EuroSDR

European Spatial Data Research

October 2019

International Benchmarking of Terrestrial Laser Scanning Approaches for Forest Inventories

joint project of EuroSDR and ISPRS

Part II: Results, Discussion and Outlooks

Xinlian Liang, Juha Hyyppä, Harri Kaartinen, Matti Lehtomäki, Jiri Pyörälä, Xiaowei Yu, Norbert Pfeifer, Gábor Brolly, Pirotti Francesco, Jan Hackenberg, Huabing Huang, Hyun-Woo Jo, Masato Katoh, Luxia Liu, Martin Mokroš, Jules Morel, Kenneth Olofsson, Jose Poveda-Lopez, Jan Trochta, Di Wang, Jinbu Wang, Zhouxi Xi, Bisheng Yang, GuangZheng, Ville Kankare, MikkoVastaranta, Yunsheng Wang

Official Publication N° 71 b

This publication was also partly published in

ISPRS Journal of Photogrammetry and Remote Sensing 115 (2016) 63–77:
“Terrestrial laser scanning in forest inventories”

Xinlian Liang [a,b], Ville Kankare [b,c], Juha Hyyppä [a,b], Yunsheng Wang [a,b,*], Antero Kukko [a,b,d], Henrik Haggrén [d], Xiaowei Yu [a,b], Harri Kaartinen [a], Anttoni Jaakkola [a,b], Fengying Guan [e], Markus Holopainen [b,c], Mikko Vastaranta [b,c]

(http://dx.doi.org/10.1016/j.isprsjprs.2016.01.006)

[a] Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, FGI (Former Finnish Geodetic Institute), 02431 Masala, Finland

[b] Centre of Excellence in Laser Scanning Research, Academy of Finland, 00531 Helsinki, Finland

[c] Department of Forest Sciences, University of Helsinki, 00014 Helsinki, Finland

[d] Department of Real Estate, Planning and Geoinformatics, Aalto University, 02150 Espoo, Finland

[e] International Centre for Bamboo and Rattan, 100102 Beijing, China

* Corresponding author at: Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, FGI (Former Finnish Geodetic Institute), 02431 Masala, Finland

Published by EuroSDR

EUROPEAN SPATIAL DATA RESEARCH

PRESIDENT 2018 – 2020:

Paul Kane, Ireland

VICE-PRESIDENT 2017 – 2019:

Fabio Remondino, Italy

SECRETARY – GENERAL 2019 – 2023:

Joep Crompvoets, Belgium

DELEGATES BY MEMBER COUNTRY:

Austria: Michael Franzen, Norbert Pfeifer
Belgium: Eric Bayers
Croatia: Ivan Landek, Željko Bačič
Cyprus: Andreas Sokratous, Georgia Papathoma, Andreas Hadjiraftis, Dimitrios Skarlatos
Denmark: Jesper Weng Haar, Tessa Anderson
Estonia: Tambet Tiits, Artu Ellmann
Finland: Juha Hyyppä, Jurkka Tuokko
France: Bénédicte Bucher, Yannick Boucher
Germany: Paul Becker, Lars Bernard
Ireland: Paul Kane, Audrey Martin
Norway: Jon Arne Trollvik, Ivar Maalen-Johansen
Poland: Krzysztof Bakuła
Slovenia: Dalibor Radovan, Peter Prešeren, Marjan Čeh
Spain: Julián Delgado Hernández
Sweden: Tobias Lindholm, Thomas Lithén, Heather Reese
Switzerland: André Streilein, François Golay
The Netherlands: Jantien Stoter, Martijn Rijsdijk
United Kingdom: Sally Cooper, Claire Ellul

ASSOCIATE MEMBERS AND THEIR REPRESENTATIVES:

Esri: Nick Land
Informatie Vlaanderen: Jo Van Valckenborgh
nFrames: Konrad Wenzel
Terratec: Leif Erik Blankenberg
Vexcel: Michael Gruber
1Spatial: Dan Warner

COMMISSION CHAIRPERSONS:

Data Acquisition: Jon Mills, United Kingdom
Modelling and Processing: Norbert Haala, Germany
Updating and Integration: Jon Arne Trollvik, Norway
Information Usage: Bénédicte Bucher, France
Business Models and Operation: Joep Crompvoets, Belgium
Knowledge Transfer: Markéta Potůčková, Czech Republic

OFFICE OF PUBLICATIONS:

Bundesamt für Eich- und Vermessungswesen
Publications Officer: Michael Franzen
Schiffamtsgasse 1-3
1020 Wien
Austria
Tel.: + 43 1 21110 825200
Fax: + 43 1 21110 82995202

CONTACT DETAILS:

Web: www.eurosdr.net
President: president@eurosdr.net
Secretary-General: secretary@eurosdr.net
Secretariat: admin@eurosdr.net

EuroSDR Secretariat
KU Leuven Public Governance Institute
Faculty of Social Sciences
Parkstraat 45 bus 3609
3000 Leuven
Belgium
Tel.: +32 16 37 98 10

The official publications of EuroSDR are peer-reviewed.

