitions

Each tree t is characterized by its coordinates t_x and t_y, as well as its height t_h. Thus, a tree can also be seen as a three-dimensional point. For a better mathematical formulation, a set of m trees shall be represented by a matrix of homogeneous coordinates $M^{m \times 4}$, such that a tree is defined by $(t_x, t_y, t_h, 1) \in M$. This representation allows a translation of M in direction $\vec{v} = (v_x, v_y, v_h)$ by applying the matrix multiplication $T_v M$, with T_v defined by Equation (1).

$$T_v = \begin{bmatrix} 1 & 0 & 0 & v_x \\ 0 & 1 & 0 & v_y \\ 0 & 0 & 1 & v_h \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

Following these definitions, each survey plot forms a data-set with a set of surveyed trees S and a set of detected trees D. A potential tree pair (s, d) with $s \in S$ and $d \in D$ is named *Matching* if s actually corresponds to d, but *Not matching* otherwise. Consistently, a set of potential tree pairs is assumed to be correct if at least 50% of the tree pairs are classified as *Matching*. The probability of a correct co-registration (GNSS error successfully identified using matching tree pairs) will be called "co-registration probability".

3.2. Digital Terrain Model (DTM) Generation

A digital terrain model (DTM) is generated for each plot based on filtered ALS points. To achieve these filtered points all points already classified as ground are selected. Then each ground point is investigated iteratively in ascending order of z-coordinate. For a given point, all points within a horizontal radius of 0.8 m are removed, except for the one with minimum z-coordinate. This procedure results in a subset of ground points which is characterized by a point spacing of 0.8 to 1.6 m. A Delaunay triangulation of this subset finally serves as DTM.

3.3. Individual Tree Detection

Since the angle counting technique prefers the selection of diameter strong (usually tall) trees, the detection of particularly dominant trees is expedient to register surveyed trees and detected trees. In the context of matching these surveyed trees to ALS detected trees, an identification of small or suppressed trees is not necessary. Thus, we decided to detect individual trees by identifying local maxima within the ALS point clouds. Hence, a point is assumed to correspond to the top of a tree if no point within a radius of 3 m has greater z-values. The threshold of 3 m seemed to be feasible due to the observed stand densities and the ALS pulse density. Unreliable tree detections might reduce the performance of the proposed algorithm, but may help to develop a robust approach. The heights of the detected trees are estimated by height normalization using the DTMs. To avoid commission errors, detections with heights below 4 m and with a horizontal distance to the plot center above 33 m are omitted.

3.4. Simulation

Usually, the exact spatial location of a survey plot center is unknown; studies focusing on co-registration [4,17] identify the most probable location by visually interpreting CHMs or aerial images. An objective expert judgment can be severely affected by the individual experience of the interpreter, data-quality, and forest characteristics. Since automated co-registration methods optimize

the accordance between the surveyed data (e.g., tree heights) and remote sensing data (e.g., CHM), an operator will unintentionally tend to the solution proposed by the algorithm. Thus, a posteriori validation might lead to a bias towards the proposed solution [17], which leads to an overestimation of the algorithm performance. If ambiguous sample plots are excluded from validation (as in [4,17]), a further source of overestimation is introduced, since the algorithm is only assessed for simple cases. Even if the correct plot location has been unambiguously identified, a comparison to a location identified by an algorithm is problematic because of remaining positional errors.

Olofsson et al. [20] already introduced an alternative and more promising approach for validating the co-registration and tree matching results by simulating forest stands. Since the corresponding tree pairs are known, the correctness of the results can be tested unambiguously. Based on the need for synthetic training and validation data, the study design illustrated in Figure 2 has been elaborated.
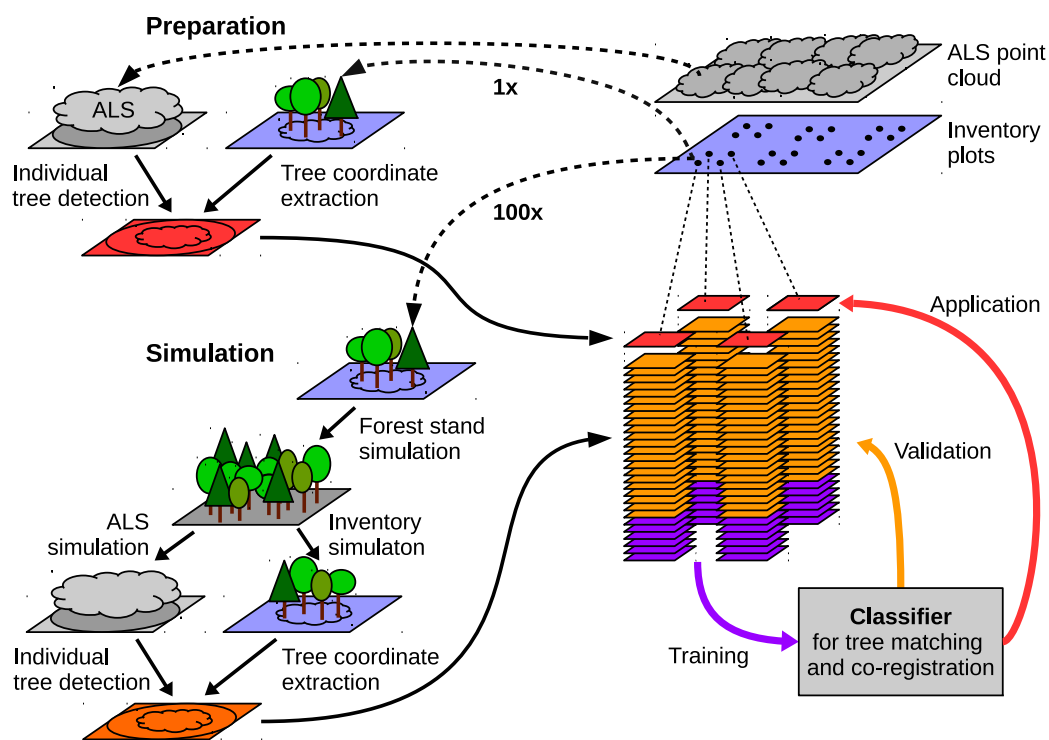
Figure 2. Study design. For each inventory plot, 100 simulations are generated which serve for algorithm training and validation. Finally, the algorithm is applied to the original inventory plots. ALS: airborne laser scanning.

3.4.1. Forest Stand Simulation

For each of the BWI sample plots within our study area, we simulated one hundred synthetic forest stands with a circular area of 0.7 ha using the *Waldplaner 2.0* [38] software. The *Waldplaner* software generates simulated forest stands based on forest inventory data using the single-tree-based forest growth model *treeGROSS* [39]. Based on the inventory data from a sample plot, *treeGROSS* estimates species-specific stem number and diameter distributions that serve as an initialization of a simulated forest stand. Subsequently, parameters such as DBH, tree height, crown base height, and width are then predicted for each individual tree using allometric functions. Since the small number of trees allows no conclusion about the actual spatial distribution, uniformly-distributed initial tree locations were assumed. The spatial location of each tree was finally modeled by taking crown competitions between the simulated tree individuals into account. For the subsequent accuracy assessment of the proposed approach, each tree was finally labeled with a unique ID.

3.4.2. Synthetic Inventory Data

We derived synthetic inventory data in each simulated stand by applying the BWI sampling technique after randomly relocating (GNSS error) the plot center. Each tree is labeled with the ID taken from the synthetic stand. Based on literature research on GNSS accuracy (Section 1.1) and a preliminary visual appraisal of GNSS errors occurring for the investigated BWI plots, we assumed normally-distributed planar GNSS errors with a standard deviation of 8 m. To reflect the specifications of [40], we simulated inaccuracies of the tree coordinates by adding normally-distributed residuals with a standard deviation of 1.5% of planar distance to the plot center, as well as compass errors with a standard deviation of 0.8°. We simulated height measurement errors by adding residuals with a standard deviation of 4% of the tree height.

3.4.3. Synthetic Tree Detections

Simulated ALS point clouds were generated to derive synthetic tree detections. Like Frazer et al. [7], we assumed vegetation surfaces impervious to the laser beam, a nadir orientation, and an infinitely small footprint. The point clouds (Figure 3) were created by modeling the light crown of each simulated tree according to relationships published by Pretzsch [41].

A tree crown is defined by the tree species-specific parameters b and c (see Table 2), the tree height h, crown base height cbh, and its crown radius r (all provided by the *Waldplaner 2.0* software, [38]). The height above ground e_i of a point with distance d_i to the tree location less than or equal to r is calculated by Equation (2). Please note that since the parameters of the Douglas fir were not given by Pretzsch [41], these were assumed to be similar to the parameters of the Norway spruce.

$$e_i = h - c \cdot (h - cbh) \cdot \left(\frac{d_i}{r}\right)^{\frac{1}{b}} \tag{2}$$

For each plot, we generated uniquely-distributed xy-coordinates with the same pulse density as the original point cloud. For each point, the z-coordinate was calculated by applying Equation (2) for all trees and adding the value of the DTM (of the original ALS point cloud) at the tree location. Finally the point is linked to the tree which resulted in the maximum z-coordinate and—for a subsequent evaluation—it is labeled with the ID of that tree.

Since the use of crowns shape functions might result in well-shaped crowns, which would allow an unambiguous co-registration, we simulated irregularities by adding residuals. After a coarse manual optimization (with the aim of realistic tree detection characteristics), we chose normally-distributed horizontal residuals with a standard deviation of 1.5 m and gamma-distributed vertical residuals with a shape of 3 and a scale of 0.3 m.

Table 2. Tree species-specific light crown model parameters according to Pretzsch [41]. Values indicated in italics were based on assumption, since they were not given by Pretzsch [41]. Reproduced with permission from Eugen Ulmer Verlag, 2017.

Species	c	b
Quercus petraea (sessile oak)	0.50	0.50
Fagus silvatica (European beech)	0.40	0.33
Abies alba (European fir)	0.50	0.50
Alnus glutinosa (European alder)	0.56	0.50
Picea abies (Norway spruce)	0.66	1.00
Pseudotsuga menziesii (Douglas fir)	*0.66*	*1.00*
Pinus silvestris (Scots pine)	0.64	0.50
Larix decidua (European larch)	0.80	0.45

(a)   (b)

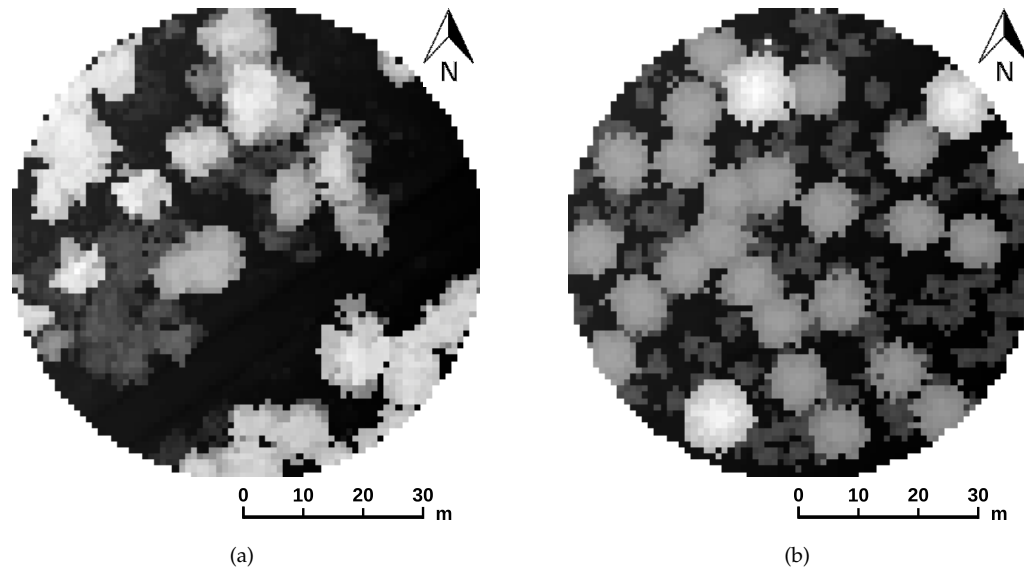**Figure 3.** (**a**) Height grids of the original airborne laser scanning (ALS) point cloud and (**b**) a corresponding synthetic point cloud with 1 m resolution. The given stand is characterized by predominant oaks, some tall spruces and young beeches. The median height of the detected trees is about 24 m.

Finally, the tree detection method presented in Section 3.3 was applied to each synthetic point cloud. To allow an accuracy assessment, each detected tree was labeled with its tree ID taken from the point cloud. If a tree was detected multiple times, only the detection with the tallest z-coordinate was labeled.

3.5. Simulation Quality Assessment

When using simulations for algorithm training and validation, it must be considered that characteristics such as the spatial distribution of the trees can have severe impacts on a successful registration and the transferability of the method. Therefore, a high similarity between the actual and simulated forest stands has to be guaranteed. Assuming full comparability between the synthetic and the original datasets, similar point statistics should be achieved. Consequently, we evaluated the number of trees, the number of detections per reference tree, and the mean nearest-neighbor distance $\overline{NND}$ (Equation (3)) for all training datasets overall and on the plot level.

$$\overline{NND_A} = \sum_{\substack{d \in NND_a \\ a \in A}} \frac{d}{|A|}$$

$$NND_a = \min_{\substack{a \in A \\ b \in A \backslash \{a\}}} \|b - a\|_2$$

(3)

To assess the suitability of the simulated datasets, various statistical analyzes of the previously mentioned variables have been performed. To decide if the values between the original and simulated datasets generally differ, the two sample Wilcoxon rank sum test (comparison of medians) and the F-test (comparison of variances) have been performed. To evaluate the correlation between the original and the simulated values on plot level, the Pearson's correlation coefficient has been calculated. To investigate the overall agreement on plot level, the RMSE has been calculated and a paired Wilcoxon rank sum test has been applied for each plot. We have chosen the nonparametric Wilcoxon test to not rely on normally distributed data.

3.6. Classification-Based Tree Matching and Co-Registration

We developed a tree-matching and co-registration method which consists of two major components.

1.   Classification-based estimation of the matching probability for each potential tree pair.
2.   Co-registration of the survey trees based on the estimated matching probabilities.

Since both components require linking surveyed trees and detected trees, we firstly define a generic point assignment process. We also define a method for pure distance based tree assignment.

3.6.1. Point Assignment Process

The generally defined point assignment process assigns unique pairs of two n-dimensional point sets A and B. To assign the point pairs hierarchically, we define some kind of weight $\omega_{a,b} \in \mathbb{R}_{\geq 0}$ for each potential point pair $(a,b) \in (A \times B)$. The point assignment is performed in ascending order of these weights, which results in a finite sequence $r^{\omega(A,B)}$ of $|r^{\omega(A,B)}|$ point tuples (Equation (4)). In particular, we define the subsequence $r_{\leq k}^{\omega(A,B)}$ (Equation (5)), which comprises at most k point pairs.

$$r^{\omega(A,B)} = \left(r_i^{\omega(A,B)} \right)_{i \in \{1,\dots,|r^{\omega(A,B)}|\}}$$

$$|r^{\omega(A,B)}| = \sum_{i=1;\ r_i^{\omega(A,B)} \neq \varnothing}^{min(|A|,|B|)} 1 \tag{4}$$

$$r_{\leq k}^{\omega(A,B)} = \left(r_i^{\omega(A,B)} \right)_{i \in \{1,\dots,k-1\}} \tag{5}$$

The elements of $r^{\omega(A,B)}$ are recursively defined by Equation (6). Note that this definition requires all weights greater zero to be strictly ordered. For simplification reasons, only the first element (in order of occurrence) is selected in the case of similar weights. Since each point $a \in A$ can only be assigned to at most one point $b \in B$ and only point pairs with weight greater than zero can get assigned, the magnitude $|r^{\omega(A,B)}|$ is at most the minimum of $|A|$ and $|B|$.

$$r_1^{\omega(A,B)} = \underset{(a,b) \in (A \times B)}{argmax}\ \omega_{a,b}$$

$$r_i^{\omega(A,B)} = \underset{(a,b) \in (A \times B)\ |\ \forall (\acute{a},\acute{b}) \in r_{\leq i-1}^{\omega(A,B)}:\ a \neq \acute{a}\ \wedge\ b \neq \acute{b}}{argmax}\ \omega_{a,b} \tag{6}$$

$$\underset{(a,b) \in C}{argmax\ \omega_{a,b}} = \left\{ (a,b) \in C \mid \omega_{a,b} > 0 \wedge \forall\, (\acute{a},\acute{b}) \in C : \omega_{\acute{a},\acute{b}} \leq \omega_{a,b} \right\}$$

3.6.2. Distance-Based Tree Assignment

To assign trees within a given distance and to combine the two-dimensional neighborhood criterion with the height criterion of e.g., [3] we define the ϵ-norm $\| \cdot \|_\epsilon$ (Equation (7)). The key idea of this ϵ-norm is that a surveyed tree $s \in S$ can only be assigned to a detected tree $d \in D$ if $\|d-s\|_\epsilon \leq 1$. From $\|d-s\|_\epsilon \leq 1$ follows $\Rightarrow |d_x - s_x| \leq \epsilon_x$, $|d_y - s_y| \leq \epsilon_y$ and $|d_h - s_h| \leq \epsilon_h$.

$$\|v\|_\epsilon = \sqrt{\left(\frac{v_x}{\epsilon_x}\right)^2 + \left(\frac{v_y}{\epsilon_y}\right)^2 + \left(\frac{v_h}{\epsilon_h}\right)^2} \tag{7}$$

Based on this ϵ-norm, we define a tree assignment method which returns a sequence of linked trees $r^{\omega^\epsilon(S,D)}$ using the weights $\omega^\epsilon(S,D)$ (Equation (8)) as presented in Section 3.6.1. This corresponds to a tree assignment in ascending order of tree pair distance.

$$\omega^\epsilon(A,B) = \left(\omega_{a,b}^\epsilon = \begin{cases} 0, & \text{if } \|b-a\|_\epsilon > 1 \\ \frac{1}{\|b-a\|_\epsilon}, & \text{if } a \neq b \\ \infty, & \text{otherwise} \end{cases} \right)_{a\in A;\ b\in B} \tag{8}$$

3.6.3. Classification-Based Tree Matching

Following the idea of feature descriptors, the proposed tree matching method derives features which have the potential to classify each potential tree pair $(s,d) \in (S \times D)$ of surveyed trees S and detected trees D as either *Matching* or *Not matching*. These features serve as input for a Random Forest [42,43] classifier to estimate the probability $p_{s,d}$ of the class *Matching* (hereafter "matching probability").

Caused by GNSS errors, a shift between the surveyed trees S and detected trees D has to be assumed. Thus, the displacement vector between both point sets S and D has to be estimated. Assuming a surveyed tree $s \in S$ corresponds to a detected tree $d \in D$, the displacement vector is approximately $d-s$, and we can perform a GNSS correction by applying $T_{d-s}S$. Since we can assume displacements similar to the GNNS errors, the probability $p_{s,d}$ should decrease with increasing values for $|d_x - s_x|$ and $|d_y - s_y|$.

Further, we assume the surveyed tree s matches the detected tree d, which results in almost similar point patters $T_{d-s}S$ and D. Thus, we can find a set of k neighbored tree pairs $r_{\leq k}^{\omega^F(T_{d-s}S,D)}$ which is characterized by point pair distances $\|r^{\omega^F(T_{d-s}S,D)}\|_2$ (Equation (10)) close to zero. Note that the weights $\omega^F(T_{d-s}S, D)$ (Equation (9)) result in a tree pair assignment in ascending order of distance. Missing tree pairs (which occur if $k > min(|S|, |D|)$) are indicated with infinite weights. Assuming s does not match d, then we would expect larger point pair distances, because here the point patterns will most likely differ. Thus, the point pair distances $\|r^{\omega^F(T_{d-s}S,D)}\|_2$ serve as a sequence of features to indicate similar point patterns.

$$\omega^F(A,B) = \left(\omega_{a,b}^F = \begin{cases} \frac{1}{\|b-a\|_2}, & \text{if } a \neq b \\ \infty, & \text{otherwise} \end{cases} \right)_{a\in A;\ b\in B} \tag{9}$$

$$\|r^{\omega(A,B)}\|_2 = \left(\|r_i^{\omega(A,B)}\|_2 \right)_{i\in\{1,\dots,|r^{\omega(A,B)}|\}}$$
$$\|r_i^{\omega(A,B)}\|_2 = \begin{cases} \infty, & \text{if } r_i^{\omega(A,B)} = \varnothing \\ \|b-a\|_2 \text{ with } (a,b) = r_i^{\omega(A,B)}, & \text{otherwise} \end{cases} \tag{10}$$

For a given sequence of matching tree pairs r, an almost one-to-one correlation between the tree heights $h_S(r)$ and $h_D(r)$ (Equation (11)) should be observable. Mismatched trees should lead to uncorrelated tree heights. Thus, the Pearson's correlation coefficient ρ of these heights serves as an additional feature of the pair (s,d).

$$h_S(r) = (s_h)_{(s,d)\in r}$$
$$h_D(r) = (d_h)_{(s,d)\in r} \tag{11}$$

In addition to these indicators for similar point patterns, additional features are derived which give information about uncertainties associated with a potential tree pair. Since the probability of identifying a correct tree pair by chance decreases with an increasing number of potential tree pair combinations, the number of trees $|S|$ and $|D|$ serve as additional features. Since close-standing trees complicate the unambiguous identification of matching tree pairs, the average nearest-neighbor distances $\overline{NND_S}$ and $\overline{NND_D}$ (Equation (3)) are calculated as an indicator of tree separability.

Based on the given features, a feature vector $f_{s,d}$ (Equation (12)) is calculated for each potential tree pair $(s,d) \in (S \times D)$. Figure 4 visualizes the workflow for deriving such a feature vector. It also illustrates how the set S is relocated by applying the translation matrix T_{d-s} to assign the neighboring tree pairs $R^{\omega^F}(T_{d-s}S, D)$.

$$
f_{s,d} \atop s\in S;\, d\in D
=
\begin{pmatrix}
\|d - s\|_2 \\
d_x - s_x \\
d_y - s_y \\
d_h - s_h \\
\|r_2^{\omega^F(T_{d-s}S,D)}\|_2 \\
\vdots \\
\|r_k^{\omega^F(T_{d-s}S,D)}\|_2 \\
|S| \\
|D| \\
\overline{NND_S} \\
\overline{NND_D} \\
\rho_{h_S\left(r_{\leq k}^{\omega^F(T_{d-s}S,D)}\right),\, h_D\left(r_{\leq k}^{\omega^F(T_{d-s}S,D)}\right)}
\end{pmatrix}
\tag{12}
$$
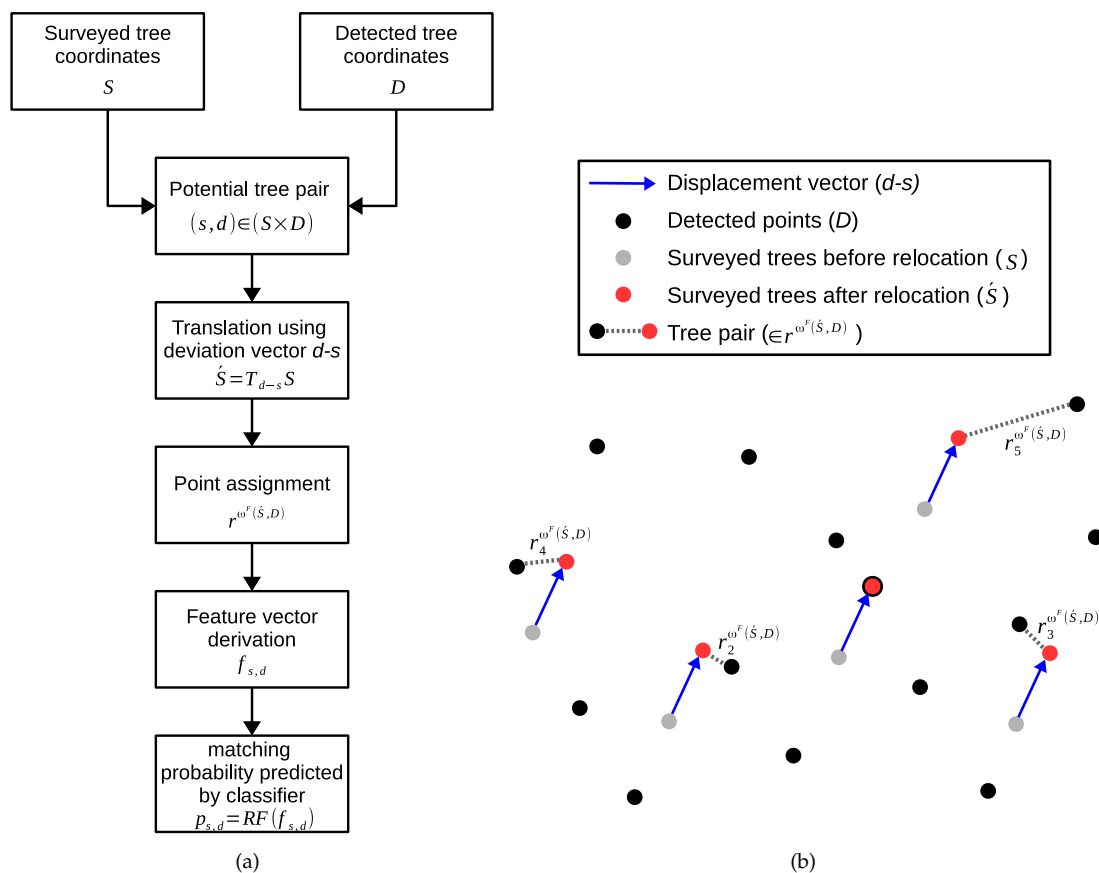
Figure 4. (**a**) Flowchart of the matching probability estimation for a given potential tree pair (s,d) and (**b**) corresponding schematic illustration of the feature vector calculation. *RF*: random forest classifier.

The classifier needs to be trained using a series of representative datasets with known matching tree pairs. For each dataset (with S and D) and each potential tree pair $(s,d) \in (S \times D)$, the feature

vector $f_{s,d}$. The class information is taken from the tree ID labels mentioned in Sections 3.4.2 and 3.4.3. Thus, a tree pair is classified as *Matching* if the ID of s corresponds to the ID of d, and as *Not matching* otherwise. Finally, all of these feature vectors and class information are used to train the random forest classifier *RF*. After this step, it is able to predict the matching probability $p_{s,d}$ (Equation (13)) for each potential tree pair $(s,d) \in (S \times D)$ for two tree sets of surveyed trees S and detected trees D. Finally the tree pair (s,d) is classified as *Matching*, if the matching probability exceeds 0.5, but *Not Matching* otherwise.

$$\underset{s \in S;\ d \in D}{p_{s,d}} = RF(f_{s,d}) \tag{13}$$

3.6.4. Co-Registration Method

Since the tree matching method presented in Section 3.6.4 only gives information about the estimated matching probability for single tree pairs, a further method is required which finally identifies the most probable set of tree pairs r_{after}.

The proposed co-registration algorithm initially selects all candidate pairs $(s,d) \in (S \times D)$ with a matching probability $p_{s,d}$ (Section 3.6.3) greater than zero. For each of these candidate pairs, the tree set S is shifted by the preliminary displacement vector $d - s$ to derive the tree pair sequence $r^{\omega^C(T_{d-s}S, D)}$. The weights $\omega^C(T_{d-s}S, D)$ (Equation (14)) correspond to a distance-based tree assignment (Section 3.6.2) in ascending order of matching probability.

$$\omega^C(A, B) = \left(\omega^C_{a,b} = \begin{cases} 0, & \text{if } \|b - a\|_\epsilon > 1 \\ p_{a,b}, & \text{otherwise} \end{cases} \right)_{a \in A;\ b \in B} \tag{14}$$

As such a tree pair sequence might include pairs with low probabilities, the pair with lowest probability is omitted iteratively until the probability $p^{correct}(r)$ (Equation (15)) of this subsequence r exceeds 0.5. If this threshold is not exceeded, the subsequence with highest probability is kept.

$$p^{correct}(r) = p^{total}_{\geq 50\%}(\{p_{a,b} \mid (a,b) \in r\})$$
$$p^{total}_{\geq 50\%}(P) = \sum_{\substack{\hat{P} \subset P \\ 2 \cdot |\hat{P}| \geq |P|}} \prod_{p \in \hat{P}} p \prod_{q \in \hat{P} \setminus Q} (1 - q) \tag{15}$$

Finally, the point pair sequence or subsequence $r_{after}(S, D)$ with highest achieved co-registration probability $p^{correct}(r_{after}(S, D))$ is returned by the algorithm. The result is classified as *Co-Registration successful*, if the co-registration probability exceeds 0.5, but *Co-Registration failed* otherwise. The final displacement vector $\vec{t}_{after}(S, D)$ (Equation (16)) is calculated by coordinate weighting, which results in a predicted GNSS error of $|\vec{t}_{after}(S, D)|$ m.

$$\vec{t}_{after}(S, D) = \sum_{(s,d) \in r_{after}(S,D)} p_{s,d} \cdot (d - s) \tag{16}$$

3.7. Application

We implemented the co-registration and point matching method using *python 2.7.6* and standard parametrization of the *scikit-learn* random forest implementation [44,45] with 50 decision trees, which enables a resolution of the predicted probabilities of 2%; 25% and 75% of the simulated datasets served for algorithm training and for accuracy assessment, respectively.

As more tree pairs do not necessarily lead to better results, we set the parameter k to 15. The suitability of this parameter has been subsequently confirmed, since not more than 14 tree pairs have been assigned. To define the ϵ-norm (Equation (7)), the distances between the synthetic inventory trees and their corresponding detections were analyzed. Based on these results and values reported in

literature (Section 1.2.2), we set ϵ_x and ϵ_y to 3.0 m and ϵ_h to 4.5 m. These parameters ensure that more than 99% of the these pairs are characterized by an ϵ-norm lower than one.

For each of these datasets (with inventory trees *S* and detected trees *D*), the co-registration method presented in Section 3.6.4 has been applied. To rate the improvement by relocating the plot centers compared to a naive tree assignment (without considering GNSS errors), we additionally applied the pure distance-based tree assignment method (as presented in Section 3.6.2) using the same ϵ-norm. This results in the sequence of potential tree pairs $r_{before}(S, D)$ (Equation (17)) and the probability for at least 50% matching tree pairs $p^{correct}(r_{before}(S, D))$. If this probability exceeds 0.5, the sequence $r_{before}(S, D)$ is classified as *correct*.

$$r_{before}(S, D) = r^{\omega^\epsilon(S,D)} \tag{17}$$

Finally, we also applied both approaches (with and without GNSS correction) to the original BWI datasets.

3.8. Methods for Accuracy Assessment

To assess the accuracy of the proposed algorithm, confusion matrices were derived using the synthetic validation datasets. Each tree pair (s, d) (assigned by the algorithm) was classified as either *Matching* or *Not Matching* (predicted class). The actual class was set to *Matching*, if *s* actually corresponds to *d*, but *Not Matching* otherwise. A co-registration result was classified by the algorithm as either *Co-Registration successful* or *Co-Registration failed* (predicted class). The actual class was set to *Co-Registration successful*, if at least 50% of the tree assigned tree pairs actually match, and *Co-Registration failed* otherwise.

If the probabilities predicted by the algorithm (Sections 3.6.3 and 3.6.4) are in-fact probabilities, we would expect a one-to-one correlation to observed probabilities. To test this, the predicted probabilities for tree matching and co-registration were derived by applying the algorithm using all synthetic validation datasets. Then, equidistant intervals were defined to calculate the average predicted probability and the average observed probability (correct classifications per number of classifications) within each interval. Finally, a regression line was fitted for both tree matching and co-registration.

3.9. Methods for the Evaluation of Feature Importance and Effects

To verify the expectations towards the features, we have analyzed the co-registration results of one randomly selected simulation round. To analyze the effect of a specific feature on the matching probabilities, we have subdivided the results of all plots into three groups. The groups have been defined by the intervals 0–0.33 (low probability), 0.33–0.67 (intermediate probability) and 0.67–1 (high probability). Based on these groups a boxplot has been generated for each feature, which allows conclusions about which values are associated with higher or lower matching probabilities.

To investigate the effect of different forest characteristics on the co-registration performance, we have evaluated the results of all validation datasets. The effect of a selected variable (e.g., tree species) on the co-registration probability has been assessed by grouping the results using equidistant intervals of the variable. Based on these groups a boxplot of the selected variable has been created and a statistical analysis has been performed to support the conclusions. To decide if the values of a given variable differ between the groups (meaning at least one group differs significantly from another group), we have applied the Kruskal–Wallis rank sum test. In case of just two groups the Wilcoxon rank sum test has been used instead, which also gives information about the direction of the effect. In case of ordered groups, we have also verified the significance of potential (linear) trends by performing an one-way ANOVA.

4. Results and Discussion

4.1. Quality of the Synthetic Datasets

Using the synthetic forest stand data, the individual tree detection method achieved an average commission error (number of unassigned detections per number of detections) of 11% and an average omission error (number of missed trees per number of reference trees) of 72%. The planar RMSE of the tree positions was about 0.75 ± 0.39 m, and the RMSE of the heights was about 1.21 ± 0.55 m. These accuracy descriptors are consistent with most of the methods benchmarked in [3,11], which indicates an appropriate modeling of the synthetic point clouds. The increased omission rate might be caused by the simplistic detection method and the bias to dominant trees.

Table 3 summarizes the results of the statistical evaluation of the synthetic training data compared to the original data (overall and on plot level).

Table 3. Evaluation results of the synthetic training data S compared to the original data O. NND: nearest-neighbor distance.

Attribute	Overall			On Plot Level		
	Characteristic Values (Minimum, Mean, Maximum)	Two Sample Wilcoxon Test	F-Test	Pearson's Correlation Coefficient	RMSE	Paired Wilcoxon Test
Amount of Surveyed Trees	$0_{min} = 2; \overline{O} = 7.6; O_{max} = 16$ $S_{min} = 2; \overline{S} = 7.6; S_{max} = 22$	$\overline{S} = \overline{O}$ $(p = 0.66)$	$\sigma_S = \sigma_O$ $(p = 0.47)$	0.98	0.69	$\overline{S - O} > 0$ $(p = 0.00)$
Amount of Detected Trees	$0_{min} = 13; \overline{O} = 46.9; O_{max} = 83$ $S_{min} = 9; \overline{S} = 46.9; S_{max} = 104$	$\overline{S} = \overline{O}$ $(p = 0.50)$	$\sigma_S > \sigma_O$ $(p = 0.00)$	0.64	15.00	$\overline{S - O} = 0$ $(p = 0.74)$
Amount of Detected Trees per Survey Tree	$0_{min} = 1.6; \overline{O} = 7.9; O_{max} = 32$ $S_{min} = 1.1; \overline{S} = 7.6; S_{max} = 49.5$	$\overline{S} = \overline{O}$ $(p = 0.69)$	$\sigma_S = \sigma_O$ $(p = 0.91)$	0.83	3.30	$\overline{S - O} = 0$ $(p = 0.33)$
Mean NND for Surveyed Trees (m)	$0_{min} = 0.5; \overline{O} = 4.2; O_{max} = 13.8$ $S_{min} = 1.1; \overline{S} = 5.6; S_{max} = 27.2$	$\overline{S} > \overline{O}$ $(p = 0.00)$	$\sigma_S > \sigma_O$ $(p = 0.01)$	0.81	1.90	$\overline{S - O} < 0$ $(p = 0.00)$
Mean NND for Detected Trees (m)	$0_{min} = 4.1; \overline{O} = 5.4; O_{max} = 8.2$ $S_{min} = 3.7; \overline{S} = 6.3; S_{max} = 13.0$	$\overline{S} > \overline{O}$ $(p = 0.00)$	$\sigma_S > \sigma_O$ $(p = 0.00)$	0.58	1.36	$\overline{S - O} < 0$ $(p = 0.00)$

The test results for the amount of trees and the amount of detections per reference tree indicate realistic stand densities. The increased average NNDs of the synthetic datasets indicate an overestimation of the tree distances. This effect might be caused by the fact that the original stands are more clustered (e.g., because of aisles or leaning trees), while the simulated trees are distributed more uniformly (with consideration of tree competition). This effect also leads to an incomplete coverage of the value range for this attribute.

As lower NNDs reduce the separability of points, the synthetic inventory plots might be easier to co-register than the original ones. Nevertheless, since the variance within the synthetic datasets is greater than within the original datasets and the value ranges of the other attributes are completely covered, the simulated datasets can be assumed to be appropriate. Thus, by training the algorithm with the synthetic datasets, it should be applicable to the original datasets without losing accuracy.

4.2. Accuracy Assessment

The algorithm for individual tree matching (Section 3.6.3) achieved an overall accuracy of 89.1% and a Cohen's kappa of 0.78 using the synthetic validation datasets (Table 4). The number of matching trees tended to be slightly underestimated, which resulted in a reduced producer's accuracy.

The proposed co-registration method achieved an overall accuracy of 82.7% and a Cohen's kappa of 0.70 using the synthetic validation datasets (Table 5).

Table 4. Confusion matrix for tree matching using synthetic validation data.

		Actual Class		Totals	Users's Accuracy
		Not Matching	Matching		
Predicted Class	Not Matching	12,747	1663	14,410	88.5%
	Matching	683	7662	8345	91.8%
	Totals	13,430	9325	22,755	
	Producer's accuracy	94.9%	82.2%		

Table 5. Confusion matrix for co-registration using synthetic validation data.

		Actual Class		Totals	Users's Accuracy
		Co-Registration Failed	Co-Registration Successful		
Predicted Class	Co-Registration failed	1731	595	2326	74.4%
	Co-Registration successful	890	5377	6267	85.8%
	Totals	2621	5972	8593	
	Producer's accuracy	66.0%	90.0%		

Figure 5 illustrates the correlation between the predicted probabilities and observed probabilities (correct classifications per number of classifications) within equidistant intervals for tree matching (Figure 5a) and co-registration (Figure 5b), respectively.

For tree matching, the observed probability was underestimated for values above 0.5, which resulted in a conservative assignment of tree pairs. This effect might be caused by involving the characteristics of almost all trees in the decision process. Since the expected one-to-one relationship was not archived, the predicted probabilities might be handled with caution. To derive real probabilities, a correction of the predicted values (e.g., by using the given regression model) would be necessary.

Although the predicted co-registration probability depends on the predicted matching probability, an almost one-to-one relationship was achieved. These results indicate that the algorithm is able to correctly estimate the reliability of a specific co-registration result.
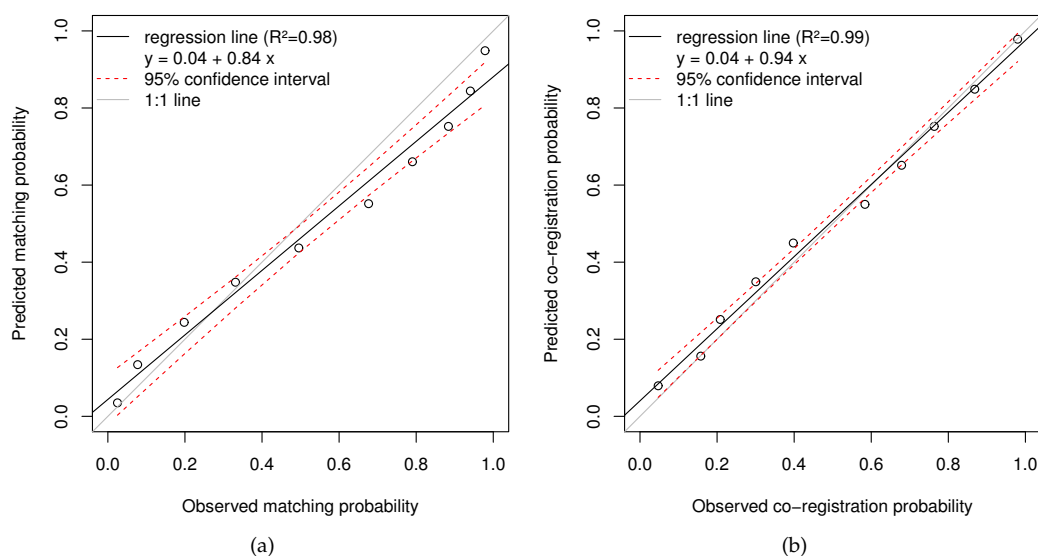
Figure 5. Correlation between predicted probability and observed probability for (**a**) tree matching and (**b**) co-registration.

4.3. Feature Importance and Effects

To analyze the meaningfulness of rules identified by the classifier, we accomplished an evaluation of the feature importance and the effects of the features on the matching probability. In general, the rules identified by the classifier aligned with our previous expectations presented in Section 3.6.3. The results of all features can be found in the supplement of the article.

The high feature importance of the 3D displacement and the vertical displacement go along with our prior expectations. Figure 6a illustrates that the highest probabilities were achieved for vertical displacements below 5 m.
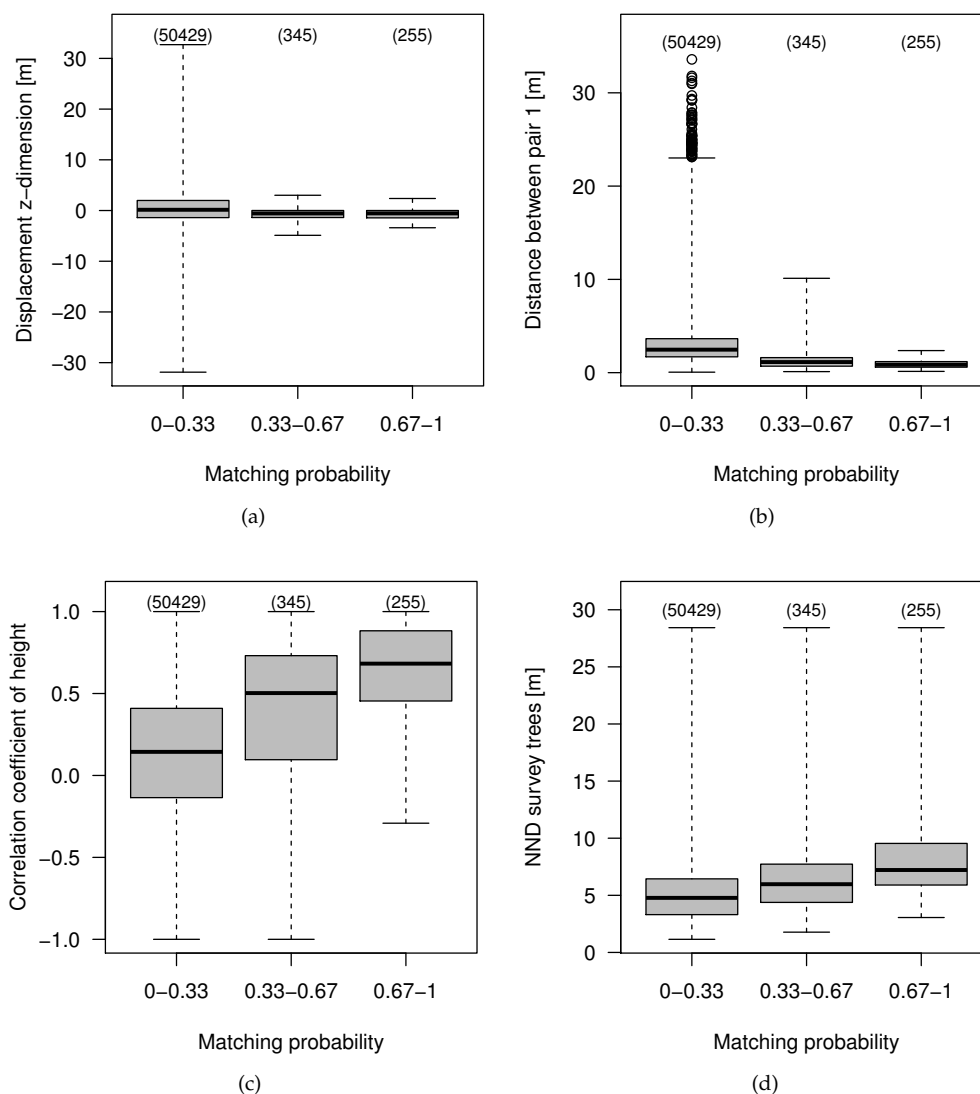
Figure 6. Effect of (**a**) the vertical displacement of the plot center; (**b**) the distance between the first point pair; (**c**) the correlation coefficient for tree height; and (**d**) the NND for surveyed trees on the matching probability (using just one simulation for each plot). The numbers in brackets correspond to the number of tree pairs. The whiskers extend to ten times the interquartile range. Outliers are marked by circles.

Since the distances between the closest tree pairs indicate the agreement of the point patterns, these form the most important group of features. Figure 6b shows that, as expected, low distances to neighbored trees result in high matching probabilities. The ascending importance from the first pair to

the fourth pair and the descending importance from the fourth to the fifteenth pair (Figure 7) can be explained by the influence of the number of linked trees. The co-registration probability increases with an increasing number of linked trees, because the certainty that the point patterns match increases with an increasing number of (probably correct) linked trees.
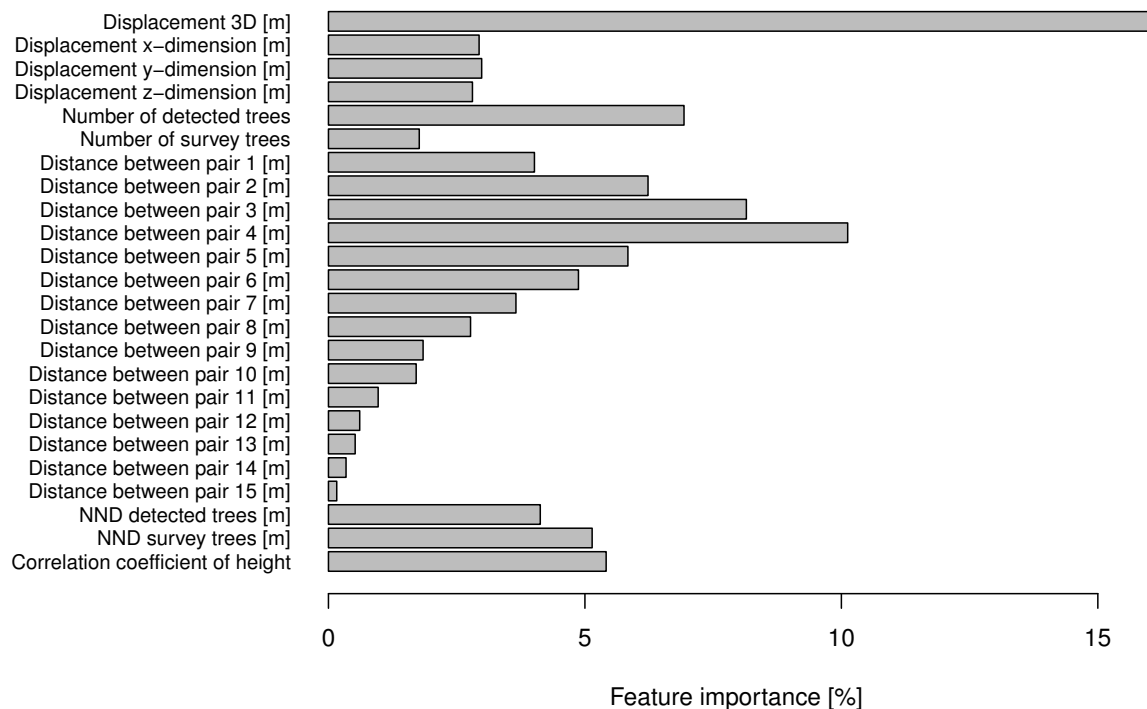
Figure 7. Feature importance derived by the random forest classifier.

The highest matching probabilities are achieved if a positive correlation between the survey tree heights and ALS tree heights occurs (Figure 6c), which goes along with our previous expectations and the heuristic used by Monnet and Mermin [17], Dorigo et al. [4] or Olofsson et al. [20]. Since the NND is an indicator of the tree separability, the matching probability slightly increases with increasing NNDs (Figure 6d).

4.4. Effect of the Data Characteristics on the Co-Registration Results

Since the number of linked trees is an indicator of matching point patterns, the probability of a successful co-registration increased with an increasing number of linked trees (Figure 8a). A saturation effect occurred for five links and more (average co-registration probability above 90%), since not every tree pair has to be investigated to decide if the tree patterns match. As desired (by choosing the given GNSS error distribution), the co-registration probability decreased with increasing horizontal displacements (Figure 8b).

Figure 9 illustrates the effect of different forest types and stand characteristics on the co-registration probability. The Kruskal–Wallis rank sum test indicates significant differences of the probabilities between different tree species ($p = 0.00$). The two-tailed Wilcoxon rank-sum test showed that coniferous trees reached significantly higher probabilities than deciduous trees ($p = 0.00$), which confirms the findings of previous studies [4,17]. Nevertheless, within the conifers, the larches reach lower probabilities than the other conifers. This effect can be explained by the different crown shapes. Cone-shaped crowns result in higher individual tree detection accuracies and higher location accuracies, which eases the co-registration compared to dome-shaped crowns.

An increasing number of tree species (as an indicator of stand variability) results in a positive effect on the co-registration result. The ANOVA of the co-registration probability grouped by the number of species within a plot (Figure 9b) indicates a significant (linear) trend ($p = 0.00$).

(a)

(b)
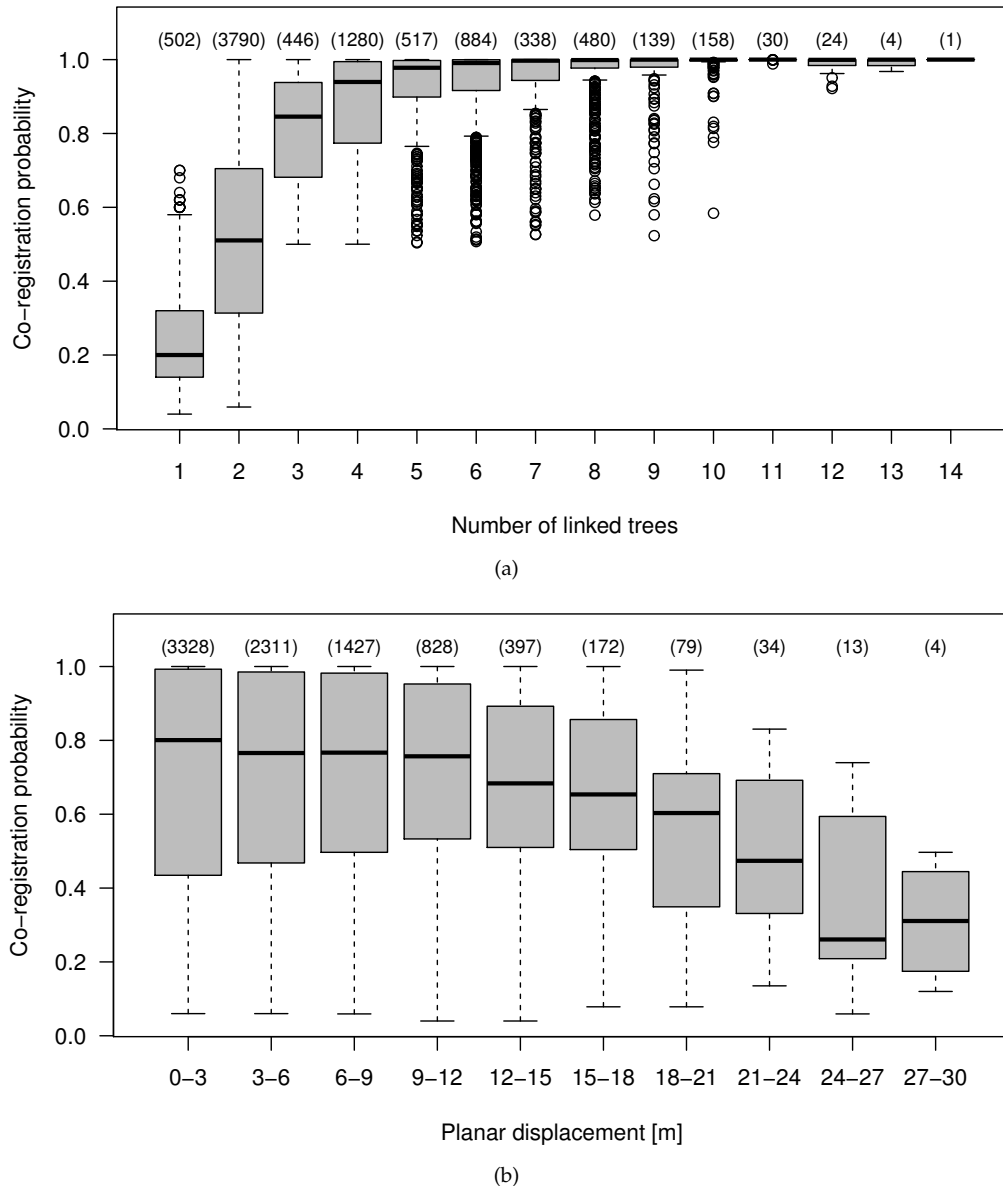
Figure 8. Effect (**a**) of the number of linked trees and (**b**) of the horizontal displacement of the plot center on the co-registration probability. The numbers in brackets correspond to the number of plots. The whiskers extend to 1.5 times the interquartile range. Outliers are marked by circles.

The ANOVA of the co-registration probability grouped by the height of the tallest tree (Figure 9c) also indicates a significant trend ($p = 0.00$). A saturation effect for tall trees can be observed. Additionally, the standard deviation of tree heights (Figure 9d) shows a significant linear trend ($p = 0.00$). With an increasing height of the tallest tree and with an increasing variability of the tree heights, the probability of a correct co-registration also increases, since large values of these variables facilitate the identification of matching tree pairs. These results go along with the findings of Monnet and Mermin [17].
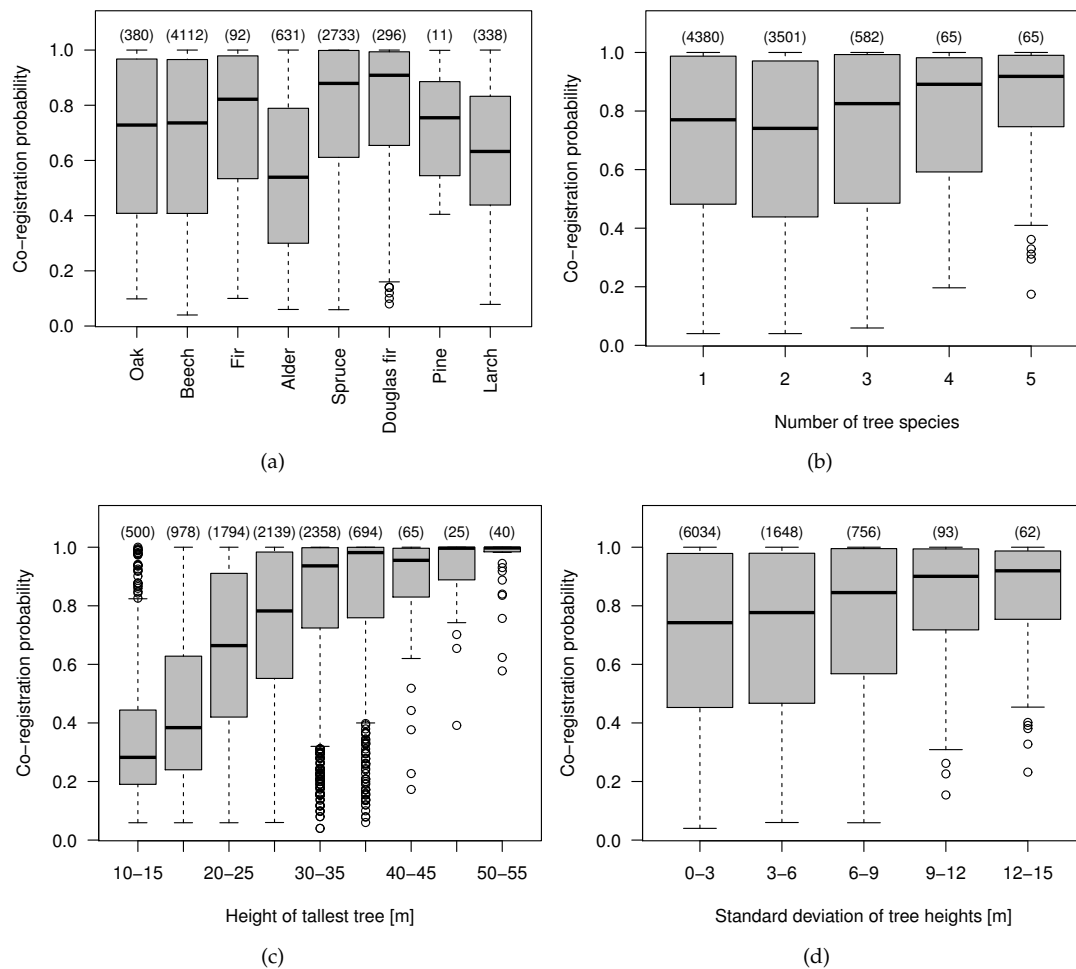
Figure 9. Effect of (**a**) the dominant tree species; (**b**) the number of tree species; (**c**) the height of the tallest tree; and (**d**) the variability of the tree heights on the probability of a correct co-registration. The numbers in brackets correspond to the number of plots. The whiskers extend to 1.5 times the interquartile range. Outliers are marked by circles.

4.5. Method Performance

The proposed method was able to successfully co-register about 69.5% of the synthetic inventory plots, while it linked 81.5% of the tree pairs correctly. By applying a pure distance-based tree assignment (ignoring GNSS errors), only 41.0% of the trees were linked correctly, and the criterion of at least 50% matching tree pairs was fulfilled for only 32.5% of the plots.

The original BWI plots showed similar results. By applying the pure distance-based tree assignment, 288 tree pairs were identified, but only 107 (37%) of these pairs were classified as matching. Only 29 of the plots (22%) fulfilled the criterion of at least 50% tree pairs classified as matching. By applying the proposed co-registration algorithm, 517 tree pairs were assigned. Of these tree pairs, 261 (50%) were classified as matching. The algorithm indicated that 80 of the 133 inventory plots (60%) were co-registered correctly, with 414 assigned tree pairs and 230 (72%) pairs classified as matching. Based on these results, the investigated original BWI plots were characterized by horizontal GNSS errors of up to 21 m, while 80% of the horizontal GNSS errors were in a range between 1.4 and 8.7 m.

Figure 10a illustrates the number of correctly co-registered original BWI plots as a function of the applied co-registration probability. Due to the GNSS correction, the number of co-registered plots increased and higher probabilities were achieved.
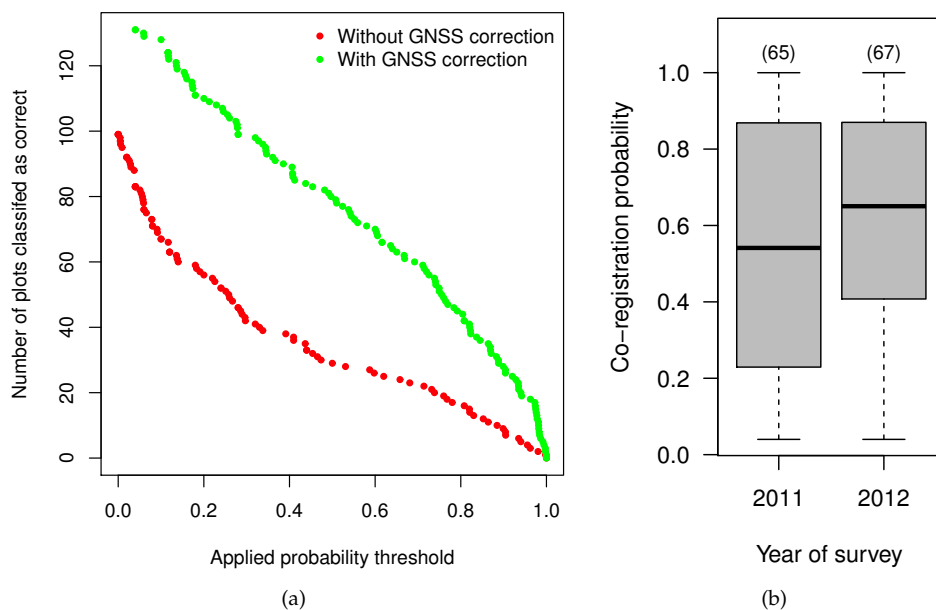
Figure 10. (**a**) Correlation between the applied probability threshold for a correct co-registration and the resulting number plots classified as correctly co-registered. (**b**) Effect of the year of survey on the co-registration probability. Both figures are based on the original BWI plots. The numbers in brackets correspond to the number of plots. GNSS: Global Navigation Satellite System.

Figure 10b illustrates the effect of the year of survey on the co-registration probability. A time-lag between the field survey and the ALS flight campaign (three to four years) might have a negative effect on the co-registration results, since trees might have grown, toppled, or been harvested in this period of time. Although the probabilities of 2011 seem to be slightly lower than the probabilities of 2012, the two-tailed Wilcoxon-test has shown no significantly lower values for 2012 ($p = 0.13$). However, these results indicate that better results could be achievable if the time lag between inventory and ALS data acquisitions is small.

4.6. Limitations

The analysis of the effect of the time-lag between the field survey and the ALS flight campaign indicates that the proposed co-registration method will be challenged if there is major changes between both dates of observation. This is particularly the case if the trees might have grown, toppled, or been harvested within the time lag between field survey and ALS data collection.

Since we needed to model synthetic training data based on various assumptions, the transfer of the probability estimations to the original datasets should be done with caution. The procedure of simulating ALS point clouds using crown shape functions is justified by the observation that the most ambiguous cases for co-registration occur in dense forest stands, where individual trees are hard to detect because of overlapping crowns. Nonetheless, non representative point clouds might affect the tree detection characteristics. Although the suitability of the simulated datasets has been tested, it cannot be ensured that the datasets are representative. This is particularly the case because the synthetic survey plots were characterized by increased NNDs compared to the original BWI data. However, the usage of simulated forest stands leads to more objective validation results compared to reference locations derived by visual interpretation.

Since the number of potential tree pairs and positional inaccuracies complicates the identification of matching tree pairs, the rudimentary tree detection method used in this study might have a negative effect on the algorithm performance. The ANOVA of the co-registration probability grouped by the number of detections per hectare (Figure 11) indicates a significant trend ($p = 0.00$). Since the probability

for a correct co-registration decreases with an increasing number of detections per hectare, commission errors of the detection method should be avoided. Since the angle count sampling technique prefers the selection of dominant trees, and a correct co-registration is possible with just a few trees assigned (Figure 8a), the reliable identification of the dominant trees is more important than the identification of all trees (which might cause increased commission errors).
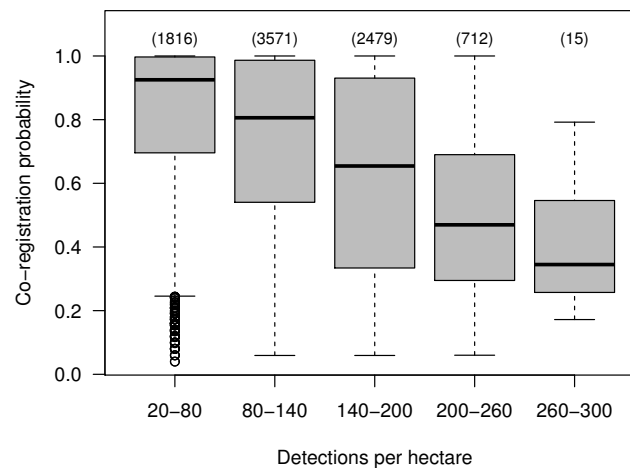

Figure 11. Effect of the number of detections per hectare on the matching probability using the simulated datasets. The numbers in brackets correspond to the number of plots. The whiskers extend to ten times the interquartile range. Outliers are marked by circles.

Another limitation of the method is that no perfect one-to-one correlation between the predicted and observed probabilities could be achieved. Thus, the predicted matching probabilities provided by the algorithm are a higher order reliability estimation but do not necessarily correspond to actual probabilities. However, since the predicted co-registration probabilities rarely differ from the one-to-one correlation, they seem to represent actual probabilities.

The parametrization of the ϵ-norm (defined to assign the tree pairs) should be optimized based on the forest stand characteristics. Since it is only used for the co-registration, it can be adapted for each plot individually after training of the classifier. Using such an optimization might lead to better results compared to the fixed parameters chosen in this study.

5. Conclusions

We performed a co-registration of 133 BWI sampling plots to ALS-derived individual tree detections of a study area in Rhineland-Palatinate, Germany as a preparatory step for a forest characterization on the individual tree level. As erroneous tree pair assignments result in a reduction of model quality (e.g., biomass estimation or tree species classification), we searched for a method for co-registration and tree matching which also rates the reliability of a match.

Since existing methods seemed to be unsuitable for this task, we developed a hybrid co-registration and tree matching algorithm. We trained the machine-learning-based method using synthetic individual tree detections and inventory data, whose representativity has been tested empirically.

The method reached an overall accuracy of 89.7% for tree matching and a user's accuracy for co-registration of 82.7% using simulated datasets. The method has been applied to the study area, and was able to successfully relocate 60% of the BWI plots, which makes these usable for a further analysis on the individual tree level. As machine-learning methods have been proven powerful to identify patterns, a similar performance of the algorithm compared to a human expert can be expected. Since the interpreter might be biased to the proposed solution, a manual post-processing is not expedient.

The derived feature importance depends on forest characteristics and the assumed GNSS error distribution. The feature importance and its effect on the matching probability has shown to be consistent with prior expectations. Thus, the machine-learning approach seems to be an appropriate alternative to heuristic methods, with the additional advantages of an automated parametrization and a robust estimation of the reliability of a single co-registration and tree matching result.

We found that with five or more linked tree pairs, at least 90% of the co-registration results can be expected to be correct. If the number of terrestrial trees is limited, heuristic methods typically cannot comply with this requirement. We found that the highest probabilities for a correct co-registration were achieved in heterogeneous stands (mixed species, differing tree heights, presence of tall trees). Stands dominated by conifers achieved significantly better co-registration results than stands dominated by deciduous trees. These findings support the results of previous studies.

To achieve better results, the local-maxima-based tree detection method used in this study should be replaced by a more advanced method (e.g., by tree stem detection). Since more accurate tree positions result in a better agreement of the point patterns, the pre-trained algorithm can directly be applied to these more reliable positions without losing explanatory power. To transfer the method to forests with different characteristics or a different sampling design, only an adaption of the (simulated) training datasets is required.

The probability estimations provided by the algorithm—as an objective indicator of the reliability of a specific result—lead to a clear added value compared to existing methods. Since it can serve as a weighting-factor for model training, for example, the proposed method is a relevant tool for gaining further knowledge in the field of forest characterization on small scales or even on the individual tree level.

Author Contributions: Andreas Hill and Sebastian Lamprecht initiated the study and conceived the study design. Andreas Hill generated the synthetic forest stands. Sebastian Lamprecht derived the simulated ALS point clouds, developed and implemented the algorithm, analyzed the data and wrote the paper. Andreas Hill, Johannes Stoffels and Thomas Udelhoven cross-checked the analysis and the manuscript. Thomas Udelhoven led the research group.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ALS	Airborne Laser Scanning
BWI	German National Forest Inventory
CHM	Canopy Height Model
DBH	Diameter at Breast Height
DTM	Digital Terrain Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
NND	Nearest-Neighbor Distance
RMSE	Root Mean Squared Error
RANSAC	RANdom SAmple Consensus
WMS	Web Map Service

References

1. Cao, L.; Coops, N.C.; Innes, J.L.; Dai, J.; Ruan, H.; She, G. Tree species classification in subtropical forests using small-footprint full-waveform LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 39–51.

2.   Zhen, Z.; Quackenbush, L.J.; Zhang, L. Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sens.* **2016**, *8*, 333.

3.   Eysn, L.; Hollaus, M.; Lindberg, E.; Berger, F.; Monnet, J.M.; Dalponte, M.; Kobal, M.; Pellegrini, M.; Lingua, E.; Mongus, D.; et al. A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space. *Forests* **2015**, *6*, 1721–1747.

4.   Dorigo, W.; Hollaus, M.; Wagner, W.; Schadauer, K. An application-oriented automated approach for co-registration of forest inventory and airborne laser scanning data. *Int. J. Remote Sens.* **2010**, *31*, 1133–1153.

5.   Baltsavias, E.P. Airborne laser scanning: Basic relations and formulas. *ISPRS J. Photogramm. Remote Sens.* **1999**, *54*, 199–214.

6.   Gobakken, T.; Næsset, E. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Can. J. For. Res.* **2009**, *39*, 1036–1052.

7.   Frazer, G.; Magnussen, S.; Wulder, M.; Niemann, K. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sens. Environ.* **2011**, *115*, 636–649.

8.   Hyyppä, J.; Yu, X.; Hyyppä, H.; Vastaranta, M.; Holopainen, M.; Kukko, A.; Kaartinen, H.; Jaakkola, A.; Vaaja, M.; Koskinen, J.; Alho, P. Advances in Forest Inventory Using Airborne Laser Scanning. *Remote Sens.* **2012**, *4*, 1190–1207.

9.   Koenig, K.; Höfle, B. Full-Waveform Airborne Laser Scanning in Vegetation Studies—A Review of Point Cloud and Waveform Features for Tree Species Classification. *Forests* **2016**, *7*, 198.

10. Vauhkonen, J.; Ene, L.; Gupta, S.; Heinzel, J.; Holmgren, J.; Pitkanen, J.; Solberg, S.; Wang, Y.; Weinacker, H.; Hauglin, K.M.; et al. Comparative testing of single-tree detection algorithms under different types of forest. *Forestry* **2011**, *85*, 27–40.

11. Kaartinen, H.; Hyyppä, J.; Yu, X.; Vastaranta, M.; Hyyppä, H.; Kukko, A.; Holopainen, M.; Heipke, C.; Hirschmugl, M.; Morsdorf, F.; et al. An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning. *Remote Sens.* **2012**, *4*, 950–974.

12. Wallace, L.; Lucieer, A.; Watson, C.S. Evaluating tree detection and segmentation routines on very high resolution UAV LiDAR data. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7619–7628.

13. Hoppus, M.; Lister, A. The status of accurately locating forest inventory and analysis plots using the Global Positioning System. In Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium, Portland, ME, USA, 3–6 October 2007.

14. Wing, M.G.; Eklund, A.; John, S.; Richard, K. Horizontal measurement performance of five mapping-grade global positioning system receiver configurations in several forested settings. *West. J. Appl. For.* **2008**, *23*, 166–171.

15. Andersen, H.E.; Clarkin, T.; Winterberger, K.; Strunk, J. An accuracy assessment of positions obtained using survey-and recreational-grade global positioning system receivers across a range of forest conditions within the Tanana Valley of interior Alaska. *West. J. Appl. For.* **2009**, *24*, 128–136.

16. Valbuena, R.; Mauro, F.; Rodriguez-Solano, R.; Manzanera, J. Accuracy and precision of GPS receivers under forest canopies in a mountainous environment. *Span. J. Agric. Res.* **2010**, *8*, 1047–1057.

17. Monnet, J.M.; Mermin, É. Cross-correlation of diameter measures for the co-registration of forest inventory plots with airborne laser scanning data. *Forests* **2014**, *5*, 2307–2326.

18. Luoma, V.; Saarinen, N.; Wulder, M.; White, J.; Vastaranta, M.; Holopainen, M.; Hyyppä, J. Assessing Precision in Conventional Field Measurements of Individual Tree Attributes. *Forests* **2017**, *8*, 38.

19. Hollaus, M.; Wagner, W.; Maier, B.; Schadauer, K. Airborne laser scanning of forest stem volume in a mountainous environment. *Sensors* **2007**, *7*, 1559–1577.

20. Olofsson, K.; Lindberg, E.; Holmgren, J. A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs. In Proceedings of the SilviLaser 2008—8th International Conference on LiDAR Applications in Forest Assessment and Inventory, Edinburgh, UK, 17–19 September 2008.

21. Mongus, D.; Žalik, B. An efficient approach to 3D single tree-crown delineation in LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 219–233.

22. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395.

23. Sattler, T.; Leibe, B.; Kobbelt, L. Fast image-based localization using direct 2d-to-3d matching. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.

24. Bienert, A.; Pech, K.; Maas, H.G. Verfahren zur Registrierung von Laserscannerdaten in Waldbeständen—Methods for registration laser scanner point clouds in forest stands. *Schweiz. Z. Forstwes.* **2011**, *162*, 178–185.

25. Besl, P.J.; McKay, N.D. A method for registration of 3-D shapes. *IEEE* **1991**, *14*; 239–256.

26. Rangarajan, A.; Chui, H.; Bookstein, F.L. The softassign procrustes matching algorithm. In *Information Processing in Medical Imaging*; Duncan, J., Gindi, G., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1997; pp. 29–42.

27. Myronenko, A.; Song, X. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2262–2275.

28. Golyanik, V.; Ali, S.A.; Stricker, D. Gravitational Approach for Point Set Registration. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

29. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

30. Castellani, U.; Cristani, M.; Fantoni, S.; Murino, V. Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Comput. Graph. Forum* **2008**, *27*, 643–652.

31. Donoser, M.; Schmalstieg, D. Discriminative feature-to-point matching in image-based localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.

32. Bitterlich, W. *The Relascope Idea. Relative Measurements in Forestry*; Commonwealth Agricultural Bureaux: Farnham Royal, UK, 1984.

33. Kublin, E. Einheitliche Beschreibung der Schaftform—Methoden und Programme -BDATPro. A Uniform Description of Stem Profiles—Methods and Programs -BDATPro. *Forstwiss. Cent.* **2003**, *122*, 183–200.

34. Kublin, E.; Breidenbach, J.; Kändler, G. A flexible stem taper and volume prediction method based on mixed-effects B-spline regression. *Eur. J. For. Res.* **2013**, *132*, 983–997.

35. Copernicus. Available online: http://image.discomap.eea.europa.eu/arcgis/services/GioLandPublic/HRL_Forest_Cover_Type_2012/MapServer/WMSServer?request=GetCapabilities&service=WMS (accessed on 12 April 2017).

36. Langley, R.B. Dilution of precision. *GPS World* **1999**, *10*, 52–59.

37. RIEGL. Available online: http://www.riegl.com/uploads/tx_pxpriegldownloads/10_DataSheet_Q560_20-09-2010_01.pdf (accessed on 27 September 2016).

38. Hansen, J.; Nagel, J. *Waldwachstumskundliche Softwaresysteme auf Basis von TreeGrOSS-Anwendung und Theoretische Grundlagen*; Niedersächsische Staats-und Universitätsbibliothek: Göttingen, Germany, 2014.

39. Nagel, J. Waldwachstumssimulation mit dem Java Software Paket TreeGross—Neuerungen, Erweiterungsmöglichkeiten und Qualitätsmanagement. In Proceedings of the 20th annual DVFFA Conference, Freiburg, Germany, 22–24 September 2016.

40. Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz. *Aufnahmeanweisung für die Dritte Bundeswaldinventur (BWI³): (2011–2012)*; BMELV: Bonn, Germany, 2011; Volume 2.

41. Pretzsch, H. *Modellierung des Waldwachstums*; Parey: Berlin, Germany, 2001.

42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

43. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 1, pp. 278–282.

44. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

45. Scikit Learn. Available online: http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (accessed on 11 November 2016).

Erratum

Erratum: A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data. *Remote Sens.* **2017**, *9*, 505

Sebastian Lamprecht [1,*], Andreas Hill [2], Johannes Stoffels [1] and Thomas Udelhoven [1]

[1]   Remote Sensing & Geoinformatics Department, Trier University, 54286 Trier, Germany; stoffels@uni-trier.de (J.S.); udelhoven@uni-trier.de (T.U.)

[2]   Department of Environmental Systems Science, ETH Zurich, 8092 Zurich, Switzerland; andreas.hill@usys.ethz.ch

*   Correspondence: lamprecht@uni-trier.de; Tel.: +49-651-201-4612

Since Equation (2) has been rearranged incorrectly during preparation for this article [1], the authors would like to correct the relevant text of Section 3.4.3 as follows:

3.4.3. Synthetic Tree Detections

A tree crown is defined by the tree species-specific parameters b and c (see Table 2), the tree height h, crown base height *cbh*, and its crown radius r (all provided by the *Waldplaner 2.0* software, [2]). The height above ground e_i of a point with distance d_i to the tree location less than or equal to r is calculated by Equation (2). Please note that since the parameters of the Douglas fir were not given by Pretzsch [3], these were assumed to be similar to the parameters of the Norway spruce.

$$e_i = h - c \cdot (h - cbh) \cdot \left(\frac{d_i}{r}\right)^{\frac{1}{b}} \tag{2}$$

For each plot, we generated uniquely-distributed xy-coordinates with the same pulse density as the original point cloud. For each point, the z-coordinate was calculated by applying Equation (2) for all trees and adding the value of the DTM (of the original ALS point cloud) at the tree location. Finally, the point is linked to the tree which resulted in the maximum z-coordinate and—for a subsequent evaluation—it is labeled with the ID of that tree.

We apologize for any inconvenience caused to the readers by these changes. The changes do not affect the scientific results. The manuscript will be updated and the original will remain online on the article webpage.

References

1.   Lamprecht, S.; Hill, A.; Stoffels, J.; Udelhoven, H. A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data. *Remote Sens.* **2017**, *9*, 505.
2.   Hansen, J.; Nagel, J. *Waldwachstumskundliche Softwaresysteme auf Basis von TreeGrOSS-Anwendung und Theoretische Grundlagen*; Niedersächsische Staats-und Universitätsbibliothek: Göttingen, Germany, 2014.
3.   Pretzsch, H. *Modellierung des Waldwachstums*; Parey: Berlin, Germany, 2001.

Chapter IV

Pyoints: A Python Package for Point Cloud, Voxel and Raster Processing

S. Lamprecht

JöSS
The Journal of Open Source Software

Pyoints: A Python package for point cloud, voxel and raster processing.

Sebastian Lamprecht[1]

1 Trier University

Summary

The evolution of automated systems like autonomous robots and unmanned aerial vehicles leads to manifold opportunities in science, agriculture and industry. Remote sensing devices, like laser scanners and multi-spectral cameras, can be combined with sensor networks to all-embracingly monitor a research object.

The analysis of such big data is based on geoinformatics and remote sensing techniques. Today, next to physically driven approaches, machine learning techniques are often used to extract relevant thematical information from the data sets. Analysis requires a fusion of the data sets, which is made difficult conceptually and technically by different data dimensions, data structures, and various spatial, spectral, and temporal resolutions.

Today, various software to deal with these different data sources is available. Software like GDAL (GDAL/OGR contributors, 2018) and OpenCV (Bradski, 2000) are intended for image processing. Libraries, like PCL (Rusu & Cousins, 2011), Open3D (Zhou, Park, & Koltun, 2018) and PDAL (PDAL contributors, 2018) focus on 3D point cloud processing. Each of these software packages provide an API specially designed to solve the problems of their field efficiently. When developing algorithms for automated processing of various types of input data, the differing APIs and programming languages of these software packages become a drawback. To support fast algorithm development and a short familiarization, a unified API would be desirable.

Pyoints is a python package to conveniently process and analyze point cloud data, voxels, and raster images. It is intended to be used to support the development of advanced algorithms for geo-data processing.

The fundamental idea of *Pyoints* is to overcome the conceptual distinction between point clouds, voxel spaces, and rasters to simplify data analysis and data fusion of variously structured data. Based on the assumption that any geo-object can be represented by a point, a data structure has been designed that provides a unified API for points, voxels, and rasters. Each data structure maintains its characteristic features, to allow for intuitive use, but all data is also considered as a two or three dimensional point cloud, providing spatial indices that are required in many applications to speed up spatial neighborhood queries.

During development, great emphasis was put on designing a powerful but simple API while also providing solutions for most common problems. *Pyoints* implements fundamental functions and some advanced algorithms for point cloud, voxel, and raster data processing, like coordinate transformation, vector algebra, point filters, and interpolation. *Pyoints* also provides a unified API for loading and saving commonly used geo-data formats.

Pyoints was designed to support research activities and algorithm development in the field of geoinformatics and remote sensing. Early versions of the software have been
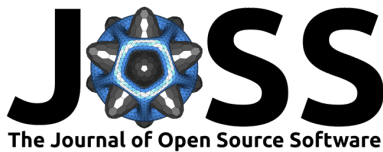
used for some pre-studies at Trier University (Lamprecht, Hill, Stoffels, & Udelhoven, 2017; Lamprecht, Stoffels, Dotzler, Haß, & Udelhoven, 2015). *Pyoints* is also used in the PANTHEON project (PANTHEON consortium, 2018) to monitor hazelnut orchards.

The source code of *Pyoints* is on GitHub (Lamprecht, 2019a). The documentation can be found on GitHub Pages (Lamprecht, 2019b).

Acknowledgements

References

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools.* Retrieved from https://opencv.org

GDAL/OGR contributors. (2018). *GDAL/OGR geospatial data abstraction software library.* Open Source Geospatial Foundation. Retrieved from http://gdal.org

Lamprecht, S. (2019a). Pyoints. doi:10.5281/zenodo.2557574

Lamprecht, S. (2019b). Pyoints documentation. Retrieved April 1, 2019, from https://laempy.github.io/pyoints/

Lamprecht, S., Hill, A., Stoffels, J., & Udelhoven, T. (2017). A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data. *Remote Sensing*, *9*(5), 505. doi:10.3390/rs9050505

Lamprecht, S., Stoffels, J., Dotzler, S., Haß, E., & Udelhoven, T. (2015). ATrunk—an als-based trunk detection algorithm. *Remote Sensing*, *7*(8), 9975. doi:10.3390/rs70809975

PANTHEON consortium. (2018). Precision farming of hazelnut orchards. Retrieved August 10, 2018, from http://www.project-pantheon.eu/

PDAL contributors. (2018). *PDAL: The point data abstraction library.* Retrieved from https://pdal.io/

Rusu, R. B., & Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). *IEEE International Conference on Robotics and Automation*, 1–4. doi:10.1109/ICRA.2011.5980567

Zhou, Q.-Y., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. *arXiv:1801.09847.*

Chapter V

ALS as Tool to Study Preferred Stem Inclination Directions

S. Lamprecht, J. Stoffels, and T. Udelhoven

remote sensing

MDPI

ALS as Tool to Study Preferred Stem Inclination Directions

Sebastian Lamprecht *, **Johannes Stoffels** and **Thomas Udelhoven**

Remote Sensing & Geoinformatics Department, Trier University, 54286 Trier, Germany; stoffels@uni-trier.de (J.S.); udelhoven@uni-trier.de (T.U.)

* Correspondence: lamprecht@uni-trier.de; Tel.: +49-651-201-4612

check for updates

Abstract: Although gravitropism forces trees to grow vertically, stems have shown to prefer specific orientations. Apart from wind deforming the tree shape, lateral light can result in prevailing inclination directions. In recent years a species dependent interaction between gravitropism and phototropism, resulting in trunks leaning down-slope, has been confirmed, but a terrestrial investigation of such factors is limited to small scale surveys. ALS offers the opportunity to investigate trees remotely. This study shall clarify whether ALS detected tree trunks can be used to identify prevailing trunk inclinations. In particular, the effect of topography, wind, soil properties and scan direction are investigated empirically using linear regression models. 299.000 significantly inclined stems were investigated. Species-specific prevailing trunk orientations could be observed. About 58% of the inclination and 19% of the orientation could be explained by the linear models, while the tree species, tree height, aspect and slope could be identified as significant factors. The models indicate that deciduous trees tend to lean down-slope, while conifers tend to lean leeward. This study has shown that ALS is suitable to investigate the trunk orientation on larger scales. It provides empirical evidence for the effect of phototropism and wind on the trunk orientation.

Keywords: LiDAR; ALS; stem detection; tree inclination; phototropism; gravitropism; wind distortion

1. Introduction

1.1. Motivation

The study by Lamprecht et al. [1] has unintentionally revealed a preferred orientation of ALS detected spruce and beech trunks within a 0.5 ha study site. This result raises the question whether ALS could be used to investigate the underlying cause of the orientation of tree trunks on larger scales. If so, tree trunk detection using ALS might be a powerful tool to address the species specific growth and inclination characteristics in response to site conditions, topography and wind.

1.2. Causes for Tree Inclination

Gravitropism describes the gravitative growth characteristics of plants. In particular, the trunk of a tree grows contrarily to the gravitative field, while the roots follow the gravitative field [2]. By asymmetric growth, trees are able to actively control their inclination. Lateral incidence of light, competition for light, continuous wind, or sudden disturbances like landslides, snow break or storms can cause a tree to deviate from gravitropism.

Phototropism causes plants to reorient their growth towards or against a light source [3,4]. Since trees try to maximize light absorption, a lateral incidence of light—as given in the mid-latitudes—is expected to result in trees inclined towards the equator. This effect has been verified for 256 Cook pines located around the globe by Reference [5].

Matsuzaki et al. [6] show that the stem inclination of trees can be actively influenced by the position of a fixed light source. In their experiment 90% of 212 investigated trees were inclined in down-slope direction, while a significant species dependent correlation between phototropism and gravitropism could be observed. The cause of this effect seems to be related to phototropism triggered by an uneven light supply in sloped terrains. In up-slope direction, the hill and the surrounding trees standing higher up the slope tend to cover the incidence of light, while in down-slope direction the lower standing trees facilitate light penetration. Ishii and Higashi [7] postulate that also understory trees adapt their trunk inclination for best possible utilization of the available light. Using a numeric model, they find a significant correlation between slope and trunk inclination for understory trees, but no effect for canopy trees at two test sites of evergreen trees with an area of 0.5 ha and 0.4 ha respectively.

In addition to phototropism, also soil movements and landslides can cause trees to incline with the slope gradient. The soil movements typically result in trees inclined downhill or uphill. Typically the trees try to recover gravitropism by asymmetric growth, resulting in bent stems with eccentric radii [8].

Wind is one of the major sources of mechanical loading on plants [9]. The complex interaction between the forest structure and the terrain results in small scale turbulences preventing a realistic prediction of the wind drag of a specific tree [9,10]. But, in general, a tree does not topple or uproot if the elastic restoring forces of the stem and roots can resist the combined wind and gravitational forces [9]. In consequence, trees grow adaptively and realign their structure to minimize wind drag, but maximize light capture. A continuous wind drag can make a tree realign its foliage permanently, resulting in windswept trees with foliage allocation in the prevailing leeward direction. Windswept crowns reduce the wind drag significantly and reduce the risk of stem or root damage [9,11]. Telewski [12] argue that the windswept growth is more likely the product of biomechanical properties than of a physiological thigmotropic growth response.

Nicoll and Ray [13] observe an adaptive growth of the root system of Sitka spruce in response to wind or site conditions. The roots of 100 randomly selected trees were characterized by an increase of structural root mass on the leeward compared to the windward side. Nicoll and Ray [13] conclude that the resulting asymmetric rooting structure reduces the tree's vulnerability to windthrow. There are also indicators that a large proportion of branches in the total tree mass increases the windthrow risk of trees significantly [14]. In consequence, species with a low allocation of branch mass should be particularly resistant to wind.

Apart from the strength of the trunk, the root plate morphology, soil type and soil moisture define the resilience of a tree to external forces [15]. Also the rooting depth controls the anchoring of a root in the soil [16]. The numeric models of Fourcaud et al. [17] indicate that in clay-like soils, the rooting depth defines the root-soil plate and consequently the anchor strength. In sand-like soils, the shape of the rooting system determines the shape and size of the soil-root plate [17]. The results also indicate that the mechanical stress introduced by wind is highest in the superficial roots and the leeward roots. The soil moisture is an important factor that controls the resistive forces of the roots [10]. Heavy rainfalls accompanying storm events can weaken the anchoring of rooting systems and make trees vulnerable to toppling [10,15].

1.3. Stem Detection Using ALS

In recent years several methods for extracting tree stems using ALS have been developed [1,18–23]. Due to differing objectives, differing ALS acquisition designs and differing investigated forest types and

structures, the accuracy of the trunk detection methods is hard to compare. In comparison to classic ALS driven tree crown delineation (see Reference [24]), all methods are able to extract vertical linear structures with a high reliability, and the precision of the tree positions is high. In general, two types of detection methods can be distinguished, heuristic methods and methods taking advantage of machine-learning techniques. With the downside of requiring training data, methods based on machine learning are expected to be transferable on stands and ALS data with various characteristics. Heuristic methods can be applied without training data, but are typically hard to parameterize and are typically not robust to differing forest characteristics.

The method of Lu et al. [18] takes advantage of the ALS intensity values to isolate points associated with trunks from leaf-off deciduous trees. The extracted points are used for a three-dimensional bottom-up trunk-growing process, while the assignment is controlled by a three dimensional and a planar distance threshold. To avoid false positives, the trunk length and the maximum height of the lowest trunk point are controlled by threshold values. Furthermore the identified trunks are used for crown delineation of 20 plots with a total area of 3.2 ha and an average point density of 10 points/m^2. The method achieves a detection rate of 84% and a precision of 97%.

Lamprecht et al. [1] identify tree trunks by creating ALS point cloud segments using a Divide & Conquer approach. Points associated with tree trunks are isolated from the canopy by estimating the crown base height and 3D-clustering. A principal axis is fitted to each cluster using a deterministic modification of the LO-RANSAC approach. The algorithm provides a vector model for each trunk detection, with attributes like root position, inclination angle and the compass direction. A validation with 109 TLS-measured trees in a plot of 0.5 ha and an average ALS point density of 7.7 points/m^2 has shown a detection rate of 84% and a precision of 95%, while an RMSE of the trunk roots of 0.78 m could be observed.

The bottom-up approach of Shendryk et al. [20] uses full-waveform ALS data to delineate individual trees using tree trunks as seeding points. In extension to Lu et al. [18], next to an intensity filter, the ALS pulse width is used to filter points associated with trunks in range of 1 and 10 m above ground. Finally, a three-dimensional Euclidean clustering is applied to identify individual trunks. Using 38 reference plots within a forest of a complex structure with a total area of 3.4 ha, the algorithm has been applied to ALS data sets of two different point densities. By doubling the point density from 12 points/m^2 to 24 points/m^2, the tree detection rate increased from 56% to 67%, while the precision diminished from 62% to 61%.

Polewski et al. [19] propose free shape context descriptors to detect dead tree trunks in ALS point clouds. After a segmentation of the point cloud, a principal axis is fitted to each cluster using the M-estimator Sample Consensus method. For each axis, 3D free shape context descriptors are derived. After training and optimizing the shape descriptors with a genetic approach on an ALS tile of 1×1 km^2 with a point density of 30 to 40 points/m^2, the algorithm is able to distinguish dead tree trunks from lining trees. The algorithm achieves an accuracy of 84.2% for 208 dead trunks and 340 living trees. The correctness of the principal axis is not discussed by the authors.

Amiri et al. [21] enhance the method of Polewski et al. [19] for identifying individual tree stems using high density ALS point clouds. With a Random Forest Classifier, points associated with tree stems are classified using 3D shape descriptors, covariance features and the normalized height. Initial linear segments are created based on points associated with the stems. These are merged to individual stems by hierarchical clustering. To describe a stem, a line is fitted by orthogonal distance regression and energy minimization. For two plots of 0.7 ha and 1.8 ha and an average point density of 300 points/m^2, the method achieved a classification accuracy of up to 86% for 196 reference stems.

Most recently, Windrim and Bryson [22] publish new approaches to automatically detect trees and reconstruct stems using supervised deep machine learning methods using high resolution ALS point clouds. After detecting individual trees, points associated with the tree stems are extracted based on

machine learning using voxel convolutions or neural networks. Finally segmented models of the main stems are fitted to gain information on the tree height, diameter, taper and sweep. The analysis has been conducted on 25 plots with an ALS point density of 300 points/m^2 up to 700 points/m^2 and 447 reference trees. The proposed methods achieve an overall tree detection accuracy of 77% up to 96% and an IoU (Intersection over Union) for stem detection of up to 52%.

As Chen et al. [23] have shown, trunk detections can be used to improve individual tree segmentation. Similarly to Lamprecht et al. [1] points associated with the tree trunks are identified by estimating the crown base height. Trunks are derived by a two-dimensional mean shift clustering using a flat kernel function with a bandwidth associated with the trunk diameter. Using the identified trunks as initial points, the mean shift individual tree segmentation could be improved significantly. The analysis has been conducted for 20 reference plots with a total area of 1.8 ha, an average ALS point density of 15 points/m^2 and 1779 reference trees.

Although the SkelTre method, developed by Bucksch et al. [25], is able to derive a point skeleton from any point cloud, it has been designed to derive the three dimensional structure of trees—inherently providing information on the stem orientation—based on high density laser scans. The algorithm creates an initial graph based on an octree [26], while a robustness criterion ensures an appropriate linkage of the graph's edges even for noisy input point clouds. By labelling each edge with its direction, the octree-graph is locally reduced to achieve a Reeb-graph [27]. Finally, the graph is embedded to the original point cloud to provide localized information on the skeleton.

1.4. Related Work

A systematic analysis of the ALS-derived trunk inclination and orientation angles is given by Razak et al. [28], which investigates the effect of landslides using high density ALS in an alpine region. As base data, they use a point cloud with an average point density of 170 points/m^2 and an approximate area of about 1.3 km^2. They delineate the tree crowns using the TreeVaW [29] algorithm, which achieves an overall accuracy of 84.8% using 560 terrestrially measured reference trees. The tree inclination θ and tree orientation ϕ have been derived with the SkelTre method [25] which tends to over-predict both variables. A regression model of the form $\theta = a + b \cdot \theta_{ALS}$ achieves an R^2 of 0.77. With a model of the form $\phi = a + b \cdot \phi_{ALS}$—which violates the cyclic character of the compass angles—they observe R^2 values of up to 0.83 at a reference height of 1.3 m. They find that trees located in landslide zones are more inclined and dispersed at different orientations compared to the control group.

2. Objectives

The studies presented in Section 1.3 focus on the detection of tree stems within study areas smaller than 2 km^2 only. To the knowledge of the authors no evaluation of the trunk inclination and orientation properties—in particular on larger scales—has been carried out so far. In the given study the following hypotheses are evaluated empirically to verify whether the ALS detected tree trunks show systematic inclination patterns, and whether these patterns match the known causes for tree inclination:

Hypothesis 1. *Preferred stem orientations can be found using ALS derived trunk vectors.*

Hypothesis 2. *Trunks preferentially lean down-slope.*

Hypothesis 3. *The tree inclination and orientation depend on tree species.*

Hypothesis 4. *Trunks preferentially lean leeward.*

Hypothesis 5. *The scan direction does not affect the observed trunk inclination and orientation.*

Hypothesis 6. *Soil properties, like soil type and moisture, affect the trunk inclination.*

3. Materials and Methods

3.1. Study Area

The study area of Hunsrück-Hochwald National Park with an area of $105 \, km^2$ is located in Rhineland-Palatinate and Saarland, Germany (Figure 1). From its southwestern extent it follows a low mountain range for about 30 km to its northeastern extent (see Figure 2). Its moderate climate of the mid-latitudes results in good conditions for sustainable forestry. The area is dominated by 44.9% European beech (*Fagus silvatica*) and 33.2% Norway spruce (*Picea abies*), 10.0% Sessile oak (*Quercus petraea*), 8.2% pine (Pinus sylvestris) and 3.6% Douglas fir (*Pseudotsuga menziesii*). Less common are European larch (*Larix decidua*), European white birch (*Betula pendula*) and other species. The soil substrate is dominated by quarzites and shales of the Devonian, and sandstone and mudstone of the Permian [30].

Figure 1. Overview map of the study area Hunsrück-Hochwald National Park.
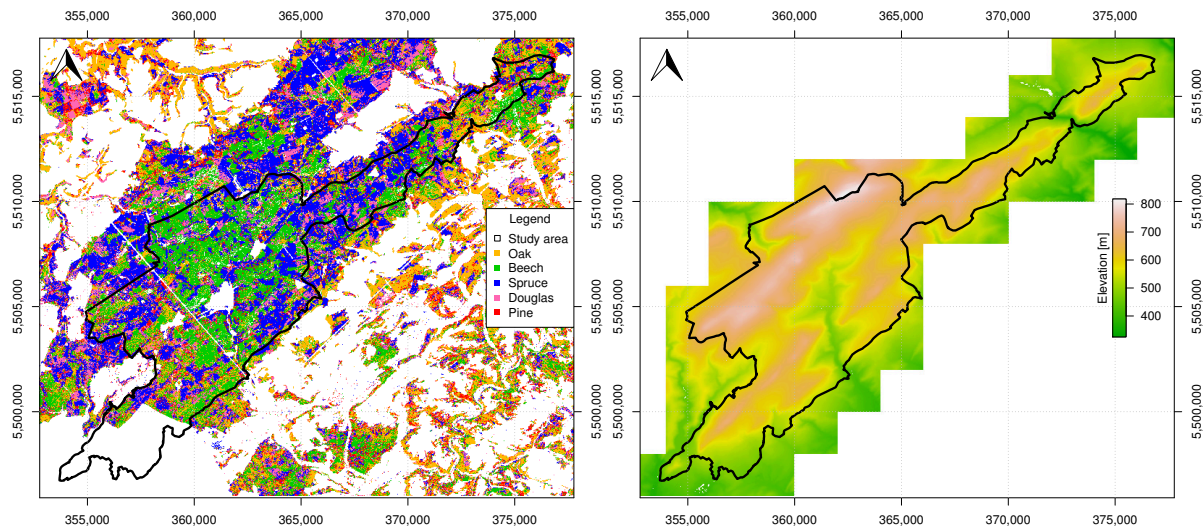
Figure 2. Tree species distribution within the study area Hunsrück-Hochwald National Park (**left**) and digital elevation model derived from ALS point clouds (**right**). The ALS point clouds have been provided by *LVermGeo*, while the tree species classification is the result of Stoffels et al. [31]. Coordinates are shown in ETRS89/UTM system (EPSG 25832).

3.2. Data

3.2.1. ALS Data

The ALS data used in this study has been acquired during nine flights from 24 March to 7 April 2015 using a *Riegl Q560* [32] under leaf off conditions. The flights have been strictly conducted in east-west, or west-east direction with an average swath width of about 200 m (Figure 3). The mean altitude has been about 600 m resulting in a footprint diameter of about 0.3 m and an average pulse density of 11.2 pulses/m^2. The 220 ALS tiles have been provided by the state surveying and geoinformation service of Rhineland-Palatinate (LVermGeo) in the form of pre-classified (ground, vegetation and other classes) *.las*-files with an extent of 1 × 1 km each. Each ALS point is labeled with its GPS timestamp expressed as second of week. For each flight the GPS tracks have been provided in *.csv* format.

3.2.2. Terrain and Canopy Models

Both a terrain and a canopy model are derived for each ALS tile by filtering points representing the surface in a first step and applying a Delaunay triangulation in a second step. The triangulation allows for a linear interpolation of the elevation within the given tile. In detail, to identify points representing the terrain, the points classified as ground are thinned with a duplicate point filter identifying local minima with a radius of 0.8 m (see *filters.surface* of Lamprecht [33]). Similarly, to identify points representing the canopy, the points classified as vegetation are thinned with the same duplicate point filter extracting local maxima with a radius of 0.8 m.

3.2.3. Scan Direction

Since the ALS points do not contain information on the scan direction, it is calculated using the supplementary GPS files. For each point of interest, the two GPS coordinates with minimum time difference are selected. Based on the time difference, the acquisition coordinate is interpolated as the linearly weighted average of both GPS coordinates. With this interpolated GPS coordinate, the line of sight (scan vector) linking the position of the laser scanner and the point of interest is derived.

3.2.4. Tree Species

To receive information on the tree species, the high resolution forest information layers of Stoffels et al. [31] are used. This data set provides information on the five most common tree species—European *beech*, *Douglas* fir, Sessile and Pedunculate *oak*, Scots *pine* and Norway *spruce*—in a spatial resolution of $5 \times 5\,m^2$. Except for the most southwestern area, the classification layer covers the study area completely, while an overall classification accuracy of about 76% is achieved.

3.2.5. Wind

Complex forests with dynamically restructuring trees in combination with hilly terrains make a deterministic prediction of the time and spatial scale dependent wind flow, eddies and turbulences practically impossible [9]. Thus, in this study, only the effect of the prevailing wind direction is considered. The daily wind direction and wind speed datasets of the TRY project [34] are used as base data. These provide information at a reference height of 10 m above ground and a spatial resolution of 10 km $\times$ 10 km. To represent the prevailing wind direction and wind speed, in this study the wind vectors are averaged over a 15 year period from January 1998 to December 2012. As Figure 4 illustrates, in the study area the wind is typically directed from West-South-West at average wind speeds below 2.1 m/s. In the Northwest of the study site, the wind turns to the Southwest.
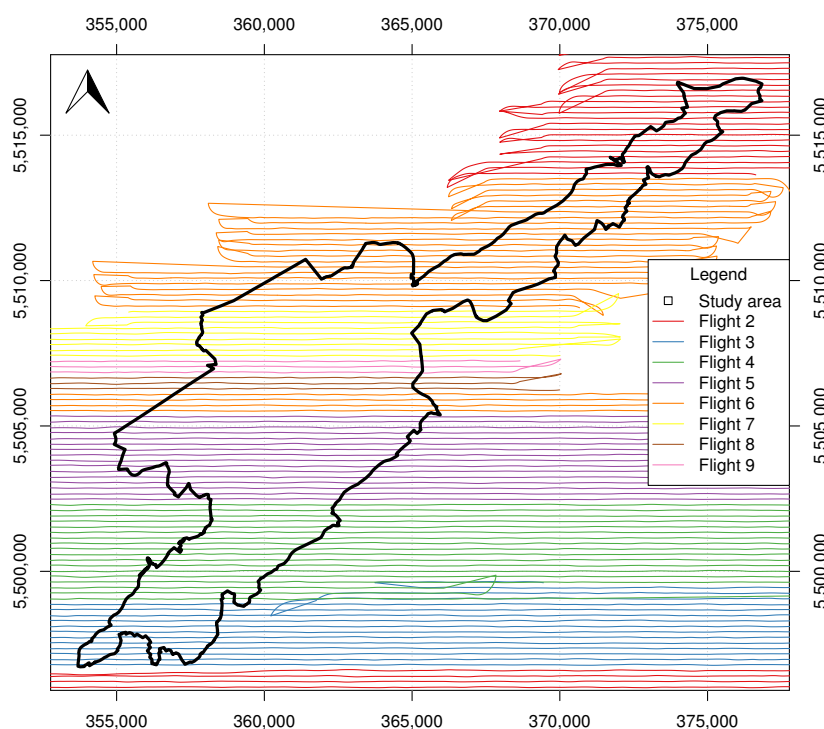
Figure 3. ALS flight lines and study area Hunsrück-Hochwald National Park. Coordinates are shown in ETRS89/UTM system (EPSG 25832).
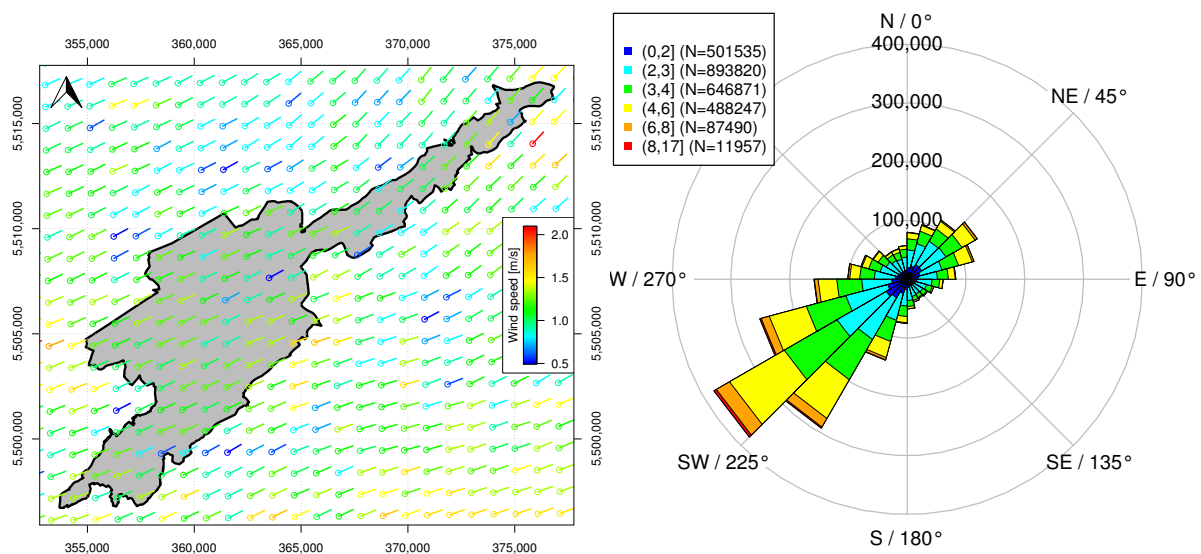
Figure 4. Fifteen year average (1989–2012) of the wind direction and wind speed in the study area (**left**) and wind rose diagram of the same area and time window (**right**). Both figures have been derived from daily wind models provided by Krähenmann et al. [34]. Coordinates are shown in ETRS89/UTM system (EPSG 25832).

3.2.6. Soil Properties

To inspect the influence of the soil, the official soil region map of Germany [35] is used. Since the given soil classification does not allow for a spatially explicit distinction between different soil types, the dominating soil substrates are extracted as illustrated in Figure 5 (left). In addition to the soil type, an ecological site classification [30]—provided by the state forest service of Rhineland-Palatinate (Landesforsten RLP)—is used to investigate the effect of soil moisture on the trunk inclination. One product of the ecological site classification is a soil moisture classification with 12 soil moisture regimes. In general, the study area is well supplied with water. The given regimes are re-sampled to three classes *dry*, *moist* and *very moist* as illustrated in Figure 5 (right).

3.3. Methods

3.3.1. Stem Detection

For the given study, a stem detection algorithm has been developed, which shall identify vertical linear structures most probably representing stems. To limit the computational effort and be independent from training data, heuristics are used to identify points associated with the stems. The algorithm is organized in three steps: *point filtering* (identifying points below canopy form vertical lines), *clustering* (identifying individual trunks) and *vector fitting* (fitting a regression vector to each trunk). A fundamental parameter of the algorithm is the radius R. Since this parameter should be in the scale of the minimum expected trunk distance, R is set to 0.9 m by expert decision.

3.3.1.1. Point Filtering

The trunk detection algorithm developed for the given study assumes that ALS points arranged in isolated vertical lines close to the ground represent tree trunks. To facilitate the isolation of linear structures, in a first step a duplicate point filter is applied to all points classified as vegetation (Figure 6, left).

In detail, the filter sweeps from top to bottom, while for each point currently observed, all points within radius R are removed. In a second step, only points with less than two neighbors within radius $2 \cdot R$ remain (Figure 6, center). In particular volumetric structures—like tree crowns—are removed by this filter, because the previously applied duplicate point filter ensures point distances of at least R. The previous filters tend to retain isolated points at the fringe of tree crowns. Thus, a vertical orientation of neighbored points needs to be ensured. For each point, the closest two neighboring points are investigated, while the z-axis is scaled by factor 2 to prefer a vertical assignment. If the horizontal extent of the three points is less than R, all three points remain (Figure 6, right).
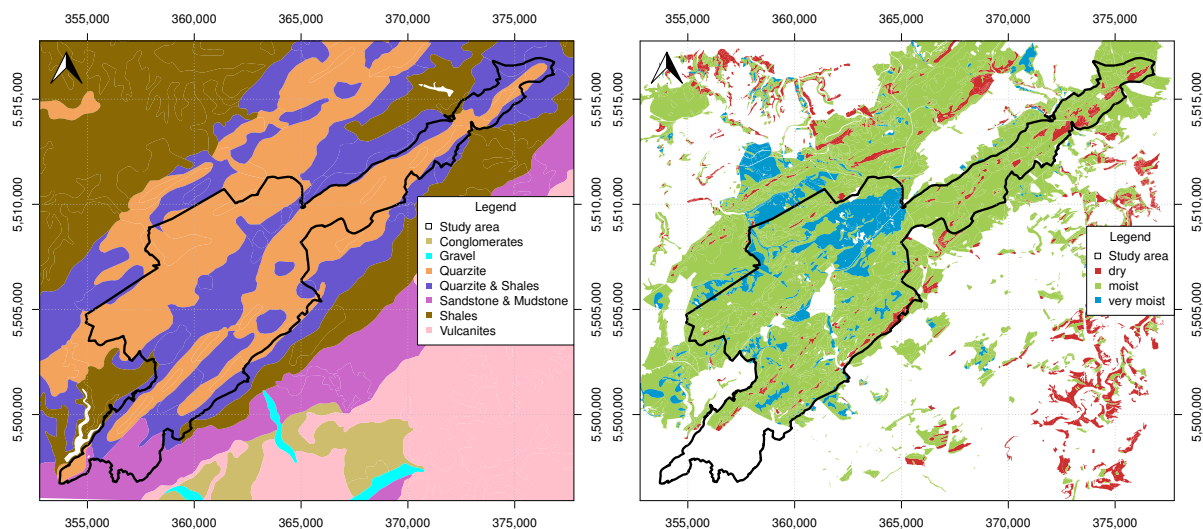
Figure 5. Dominating soil substrates (**left**) and soil moisture regimes (**right**) in the study area. Coordinates are shown in ETRS89/UTM system (EPSG 25832).
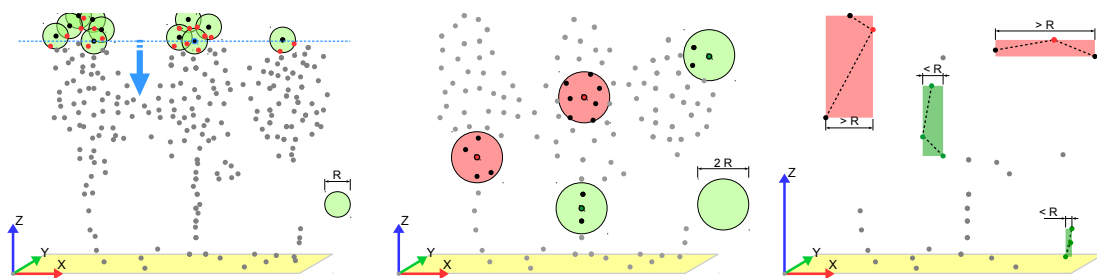
Figure 6. A duplicate point filter with filtering radius R is applied to points associated with vegetation (**left**). The resulting filtered point cloud serves as the input for a density filter (**center**). Points with more than 3 neighboring points within radius R are omitted. Working principle of the vertical line filter with maximum horizontal distance R and z-axis scaled by factor 4 (**right**).

3.3.1.2. Clustering

To identify isolated lines in the previously filtered point clouds, the points are clustered. The clustering algorithm sweeps from bottom to top (Figure 7, left). A point is assigned to the most frequent cluster of its neighboring points within radius $1.5 \cdot R$. If no neighboring point has already been assigned to a cluster, the point defines a new cluster. To take into account the vertical structure of the trunks, the z-axis is scaled by factor 1/4 before applying the clustering. To remove remaining linear structures in the top of the trees with no connection to the ground, only clusters with a gap to the ground $min(h)$ smaller $0.6 \cdot max(h)$ are retained, while h represents the height above ground for each point of the cluster. To use the full number

of available trunk points, all original ALS points with distance less than *R* to any of the cluster points define the final cluster.

3.3.1.3. Vector Fitting

A regression vector is fitted to each previously identified trunk cluster. Since the clustering algorithm tends to retain dispersed points at the bottom of the tree crowns, these points might distort the vector orientation. Thus, a consensus set is created by applying the MSAC [36] approach with threshold *R* to all point pair combinations of a cluster. The first Eigenvector of the consensus set represents the desired regression vector (Figure 7, center). Consensus sets with less than 4 points are omitted.

Based on the regression vector, a local coordinate system with the axes *v*, *z* and *a* is defined (Figure 7, right). The origin of the coordinate system corresponds to the average coordinate of the trunk points, while the Eigenvector defines the *v*-axis. The *a*-axis is oriented perpendicular to the regression vector and the global z-axis. Finally, the *z*-axis is oriented perpendicular to the global x- and z-axis.

The trunk root is defined as the intersection point between the regression vector and the terrain model. Each trunk vector provides its *inclination* or *zenith* angle t_θ, referring to the angular deviation of the stem from the vertical orientation ($0°$). The *orientation* or *azimuth* t_ϕ refers to the compass direction of the stem, defined as the clockwise planar deviation from the North ($0°$).
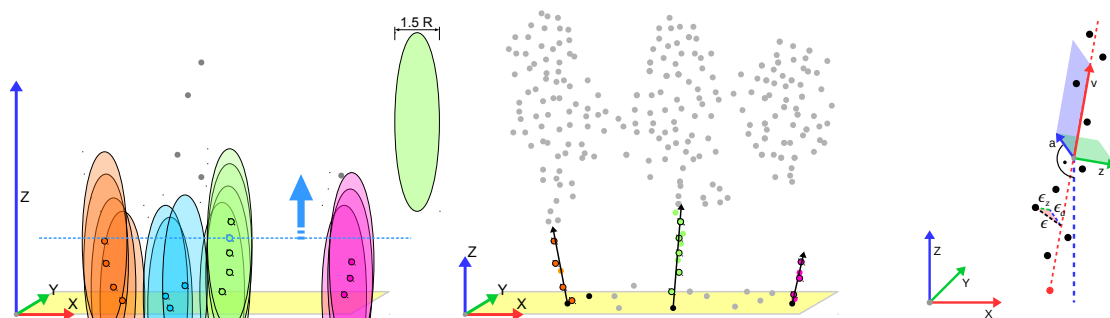
Figure 7. The filtered point cloud serves as the input for the clustering algorithm (**left**). A point is assigned to the most frequent class of its neighboring points within radius 1.5·*R*. To prefer linear structures, the z-axis is scaled by factor 1/4. A regression vector is fitted to each trunk cluster (**center**). Implausible vectors are omitted from further analyses. Each regression vector defines a local coordinate system, with axes *v*, *z* and *a* (**right**). The black dots represent the trunk points, while the red dot corresponds to the trunk root.

3.3.1.4. Vector Uncertainty

Using its local coordinate system, the residuals ϵ of a trunk vector fitted to *n* points are split up in a vertical ϵ_z and a horizontal ϵ_a component (see Figure 7, right). Similar to a two dimensional regression line, the standard error of the inclination se_z is calculated with Equation (1). Accordingly, the standard error se_a gives information on the uncertainty of the trunk azimuth.

$$SE = \frac{\frac{1}{n-2} \sum_{i=1}^{n} \epsilon_i^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \tag{1}$$

Similar to the two dimensional regression, the inclination t_θ is assumed to follow the Student's t-distribution with $t_n - 2$ degrees of freedom. The *p*-value t_p is calculated according to Equation (2).

The function $F_\nu(x)$ corresponds to the cumulative distribution function of the Student's t-distribution with test value x and ν degrees of freedom.

$$t_p = 2 \cdot \mathrm{P}(T_\theta \leq t_\theta) = 2 \cdot F_{t_n-2}\left(-\frac{\sin(t_\theta)}{t_{se_z}}\right). \tag{2}$$

3.3.2. Trunk Features

To evaluate the hypotheses, each detected trunk is labeled with its inherent vector features and the attributes of the data layers. To increase readability, a set of an arbitrary higher order property is derived by a function $\mathrm{f}(t)$ and $t \in T$ according to Equation (3). Each trunk t is characterized by the inclination angle t_θ and the compass orientation t_ϕ of its three-dimensional regression vector. Its root is defined by the intersection point between its vector and the DEM. The attributes t_{se_z} and t_{se_a} give information on the uncertainty of the vector in inclination direction and compass direction respectively (see Section 3.3.1.4). Using t_{se_z} the p-value t_p of the trunk inclination is derived (see Equation (2)).

$$T_{\mathrm{f}(t)} = \{\mathrm{f}(t) \mid \forall\, t \in T\}. \tag{3}$$

Features related to the ALS properties are aggregated from the points used to fit the trunk vectors. In particular, the scan vectors are averaged and split up into a vertical scan angle *scan_zenith* and a horizontal scan direction *scan_azimuth*. In addition, the horizontal (*scan_horizontal_std*) and vertical (*scan_vertical_std*) standard deviations of the scan vector are calculated.

The tree species, soil moisture and soil substrate classes are assigned by intersecting the trunk root coordinates with their raster cell or polygon. The wind direction and wind speed for each trunk is determined by a Cressman interpolation [37] of the wind vector rasters with a radius of 1 km.

The site *aspect*, *slope* and *elevation* are derived for each trunk by inspecting the seven terrain points closest to the trunk root. Due to the planar character of the ground points, their third Eigenvector provides the aspect and slope, while the elevation is extracted from their average z-coordinate.

As a measure of phototropism, the relative direct solar radiation *solar_direct* is calculated for each tree according to Gassel [38] using Equation (A1) (see Appendix A). The radiation is calculated for each tree based on its geographic latitude, site aspect and slope assuming solar noon at mid of summer. The effect of the site aspect and slope on the relative direct solar radiation is illustrated in Figure 8.

3.3.3. Regression Models

To analyze underlying causes for stem inclination, for each trunk t its inclination t_θ is predicted using weighted least squares regression. Weighted regression is used to suppress biases caused by very frequent trunk inclinations and orientations. Only attributes potentially causally related to the inclination are considered to derive the predicted inclination $t_{\hat{\theta}}$. Next to the variables of interest (slope, tree height, tree species, wind speed, soil moisture, soil substrate and scan angle), attributes associated with the uncertainty of the trunk vectors (t_{se_z}, t_{se_a} and t_n) are considered since they might relate to random effects.

To make qualitative statements on potential underlying causes for specific inclination directions, the trunk orientation t_ϕ is assumed to be a linear combination of two dimensional explanatory vectors. To implement this concept, each directed variable defines an explanatory vector as illustrated in Equation (4), while for each trunk its orientation is modeled by creating pairwise coupled equations according to Equation (5). By multiplying scalar variables with these vectors, interaction terms can also be considered. Solving the given equation system using weighted least squares regression provides the desired regression coefficients which are used to calculate the predicted trunk orientation $t_{\hat{\phi}}$.
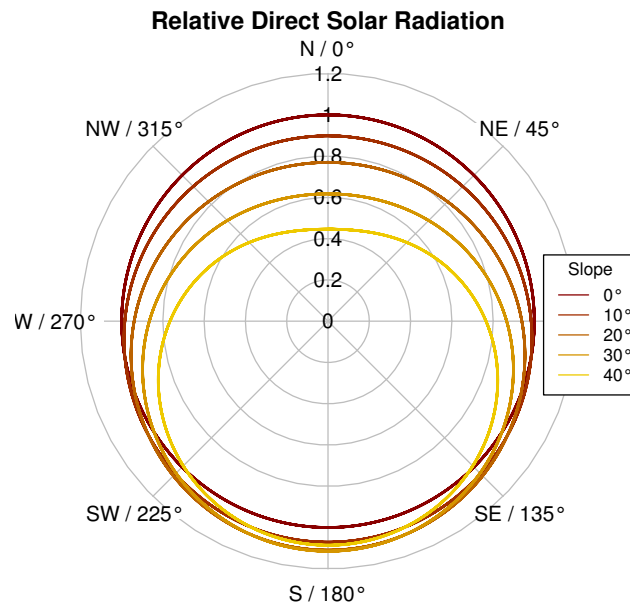
Figure 8. Relative direct solar radiation depending on site aspect for differing site slopes at solar noon in summer at a geographic latitude of 50°N.

Please note that the signs of the regression coefficients give information on the direction of the underlying effects. The coefficient $-c_{south}$—corresponding to a directed intercept—shall allow for a systematic inclination to the South.

$$
\begin{aligned}
terrain_x &= \sin(slope) \cdot \sin(aspect) & terrain_y &= \sin(slope) \cdot \cos(aspect) \\
solar_x &= solar\_direct \cdot \sin(aspect) & solar_y &= solar\_direct \cdot \cos(aspect) \\
wind_x &= wind\_speed \cdot \sin(wind\_direction) & wind_y &= wind\_speed \cdot \cos(wind\_direction) \\
scan_x &= \cos(scan_\theta) \cdot \sin(scan_\phi) & scan_y &= \cos(scan_\theta) \cdot \cos(scan_\phi)
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
\sin(\phi) &= & c_{terrain} \cdot terrain_x + c_{solar} \cdot solar_x &+ c_{wind} \cdot wind_x + c_{scan} \cdot scan_x + \epsilon_x \\
\cos(\phi) &= -c_{south} &+ c_{terrain} \cdot terrain_y + c_{solar} \cdot solar_y &+ c_{wind} \cdot wind_y + c_{scan} \cdot scan_y + \epsilon_y.
\end{aligned}
\tag{5}
$$

To avoid biases caused by the uneven frequency of the observed stem orientations, the equations of the regression models are weighted by the p-value of the trunk inclination t_p times the inverse circular density of the trunk orientation. The density of the trunk orientation is determined by a kernel density estimation using the bandwidth selection procedure of Sheather and Jones [39].

3.3.4. Feature Selection

To pre-select relevant features, the trunk inclination as well as the orientation are modeled with Random Forest regressors of 500 trees and 50,000 randomly selected trunks. The default parametrization of the *ranger* package for *R* [40] is used. Missing numeric values are filled by mean imputation, while categorical values are imputed by random sampling of the non-missing values. Attributes with low feature importance are neglected for further analyses. After identifying the most relevant features, several knowledge driven linear models are fitted, while equations with missing values are omitted

(see Section 4.5). This iterative procedure shall guarantee the selection of robust knowledge driven models to reveal potential underlying causes for the trunk inclination.

Due to the circular character of the orientation, the model residuals tend to be bimodally distributed. Also the inclination model tends to violate the homoscedasticity assumption. To still be able to make qualitative statements on the effects, the confidence intervals of the regression coefficients are empirically determined by bootstrapping. Each regression model is fitted 10,000 times to differing random samples of 1% of the training dataset. Finally the inclination and orientation models are fitted to the full training dataset. Model coefficients not significantly differing from zero (99% significance level, respectively 98% confidence interval) are discarded, or interaction terms are added. Factorial variables are omitted, if the coefficients of at least two categories do not differ significantly (99% significance level, respectively 80% confidence interval of each coefficient).

3.3.5. Training and Validation

The trunk detection algorithm of Section 3.3.1 is applied to all ALS tiles of the study area. Since border effects might affect the detection of trees at the edges of the tiles, all trunks with distance below 5 m to the tile extends are disregarded for further analyses. Trunk vectors with a high vector uncertainty can lead to an inaccurate inclination and orientation. If the inclination error exceeds the inclination angle, the trunk orientation can even topple. These effects have the potential to cover systematic orientation patterns, since they appear as noise. For this reason only significantly inclined trunks—t_θ significantly different from zero—are further investigated. The remaining trunk detections form the set T. Of the data set, 70% is used for training T_t and 30% for validation T_v.

The accuracy of the trunk inclination prediction $t_{\hat{\theta}}$ is assessed with the ordinary coefficient of determination R^2. To adress the accuracy of the modelled orientation $t_{\hat{\phi}}$—in addition to the ordinary R^2 provided by the regression model—the angular residuals $T_{\epsilon_\phi} := T_{\epsilon_\circ(t_\phi, t_{\hat{\phi}})}$ (see Equation (6)) are examined. The mean direction $\overline{T_{\epsilon_\phi}}$ [41], as well as the circular standard deviation $\sigma_{T_{\epsilon_\phi}}$ [41] describe the location and dispersion of the circular data. Based on the orientation vector $\vec{\phi} = (\sin(\phi), \cos(\phi))$ (cf. Equation (5)), the average variance explained by each directed explanatory variable is calculated to assess its explanatory power. The explained variance of a directed explanatory variable $\vec{E} = (E_x, E_y)$ with regression coefficient c_E corresponds to the scalar product $c_E \vec{E} \cdot \vec{\phi}$.

$$\epsilon_\circ(\alpha, \beta) = (\alpha - \beta + 180) \, \% \, 360 - 180. \tag{6}$$

4. Results

4.1. Stem Detection

In total 1,147,633 stems have been detected, while 846,851 stems have been located within the study area. The partial coverage of the ground data results in missing data within the borders of the study area for soil moisture (10.6%) and tree species (11.5%). Since the given study shall provide a proof of concept only, only the plausibility is checked instead of an accuracy assessment of the stem detection. In particular, a strongly species-dependent detection rate might affect the results, since it would indicate problems of extracting the linear structures under certain stand characteristics.

Table 1 illustrates the percentage of area covered by each tree species (see Section 3.2.4) compared to the amount resulting from the stem detection. A chisquared goodness-of-fit test indicates that the detected species differ significantly from the species distribution provided by the classification layer. In particular, the amount of spruce and beech detections are higher than the area, while Douglas, oak and pine are under-represented by the stem detection. Since the surface area does not provide information on the actual

stand density, no final conclusion on detection rates can be drawn. But, since the algorithm seems not to fail entirely for a specific species, the detected trunks seem to be suitable for further analyses.

Table 1. Comparison of surface area with the amount of detected stems.

	Beech	Douglas	Oak	Pine	Spruce
Area of classification layer	36.8%	11.4%	13.9%	9.8%	28.1%
Number of detected stems	40.7%	6.5%	11.5%	8.3%	32.9%

4.2. Selection of Trunks

Figure 9 illustrates the effect of the *p*-value t_p on the trunk inclination and orientation. By selecting significantly inclined trees only, the probability of an erroneous trunk orientation—due to randomly inclined observations—is reduced. Per definition, with decreasing significance levels, preferably trunks with high inclination angles are selected (Figure 9, left). Also with decreasing significance levels, a preferred South orientation of the trunks is revealed (Figure 9, right). To gain clean data, the authors have decided to choose a significance level of 1% resulting in a data set *T* of 299.000 trunks.
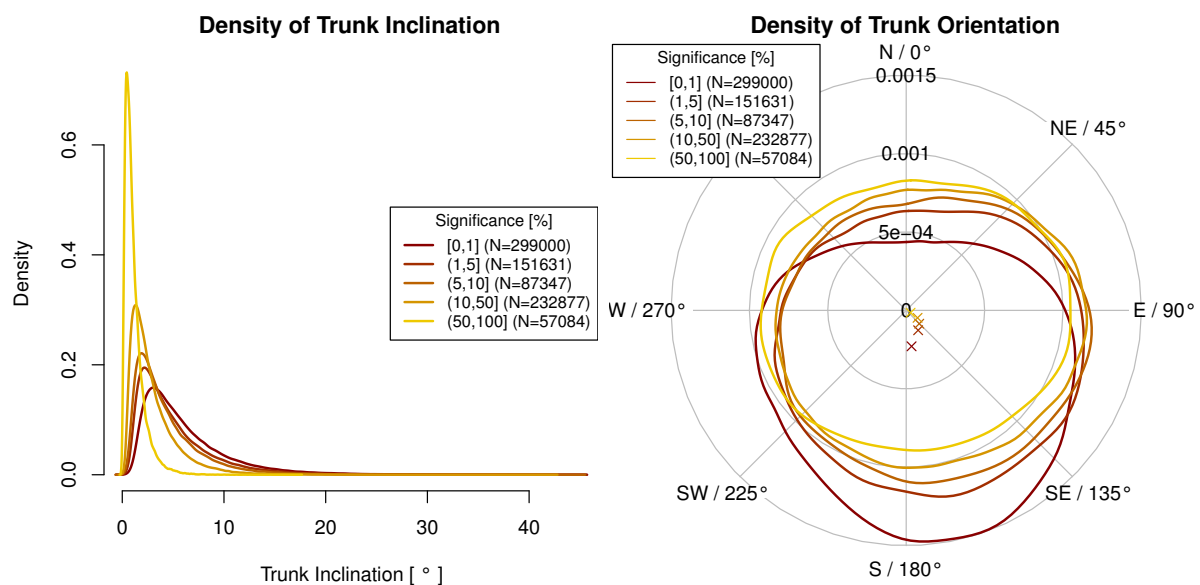
Figure 9. Density plots of the trunk inclination (**left**) and trunk orientation (**right**) grouped by differing significance levels of the trunk inclination. For each group, the mean trunk orientation is marked with a cross.

4.3. Stand Characteristics

Figure 10 illustrates the site aspect, scan direction and observed trunk inclination for all significantly inclined trunks. Due to the orientation of the mountain range, most sites are oriented to South-East or North-West. With less than 3%, trees on slopes with an inclination greater than 25° are rare. The regular ALS flight pattern causes the trunks to mostly be scanned from the North or the South. 95% of the trunks are inclined less than 15°. The trunks preferably lean to South-Southeast, with an average orientation of about 171.7° and a circular standard deviation of 94.2°. A Watson's test for circular uniformity [42,43] confirms that the trunk orientation differs highly significantly from a uniform distribution, supporting Hypothesis 1.
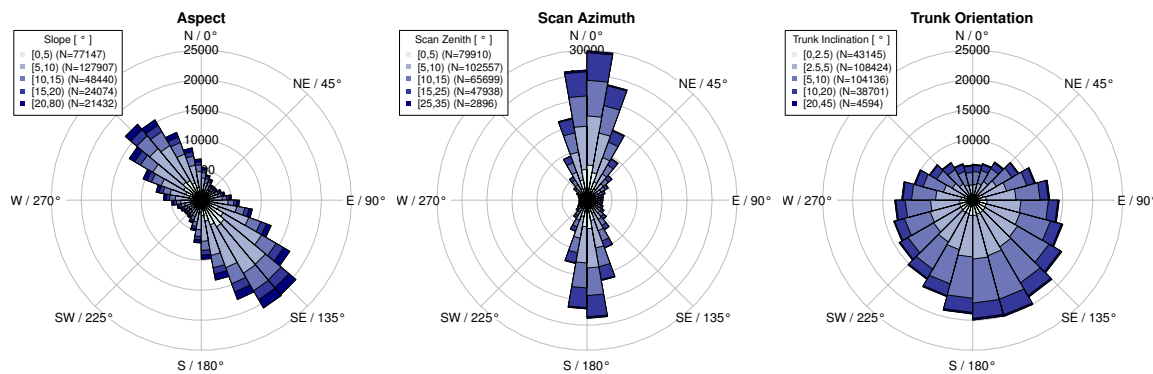
Figure 10. Site aspect in relation to slope (**left**), scan compass direction in relation to the scan zenith angle (**center**) and trunk orientation in relation to the trunk inclination (**right**).

Table 2 shows that regardless of tree species about 67% to 74% of the detected trunks are located on *Quarzite* sites and about 26% to 32% are located on mixed *Quarzite & Shales* sites. The portion of trunks located on pure *Shales* is minor. 82% to 86% of the trunks are located on *moist* sites. About 15% of the beech and spruce can be found on *very moist* sites, while about 10% of the Douglas and pine are located on *very moist* sites. With 10%, oak can preferably be found on *dry* sites rather than on *very moist* sites.

Table 2. Distribution of soil properties by tree species.

		Tree Species				
Soil Property		**Beech**	**Douglas**	**Oak**	**Pine**	**Spruce**
Soil substrate	Quarzite	73.5%	72.2%	70.7%	68.8%	67.4%
	Quarzite & Shales	26.2%	27.5%	27.8%	30.5%	32.4%
	Shales	0.3%	0.3%	1.5%	0.6%	0.2%
Soil moisture	dry	3.1%	5.4%	10.0%	5.6%	3.7%
	moist	82.2%	84.3%	85.8%	83.4%	81.8%
	very moist	14.7%	10.3%	4.2%	11.0%	14.4%

Figure 11 shall highlight potential species specific site characteristics, which might affect the modeling procedure. Although the local wind regime is characterized by topography and forest structure, the mean wind direction indicates an almost similar general wind exposition for all tree species. The spatial distribution of the trees lead to a species dependent site aspect. In general, the orientation of the mountain range results in most of the trees facing North-West, respectively South-East. Beech and Douglas tend to be located on sites facing South-East, while spruce and pine can mostly be found on sites facing North-West.

4.4. Preliminary Analysis

Figure 12 illustrates the Pearson's product moment correlation coefficients of the trunk features. For the directed variables, the circular variant [42,43] is used. A strong correlation between the trunk inclination (*zenith*) and variables associated with the vector uncertainty is given. With increasing uncertainty—high se_z, high se_a, low n—high inclination angles can occur at random. The negative correlation of the tree height with the trunk inclination might be caused by the increased vector uncertainty. The trunk orientation is positively correlated with the site aspect. A slight negative correlation with the wind direction and a slight positive correlation with the scan direction can be found. But, these apparent correlations could be caused by coincidence due to the orientation of the mountain range and the stripe-wise ALS scanning pattern.
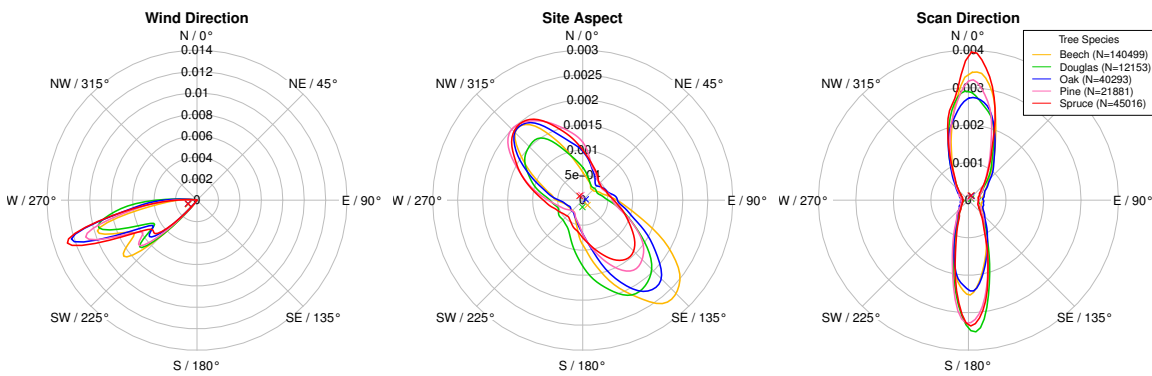
Figure 11. Density of the tree trunks for the explanatory variables wind direction (**left**), site aspect (**center**) and scan direction (**right**) grouped by tree species.
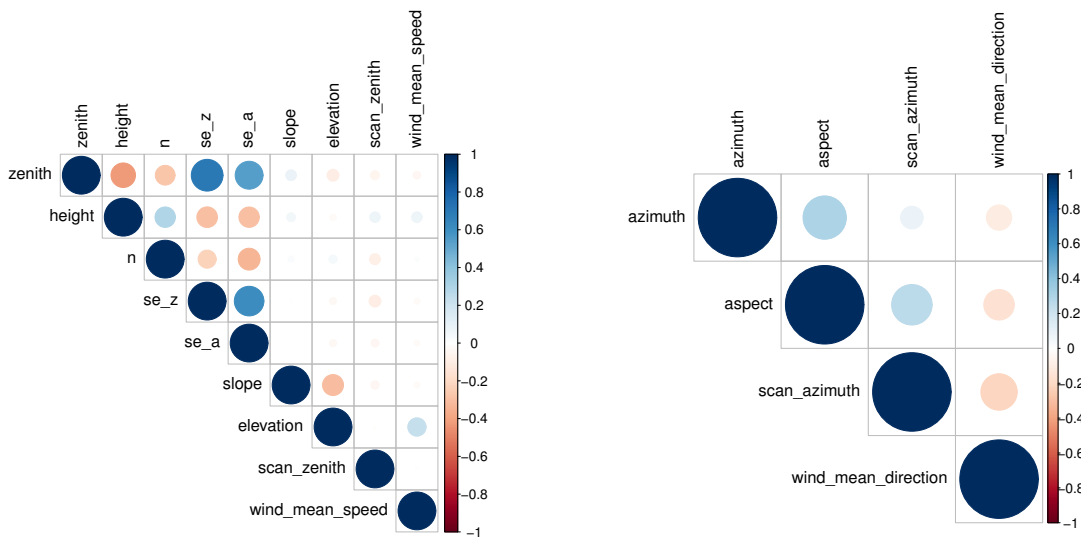

Figure 12. Pearson's product moment correlation coefficients for scalar variables (**left**) and directed variables (**right**).

As Figure 13 and Table 3 illustrate, the observed trunk inclination and orientation depend on the tree species. A Kruskal-Wallis rank sum test indicates significant differences of the average inclination between the species. More interestingly, the trunk orientations between the species differ highly significantly. In particular, deciduous trees tend to lean to the South, while conifers tend to lean to the East. The first observation is an indicator for phototropism, the latter might be caused by wind.

These observations are in accordance with Table 4, which summarizes the amount of trunks inclined in the explanatory directions. About 70% of the deciduous trunks are inclined to the South, compared to 54% of the conifers. About 64% of the conifers lean leeward, compared to 41% of the deciduous trees (cf. Hypothesis 4).

Table 3. Inclination and orientation of the significantly inclined trunks for different species.

	Beech	Douglas	Oak	Pine	Spruce	Average
$\overline{T_\theta}$	$6.14° \pm 0.01°$	$7.11° \pm 0.05°$	$5.71° \pm 0.02°$	$6.28° \pm 0.03°$	$5.77° \pm 0.02°$	$6.14° \pm 0.01°$
$\overline{T_\phi}$	$187.7° \pm 0.3°$	$115.1° \pm 1.6°$	$182.9° \pm 0.8°$	$112.9° \pm 1.4°$	$99.0° \pm 0.6°$	$171.7° \pm 0.3°$
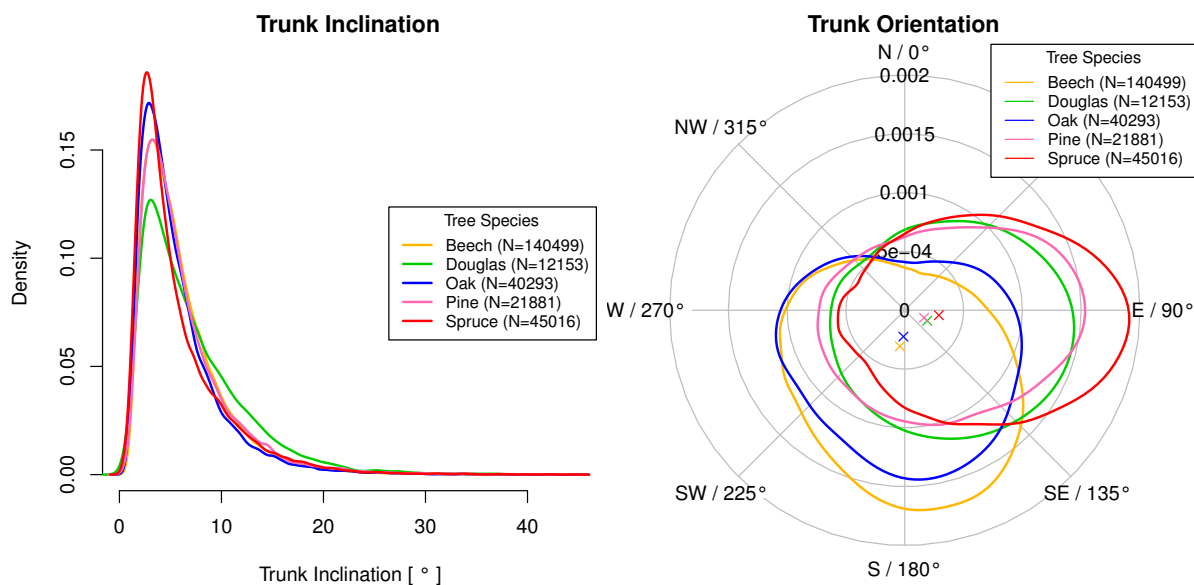
Figure 13. Influence of the tree species on the trunk inclination (**left**) and trunk orientation (**right**). The deciduous trees tend to lean to the South, while the conifers tend to lean to the East. The orientation follows the site aspect, while a general trend to the South can be observed.

Table 4. Amount of trees inclined with a specific explanatory variable for different groups.

Group		Inclination Direction			
		Southward	Down-Slope	Leeward	in Scan Direction
Species	Beech	70.8%	77.2%	39.9%	40.3%
	Douglas	56.2%	60.5%	61.0%	50.4%
	Oak	65.5%	74.1%	44.5%	42.9%
	Pine	54.5%	63.4%	58.8%	43.1%
	Spruce	53.6%	51.8%	67.5%	53.7%
Species Type	Conifer	54.3%	56.3%	64.1%	50.2%
	Deciduous	69.6%	76.5%	40.9%	40.9%
Root System	Flat	53.6%	51.8%	67.5%	53.7%
	Heart	69.7%	75.9%	41.6%	41.1%
	Tap	61.6%	70.3%	49.6%	43.0%
	Total	66.0%	71.3%	46.8%	43.3%

Supporting Hypothesis 2, Table 4, shows that 71 % of the trees are inclined down-slope. The effect of the terrain is highlighted in Figure 14. In general, increasing inclinations of slopes result in increased inclination angles, while the orientation follows the site aspect systematically with a general trend to the South. This observation is in accordance with the interaction between the site aspect and phototropism, since a site facing to the sun facilitates light capture compared to a site facing to the North.

Figure 15 illustrates that this effect is highly species dependent. Beech and oak seem to be highly affected by the interaction between phototropism and the terrain, while the site aspect has almost no influence on the orientation of spruce.
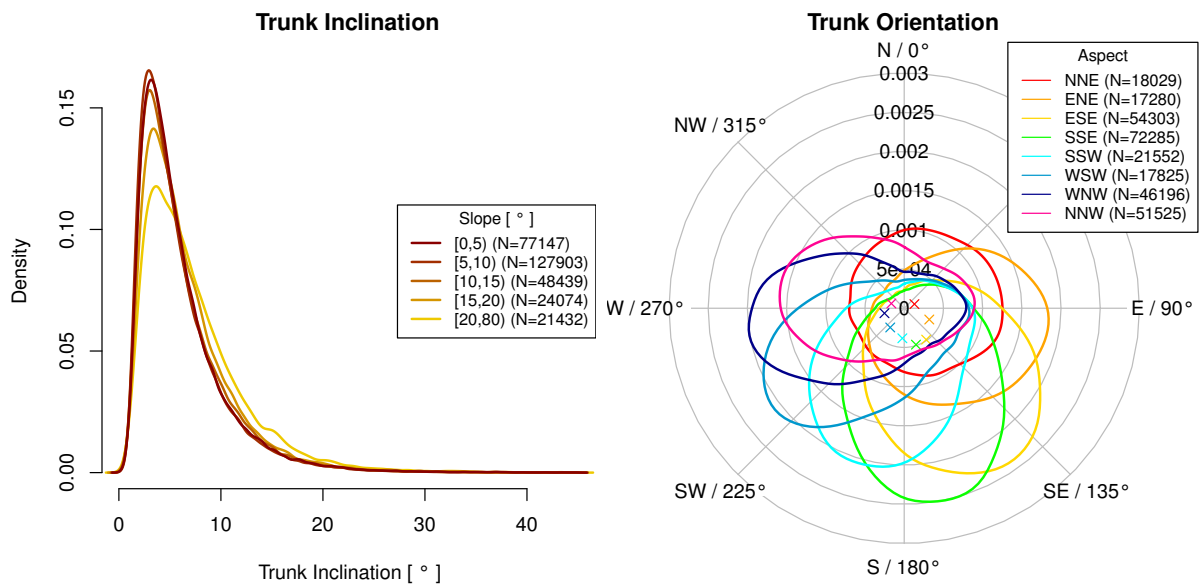
Figure 14. Influence of the slope on the tree inclination (**left**) and effect of the aspect on the trunk orientation (**right**).
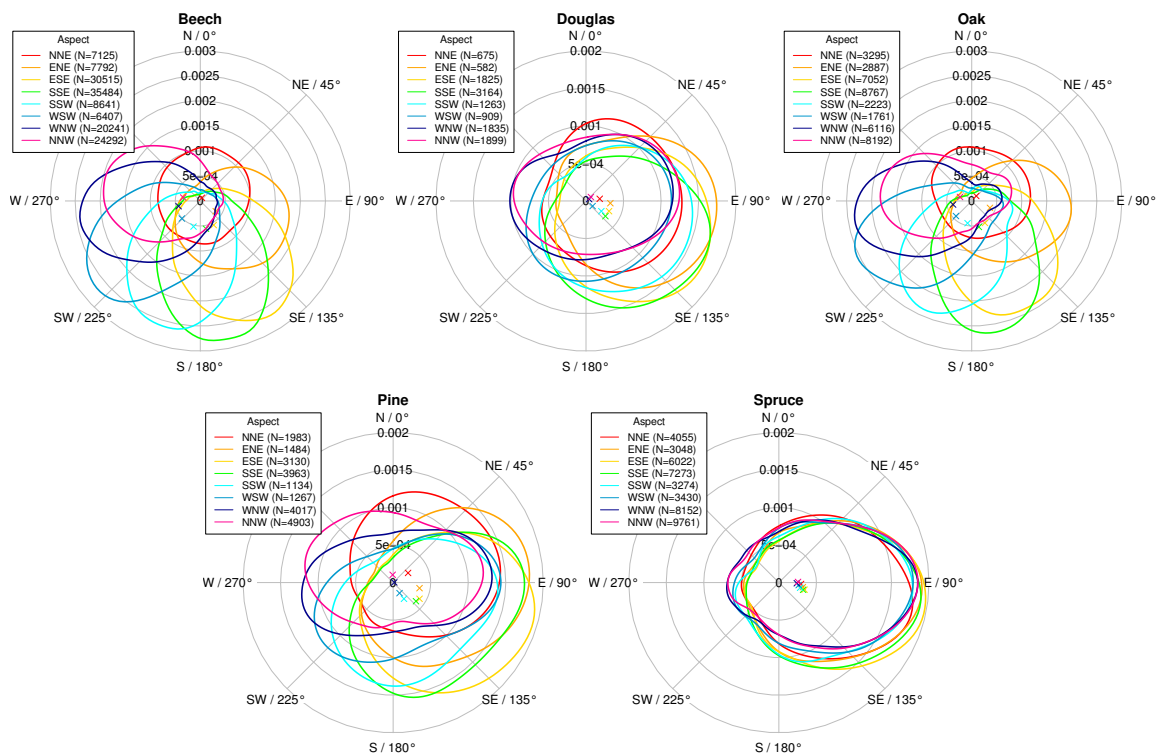
Figure 15. Effect of the site aspect on differing tree species.

Table 4 also indicates an effect of the scan direction on the observed trunk orientation (Hypothesis 5). But, the regular ALS flight pattern in combination with the uneven distribution of the tree species in the mountain range makes an autocorrelation with the scan direction highly probable. The same is true for a potential autocorrelation between the other explanatory variables. In consequence, for final conclusions the regression models are investigated.

4.5. Regression Models

4.5.1. Trunk Inclination

Figure 16 illustrates the feature selection procedure with the feature importances of the Random Forest regressor and six linear regression models with differing features and interaction terms. Table 5 provides information on the partial and total R^2 of all six linear models. With an out of the bag R^2 of 0.594, the Random Forest regressor performs slightly better than the linear models, with R^2 values ranging from 0.570 to 0.590.
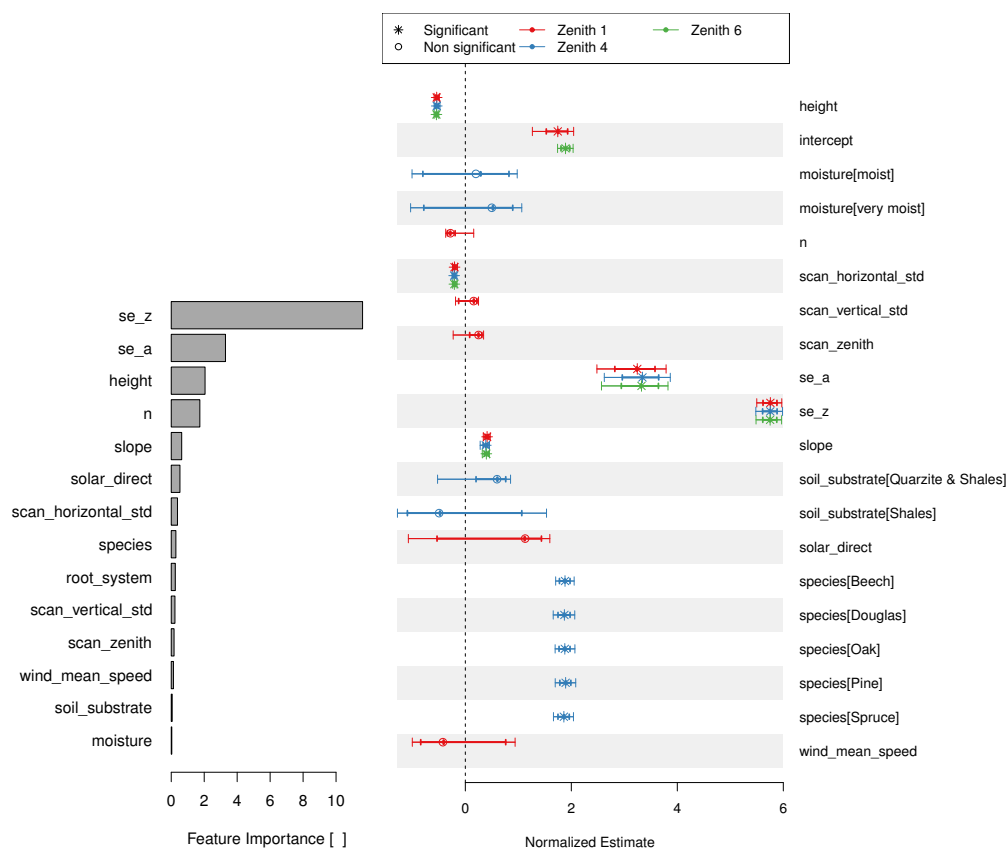
Figure 16. Permutation feature importance of the Random Forest regressor **(left)** and regression coefficients of selected linear trunk inclination models (**right**). The coefficients are scaled by $\sqrt[3]{}$ to display them in the same plot. The given 98% and 80% confidence intervals are based on bootstrapping.

The coefficients of the control variables se_z and se_z—both associated with the uncertainty of the trunk vectors—are significantly different from zero for all linear models, and dominate the models by their explanatory power. The variable se_z is positively correlated with the trunk zenith angle, since with increasing uncertainty large zenith angles can occur at random. The tree height is negatively correlated to the trunk inclination, which might either be caused by the uncertainty of the vectors or by the need of tall trees to align vertically to avoid toppling.

Despite its low explanatory power with just about 2% of variance explained, the slope inclination is positively correlated with the trunk inclination. Supporting Hypothesis 2, this is a first indicator for a terrain dependent inclination of the trunks. In contrast to Hypothesis 5, a slight, but significant negative correlation between the scan direction and the inclination angle can be observed.

The models imply that none of the factors tree species, soil substrate or soil moisture affect the trunk inclination significantly. Based on the given data, the mean wind speed as well as the direct solar radiation seem not to influence the trunk inclination significantly. In addition, the control variables *n* and *scan_vertical_std* seem not to affect the trunk inclination.

Table 5. Total R^2 and partial R^2 per regression coefficient of the linear trunk inclination models. Coefficients significantly different from zero are marked in bold. Significant inter group differences are marked in italics.

	Zenith 1	Zenith 2	Zenith 3	Zenith 4	Zenith 5	Zenith 6
(Intercept)	**0.010**	0.000	**0.000**			**0.096**
se_z	**0.308**	**0.314**	**0.312**	**0.312**	**0.311**	**0.307**
se_a	**0.020**	**0.021**	**0.025**	**0.025**	**0.025**	**0.024**
n	0.002	0.003				
scan_horizontal_std	**0.009**	**0.010**	**0.011**	**0.011**	**0.011**	**0.010**
scan_vertical_std	0.000	0.000				
scan_zenith	0.001	0.001				
wind_mean_speed	0.000	0.000				
solar_direct	0.001	0.001				
height	**0.076**	**0.067**	**0.074**	**0.074**	**0.075**	**0.083**
slope	**0.018**	**0.013**	**0.013**	**0.013**	**0.013**	**0.018**
species:moisture:soil_substrate		*0.004*	*0.004*			
species				0.001	0.098	
moisture				0.000		
soil_substrate				0.001		
Total R^2	0.573	0.590	0.586	0.581	0.578	0.570

4.5.2. Trunk Orientation

Figure 17 as well as Tables 6 and 7 illustrate the feature selection procedure for modeling the trunk orientation. The Random Forest regressor achieves an adjusted R^2 of about 0.27 while the circular standard deviation is about 62°. It identifies the *terrain* and *solar*—as the composites of the site aspect and the slope inclination, respectively direct solar radiation—as the driving factors for the trunk orientation. It also identifies the *wind*—as the combination of the prevailing wind direction and speed—and the *scan* direction as important explanatory variables. Except for the tree species, the non-directed parameters achieve a lower feature importance, while the soil moisture as well as the soil substrate have a minor contribution to the model. With adjusted R^2 values ranging from 0.13 to 0.18, as well as circular standard deviations ranging from 76° to 83°, the linear models achieve less accurate results than the Random Forest regressor. But, up to about 19% of the variance can be explained by the linear models.

In accordance with the Random Forest regressor, with up to 14.7% of the variance explained by interactions with the vectors *terrain* and *solar* (see *Azimuth 6*), the linear models identify the aspect as the variable with most explanatory power. As Figure 17 illustrates, the trees tend to lean down-slope, while significant species dependent differences can be found (*terrain:species*, *solar:species*). In detail, deciduous trees gain a positive *solar* coefficient, while the coefficient for conifers does not differ significantly from zero (see *solar:species_type*). Although a general tendency of the trunk orientation to the South could be found in Section 4.4, the effect is not significant. Also a grouping by tree species (*south:species*) does not reveal any significant effects. In consequence, the observed South inclination of the trunks (compare Figure 14) is probably the result of the interactions with the site aspect. The high explanatory power of the vectors *solar*—representing the relative direct solar radiation—and *terrain* are indicators for the interaction between phototropism and the terrain as the driving factor for the observed South-inclination of the trunks.

With up to 3.2% variance explained, the prevailing wind direction and speed (*wind*) have a significant species dependent effect on the trunk orientation (*wind:species*). In particular, conifers tend to lean leeward, while the models imply that beech tend to lean windward.

In accordance with the results of the trunk inclination modeling (Section 4.5.1), the tree height has a significant effect on the trunk orientation. In particular, it intensifies the species depending leeward or windward orientation (*wind:species:height*) of the trunks.

With 0.8% to 0.9% of variance explained by the vector *scan*, the LiDAR scan direction has a minor but significant and robust effect on the trunk orientation for all six linear models. The model coefficients imply that the trunk vectors are systematically inclined towards the direction of acquisition. But, this effect might still be an artifact of the systematic scanning pattern.
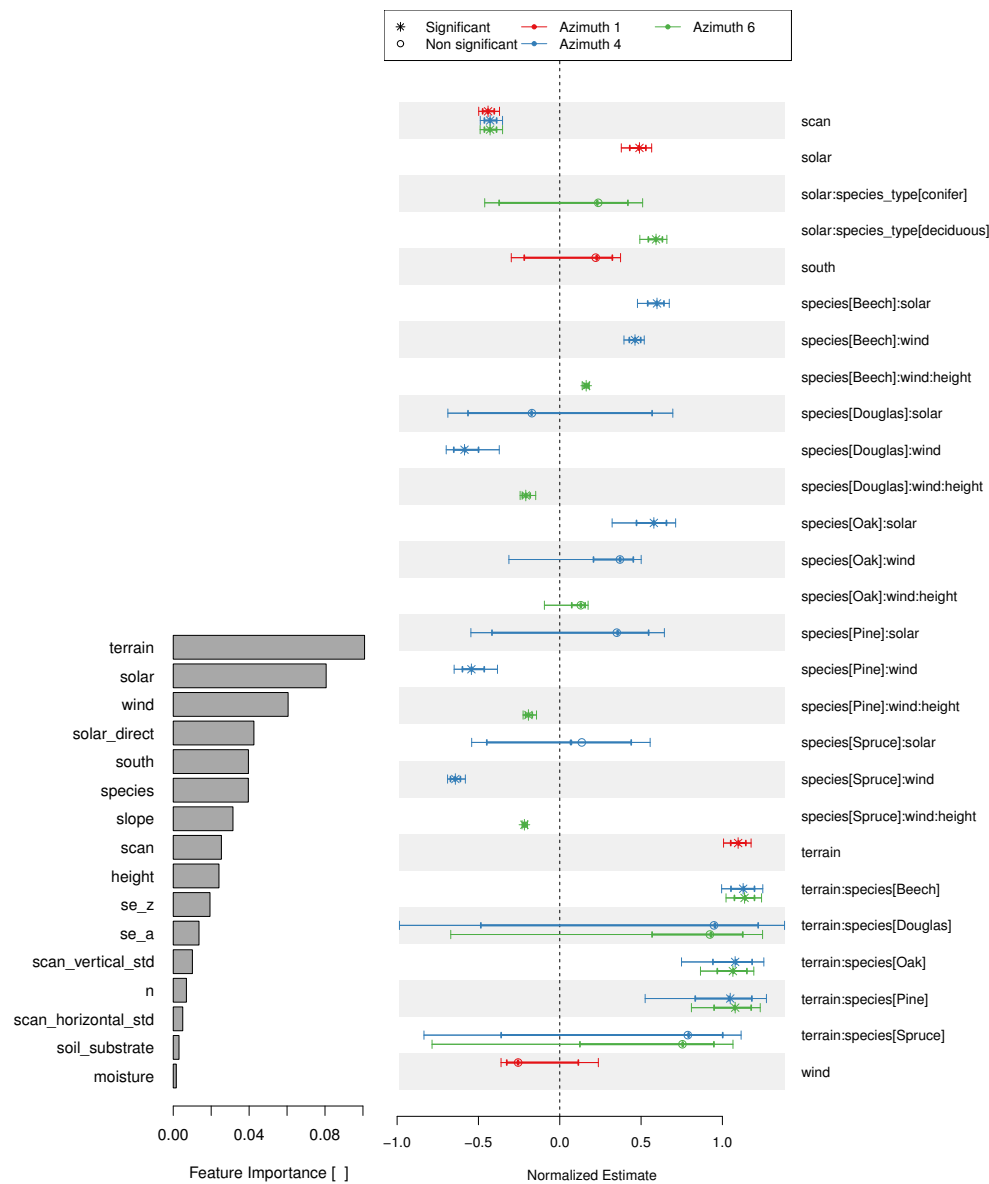
Figure 17. Permutation feature importance of the Random Forest regressor (**left**) and regression coefficients of selected linear trunk orientation models (**right**). The coefficients are scaled by $\sqrt[3]{}$ to display them in the same plot. The given 98% and 80% confidence intervals are based on bootstrapping.

Table 6. Average explained variance per regression coefficient for the linear trunk orientation models. Coefficients significantly different from zero are marked in bold. Significant inter group differences are marked in italics. Measures are based on the training data set T_t.

	Azimuth 1	Azimuth 2	Azimuth 3	Azimuth 4	Azimuth 5	Azimuth 6
terrain	**8.9**%					
terrain:species		*8.5%*	*8.3%*	*8.4%*	*8.5%*	*8.5%*
solar	**4.1**%					
solar:species		*6.2%*	*6.3%*	*6.3%*	*6.1%*	
solar:species_type						*6.2%*
wind	−0.1%					
wind:species		*2.9%*	*3.2%*	*3.2%*		
wind:species:height					*3.6%*	*3.6%*
south	0.3%		−0.1%			
south:species		*0.5%*				
scan	**0.9**%	**0.8**%	**0.8**%	**0.8**%	**0.8**%	**0.8**%
Total	**14.0**%	**18.8**%	**18.5**%	**18.6**%	**19.1**%	**19.1**%

Table 7. Accuracy metrics of the trunk orientation prediction $t_{\hat{\phi}}$ (coefficient of determination R^2, mean angular deviation $\overline{T_{\epsilon_\phi}}$, circular standard deviation $\sigma_{T_{\epsilon_\phi}}$ and the amount of trunks inclined with the model $T_{|\epsilon_\phi|<90°}$). The metrics are based on the training data set T_t.

	RF	Azimuth 1	Azimuth 2	Azimuth 3	Azimuth 4	Azimuth 5	Azimuth 6
R^2	0.271	0.127	0.178	0.177	0.178	0.184	0.181
$\overline{T_{\epsilon_\phi}}$	0.1°	−1.0°	−3.6°	−1.2°	−0.4°	0.1°	−0.1°
$\sigma_{T_{\epsilon_\phi}}$	61.8°	82.7°	77.3°	76.9°	76.7°	76.4°	76.4°
$T_{|\epsilon_\phi|<90°}$	82.8%	71.9%	74.7%	75.0%	75.0%	75.3%	75.3%

Table 8 highlights that the explanatory power of the vectors strongly depends on tree species. In particular, with 22.6% of the variance explained, the orientation of deciduous trees can be modeled more accurately than the orientation of the conifers with only 10.7%. In detail, the orientation of the deciduous trees (beech and oak) is driven by interactions with the site aspect resulting in about 20% of the variance explained by *terrain:species* and *solar:species_type*, while with about 2% the wind has a minor, but significant contribution.

In contrast, with 11.1% the orientation of the spruce is driven by the interaction of the tree height with *wind*, while the site aspect has no significant effect. In general, the effect of *solar* is not significant for conifers, while the *terrain* is only significant for pine. For Douglas and pine, the interaction of tree height with *wind* has a considerable effect on the trunk orientation.

Table 8. Average explained variance per regression coefficient of the regression model *Azimuth 6* depending on tree species. The measures are based on the validation data set T_v. If applicable, coefficients significantly different from zero are highlighted in bold.

	Beech	Douglas	Oak	Pine	Spruce	Conifer	Deciduous	Total
terrain:species	**11.4**%	3.2%	**9.8**%	**5.7**%	0.3%	2.2%	11.1%	8.4%
solar:species_type	9.0%	0.2%	8.0%	0.3%	0.0%	0.1%	**8.8**%	6.1%
wind:species:height	**2.0**%	**5.7**%	0.6%	**4.1**%	**11.1**%	8.3%	1.7%	3.7%
scan	1.2%	0.0%	0.8%	0.8%	−0.4%	−0.0%	1.1%	**0.8**%
Total	23.7%	9.1%	19.2%	10.9%	11.0%	10.7%	22.7%	19.0%

Although the orientation of the trunks has been used to weight the linear equations, Figure 18 reveals that the predicted orientation $t_{\hat{\phi}}$ is not fully independent from the observed orientation t_ϕ. In particular,

trunks inclined to the South-South-East and the East-North-East axes show the lowest residuals. In addition, the accuracy of the model depends on the tree species and site aspect, while the effect of the scan direction is minor.
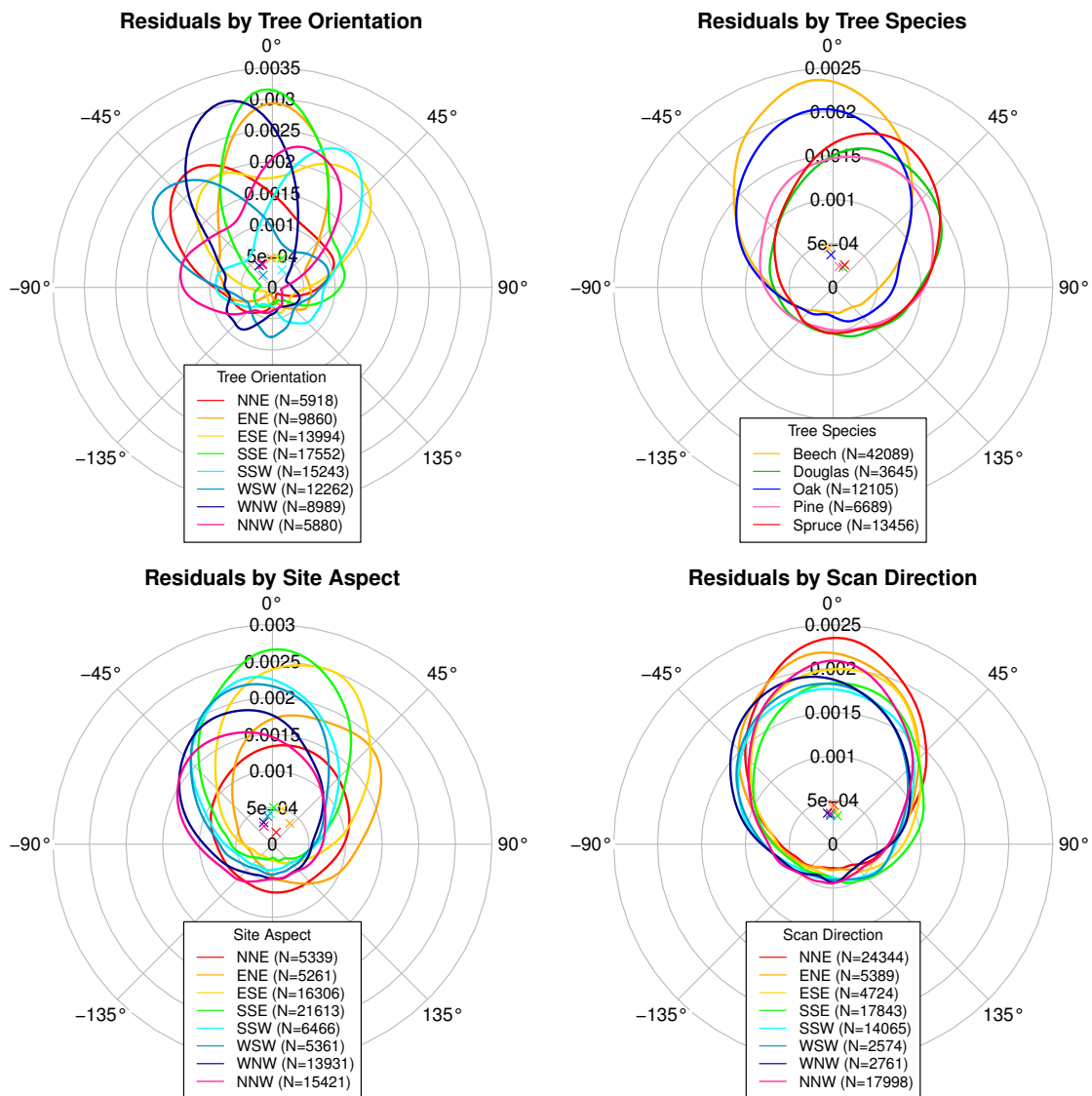

Figure 18. Residuals of the linear trunk orientation model *Azimuth 6* depending on observed trunk orientation (**top-left**), tree species (**top-right**), site aspect (**bottom-left**) and scan direction (**bottom-right**).

5. Discussion

The given study demonstrates that ALS detected tree trunks show preferred vector orientations (Hypothesis 1). In particular, most of the trunks lean down-slope (Hypothesis 2) and a general tendency to the South can be observed. The linear regression models imply that this interaction is highly species specific (Hypothesis 3) and strongest for sites facing to the South, but diminishes for sites directed away from the sun. Almost 20% of the predicted orientation of the beech is related to the site aspect, compared to about 0% for spruce. These observations are in alignment with the expected species dependent interaction between phototropism and tilted terrains as investigated by Matsuzaki et al. [6]. In the given study,

the interaction between the site aspect and the direct solar radiation is suitable to explain the observed South orientation of the trunks.

With 11.1% (spruce), 5.7% (Douglas) and 4.1% (pine) of the variance explained by the interaction between tree species and tree height with the dominant wind direction and speed, the conifers tend to lean leeward (Hypothesis 4). With only 2.0% for beech and 0.6% for oak, this effect is minor for deciduous trees. These results are in alignment with Gardiner et al. [9] and Fourcaud et al. [17], since in particular spruce are highly vulnerable to wind drag due to their compact crown shape and their flat roots. By adaptive growth of the roots [9,13] and reducing the wind-drag by realigning the crown shape, the wind resistance is increased [9,11]. Both effects might cause the observed leeward orientation of the stems.

No significant effect of the soil substrate as well as the soil moisture on the tree inclination can be found in the given study. Since soil characteristics [16] and soil moisture [9,44] have been expected to influence the resistance against toppling of trees, the given results are contrary to Hypothesis 6.

As a confounding variable, the ALS acquisition direction shows a minor, but significant effect on the observed trunk inclination and orientation. Since the systematic West-East flight pattern in combination with the Southeast-Northwest orientation of the mountain range makes an auto-correlation with the preferred South inclination highly probable, no final conclusion on Hypothesis 5 can be drawn in this study.

5.1. Strengths and Limitations

The high uncertainty of ALS detected trunk vectors results in hidden patterns. Even for the significantly inclined trunks, more than half of the observed trunk inclination is associated with attributes expressing the uncertainty of the vectors. As a consequence, a huge number of tree trunks need to be identified to make qualitative statements on the underlying causes for trunk inclination and orientation.

To address this issue, the trunk detection method developed for this study is designed to rapidly identify a majority of doubtless linear structures based on heuristics. Its simplicity allows for a full analysis of the study area covering hundreds of square kilometers within a reasonable amount of time (in the magnitude of several hours). By excluding all trunk vectors with an uncertain orientation, hidden patterns can be revealed. Since the algorithm provides a biased sample of dominant trees—for example, by missing trees of the understory—effects like those postulated by Ishii and Higashi [7] cannot be investigated by this approach. To gain a deeper insight, an accuracy assessment of the trunk detection algorithm and an analysis of the detection rates for different habitats should be applied for future studies. Additionally, in this study the significance of the conclusions is limited to the tree trunks only. Statements on the influence of the underlying causes on the shape and orientation of the tree crowns are inadmissible.

By predicting the trunk orientation using a linear combination of explanatory vectors, qualitative statements on the effect directions can be drawn. The strength of this knowledge-driven approach lies in its simplicity, since the orientation of the trunks is modeled as the sum of the explanatory vectors only. In consequence the effects postulated by Gardiner et al. [9] and Matsuzaki et al. [6] can be empirically confirmed in this study. A limitation of using a linear combination of vectors is given in its inability to describe effects oblique to the vector directions. In consequence such effects can be described by an interaction of the vectors only. To incorporate non linear interaction terms, a detailed prior knowledge on the underlying causes is required. In addition to these conceptual limitations, the ordinary least squares regression cannot achieve the mathematically optimal solution, since the model minimizes the residuals $\epsilon_x + \epsilon_y$ instead of $\epsilon = \sqrt{\epsilon_x^2 + \epsilon_y^2}$ (see Equation (5)).

As an empirical pre-study to investigate the usefulness of ALS data to study the tree trunk inclination and orientation, the input data is not optimal for a detailed thematic analysis. In particular, the geometric resolution of the wind dataset used in this study is much too coarse to inspect the small-scale variation of the wind direction and speed and their effect on the inclination of trees. But, due to the complex interaction

between forests and the terrain [9], a sufficient accuracy on the individual tree level might still not be achievable. Due to the low variability of the wind directions in the study area, it might correlate with the trunk orientation by coincidence. Other factors influencing the wind-drag, like the forest structure, tree spacing and site conditions [9] could not be considered in the given study.

The tree species classification map used in the given study shows some weaknesses in the investigated area. In consequence, species dependent effects might have been covered due to miss-classified tree trunks. The explanatory power of the soil moisture is limited due to the spatial resolution of the soil moisture map. Also due to the spatial auto-correlation between slope inclination, soil moisture and tree species, potential effects might have been covered. The soil substrate has not shown any significant effect on the trunk inclination. This is probably caused by the low spatial resolution and the coarse class division, which give poor information on the actual structural integrity of the soils. Again, more suitable input data and a detailed geo-spatial analysis could reveal hidden effects.

Despite the weaknesses of the input data, for the first time—to the knowledge of the authors—the given study can empirically confirm systematic trunk inclination patterns using ALS data. In particular, several hypotheses on the trunk orientation can be addressed. Effects like the species dependent interaction between phototropism in tilted terrains—as postulated by Matsuzaki et al. [6]—can be confirmed.

5.2. Future Research

Having shown that the ALS derived trunk orientations are in alignment with the current state of knowledge on tree inclination, a transfer to other study areas seems promising. Trunk detection might serve as a tool to empirically confirm assumptions on predominant tree orientations on large scales. With more robust detection methods, the number of identified trunks and the accuracy of the trunk vectors may be increased, allowing for more detailed analyses. In general, a focus should be set on the geo-spatial analysis of the trunk inclination and orientation while comparing this information with terrestrial reference data. In this manner, the investigation of the ALS derived trunk inclination and orientation may be used to identify landslide areas and to estimate the potential wind drag to assess the risk of wind throw. In this context, a supplementary detection of fallen trees and the investigation of storm events would provide valuable information.

6. Conclusions

This study has shown that tree trunks detected using airborne laser scanning can provide information on preferred stem inclination angles and directions. Although a significant proportion of the trunk inclination variance is associated with the uncertainty of the fitted trunk vectors, an empirical analysis has shown that the observed orientation of the significantly inclined trunks is in alignment with today's state of knowledge. In particular, weighted least squares regression models—describing the trunk orientation as a linear combination of directed explanatory variables—have confirmed a highly species specific down-slope inclination of the trunks. While the orientation of beech and oak is dominated by the site aspect—associated with the interaction of phototropism and tilted terrains—the spruce tend to lean leeward. The orientation of the conifers Douglas and pine seems to be driven by both, aspect and wind.

Given these results, the detection of tree trunks using airborne laser scans might be a promising tool to investigate the effect of various influence factors on the tree inclination and orientation. In addition to empirically confirming findings of fundamental research, the presented methods might be used to identify landslides or to address the risk of wind throw on large scales.

Abbreviations

The following abbreviations are used in this manuscript:

ALS	Airborne Laser Scanning
GPS	Global Positioning System
IoU	Intersection over Union
LiDAR	Light Detection And Ranging
LVermGeo	Landesamt für Vermessung und Geobasisinformation Rheinland-Pfalz
NNE	North-North-East
ENE	East-North-East
ESE	East-South-East
SSE	South-South-East
SSW	South-South-West
WSW	West-South-West
WNW	West-North-West
NNW	North-North-West
I	Relative direct solar radiation.
δ	Solar declination angle in °.
ϕ_g	Geographic latitude in °.
τ	Hour of the day.
β	Slope of the surface in °.
γ_s	Site aspect in °.
θ	Declination angle °.
h	True sun height °.
$\hat{h}$	Apparent sun height °.
DOY	Day of year.

Appendix A. Solar Radiation

In the given study, the relative direct solar radiation at a specific site I is calculated according to Reference [38] using Equations (A1)–(A6):

$$I = \frac{\sin(\theta)}{\sin(\hat{h})} \tag{A1}$$

$$\theta = \sin^{-1}\left(\sin(h) \cdot \cos(\beta) + \cos(h) \cdot \cos(\gamma_s - \gamma) \cdot \sin(\beta)\right) \tag{A2}$$

$$h = \sin^{-1}\left(\sin(\phi_g) \cdot \sin(\delta) - \cos(\phi_g) \cdot \cos(\delta) \cdot \cos\left(\frac{2\pi}{24}\tau\right)\right) \tag{A3}$$

$$\hat{h} = h + \frac{1.4705}{3.0427 + h} - 0.0158 \tag{A4}$$

$$\gamma = \sin^{-1}\left(\frac{\cos(\delta) + \sin\left(\frac{2\pi}{24}\tau\right)}{\cos(h)}\right) \tag{A5}$$

$$\begin{aligned}
\delta = 0.7896 & \\
& - 23.2559 \cdot \cos\left(\frac{2\pi \cdot \text{DOY}}{365} + 0.1582\right) \\
& - 0.3915 \cdot \cos\left(\frac{4\pi \cdot \text{DOY}}{365} + 0.0934\right) \\
& - 0.1764 \cdot \cos\left(\frac{6\pi \cdot \text{DOY}}{365} + 0.4539\right)
\end{aligned} \tag{A6}$$

References

1. Lamprecht, S.; Stoffels, J.; Dotzler, S.; Haß, E.; Udelhoven, T. aTrunk—An ALS-Based Trunk Detection Algorithm. *Remote Sens.* **2015**, *7*, 9975. [CrossRef]
2. Chen, R.; Rosen, E.; Masson, P.H. Gravitropism in Higher Plants. *Plant Physiol.* **1999**, *120*, 343–350. [CrossRef]
3. Christie, J.M.; Murphy, A.S. Shoot phototropism in higher plants: New light through old concepts. *Am. J. Bot.* **2013**, *100*, 35–46. [CrossRef]
4. Iino, M. Phototropism in higher plants. In *Photomovement*; Elsevier: Amsterdam, The Netherlands, 2001; pp. 659–811. [CrossRef]
5. Johns, J.W.; Yost, J.M.; Nicolle, D.; Igic, B.; Ritter, M.K. Worldwide hemisphere-dependent lean in Cook pines. *Ecology* **2017**, *98*, 2482–2484. [CrossRef] [PubMed]
6. Matsuzaki, J.; Masumori, M.; Tange, T. Stem Phototropism of Trees: A Possible Significant Factor in Determining Stem Inclination on Forest Slopes. *Ann. Bot.* **2006**, *98*, 573–581. [CrossRef] [PubMed]
7. Ishii, R.; Higashi, M. Tree coexistence on a slope: an adaptive significance of trunk inclination. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **1997**, *264*, 133–139. [CrossRef]
8. Wistuba, M.; Malik, I.; Gärtner, H.; Kojs, P.; Owczarek, P. Application of eccentric growth of trees as a tool for landslide analyses: The example of Picea abies Karst. in the Carpathian and Sudeten Mountains (Central Europe). *CATENA* **2013**, *111*, 41–55. [CrossRef]
9. Gardiner, B.; Berry, P.; Moulia, B. Review: Wind impacts on plant growth, mechanics and damage. *Plant Sci.* **2016**, *245*, 94–118. [CrossRef]
10. Schindler, D.; Bauhus, J.; Mayer, H. Wind effects on trees. *Eur. J. For. Res.* **2011**, *131*, 159–163. [CrossRef]
11. Telewski, F.W.; Jaffe, M.J. Thigmomorphogenesis: Field and laboratory studies of Abies fraseri in response to wind or mechanical perturbation. *Physiol. Plant.* **1986**, *66*, 211–218. [CrossRef]
12. Telewski, F.W. Is windswept tree growth negative thigmotropism? *Plant Sci.* **2012**, *184*, 20–28. [CrossRef] [PubMed]
13. Nicoll, B.C.; Ray, D. Adaptive growth of tree root systems in response to wind action and site conditions. *Tree Physiol.* **1996**, *16*, 891–898. [CrossRef] [PubMed]
14. Planck, N.R.V.; MacFarlane, D.W. Branch mass allocation increases wind throw risk for Fagus grandifolia. *For. Int. J. For. Res.* **2019**, *92*, 490–499. [CrossRef]
15. Gardiner, B.; Byrne, K.; Hale, S.; Kamimura, K.; Mitchell, S.J.; Peltola, H.; Ruel, J.C. A review of mechanistic modelling of wind damage risk to forests. *Forestry* **2008**, *81*, 447–463. [CrossRef]
16. Smiley, E.T.; Key, A.; Greco, C.; others. Root barriers and windthrow potential. *J. Arboric.* **2000**, *26*, 213–217.
17. Fourcaud, T.; Ji, J.N.; Zhang, Z.Q.; Stokes, A. Understanding the Impact of Root Morphology on Overturning Mechanisms: A Modelling Approach. *Ann. Bot.* **2007**, *101*, 1267–1280. [CrossRef]
18. Lu, X.; Guo, Q.; Li, W.; Flanagan, J. A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *94*, 1–12. [CrossRef]

19. Polewski, P.; Yao, W.; Heurich, M.; Krzystek, P.; Stilla, U. Free shape context descriptors optimized with genetic algorithm for the detection of dead tree trunks in als point clouds. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *1*, 41–48. [CrossRef]

20. Shendryk, I.; Broich, M.; Tulbure, M.G.; Alexandrov, S.V. Bottom-up delineation of individual trees from full-waveform airborne laser scans in a structurally complex eucalypt forest. *Remote Sens. Environ.* **2016**, *173*, 69–83. [CrossRef]

21. Amiri, N.; Polewski, P.; Yao, W.; Krzystek, P.; Skidmore, A.K. Detection of Single Tree Stems in Forested Areas from High Density Als Point Clouds Using 3D Shape Descriptors. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *IV-2/W4*, 35–42. [CrossRef]

22. Windrim, L.; Bryson, M. Detection, Segmentation, and Model Fitting of Individual Tree Stems from Airborne Laser Scanning of Forests Using Deep Learning. *Remote Sens.* **2020**, *12*, 1469. [CrossRef]

23. Chen, W.; Hu, X.; Chen, W.; Hong, Y.; Yang, M. Airborne LiDAR Remote Sensing for Individual Tree Forest Inventory Using Trunk Detection-Aided Mean Shift Clustering Techniques. *Remote Sens.* **2018**, *10*, 1078. [CrossRef]

24. Zhen, Z.; Quackenbush, L.J.; Zhang, L. Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sens.* **2016**, *8*, 333. [CrossRef]

25. Bucksch, A.; Lindenbergh, R.; Menenti, M. SkelTre. *Vis. Comput.* **2010**, *26*, 1283–1300. [CrossRef]

26. Chen, H.H.; Huang, T.S. A survey of construction and manipulation of octrees. *Comput. Vis. Graph. Image Process.* **1988**, *43*, 409–431. [CrossRef]

27. Reeb, G. Sur les points singuliers d'une forme de Pfaff compltèment intégrable ou d'une fonction numérique. *C. R. Hebd. Sances Acad. Sci.* **1946**, *222*, 847–849.

28. Razak, K.A.; Bucksch, A.; Straatsma, M.; Westen, C.J.V.; Bakar, R.A.; de Jong, S.M. High density airborne lidar estimation of disrupted trees induced by landslides. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium—IGARSS, Melbourne, Australia, 21–26 July 2013. [CrossRef]

29. Kini, A.U.; Popescu, S.C. *TreeVaW: A Versatile Tool for Analyzing Forest Canopy LIDAR Data: A Preview with an Eye towards Future*; ASPRS: Kansas City, MO, USA, 2004.

30. MUEEF. Daten der Forstlichen Standortkartierung—InGrid-Portal. 2013. Available online: https://www.portalu.rlp.de/trefferanzeige?docuuid=54C251AF-542E-11D7-B776-0002A5CE70F9 (accessed on 23 July 2020).

31. Stoffels, J.; Hill, J.; Sachtleber, T.; Mader, S.; Buddenbaum, H.; Stern, O.; Langshausen, J.; Dietz, J.; Ontrup, G. Satellite-Based Derivation of High-Resolution Forest Information Layers for Operational Forest Management. *Forests* **2015**, *6*, 1982–2013. [CrossRef]

32. RIEGL. LMS-Q560. Available online: http://www.riegl.com/uploads/tx_pxprriegldownloads/10_DataSheet_Q560_20-09-2010_01.pdf (accessed on 27 September 2016)

33. Lamprecht, S. Pyoints: A Python package for point cloud, voxel and raster processing. *J. Open Source Softw.* **2019**, *4*, 990. [CrossRef]

34. Krähenmann, S.; Walter, A.; Brienen, S.; Imbery, F.; Matzarakis, A. Monthly, daily and hourly grids of 12 commonly used meteorological variables for Germany estimated by the Project TRY Advancement. *Theor. Appl. Climatol.* **2016**. [CrossRef]

35. Landesamt für Geologie und Bergbau Rheinland-Pfalz. Bodenübersichtskarte 1:200.000 (BÜK200)—CC6302 Trier. 2019. Available online: https://www.lgb-rlp.de/karten-und-produkte/online-karten/online-bodenkarten.html (accessed on 17 December 2019).

36. Torr, P.; Zisserman, A. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Comput. Vis. Image Underst.* **2000**, *78*, 138–156. [CrossRef]

37. Cressman, G.P. An Operational Objective Analysis System. *Mon. Weather Rev.* **1959**, *87*, 367–374. [CrossRef]

38. Gassel, A. Beiträge zur Berechnung Solarthermischer und Exergieeffizienter Energiesysteme. Ph.D. Thesis, Fakultät Maschinenwesender Technischen Universität Dresden, Dresden, Germany, 1996.

39. Sheather, S.J.; Jones, M.C. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *J. R. Stat. Soc. Ser. B* **1991**, *53*, 683–690. [CrossRef]

40. Wright, M.N.; Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*. [CrossRef]

41. Mardia, K.V. *Statistics of Directional Data*; Academic Press: London, UK, 1972. [CrossRef]

42. Agostinelli, C.; Lund, U. *R Package Circular: Circular Statistics (Version 0.4-93)*; Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University: Venice, Italy; Department of Statistics, California Polytechnic State University: San Luis Obispo, CA, USA, 2017.

43. Jammalamadaka, S.R.; SenGupta, A. *Topics in Circular Statistics*; World Scientific: Singapore, 2001. [CrossRef]

44. Gardiner, B.; Schuck, A.; Schelhaas, M.; Orazio, C.; Blennow, K.; Nicoll, B. (Eds.) *Living with Storm Damage to Forests*; Number 3 in What Science Can Tell Us; European Forest Institute: Joensuu, Finland, 2013.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter VI

Synthesis

VI.1 Summary

The research tasks of the given thesis have been addressed in four major objectives. A particular focus has been set on the potential of ALS detected tree stems for forest applications. Next to the development of methods for stem detection, the accuracy of the trunk positions (Objective I) as well as the trunk inclination characteristics (Objective II) have been explored in detail. To be able to address these overarching tasks and to exploit the potential of ALS detected trunks, the elaboration of tools for the processing and analysis of ALS- and TLS-derived point clouds (Objective IV), as well as data fusion for forest applications (Objective III) have received considerable attention. In the following, the main findings of the four objectives are briefly summarized and brought into a broader context. In particular, the potential of stem detection for forest research will be discussed.

VI.1.1 Objective I: Tree Trunk Detection Using ALS

To address Objective I, two algorithms for tree stem detection using discrete ALS data have been developed. Both algorithms exploit the structure of the point clouds and do not rely on a prior identification of tree crowns. To overcome some issues with the parametrization of the approach presented in Chapter II, the method of Chapter V uses simplified heuristics with just one core parameter to ease its application in large scale study areas.

Chapters II and V have shown that ALS is suitable to identify individual tree trunks in managed mixed-species forest stands of the mid-latitudes. The most relevant findings of Chapter II are that the stem detections are very reliable (commission error below 5 %) and accurate in positioning (RMSE less than 0.8 m). In comparison, typical methods for tree crown detection achieve commission errors ranging from about 10 % up to above 110 % (Eysn et al., 2015) and an RMSE of the displacement ranging from 0.7 m to 3 m (Eysn et al., 2015; Kaartinen et al., 2012; Edson and Wing, 2011). Since the trunk detection relies on the laser beam hitting the stems, the detection rates might strongly vary,

depending on stand characteristics, flight pattern and ALS pulse density. But in general, omission errors of stem detection are expected to be much higher compared to crown detection.

Chapter II confirms the observation of Reitberger et al. (2009) that by combining tree crown detection with stem detection, the accuracy of the determined tree positions can be improved significantly. In addition, by a mutual confirmation of the tree positions by a detection of the stems and the crowns, the commission errors can be decreased.

VI.1.2 Objective II: Causes for Tree Trunk Inclination

The studies presented in Chapters II and V have shown that prevailing trunk inclinations of ALS detected tree stems can be observed. To best knowledge of the author, the later study is the first to show that ALS detected tree stems are suitable to study prevailing trunk inclination angles and directions. In addition, the effect of topography, wind, soil properties and scan direction could be investigated using linear regression models. In particular, the results are in line with today's state of knowledge, while a highly species-specific down-slope inclination of the trunks could be confirmed. In detail, the orientation of beech and oak is dominated by the site aspect, while spruce tend to lean leeward. The orientation of Douglas fir and pine seems to be driven by both, aspect and wind.

Although Chapter V could show the suitability of ALS detected tree stems to study species-specific inclination characteristics over large scales, a significant proportion of the trunk inclination variance is associated with the uncertainty of the fitted trunk vectors. Hence, meaningful information on the inclination can only be extracted for groups of trees rather than at the individual tree level.

VI.1.3 Objective III: Methods for Data Fusion

Several methods for data fusion—1) linking individual trees from various data sources; 2) geo-referencing of forest survey data; 3) marker-less alignment of point clouds—have been developed and implemented. Synergies between the algorithmically related tasks have been exploited and integrated in the *Pyoints* package as presented in Chapter IV.

A post-processing strategy for geo-referencing forest inventory plots using ALS detected trees has been elaborated on in Chapter III. By generating synthetic forest stands using the characteristics of the inventory plots, a random forest classifier has been trained to predict the matching probabilities of field measured and ALS determined tree positions. With this method probably 60 % of the

investigated forest inventory plots could be geo-referenced. In the same study, methods for assigning individual trees of differing data sets have been developed and successfully tested. A further finding is that the highest probabilities for a correct co-registration have been achieved in heterogeneous stands, which is in line with the findings of other authors. In addition, the results indicate that the GNSS displacement of the inspected BWI plots can reach up to 21 m, while 80 % of the plots are characterized by errors between 1.4 m and 8.7 m.

Furthermore, algorithms for a marker-less alignment of several point clouds have been developed. An early version of the method has been used in Chapter II to support the alignment of TLS detected tree stems and to generate a common data set of terrestrial reference positions. A further developed algorithm, based on the methods presented in Appendix B, has been applied successfully in the EU Horizon 2020 project PANtHEOn (grant agreement 774571) to align TLS scans of hazelnut trees. In this context, relative alignment errors below 3 mm could be observed (Lamprecht, 2018).

Since the presented strategy for geo-referencing forest inventory plots is transferable to other study sites and the method provides an objective indicator of the reliability of each result, it is particularly suited to integrate the ITA into forest inventories. Along with methods for assigning trees of different data sets and algorithms for the alignment of multi-dimensional point clouds, a set of tools for data fusion with the focus on forest applications is provided.

VI.1.4 Objective IV: Tools for Point Cloud Processing in Forestry

To address Objective IV, a Python package has been successfully elaborated as presented in Chapter IV. The *Pyoints* package meets the demands defined in Section I.4.1.4 as follows:

- **Geospatial Operations** A common data structure (*GeoRecords*) for point cloud, voxel and raster data has been elaborated. For this purpose, each spatial object (like a point, voxel or raster cell) is understood as a point entity, while the original 1D (unorganized data), 2D (raster) or 3D (voxel) matrix structure of the original data is maintained using a *numpy* (Harris et al., 2020) record array. Each entity is represented by its coordinates and attributes. Next to the records, the spatial projection as well as the extrinsic orientation (roto-translation matrix) are stored to define the spatial reference of the geo-objects. To perform multi-dimensional spatial queries efficiently, the spatial indices *KDTree* (Bentley, 1975) and *Rtree* (Guttman, 1984) are initialized on demand.

- **Point Cloud Processing**   A module for point filtering is provided implementing several algo-
rithms for sub-sampling point clouds. Next to down-sampling the amount of data, filters have
been designed to identify features like local maxima and minima, surfaces, or linear structures.
Such filtered point clouds serve as input for the generation of surface models, interpolation,
alignment of point clouds, clustering or primitive fitting. In addition, common clustering algo-
rithms have been integrated or implemented.

- **Model Fitting**   A module providing several spatial interpolation strategies (e.g., nearest neigh-
bor, linear, polynomial or Cressman interpolation) has been implemented for two or more di-
mensions. A common class for the representation of surface models (DTM, DSM or CHM) has
been implemented. These receive a filtered point cloud as an input, while height predictions are
based on interpolation. Geometric primitives, like vectors, spheres or cylinders, are typically
fitted by mathematical or numerical optimization.

- **Data Alignment**   Algorithms for an automated rigid alignment of point clouds have been de-
veloped and implemented. The well known ICP algorithm has been adapted for a marker-less
automated alignment of several 2D or 3D point clouds as presented in Appendix B.

In summary, the *Pyoints* package provides fundamental, well documented tools for the pro-
cessing and fusion of ALS, TLS and rasterized data with a particular focus on forest applications. Not
only has it been the basis for Chapter V, but also for the development of processing pipelines for the
analysis of a hazelnut orchard in the PANtHEOn project.

VI.2   Conclusions & Outlook

VI.2.1   Future Research on Stem Detection

The complexity of ALS point clouds of heterogeneous forests makes it hard to parameterize
heuristic methods for identifying individual tree stems. The works of Amiri et al. (2017) and Windrim
and Bryson (2020) have shown that methods based on machine learning might have a higher poten-
tial to identify individual stems than heuristic methods. Thus, future algorithms can be expected to
utilize the capabilities of modern machine learning techniques. For high density point clouds (e.g.,
ULS) voxelized methods might prevail, due to the simplicity of the approach and the computational
efficiency. For lower resolution point clouds (e.g., ALS), the low number of points associated with
the stems might still require a full analysis of the 3D data.

Due to their high reliability and their remarkable positional accuracy, ALS detected stems can be seen as complementary to crown detections. Thus, they are particularly suited to improve ITCD by providing accurate positions and by mutually confirming the tree detections. Today, some algorithms extract tree stems after tree segmentation, but also methods using trunk detection-aided tree segmentation are developing (Windrim and Bryson, 2020; Chen et al., 2018). Because of the thematic similarity of tree crown and stem detection, future ITCD algorithms might exploit both at once, features of the tree crowns and the stems. This might be the most promising solution to combine ITCD and stem detection, since assigning stems to tree crowns subsequently is a non trivial task. In summary, the field of ITCD might benefit from more research on individual stem detection.

To assess carbon dynamics and biodiversity, as well as to maintain and manage forests, knowledge on the spatial distribution of dead wood is required (Polewski et al., 2015a). In consequence, another prominent—but rarely addressed—field of research is the detection of standing dead trees (Polewski et al., 2015b). Due to algorithmical synergies between the detection of dead and living stems (c.f. Amiri et al., 2017), more hybrid algorithms can be expected in the future.

Based on the preliminary considerations presented in Appendix A, the commonly used see-saw or parallel line scanning patterns are not optimal for stem detection. If priority is set on stem identification, a Palmer scanner or a scanner inclined toward the flight direction would increase the probability of the laser beam intersecting with the vertically oriented stems.

VI.2.2   Potential of Trunk Inclination Analysis

The given thesis has shown that ALS detected stems are suitable for a large scale analysis of prevailing trunk inclination angles and orientations. In consequence, inclination characteristics might be used to study the growth behavior of various tree species under real-world conditions. This approach could gain fundamental knowledge on the species-specific interaction between gravitropism and phototropism, as well as the response to wind or other factors. Based on the observed preferred leeward orientation of some tree species, the stem inclination might also be suitable to assess the risk of windthrow.

Landslides cause trees to incline in various directions (Razak et al., 2013), contrary to their natural tendency to incline down-slope. After a while the trees reorient their inclination by excentric growth (Wistuba et al., 2013), resulting in bent stems. Both effects, abnormal inclination and bent

trunks, could be studied using ALS detected stems. In consequence ALS, or ULS detected stems might have the potential to serve as indicators for landslides in mountainous terrains.

VI.2.3   Data Processing for Forest Applications

Pyoints has proven to be well suited for TLS-driven forest applications in the context of the PANtHEOn project. Its modular design has shown to be flexible enough to create customized and efficient algorithms for point cloud and image processing. However, Roussel et al. (2020) note that after a major release in 2019, the *lidR* software has achieved a broad acceptance in the research community. Since it has been designed with a particular focus on ALS driven forest applications and increasingly supports TLS, UAV and DAP point clouds, it is a logical and well maintained alternative to *Pyoints*.

As the algorithms for a marker-less scan alignment have been successfully applied to complex multi-stemmed hazelnut trees in context of the PANtHEOn project, the methods can be expected to be transferable to other environments. To co-register multiple scans using this method, the approximate scan location and the approximate orientation (e.g., measured by GNSS and a customary compass) need to be known. Since this can be achieved in a typical managed forest of the mid-latitudes, the given method—as well as alternative state-of-the-art methods—might be usable operationally in forest surveys.

In the future, the assessment of the GNSS errors might influence the design of ALS-driven forest inventories. In the context of timber volume estimation, Hill, Buddenbaum, and Mandallaz (2018) find that in case of positional mismatches the prediction accuracy is increasingly diminished the more the spatial resolution of the ALS features match the size of the inventory plots. Hence, the proposed geo-referencing strategy might enhance the results of ITCD-driven forest inventories. This would particularly be true if the estimated reliability were taken into account when calibrating the statistical models. By introducing the highly reliable and accurate stem locations, the reliability of the tree matching as well as the geo-referencing could further be improved.

Appendix A

Preliminary Considerations on ALS-based Stem Detection

To assess if common line scanning ALS is suited to identify individual tree stems, its sampling characteristics need to be reviewed. Assuming a typical stem can be considered as an almost straight line, at least two (differing) points need to be sampled to characterize its position and orientation. To distinguish a trunk from other objects and to reduce inaccuracies of the determined orientation, more points are of course required.

Since a typical stem can be assumed to be oriented almost vertically, the expected vertical point spacing is of particular interest to assess if a detection is reasonable. In addition, the probability of a laser beam hitting a trunk of a given diameter depends on the ALS's line spacing b and the laser's instantaneous field of view. This theoretical probability is further reduced due to transmission losses and occlusion effects caused by the upper canopy.

A.1   Vertical Point Spacing

The vertical point spacing is presumed to have a significant effect on the possibility of sampling a tree stem. Figure A.1 illustrates the resulting vertical point spacing on a scanned tree trunk assuming the aircraft is located beside the tree.

The relevant geometrical scanning characteristics depend on the aircraft's altitude h, its scanning range $\Theta \in]0, 180[$ as well as the scanning angles $\{\theta_1, \theta_2, ..., \theta_n\}$ with $\theta_i \in \left[-\frac{\Theta}{2}, \frac{\Theta}{2}\right]$ $\forall$ $i \in \{1, 2, ..., n\}$ and $\theta_i < \theta_j$ $\forall$ $j > i$. Consequently, the intra-line distance between the aircraft and the point where the laser beam hits the ground s is determined by Equation A.1. As discussed in Section I.2, the distribution of the intra-line distances $\{s_1, s_2, ..., s_n\}$ depends on the device. But in general, $|s_i - s_{i-1}| \approx |s_{i+1} - s_i|$ $\forall$ $i \in \{2, 3, ..., n-1\}$ can be assumed, meaning the spacing between neighbored ground points on the same scanning line is almost equal.
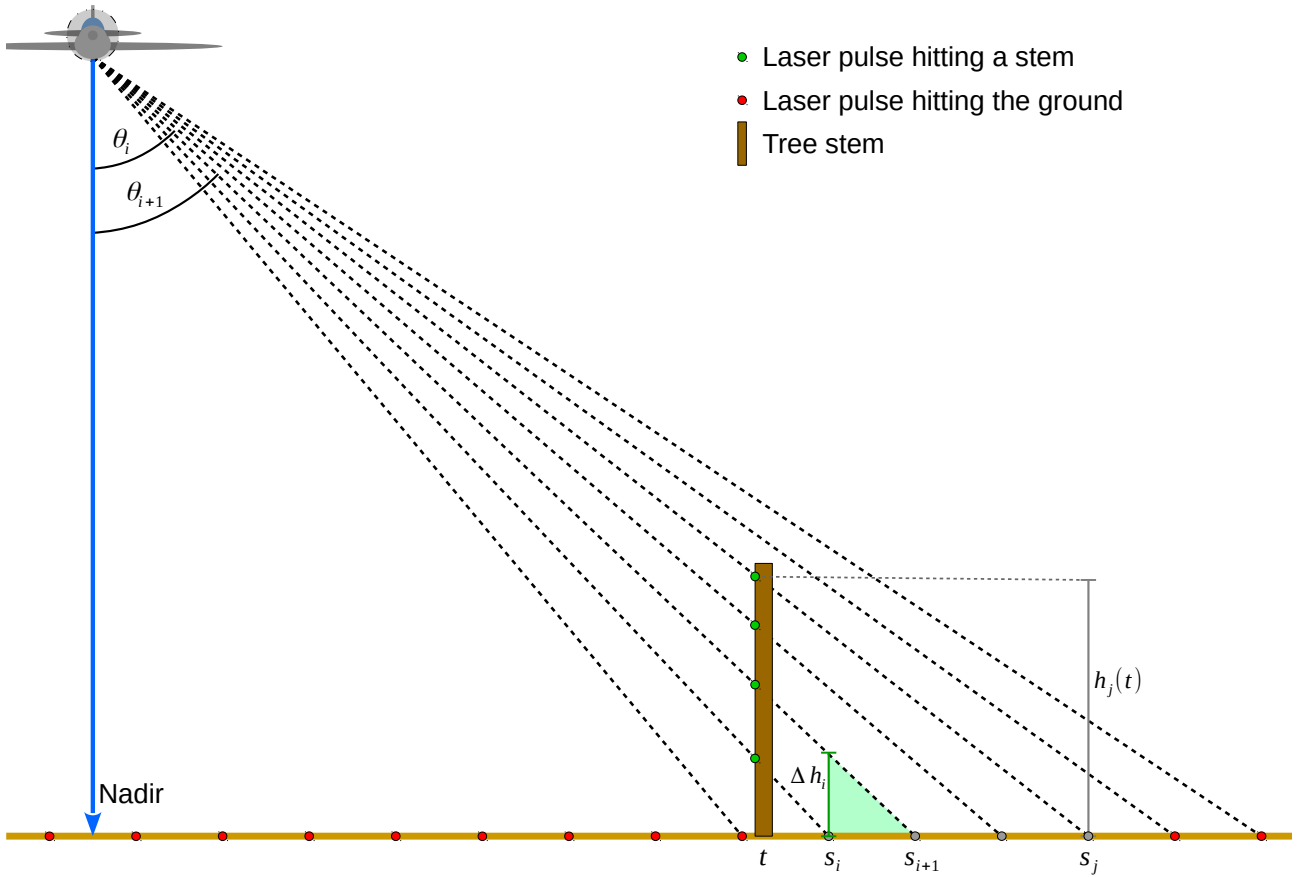
FIGURE A.1: Schematic concept to estimate the vertical point spacing Δh_i of laser pulses reflected by a stem.

Figure A.1 shows a tree stem with intra-line distance t which is hit by several laser pulses $\{i, i+1, ..., j\}$. The height above ground $h_k(t)$ of each of these pulses can be determined by Equation A.2 assuming vertically oriented stems on a flat terrain and $t < s_k \ \forall \ k \in \{i, i+1, ..., j\}$. The vertical point spacing can be approximated by Δh_i using Equations A.3 and A.4, because again almost equal distances of neighbored points can be assumed within a scan line.

$$s_i = h \cdot \tan(\theta_i) \tag{A.1}$$

$$h_i(t) = \frac{s_i - t}{\tan(\theta_i)} \tag{A.2}$$

$$\Delta h_i := h_{i+1}(s_i) \tag{A.3}$$

$$h_j(t) - h_{j-1}(t) \approx \Delta h_i \qquad\qquad (A.4)$$

Based on these thoughts, the vertical point spacing Δh within a stripe of several ALS tracks can be examined. Figure A.2 illustrates the distribution of Δh within a composite of several tracks assuming an altitude h of 600 m, a scanning range Θ of 60° and a constant point spacing for a single track of 0.6 m. The green line indicates the expected vertical point spacing if a stem at a given intra-line distance is captured by the optimal track, which is the track with minimum Δh. For the given parameters, the expected vertical point spacing ranges from about 1 m to 2 m resulting in point densities of about 0.5 points/m to 1 points/m.
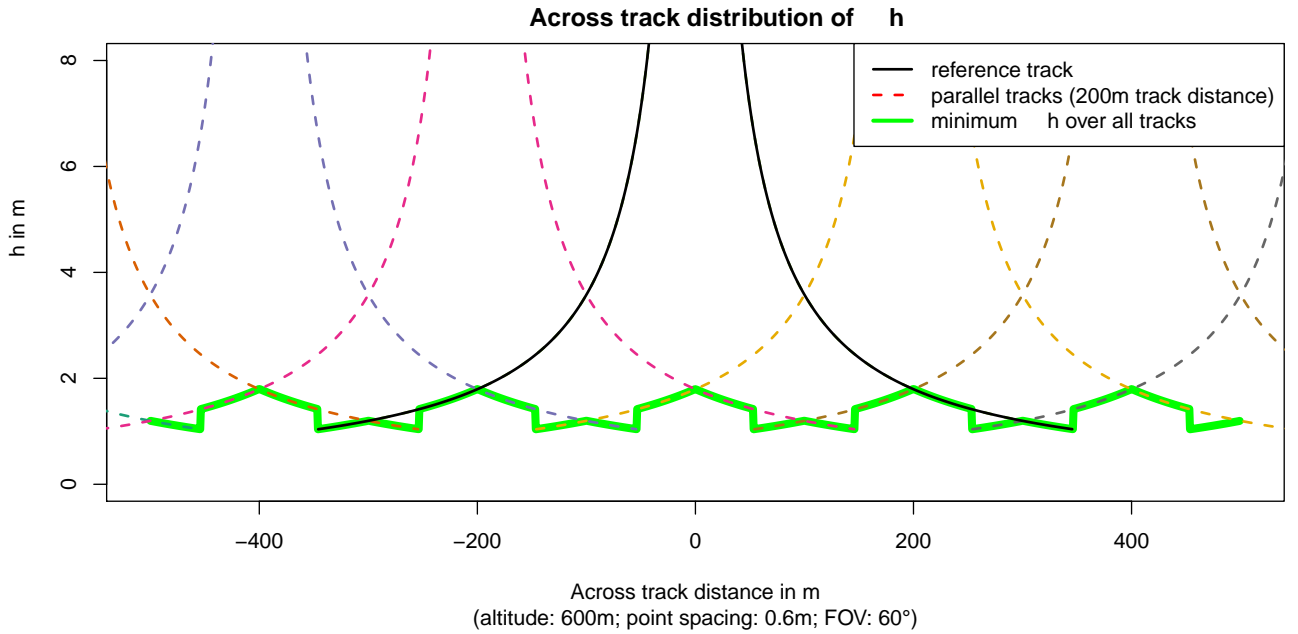
FIGURE A.2: Estimation of the vertical point spacing Δh_i for laser beams hitting a stem from several tracks.

A.2   Line Spacing

Assuming a flat terrain is scanned using a constant scan line distance b, the probability P_I that a (non inclined) isolated trunk of diameter d intersects with a scan line is predicted using Equation A.5. This probability depends on the effective diameter of the laser's footprint δ_θ, which is determined by Equation A.6 using the beam divergence β and the the scanning angle θ. Utilizing the parameter $v \in [0,1]$ the equation also assumes that a minimum proportion of the laser beam needs to be reflected to produce a discrete echo.

$$P_I = \min\left(1, \frac{\delta_\theta + d}{b}\right) \tag{A.5}$$

$$\delta_\theta = \frac{h}{\cos(\theta)} \cdot 2\sin(\beta \cdot v) \tag{A.6}$$

Figure A.3 illustrates the probability that a stem intersects with a scan line of a single track for various scan line spacings under the assumption of δ_θ to be fixed. The figure shows that the probability that a stem intersects with a scan line increases linearly with the stem's diameter. Next to this, the probability increases linearly by decreasing the scan line spacing.
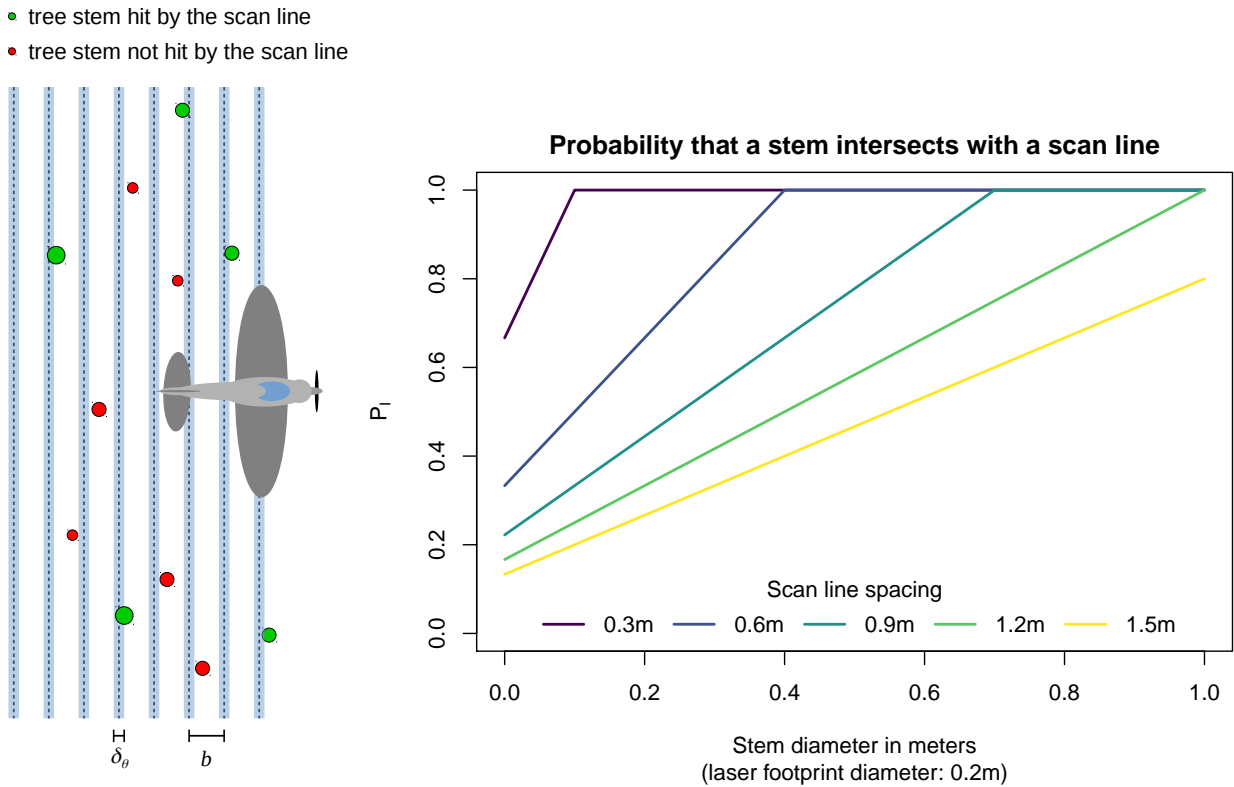
FIGURE A.3: Estimation of the probability that a stem intersects with a scan line P_I assuming a fixed scan line distance b and a fixed diameter of the laser's footprint δ_θ.

A.3 Forest Structure

The influence of the forest structure on the probability of detecting a trunk is hard to predict. In general, due to transmission losses, the higher the density of the canopy, the fewer laser pulses are expected to hit understory trees or a trunk (Aubry-Kientz et al., 2019; Koch et al., 2013; Korpela, Hovi,

and Morsdorf, 2012). Since the density of the canopy depends on the tree species and the structure of the forest, the actual detection rates are expected to be dependent on the species, forest type and season. In general, compared to conifers, a higher portion of laser beams is expected to hit the trunks of leaf-off broad-leaved trees.

Ignoring the vegetation below the canopy, after passing the canopy, a laser beam hits either a trunk or the ground. In consequence, the available amount of laser beams hitting the stems can be estimated to be similar to that hitting the ground.

A.4   Initial Conclusions

The ALS data used in the studies which these considerations are applied to (see Lamprecht, Stoffels, and Udelhoven, 2020 and Lamprecht et al., 2015) are characterized by a footprint diameter of about 0.3 m, an average distance between the scan lines of 0.6 m and a point spacing of about 0.6 m. For all investigated ALS tiles at least 41 % of the pulses reach the ground. In the study area of the given studies 75 % of the inventory trees are characterized by a DBH larger than 23.8 cm and 75 % of the inventory trees are taller than 19.8 m (cf. Lamprecht et al., 2017a). Based on this information and the previous considerations, a detection of tree stems using discrete ALS data seems reasonable, although a large proportion of omissions can be expected.

In detail, if an effective laser's footprint of 0.2 m is assumed (portion of the laser's footprint where the scanner discretization settings would be sensitive to stems), the probability that a stem intersects with a scan line can be assumed to be greater than 70 % (see Figure A.3) for the majority of the dominant trees. Assuming a vertical point spacing of 1 m to 2 m (see Section A.1), a stem with a length of 10 m is expected to be hit 5 to 10 times if it is scanned by the optimal ALS track and the laser beams are not disturbed by the upper canopy. Of course, a stem might be scanned by a non-optimal track, resulting in a lower number of hits, or scanned by several tracks, resulting in additional hits. However, the ability to identify individual stems using the given ALS data will probably be mostly driven by the forest structure and canopy density (see Section A.3). Thus, there is a good chance that trunks can be hit by the laser beams.

In conclusion, a trunk identification seems to be realistic for the given ALS data and study area, but only mature trees are expected to be detectable. In addition, the point densities at the stems

are expected to be low. For the development of algorithms, this means that a stem detection based on the ALS points is expected to be more promising than a detection via voxels.

Appendix B

Point Cloud Alignment

To align several 3D point clouds (intended to be recorded by a TLS), a variant of the iterative closest point (ICP) algorithm (Besl and McKay, 1992) has been developed. As illustrated in Section B.2, it uses the main idea of the normal iterative closest point (NICP) algorithm (Serafin and Grisetti, 2014) to take the point distance into account but also account for the orientation of a point. After identifying matching point pairs between the point clouds, the roto-translation parameters of all point clouds are derived simultaneously by solving an equation system, like presented in Section B.3. Using these concepts, a method to align several terrestrial laser scans recorded in a forested environment is developed in Section B.4.

B.1 Definitions

Having a three-dimensional point cloud $P \subset \mathbb{R}^3$, we can assume a positional error (e.g., caused by GNNS) of $t_P = (t_{Px}, t_{Py}, t_{Pz})$ and a slight twist (e.g., caused by compass errors) of $\omega_P = (\omega_{Px}, \omega_{Py}, \omega_{Pz})$. Thus, the true world coordinates $p' = (p'_x, p'_y, p'_z)$ of an arbitrary point $p \in P$ can be derived by Equation B.1 with R_{ω_P} as defined in Equation B.2:

$$p' = t_P + R_{\omega_P} \cdot p \tag{B.1}$$

$$R_{\omega_P} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega_{Px} & -\sin\omega_{Px} \\ 0 & \sin\omega_{Px} & \cos\omega_{Px} \end{bmatrix} \times \begin{bmatrix} \cos\omega_{Py} & 0 & \sin\omega_{Py} \\ 0 & 1 & 0 \\ -\sin\omega_{Py} & 0 & \cos\omega_{Py} \end{bmatrix} \times \begin{bmatrix} \cos\omega_{Pz} & -\sin\omega_{Pz} & 0 \\ \sin\omega_{Pz} & \cos\omega_{Pz} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{B.2}$$

Assuming almost infinitesimal small rotations, the rotation matrix can be approximated by Equation B.3 (cf. Venuti, 2009):

$$R_{\omega_P} \approx \begin{bmatrix} 1 & -\omega_{Pz} & \omega_{Py} \\ \omega_{Pz} & 1 & -\omega_{Px} \\ -\omega_{Py} & \omega_{Px} & 1 \end{bmatrix} \tag{B.3}$$

B.2 ICP Variant

The initial idea of the NICP algorithm is simplified by disregarding the explicit definition of an objective function and by introducing additional dimensions to the point clouds based on the surface normals instead. In case of 2D point clouds, two dimensions are added, while for 3D point clouds three dimensions are added. By doing so, the algorithm identifies probably matching points based on six-dimensional distances in case of 3D point clouds and four-dimensional distances for 2D point clouds respectively. In summary the roto-translation parameters of two or many similar point clouds are refined iteratively as follows:

While the accuracy is not satisfying:

1. Transform all point clouds using the latest roto-translation parameters.

2. Identify all probably matching point pairs using point normals.

3. Refine the initial roto-translation parameters based on the previously identified point pairs.

To calculate the surface normal of a point, the three characteristic vectors of the k closest points are calculated. If a planar distribution of the points is assumed, the principal vector explaining the smallest proportion of variance represents the point normal. Since a terrestrial laser scanner captures the front of an object only, all point normals facing backwards to the scanner are inverted.

To identify matching point pairs of two 3D point clouds A and B, an epsilon environment $A_{\delta(b)}$ is defined according to Equation B.4 with $d = 6$ so that the resolution δ defines the maximum difference between two points $a \in A$ and $b \in B$ to let the algorithm assume a match between a and b. For example, by setting δ to $\left(0.5, 0.5, 0.1, \sqrt[3]{\sin(10°)}, \sqrt[3]{\sin(10°)}, \sqrt[3]{\sin(10°)}\right)$, point a is assumed to match b, if a horizontal distance of $0.5\,\mathrm{m}$, a vertical difference of $0.1\,\mathrm{m}$ and a normal difference of $10°$ is not exceeded. Please note that the given procedure is applied for all combinations of two points taken from the two given point clouds.

$$a \in A_{\delta(b)} :\Leftrightarrow a_i - b_i \le \delta_i \; \forall \, i \in \{1, 2, \ldots, d\} \text{ and } b \in B \tag{B.4}$$

B.3   Designing an Equation System

An equation system is designed using the basic idea of Venuti (2009). If an object is recorded twice by two laser scans A and B, one or many matching point pairs (a, b) with $a \in A$ and $b \in B$ are created. By roto-translating both point clouds we can assume $a' = b'$ and by using Equation B.1 we receive Equation B.5:

$$t_A + R_{\omega_A} \cdot a = t_B + R_{\omega_B} \cdot b \tag{B.5}$$

If small rotations ω_A and ω_B can be assumed (as typical for terrestrial laser scans), we receive Equation B.6 by compiling Equations B.3 and B.5:

$$\begin{bmatrix} t_{Ax} \\ t_{Ay} \\ t_{Az} \end{bmatrix} + \begin{bmatrix} 1 & -\omega_{Az} & \omega_{Ay} \\ \omega_{Az} & 1 & -\omega_{Ax} \\ -\omega_{Ay} & \omega_{Ax} & 1 \end{bmatrix} \cdot \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} t_{Bx} \\ t_{By} \\ t_{Bz} \end{bmatrix} + \begin{bmatrix} 1 & -\omega_{Bz} & \omega_{By} \\ \omega_{Bz} & 1 & -\omega_{Bx} \\ -\omega_{By} & \omega_{Bx} & 1 \end{bmatrix} \cdot \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} \tag{B.6}$$

Equation B.6 can be rearranged to Equation B.7 and finally to Equation B.8:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -a_z & a_y \\ 0 & 1 & 0 & a_z & 0 & -a_x \\ 0 & 0 & 1 & -a_y & a_x & 0 \end{bmatrix} \cdot \begin{bmatrix} t_{Ax} \\ t_{Ay} \\ t_{Az} \\ \omega_{Ax} \\ \omega_{Ay} \\ \omega_{Az} \end{bmatrix} + \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -b_z & b_y \\ 0 & 1 & 0 & b_z & 0 & -b_x \\ 0 & 0 & 1 & -b_y & b_x & 0 \end{bmatrix} \cdot \begin{bmatrix} t_{Bx} \\ t_{By} \\ t_{Bz} \\ \omega_{Bx} \\ \omega_{By} \\ \omega_{Bz} \end{bmatrix} + \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} \tag{B.7}$$

$$
\begin{bmatrix} b_x - a_x \\ b_y - a_y \\ b_z - a_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -a_z & a_y & -1 & 0 & 0 & 0 & b_z & -b_y \\ 0 & 1 & 0 & a_z & 0 & -a_x & 0 & -1 & 0 & -b_z & 0 & b_x \\ 0 & 0 & 1 & -a_y & a_x & 0 & 0 & 0 & -1 & b_y & -b_x & 0 \end{bmatrix} \cdot \begin{bmatrix} t_{Ax} \\ t_{Ay} \\ t_{Az} \\ \omega_{Ax} \\ \omega_{Ay} \\ \omega_{Az} \\ t_{Bx} \\ t_{By} \\ t_{Bz} \\ \omega_{Bx} \\ \omega_{By} \\ \omega_{Bz} \end{bmatrix}
\tag{B.8}
$$

Based on Equation B.8, the equation system defined by Equation B.9 is built using at least $m \geq 3$ matching point pairs (a_i, b_i) with $a_i \in A$ and $b_i \in B$ for all $i \in \{1, \dots m\}$.

$$
\begin{bmatrix} b_{1x} - a_{1x} \\ b_{1y} - a_{1y} \\ b_{1z} - a_{1z} \\ \vdots \\ b_{mx} - a_{mx} \\ b_{my} - a_{my} \\ b_{mz} - a_{mz} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -a_{1z} & a_{1y} & -1 & 0 & 0 & 0 & b_{1z} & -b_{1y} \\ 0 & 1 & 0 & a_{1z} & 0 & -a_{1x} & 0 & -1 & 0 & -b_{1z} & 0 & b_{1x} \\ 0 & 0 & 1 & -a_{1y} & a_{1x} & 0 & 0 & 0 & -1 & b_{1y} & -b_{1x} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & -a_{mz} & a_{my} & -1 & 0 & 0 & 0 & b_{mz} & -b_{my} \\ 0 & 1 & 0 & a_{mz} & 0 & -a_{mx} & 0 & -1 & 0 & -b_{mz} & 0 & b_{mx} \\ 0 & 0 & 1 & -a_{my} & a_{mx} & 0 & 0 & 0 & -1 & b_{my} & -b_{mx} & 0 \end{bmatrix} \cdot \begin{bmatrix} t_{Ax} \\ t_{Ay} \\ t_{Az} \\ \omega_{Ax} \\ \omega_{Ay} \\ \omega_{Az} \\ t_{Bx} \\ t_{By} \\ t_{Bz} \\ \omega_{Bx} \\ \omega_{By} \\ \omega_{Bz} \end{bmatrix}
\tag{B.9}
$$

For a better readability, parts of the matrices in Equation B.9 are renamed to receive Equation B.10:

$$
\begin{bmatrix} \widehat{AB} \end{bmatrix} = \begin{bmatrix} \widehat{A_B} & \widehat{B_A} \end{bmatrix} \cdot \begin{bmatrix} \widehat{A_{t\omega}} \\ \widehat{B_{t\omega}} \end{bmatrix}
\tag{B.10}
$$

Finally, to align several point clouds, the equation system as defined in Equation B.10 can be extended as illustrated in Equation B.11 for three point clouds A, B and C:

$$\begin{bmatrix} \widehat{AB} \\ \widehat{AC} \\ \widehat{BC} \end{bmatrix} = \begin{bmatrix} \widehat{A_B} & \widehat{B_A} & 0 \\ \widehat{A_C} & 0 & \widehat{C_A} \\ 0 & \widehat{B_C} & \widehat{C_B} \end{bmatrix} \cdot \begin{bmatrix} \widehat{A_{t\omega}} \\ \widehat{B_{t\omega}} \\ \widehat{C_{t\omega}} \end{bmatrix} \tag{B.11}$$

B.4   Alignment of TLS Scans in Forest Surveys

To align TLS scans of a forest, characteristic object geometries need to be identified. As such, trunk sections and terrain characteristics are particularly suited geometries, since they provide information on the horizontal and vertical displacement respectively. In contrast, the understory, twigs and leaves are less suited, since their representation in the point clouds highly depends on the scanning perspective.

In this thesis, a robust method for the alignment of TLS scans has been developed, by splitting the alignment task into three sequential steps: 1) horizontal alignment; 2) vertical alignment; 3) refinement. Since the method should be adapted to the scanner characteristics and the specific environmental conditions, only the rough idea is presented here.

As points associated with tree stems hold much information on the horizontal displacement of terrestrial scans, trunk sections are selected. To limit the computational effort, a duplicate point filter of radius ϵ (e.g., 0.01 m) is applied, and points with only a few neighbors are omitted. To compensate for a vertical displacement, the relative heights above ground (derived from a DEM) are used as z-coordinates. Finally, the ICP variant presented in section B.2 is applied. To avoid the effect of erroneous point matches and to limit the degree of freedom, the procedure is repeated several times with decreasing size of the epsilon environment δ. For example, beginning with δ set to $\left(2, 2, 0.1, \sqrt[3]{\sin(15°)}, \sqrt[3]{\sin(15°)}\}, \sqrt[3]{\sin(15°)}\right)$ an approximate GNSS error of up to 2 m and compass errors of up to $15°$ could be corrected. Afterwards, a fine horizontal registration could be achieved, with δ set to $\left(0.05, 0.05, 0.1, \sqrt[3]{\sin(5°)}, \sqrt[3]{\sin(5°)}, \sqrt[3]{\sin(5°)}\right)$.

To align the scans vertically, a random subset of points associated with the ground is used. After applying the roto-translations (estimated by the horizontal alignment) to each scan, the closest ground point in each neighboring scan is selected. These selected points form a subset of points used for the vertical alignment. Based on these subsets of points, the vertical displacements are estimated

by applying the ICP algorithm once, while using the roto-translations estimated by the horizontal alignment as initial values.

To diminish remaining alignment errors (caused by the separate horizontal and vertical alignment) the point subsets sampled for the horizontal and vertical alignment are merged to a joint subset. Finally, the ICP algorithm is applied a last time, using the roto-translations estimated from the vertical alignment as initial values, providing the refined roto-translation parameters for each input point cloud.

References

Amiri, N., P. Polewski, W. Yao, P. Krzystek, and A. K. Skidmore (2017). "Detection of single tree stems in forested areas from high density ALS point clouds using 3D shape descriptors". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W4, 35–42. DOI: 10.5194/isprs-annals-iv-2-w4-35-2017.

Andersen, H.-E., T. Clarkin, K. Winterberger, and J. Strunk (2009). "An accuracy assessment of positions obtained using survey- and recreational-grade global positioning system receivers across a range of forest conditions within the Tanana valley of interior Alaska". *Western Journal of Applied Forestry* 24 (3), 128–136. DOI: 10.1093/wjaf/24.3.128.

Armston, J., P. Bunting, N. Flood, and S. Gillingham (2015). *PyLidar*. URL: http://www.pylidar.org.

ASPRS (2013). *LAS specification version 1.4 – R13*. The American Society for Photogrammetry & Remote Sensing. 5410 Grosvenor Lane, Suite 210 Bethesda, Maryland 20814-2160.

Atkins, J. W., G. Bohrer, R. T. Fahey, B. S. Hardiman, T. H. Morin, A. E. L. Stovall, N. Zimmerman, and C. M. Gough (2018). "Quantifying vegetation and canopy structural complexity from terrestrial LiDAR data using the FORESTR R package". *Methods in Ecology and Evolution* 9 (10), 2057–2066. DOI: 10.1111/2041-210x.13061.

Aubry-Kientz, M., R. Dutrieux, A. Ferraz, S. Saatchi, H. Hamraz, J. Williams, D. Coomes, A. Piboule, and G. Vincent (2019). "A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests". *Remote Sensing* 11 (9), 1086. DOI: 10.3390/rs11091086.

Baltsavias, E. (1999). "Airborne laser scanning: basic relations and formulas". *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 199–214. DOI: 10.1016/S0924-2716(99)00015-5.

Beland, M., G. Parker, B. Sparrow, D. Harding, L. Chasmer, S. Phinn, A. Antonarakis, and A. Strahler (2019). "On promoting the use of LiDAR systems in forest ecosystem research". *Forest Ecology and Management* 450, 117484. DOI: 10.1016/j.foreco.2019.117484.

Bentley, J. L. (1975). "Multidimensional binary search trees used for associative searching". *Communications of the ACM* 18 (9), 509–517. DOI: 10.1145/361002.361007.

Besl, P. and N. D. McKay (1992). "A method for registration of 3-D shapes". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), 239–256. DOI: 10.1109/34.121791.

References

Bienert, A., K. Pech, and H.-G. Maas (2011). "Verfahren zur Registrierung von Laserscannerdaten in Waldbeständen". *Schweizerische Zeitschrift für Forstwesen* 162 (6), 178–185. DOI: `10.3188/szf.2011.0178`.

Bitterlich, W. (1984). *The relascope idea. Relative measurements in forestry*. Commonwealth Agricultural Bureaux.

BMELV (2011). *Aufnahmeanweisung für die dritte Bundeswaldinventur (BWI³): (2011-2012)*. Ed. by L. u. V. Bundesministerium für Ernährung. Vol. 2. BMELV, Bonn.

Breidenbach, J. and R. Astrup (2013). "The semi-individual tree crown approach". In: *Forestry Applications of Airborne Laser Scanning*. Springer Netherlands, 113–133. DOI: `10.1007/978-94-017-8663-8_6`.

Burt, A., M. Disney, and K. Calders (2019). "Extracting individual trees from LiDAR point clouds using treeseg". *Methods in Ecology and Evolution* 10 (3), 438–445. DOI: `10.1111/2041-210x.13121`.

Cabello-Leblic, A. (2015). "Tree crown delineation". In: Held, A., S. Phinn, M. Soto-Berelov, and S. Jones. *AusCover Good Practice Guidelines: A technical handbook supporting calibration and validation activities of remotely sensed data product*. Ed. by A. Held, S. Phinn, M. Soto-Berelov, and J. S. Vol. 1.2. TERN AusCover. Chap. 11, pp. 191–201.

Calders, K., G. Newnham, A. Burt, S. Murphy, P. Raumonen, M. Herold, D. Culvenor, V. Avitabile, M. Disney, J. Armston, and M. Kaasalainen (2014). "Nondestructive estimates of above-ground biomass using terrestrial laser scanning". *Methods in Ecology and Evolution* 6 (2), 198–208. DOI: `10.1111/2041-210x.12301`.

Chen, R., E. Rosen, and P. H. Masson (1999). "Gravitropism in higher plants". *Plant Physiology* 120 (2), 343–350. DOI: `10.1104/pp.120.2.343`.

Chen, W., X. Hu, W. Chen, Y. Hong, and M. Yang (2018). "Airborne LiDAR remote sensing for individual tree forest inventory using trunk detection-aided mean shift clustering techniques". *Remote Sensing* 10 (7), 1078. DOI: `10.3390/rs10071078`.

Christie, J. M. and A. S. Murphy (2013). "Shoot phototropism in higher plants: New light through old concepts". *American Journal of Botany* 100 (1), 35–46. DOI: `10.3732/ajb.1200340`.

Dalponte, M. and D. A. Coomes (2016). "Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data". *Methods in Ecology and Evolution* 7 (10), 1236–1245. DOI: `10.1111/2041-210x.12575`.

Date, H., E. Wakisaka, Y. Moribe, and S. Kanai (2019). "TLS point cloud registration based on ICP algorithm using point quality". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W13, 963–968. DOI: `10.5194/isprs-archives-xlii-2-w13-963-2019`.

References

Disney, M., A. Burt, K. Calders, C. Schaaf, and A. Stovall (2019). "Innovations in ground and airborne technologies as reference and for training and validation: Terrestrial laser scanning (TLS)". *Surveys in Geophysics* 40 (4), 937–958. DOI: 10.1007/s10712-019-09527-x.

Dorigo, W., M. Hollaus, W. Wagner, and K. Schadauer (2010). "An application-oriented automated approach for co-registration of forest inventory and airborne laser scanning data". *International Journal of Remote Sensing* 31 (5), 1133–1153. DOI: 10.1080/01431160903380581.

Duncanson, L., J. Armston, M. Disney, V. Avitabile, N. Barbier, K. Calders, S. Carter, J. Chave, M. Herold, N. MacBean, R. McRoberts, D. Minor, K. Paul, M. Réjou-Méchain, S. Roxburgh, M. Williams, C. Albinet, T. Baker, H. Bartholomeus, J. F. Bastin, D. Coomes, T. Crowther, S. Davies, S. de Bruin, M. De Kauwe, G. Domke, R. Dubayah, M. Falkowski, L. Fatoyinbo, S. Goetz, P. Jantz, I. Jonckheere, T. Jucker, H. Kay, J. Kellner, N. Labriere, R. Lucas, E. Mitchard, F. Morsdorf, E. Næsset, T. Park, O. L. Phillips, P. Ploton, S. Puliti, S. Quegan, S. Saatchi, C. Schaaf, D. Schepaschenko, K. Scipal, A. Stovall, C. Thiel, M. A. Wulder, F. Camacho, J. Nickeson, M. Román, and H. Margolis (2021). *Aboveground woody biomass product validation: Good practices protocol*. Ed. by L. Duncanson, M. Disney, J. Armston, J. Nickeson, D. Minor, and F. Camacho. Land Product Validation Subgroup (WGCV/CEOS). DOI: 10.5067/DOC/CEOSWGCV/LPV/AGB.001.

Edson, C. and M. G. Wing (2011). "Airborne light detection and ranging (LiDAR) for individual tree stem location, height, and biomass measurements". *Remote Sensing* 3 (11), 2494–2528. DOI: 10.3390/rs3112494.

Einstein, A. (1916). "Strahlungs-Emission und Absorption nach der Quantentheorie". *Deutsche Physikalische Gesellschaft* 18, 318–323.

— (1979). *Albert Einstein, the human side: New glimpses from his archives*. Princeton University Press.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise." In: *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 226–231.

Eysn, L., M. Hollaus, E. Lindberg, F. Berger, J.-M. Monnet, M. Dalponte, M. Kobal, M. Pellegrini, E. Lingua, D. Mongus, and N. Pfeifer (2015). "A benchmark of Lidar-based single tree detection methods using heterogeneous forest data from the alpine space". *Forests* 6 (5), 1721–1747. DOI: 10.3390/f6051721.

Fernandez-Diaz, J., W. Carter, C. Glennie, R. Shrestha, Z. Pan, N. Ekhtari, A. Singhania, D. Hauser, and M. Sartori (2016). "Capability assessment and performance metrics for the titan multispectral mapping Lidar". *Remote Sensing* 8 (11), 936. DOI: 10.3390/rs8110936.

Ferraz, A., S. Saatchi, C. Mallet, and V. Meyer (2016). "Lidar detection of individual tree size in tropical forests". *Remote Sensing of Environment* 183, 318–333. DOI: 10.1016/j.rse.2016.05.028.

References

Frazer, G., S Magnussen, M. Wulder, and K. Niemann (2011). "Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass". *Remote Sensing of Environment* 115 (2), 636–649. DOI: 10.1016/j.rse.2010.10.008.

Gardiner, B., P. Berry, and B. Moulia (2016). "Review: Wind impacts on plant growth, mechanics and damage". *Plant Science* 245, 94–118. DOI: 10.1016/j.plantsci.2016.01.006.

Gatziolis, D. and H.-E. Andersen (2008). "A guide to LIDAR data acquisition and processing for the forests of the Pacific Northwest." *US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Portland, OR, USA*, 32. DOI: 10.2737/pnw-gtr-768.

Ge, X. and Q. Zhu (2021). "Target-based automated matching of multiple terrestrial laser scans for complex forest scenes". *ISPRS Journal of Photogrammetry and Remote Sensing* 179, 1–13. DOI: 10.1016/j.isprsjprs.2021.06.019.

Gobakken, T. and E. Næsset (2009). "Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data". *Canadian Journal of Forest Research* 39 (5), 1036–1052. DOI: 10.1139/X09-025.

Guttman, A. (1984). "R-trees: A dynamic index structure for spatial searching". In: *SIGMOD '84: Proceedings of the 1984 ACM SIGMOD international conference on Management of data*. Vol. 14. 2. ACM SIGMOD Record, 47–57. DOI: 10.1145/602259.602266.

Hackenberg, J., H. Spiecker, K. Calders, M. Disney, and P. Raumonen (2015). "SimpleTree —An efficient open source tool to build tree models from TLS clouds". *Forests* 6 (12), 4245–4294. DOI: 10.3390/f6114245.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant (2020). "Array programming with NumPy". *Nature* 585 (7825), 357–362. DOI: 10.1038/s41586-020-2649-2.

Hauglin, M., V. Lien, E. Næsset, and T. Gobakken (2014). "Geo-referencing forest field plots by co-registration of terrestrial and airborne laser scanning data". *International Journal of Remote Sensing* 35 (9), 3135–3149. DOI: 10.1080/01431161.2014.903440.

Heinzel, J. and B. Koch (2012). "Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation". *International Journal of Applied Earth Observation and Geoinformation* 18, 101–110. DOI: 10.1016/j.jag.2012.01.025.

Hill, A., H. Buddenbaum, and D. Mandallaz (2018). "Combining canopy height and tree species map information for large-scale timber volume estimations under strong heterogeneity of auxiliary

data and variable sample plot sizes". *European Journal of Forest Research* 137 (4), 489–505. DOI: 10.1007/s10342-018-1118-z.

Hill, A., D. Mandallaz, and J. Langshausen (2018). "A double-sampling extension of the German national forest inventory for design-based small area estimation on forest district levels". *Remote Sensing* 10 (7), 1052. DOI: 10.3390/rs10071052.

Hill, A., A. Massey, and D. Mandallaz (2021). "The R package forestinventory: Design-based global and small area estimations for multiphase forest inventories". *Journal of Statistical Software* 97 (4), 1–40. DOI: 10.18637/jss.v097.i04.

Holmgren, J., Å. Persson, and U. Söderman (2008). "Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images". *International Journal of Remote Sensing* 29 (5), 1537–1552. DOI: 10.1080/01431160701736471.

Hoppus, M. and A. Lister (2007). "The status of accurately locating forest inventory and analysis plots using the global positioning system". In: *2005 Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium*. October 3-6, 2005; Portland, 179–184.

Hug, C., P. Krzystek, and W. Fuchs (2004). "Advanced LiDAR data processing with LasTools". In: *ISPRS Archives* (Istanbul, Turkey, July 12, 2004–July 23, 2004). Ed. by O. Altan. Vol. 35.B2. XXXV vols. Istanbul, Turkey, 832–837.

Hyyppä, J., H. Hyyppä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo (2008a). "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. DOI: 10.1080/01431160701736489.

Hyyppä, J., H. Hyyppä, X. Yu, H. Kaartinen, A. Kukko, and M. Holopainen (2008b). "Forest inventory using small-footprint airborne LiDAR". In: *Topographic Laser Ranging and Scanning*. CRC Press, 335–370. DOI: 10.1201/9781420051438.ch12.

Iino, M. (2001). "Phototropism in higher plants". In: *Comprehensive Series in Photosciences*. Elsevier. Chap. 23, 659–811. DOI: 10.1016/s1568-461x(01)80027-2.

Ishii, R. and M. Higashi (1997). "Tree coexistence on a slope: An adaptive significance of trunk inclination". *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264 (1378), 133–139. DOI: 10.1098/rspb.1997.0020.

Jelalian, A. V. (1992). *Laser Radar Systems*. Norwood, South Australia, Australia: Artech House Publishers.

Johns, J. W., J. M. Yost, D. Nicolle, B. Igic, and M. K. Ritter (2017). "Worldwide hemisphere-dependent lean in cook pines". *Ecology* 98 (9), 2482–2484. DOI: 10.1002/ecy.1850.

Kaartinen, H., J. Hyyppä, X. Yu, M. Vastaranta, H. Hyyppä, A. Kukko, M. Holopainen, C. Heipke, M. Hirschmugl, F. Morsdorf, E. Næsset, J. Pitkänen, S. Popescu, S. Solberg, B. M. Wolf, and J.-C. Wu

(2012). "An international comparison of individual tree detection and extraction using airborne laser scanning". *Remote Sensing* 4 (4), 950–974. DOI: `10.3390/rs4040950`.

Khorram, S., C. F. van der Wiele, F. H. Koch, S. A. C. Nelson, and M. D. Potts (2016a). "Data processing tools". In: *Principles of Applied Remote Sensing*. Springer International Publishing, 69–124. DOI: `10.1007/978-3-319-22560-9_3`.

— (2016b). "Terrestrial applications of remote sensing". In: *Principles of Applied Remote Sensing*. Springer International Publishing, 125–176. DOI: `10.1007/978-3-319-22560-9_4`.

Koch, B., T. Kattenborn, C. Straub, and J. Vauhkonen (2013). "Segmentation of forest to tree objects". In: *Forestry Applications of Airborne Laser Scanning*. Springer Netherlands, 89–112. DOI: `10.1007/978-94-017-8663-8_5`.

Koenig, K. and B. Höfle (2016). "Full-waveform airborne laser scanning in vegetation studies—A review of point cloud and waveform features for tree species classification". *Forests* 7 (9), 198. DOI: `10.3390/f7090198`.

Korpela, I., A. Hovi, and F. Morsdorf (2012). "Understory trees in airborne LiDAR data — Selective mapping due to transmission losses and echo-triggering mechanisms". *Remote Sensing of Environment* 119, 92–104. DOI: `10.1016/j.rse.2011.12.011`.

Korpela, I., H. Ørka, M. Maltamo, T. Tokola, and J. Hyyppä (2010). "Tree species classification using airborne LiDAR – Effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type". *Silva Fennica* 44 (2), 319–339. DOI: `10.14214/sf.156`.

Kumar, A. and L. Dasgupta (2006). "Development of multiwave lidar at LASTEC". In: *Lidar Remote Sensing for Environmental Monitoring VII*. Ed. by U. N. Singh, T. Itabe, and D. N. Rao. SPIE, 64090W. DOI: `10.1117/12.693992`.

Lamprecht, S. (2018). *Project deliverable 4.1: Multispectral LiDAR point clouds*. Research rep. Version 1.0. European Community Horizon 2020 project PANTHEON (grant agreement 774571).

— (2019a). "Pyoints: A Python package for point cloud, voxel and raster processing". *Journal of Open Source Software* 4 (36), 990. DOI: `10.21105/joss.00990`.

— (2019b). *Pyoints*. [Online; accessed 18. Jun. 2021]. URL: `https://github.com/laempy/pyoints`.

— (2019c). *Pyoints: A Python package for point cloud, voxel and raster processing*. Zenodo. DOI: `10.5281/ZENODO.2619945`.

Lamprecht, S., A. Hill, J. Stoffels, and T. Udelhoven (2017a). "A machine learning method for co-registration and individual tree matching of forest inventory and airborne laser scanning data". *Remote Sensing* 9 (5), 505. DOI: `10.3390/rs9050505`.

Lamprecht, S., A. Hill, J. Stoffels, and T. Udelhoven (2017b). "Erratum: a machine learning method for co-registration and individual tree matching of forest inventory and airborne laser scanning data. remote sensing 2017, 9, 505". *Remote Sensing* 9 (7), 692. DOI: `10.3390/rs9070692`.

Lamprecht, S., J. Stoffels, S. Dotzler, E. Haß, and T. Udelhoven (2015). "aTrunk—An ALS-based trunk detection algorithm". *Remote Sensing* 7 (8), 9975–9997. DOI: `10.3390/rs70809975`.

Lamprecht, S., J. Stoffels, and T. Udelhoven (2015). "VecTree – Konzepte zur 3D Modellierung von Laubbäumen aus terrestrischem Lidar". *Photogrammetrie - Fernerkundung - Geoinformation* 2015 (3), 241–255. DOI: `10.1127/pfg/2015/0266`.

— (2020). "ALS as tool to study preferred stem inclination directions". *Remote Sensing* 12 (22), 3744. DOI: `10.3390/rs12223744`.

Li, R., G. Bu, and P. Wang (2017). "An automatic tree skeleton extracting method based on point cloud of terrestrial laser scanner". *International Journal of Optics* 2017, 1–11. DOI: `10.1155/2017/5408503`.

Liang, X., V. Kankare, J. Hyyppä, Y. Wang, A. Kukko, H. Haggrén, X. Yu, H. Kaartinen, A. Jaakkola, F. Guan, M. Holopainen, and M. Vastaranta (2016). "Terrestrial laser scanning in forest inventories". *ISPRS Journal of Photogrammetry and Remote Sensing* 115, 63–77. DOI: `10.1016/j.isprsjprs.2016.01.006`.

Lindberg, E., J. Holmgren, K. Olofsson, and H. Olsson (2012). "Estimation of stem attributes using a combination of terrestrial and airborne laser scanning". *European Journal of Forest Research* 131 (6), 1917–1931. DOI: `10.1007/s10342-012-0642-5`.

Lindberg, E., J. Holmgren, K. Olofsson, J. Wallerman, and H. Olsson (2013). "Estimation of tree lists from airborne laser scanning using tree model clustering and k-MSN imputation". *Remote Sensing* 5 (4), 1932–1955. DOI: `10.3390/rs5041932`.

Lloyd, S. (1982). "Least squares quantization in PCM". *IEEE Transactions on Information Theory* 28 (2), 129–137. DOI: `10.1109/tit.1982.1056489`.

Lohani, B. and S. Ghosh (2017). "Airborne LiDAR technology: A review of data collection and processing systems". *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* 87 (4), 567–579. DOI: `10.1007/s40010-017-0435-9`.

Lovell, J. L., D. L. B. Jupp, D. S. Culvenor, and N. C. Coops (2003). "Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests". *Canadian Journal of Remote Sensing* 29 (5), 607–622. DOI: `10.5589/m03-026`.

Lu, X., Q. Guo, W. Li, and J. Flanagan (2014). "A bottom-up approach to segment individual deciduous trees using leaf-off lidar point cloud data". *ISPRS Journal of Photogrammetry and Remote Sensing* 94, 1–12. DOI: `10.1016/j.isprsjprs.2014.03.014`.

Maiman, T. H. (1960). "Stimulated optical radiation in ruby". *Nature* 187 (4736), 493–494. DOI: 10 . 1038/187493a0.

Maltamo, M., E. Næsset, and J. Vauhkonen, eds. (2014). *Forestry applications of airborne laser scanning*. Vol. 1. Springer Netherlands. DOI: 10.1007/978-94-017-8663-8.

Matsuzaki, J., M. Masumori, and T. Tange (2006). "Stem phototropism of trees: A possible significant factor in determining stem inclination on forest slopes". *Annals of Botany* 98 (3), 573–581. DOI: 10.1093/aob/mcl127.

McGaughey, R. (2016). "FUSION/LDV: Software for LiDAR data analysis and visualization - V3.10". *USDA Forest Service*.

McRoberts, R. E., H.-E. Andersen, and E. Næsset (2013). "Using airborne laser scanning data to support forest sample surveys". In: *Forestry Applications of Airborne Laser Scanning*. Springer Netherlands, 269–292. DOI: 10.1007/978-94-017-8663-8_14.

McRoberts, R. E. and E. O. Tomppo (2007). "Remote sensing support for national forest inventories". *Remote Sensing of Environment* 110 (4), 412–419. DOI: 10.1016/j.rse.2006.09.034.

Monnet, J.-M. and É. Mermin (2014). "Cross-correlation of diameter measures for the co-registration of forest inventory plots with airborne laser scanning data". *Forests* 5 (9), 2307–2326. DOI: 10 . 3390/f5092307.

Nelson, R. (2013). "How did we get here? An early history of forestry lidar". *Canadian Journal of Remote Sensing* 39 (sup1), 6–17. DOI: 10.5589/m13-011.

Newnham, G. J., J. D. Armston, K. Calders, M. I. Disney, J. L. Lovell, C. B. Schaaf, A. H. Strahler, and F. M. Danson (2015). "Terrestrial laser scanning for plot-scale forest measurement". *Current Forestry Reports* 1 (4), 239–251. DOI: 10.1007/s40725-015-0025-5.

Nicoll, B. C. and D. Ray (1996). "Adaptive growth of tree root systems in response to wind action and site conditions". *Tree Physiology* 16 (11-12), 891–898. DOI: 10.1093/treephys/16.11-12.891.

Olofsson, K., E Lindberg, and J Holmgren (2008). "A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs". In: *Proceedings of SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory*. Ed. by R. A. Hill, J. Rosette, and J. Suárez. SilviLaser 2008, Sept. 17-19, 2008 – Edinburgh, UK. Heriot-Watt University, Edinburgh, UK: SilviLaser 2008 Organizing Committee, 95–104.

Othmani, A., A. Piboule, M Krebs, C Stolz, and L. L. Y. Voon (2011). "Towards automated and operational forest inventories with T-Lidar". In: *Proceedings of SilviLaser 2011, 11th international conference on LiDAR applications for assessing forest ecosystems*.

PDAL contributors (2018). *PDAL point data abstraction library*. Zenodo. DOI: `10.5281/zenodo.2556738`.

Pfennigbauer, M. and A. Ullrich (2011). "Multi-wavelength airborne laser scanning". In: *Proceedings of the International Lidar Mapping Forum, ILMF 2011, New Orleans*.

Polewski, P., W. Yao, M. Heurich, P. Krzystek, and U. Stilla (2015a). "Active learning approach to detecting standing dead trees from ALS point clouds combined with aerial infrared imagery". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 10–18. DOI: `10.1109/CVPRW.2015.7301378`.

— (2015b). "Free Shape Context descriptors optimized with genetic algorithm for the detection of dead tree trunks in ALS point clouds". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, 41–48. DOI: `10.5194/isprsannals-ii-3-w5-41-2015`.

Popescu, S. C. and M. Hauglin (2013). "Estimation of biomass components by airborne laser scanning". In: *Forestry Applications of Airborne Laser Scanning*. Springer Netherlands, 157–175. DOI: `10.1007/978-94-017-8663-8_8`.

Pueschel, P., G. Newnham, G. Rock, T. Udelhoven, W. Werner, and J. Hill (2013). "The influence of scan mode and circle fitting on tree stem detection, stem diameter and volume extraction from terrestrial laser scans". *ISPRS Journal of Photogrammetry and Remote Sensing* 77, 44–56. DOI: `10.1016/j.isprsjprs.2012.12.001`.

Quadros, N. and J. Keysers (2015). "Airborne LiDAR acquisition and validation". In: Held, A., S. Phinn, M. Soto-Berelov, and S. Jones. *AusCover Good Practice Guidelines: A technical handbook supporting calibration and validation activities of remotely sensed data product*. Ed. by A. Held, S. Phinn, M. Soto-Berelov, and J. S. Vol. 1.2. TERN AusCover. Chap. 16, pp. 291–293.

Åkerblom, M. (2020). *InverseTampere/TreeQSM: Version 2.4.0*. Zenodo. DOI: `10.5281/ZENODO.3989883`.

Raumonen, P., E. Casella, K. Calders, S. Murphy, M. Åkerblom, and M. Kaasalainen (2015). "Massive-scale tree modelling from TLS data". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W4, 189–196. DOI: `10.5194/isprsannals-II-3-W4-189-2015`.

Raumonen, P., M. Kaasalainen, M. Åkerblom, S. Kaasalainen, H. Kaartinen, M. Vastaranta, M. Holopainen, M. Disney, and P. Lewis (2013). "Fast automatic precision tree models from terrestrial laser scanner data". *Remote Sensing* 5 (2), 491–520. DOI: `10.3390/rs5020491`.

Razak, K. A., A. Bucksch, M. Straatsma, C. J. V. Westen, R. A. Bakar, and S. M. de Jong (2013). "High density airborne LiDAR estimation of disrupted trees induced by landslides". In: *IEEE International Geoscience and Remote Sensing Symposium - IGARSS 2013*. IEEE, 2617–2620. DOI: `10.1109/igarss.2013.6723359`.

References

Reitberger, J., C. Schnörr, P. Krzystek, and U. Stilla (2009). "3D segmentation of single trees exploiting full waveform LIDAR data". *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (6), 561–574. DOI: `10.1016/j.isprsjprs.2009.04.002`.

Roncat, A., F. Morsdorf, C. Briese, W. Wagner, and N. Pfeifer (2013). "Laser pulse interaction with forest canopy: Geometric and radiometric issues". In: *Forestry Applications of Airborne Laser Scanning*. Springer Netherlands, 19–41. DOI: `10.1007/978-94-017-8663-8_2`.

Roussel, J.-R., D. Auty, N. C. Coops, P. Tompalski, T. R. Goodbody, A. S. Meador, J.-F. Bourdon, F. de Boissieu, and A. Achim (2020). "lidR: An R package for analysis of airborne laser scanning (ALS) data". *Remote Sensing of Environment* 251, 112061. DOI: `10.1016/j.rse.2020.112061`.

Rusu, R. B. and S. Cousins (2011). "3D is here: Point Cloud Library (PCL)". *2011 IEEE International Conference on Robotics and Automation*, 1–4. DOI: `10.1109/ICRA.2011.5980567`.

Serafin, J. and G. Grisetti (2014). "Using augmented measurements to improve the convergence of ICP". In: *International Conference on Simulation, Modeling, and Programming for Autonomous Robots*. Springer, 566–577. DOI: `10.1007/978-3-319-11900-7_48`.

Shendryk, I., M. Broich, M. G. Tulbure, and S. V. Alexandrov (2016). "Bottom-up delineation of individual trees from full-waveform airborne laser scans in a structurally complex eucalypt forest". *Remote Sensing of Environment* 173, 69–83. DOI: `10.1016/j.rse.2015.11.008`.

Strîmbu, V. F. and B. M. Strîmbu (2015). "A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data". *ISPRS Journal of Photogrammetry and Remote Sensing* 104, 30–43. DOI: `10.1016/j.isprsjprs.2015.01.018`.

Tan, S. and R. M. Narayanan (2004). "Design and performance of a multiwavelength airborne polarimetric LiDAR for vegetation remote sensing". *Applied Optics* 43 (11), 2360–2368. DOI: `10.1364/ao.43.002360`.

Telewski, F. W. and M. J. Jaffe (1986). "Thigmomorphogenesis: Field and laboratory studies of Abies fraseri in response to wind or mechanical perturbation". *Physiologia Plantarum* 66 (2), 211–218. DOI: `10.1111/j.1399-3054.1986.tb02411.x`.

Tomppo, E., T. Gschwantner, M. Lawrence, and R. E. McRoberts, eds. (2010). *National forest inventories*. Vol. 1. Springer Netherlands. DOI: `10.1007/978-90-481-3233-1`.

Torabzadeh, H., F. Morsdorf, and M. E. Schaepman (2014). "Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems – A review". *ISPRS Journal of Photogrammetry and Remote Sensing* 97, 25–35. DOI: `10.1016/j.isprsjprs.2014.08.001`.

Trochta, J., M. Krůček, T. Vrška, and K. Král (2017). "3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR". *PLoS One* 12 (5), e0176871. DOI: `10.1371/journal.pone.0176871`.

References

Ullrich, A. (2013). "Sampling the world in 3D by airborne LIDAR – Assessing the information content of LIDAR point clouds". In: *Proceedings of the Photogrammetric Week '13*. Ed. by D. Fritsch. Wichmann/VDE Verlag Berlin, Germany, 247–259.

Valbuena, R. (2013). "Integrating airborne laser scanning with data from global navigation satellite systems and optical sensors". In: *Forestry Applications of Airborne Laser Scanning*. Springer Netherlands, 63–88. DOI: 10.1007/978-94-017-8663-8_4.

Valbuena, R., F. Mauro, R Rodriguez-Solano, and J. Manzanera (2010). "Accuracy and precision of GPS receivers under forest canopies in a mountainous environment". *Spanish Journal of Agricultural Research* 8 (4), 1047–1057.

Vauhkonen, J., L. Ene, S. Gupta, J. Heinzel, J. Holmgren, J. Pitkänen, S. Solberg, Y. Wang, H. Weinacker, K. M. Hauglin, V. Lien, P. Packalén, T. Gobakken, B. Koch, E. Næsset, T. Tokola, and M. Maltamo (2011). "Comparative testing of single-tree detection algorithms under different types of forest". *Forestry* 85 (1), 27–40. DOI: 10.1093/forestry/cpr051.

Vauhkonen, J., M. Maltamo, R. E. McRoberts, and E. Næsset (2013). "Introduction to forestry applications of airborne laser scanning". In: *Forestry Applications of Airborne Laser Scanning*. Springer Netherlands, 1–16. DOI: 10.1007/978-94-017-8663-8_1.

Vega, C., A. Hamrouni, S. El Mokhtari, J. Morel, J. Bock, J.-P. Renaud, M. Bouvier, and S. Durrieu (2014). "PTrees: A point-based approach to forest tree extraction from lidar data". *International Journal of Applied Earth Observation and Geoinformation* 33, 98–108. DOI: 10.1016/j.jag.2014.05.001.

Venuti, G. (2009). *Roto-translations*. [Online; accessed 29. Dec. 2018]. URL: http://geomatica.como.polimi.it/corsi/def_monitoring/roto-translationsb.pdf.

Wagner, W., A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka (2006). "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner". *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (2), 100–112. DOI: 10.1016/j.isprsjprs.2005.12.001.

Wang, Y., J. Hyyppa, X. Liang, H. Kaartinen, X. Yu, E. Lindberg, J. Holmgren, Y. Qin, C. Mallet, A. Ferraz, H. Torabzadeh, F. Morsdorf, L. Zhu, J. Liu, and P. Alho (2016). "International benchmarking of the individual tree detection methods for modeling 3-D canopy structure for silviculture and forest ecology using airborne laser scanning". *IEEE Transactions on Geoscience and Remote Sensing* 54 (9), 5011–5027. DOI: 10.1109/tgrs.2016.2543225.

Wang, Y., H. Weinacker, and B. Koch (2008). "A Lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest". *Sensors* 8 (6), 3938–3951. DOI: 10.3390/s8063938.

References

Wehr, A. and U. Lohr (1999). "Airborne laser scanning—An introduction and overview". *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3), 68–82. DOI: 10.1016/S0924-2716(99)00011-8.

White, J. C., N. C. Coops, M. A. Wulder, M. Vastaranta, T. Hilker, and P. Tompalski (2016). "Remote sensing technologies for enhancing forest inventories: A review". *Canadian Journal of Remote Sensing* 42 (5), 619–641. DOI: 10.1080/07038992.2016.1207484.

White, J. C., M. A. Wulder, A. Varhola, M. Vastaranta, N. C. Coops, B. D. Cook, D. Pitt, and M. Woods (2013). "A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach". *The Forestry Chronicle* 89 (6), 722–723. DOI: 10.5558/tfc2013-132.

Wilkes, P., A. Lau, M. Disney, K. Calders, A. Burt, J. G. de Tanago, H. Bartholomeus, B. Brede, and M. Herold (2017). "Data acquisition considerations for terrestrial laser scanning of forest plots". *Remote Sensing of Environment* 196, 140–153. DOI: 10.1016/j.rse.2017.04.030.

Williams, J., C.-B. Schonlieb, T. Swinfield, J. Lee, X. Cai, L. Qie, and D. A. Coomes (2020). "3D segmentation of trees through a flexible multiclass graph cut algorithm". *IEEE Transactions on Geoscience and Remote Sensing* 58 (2), 754–776. DOI: 10.1109/tgrs.2019.2940146.

Windrim, L. and M. Bryson (2020). "Detection, segmentation, and model fitting of individual tree stems from airborne laser scanning of forests using deep learning". *Remote Sensing* 12 (9), 1469. DOI: 10.3390/rs12091469.

Wing, M. G., A. Eklund, J. Sessions, and R. Karsky (2008). "Horizontal measurement performance of five mapping-grade global positioning system receiver configurations in several forested settings". *Western Journal of Applied Forestry* 23 (3), 166–171. DOI: 10.1093/wjaf/23.3.166.

Wistuba, M., I. Malik, H. Gärtner, P. Kojs, and P. Owczarek (2013). "Application of eccentric growth of trees as a tool for landslide analyses: The example of Picea abies Karst. in the Carpathian and Sudeten mountains (Central Europe)". *Catena* 111, 41–55. DOI: 10.1016/j.catena.2013.06.027.

Young, M. (1986). *Optics and Lasers*. Vol. 3. Berlin, Germany: Springer-Verlag. DOI: 10.1007/978-3-540-37456-5.

Zhang, W., J. Shao, S. Jin, L. Luo, J. Ge, X. Peng, and G. Zhou (2021). "Automated marker-free registration of multisource forest point clouds using a coarse-to-global adjustment strategy". *Forests* 12 (3), 269. DOI: 10.3390/f12030269.

Zhen, Z., L. J. Quackenbush, and L. Zhang (2016). "Trends in automatic individual tree crown detection and delineation—Evolution of LiDAR data". *Remote Sensing* 8 (4), 333. DOI: 10.3390/rs8040333.

Zhou, Q.-Y., J. Park, and V. Koltun (2018). "Open3D: A modern library for 3D data processing". *arXiv*: 1801.09847.

Curriculum Vitae

Sebastian Lamprecht

Address
Bergstraße 15
56859 Bullay, Germany

Date of Birth   17/02/1989

Place of Birth   Wittlich, Germany

Professional Experience

04/2015 – today
Trier University, Trier, Germany
Department: Environmental Remote Sensing & Geoinformatics
Scientific staff (Projects: TriCCS and PANtHEOn)

04/2013 – 03/2015
Trier University, Trier, Germany
Department: Environmental Remote Sensing & Geoinformatics
Graduate assistant (Projects: Sentinel 2, Yangui 2 and TriCSS)

04/2011 – 03/2014
Trier University, Trier, Germany
Department: Mathematics
Graduate assistant (Project TriCSS)

11/2012 – 12/2012
NetGis GdbR, Trier, Germany
Freelancer (software development)

07/2012 – 09/2012
NetGis GdbR, Trier, Germany
Internship

01/2009 – 03/2009
Benninghoven, Mühlheim, Germany
Labourer for metal processing

04/2008 – 12/2008
Basic military service, Germersheim & Büchel, Germany

Academic Career

03/2015 – today **Trier University**, Trier, Germany
Department: Environmental Remote Sensing & Geoinformatics
Ph.D candidacy

04/2013 – 03/2015 **Trier University**, Trier, Germany
Master in Applied Geoinformatics – overall grade: 1.1
Master thesis titled: 'aTrunk – An ALS-based Trunk Detection Algorithm'
Supervisors: Prof. Dr. Thomas Udelhoven, Dr. Johannes Stoffels

10/2010 – 03/2013 **Trier University**, Trier, Germany
Bachelor in Applied Geoinformatics – overall grade: 1.9
Bachelor thesis titled: 'Vektorielle Modellierung von Forstbeständen aus terrestrischen Laserscanner-Daten'
Supervisors: Prof. Dr. Thomas Udelhoven, Dr. Sebastian Mader

04/2009 – 09/2010 **Trier University**, Trier, Germany
Study of Applied Mathematics with focus earth sciences (no graduation)

08/1999 – 03/2008 **Secondary school (Cusanus-Gymnasium)**, Wittlich, Germany

08/1995 – 07/1999 **Primary school**, Wittlich-Wengerohr, Germany

Extra Curricular Activities

3/2021 – today **ProMarienburg**, Zell (Mosel), Germany
Management member of an initiative for the preservation of the Marienburg Zell

09/2020 – today **Schachclub Wittlich 1947 e.V.**, Wittlich, Germany
Board member (kit manager)

11/2016 – 3/2021 **Kirche der Jugend Marienburg**, Zell (Mosel), Germany
Online editor

01/2016 – 6/2021 **Förderverein Lebens(T)räume Marienburg e.V.**, Zell (Mosel), Germany
Board member

02/2010 – 2/2012 **Schachclub Wittlich 1947 e.V.**, Wittlich, Germany
Board member (kit manager, youth development officer)

02/2009 – 2/2010 **Schachclub Wittlich 1947 e.V.**, Wittlich, Germany
Auditor

Publications (peer-reviewed)

Lamprecht, S., J. Stoffels, and T. Udelhoven (2020). "ALS as tool to study preferred stem inclination directions". *Remote Sensing* 12 (22), 3744. DOI: `10.3390/rs12223744`.

Lamprecht, S. (2019a). "Pyoints: A Python package for point cloud, voxel and raster processing". *Journal of Open Source Software* 4 (36), 990. DOI: `10.21105/joss.00990`.

Lamprecht, S., A. Hill, J. Stoffels, and T. Udelhoven (2017a). "A machine learning method for co-registration and individual tree matching of forest inventory and airborne laser scanning data". *Remote Sensing* 9 (5), 505. DOI: `10.3390/rs9050505`.

Lamprecht, S., J. Stoffels, S. Dotzler, E. Haß, and T. Udelhoven (2015). "aTrunk—An ALS-based trunk detection algorithm". *Remote Sensing* 7 (8), 9975–9997. DOI: `10.3390/rs70809975`.

Lamprecht, S., J. Stoffels, and T. Udelhoven (2015). "VecTree – Konzepte zur 3D Modellierung von Laubbäumen aus terrestrischem Lidar". *Photogrammetrie - Fernerkundung - Geoinformation* 2015 (3), 241–255. DOI: `10.1127/pfg/2015/0266`.