Table of Contents

INTERNATIONAL BENCHMARKING OF TERRESTRIAL LASER SCANNING APPROACHES FOR FOREST INVENTORIES

joint project of EuroSDR and ISPRS

PART II: RESULTS, DISCUSSION AND OUTLOOKS

With 21 figures and 2 tables

Xinlian Liang [1,2], Juha Hyyppä [1,2], Harri Kaartinen [1], Matti Lehtomäki [1,2], Jiri Pyörälä [1,2], Xiaowei Yu [1,2], Norbert Pfeifer [3], Gábor Brolly [4], Pirotti Francesco [5], Jan Hackenberg [6,7], Huabing Huang [8], Hyun-Woo Jo [9], Masato Katoh [10], Luxia Liu [11], Martin Mokroš [12], Jules Morel [13], Kenneth Olofsson [14], Jose Poveda-Lopez [15], Jan Trochta [16], Di Wang [3], Jinbu Wang [17], Zhouxi Xi [18], Bisheng Yang [19], Guang Zheng [20], Ville Kankare [2,21], Mikko Vastaranta [2,21], Yunsheng Wang * [1,2]

[1] Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, FGI (former Finnish Geodetic Institute), 02431 Masala, Finland

[2] Centre of Excellence in Laser Scanning Research, Academy of Finland, 00531 Helsinki, Finland

[3] Department of Geodesy and Geoinformation, Technische Universität Wien, 1040 Vienna, Austria

[4] Institute of Geomatics and Civil Engineering, Faculty of Forestry, University of Sopron (former University of West Hungary), H-9400 Sopron, Hungary

[5] CIRGEO—Interdepartment Research Center of Geomatics, University of Padova, 35020 Legnaro, PD, Italy

[6] Biogéochimie des Ecosystèmes Forestiers, INRA, 54280 Nancy, France

[7] Laboratoire d'Inventaire Forestier, Institut géographique national (IGN), 54250 Nancy, France

[8] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. 100101 Beijing, China

[9] Environmental GIS/RS Lab., Division of Environmental Science & Ecological Engineering, Korea University, 02855 Seoul, South Korea

[10] Shinshu University, Japan

[11] Chinese Academy of Forestry, China

[12] Department of Forest Management and Geodesy, Technical University in Zvolen, 96053 Zvolen, Slovakia

[13] Institut Français de Pondichéry – Laboratoire des Sciences de l'Information et des Systèmes, India/France

[14] Department of Forest Resource Management, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden

[15] Treemetrics, Irland

[16] Department of Forest Ecology and Department of Landscape Ecology and Geoinformatics, Silva Tarouca Research Institute for Landscape and Ornamental Gardening, 602 00 Brno, Czech Republic

[17] Department of Geoscience and Remote Sensing, Delft University of Technology, 2628CN Delft, The Netherlands

[18] Department of Geography, University of Lethbridge, T1K3M4 Lethbridge, Canada

[19] University of Wuhan, China

[20] International Institute for Earth System Science, Nanjing University, 210023 Nanjing, China

[21] Department of Forest Sciences, University of Helsinki, 00014 Helsinki, Finland

* corresponding author

Index of Figures

Index of Tables

ABSTRACT PART II

The past two decades have witnessed tremendous research efforts and significant progress in the application of terrestrial laser scanning (TLS) in forest inventories. To clarify the strengths and weaknesses of TLS as a measure of forest digitization, as well as the capability of recent algorithms to extract attribute of trees in forests, an international benchmarking of TLS approaches for forest inventories was launched in 2014 by the European Spatial Data Research Organization (EuroSDR) and coordinated by the Finnish Geospatial Research Institute (FGI). Twenty-four groups worldwide initially indicated that they would participate, and eighteen groups successfully processed the data and submitted their results for evaluation. Partners processed identical TLS datasets and delivered a common set of results, including the digital terrain model (DTM), the tree map, the tree height, the diameter at the breast height (DBH) and the stem curve of each individual tree at the plot level. In addition, the stem volume and tree biomass were calculated based on these parameters and on local allometric models. The outcomes from the partners were evaluated with a standard evaluation procedure. This paper reports the results of this benchmarking project and discusses the findings, which clarify the status quo of TLS-based forest investigations. With single-scan data, i.e., one hemispherical scan per plot, most of the recent algorithms are capable of achieving stem detection with approximately 75% completeness and 90% correctness in the easy forest stands (easy plots: 600 stems/ha, 20 cm mean DBH). The detection rate decreases when the stem density increases and the average DBH decreases, i.e., 60% completeness with 90% correctness (medium plots: 1000 stem/ha, 15 cm mean DBH) and 30% completeness with 90% correctness (difficult plots: 2000 stems/ha, 10 cm mean DBH). The application of the multi-scan approach, i.e., five scans per plot at the center and four quadrant angles, is more effective in complex stands, increasing the completeness to approximately 90% for medium plots and to approximately 70% for difficult plots, with almost 100% correctness. The results of this benchmarking also showed that the TLS-based approaches can provide the estimates of the DBH and the stem curve that are close to what is required in practical applications, e.g., national forest inventories (NFIs). In terms of algorithm development, a high level of automation is a commonly shared standard, but a bottleneck occurs at stem detection and tree height estimation, especially in multilayer and dense forest stands. The greatest challenge is that even with the multi-scan approach, it is still hard to completely and accurately record all trees in a plot due to the occlusion effects of the trees and bushes in forests. Future development must address the redundant yet incomplete point clouds of forest sample plots and recognize trees more accurately and efficiently. It is worth noting that TLS currently provides the best quality terrestrial point clouds in comparison with all other technologies, meaning that all the benchmarks labelled in this paper can also serve as a reference for other terrestrial point clouds sources.

Keywords: forest inventory, point cloud, terrestrial laser scanning, TLS, benchmarking

1 INTRODUCTION

Terrestrial laser scanning (TLS) is a laser-based instrument that measures its surroundings three-dimensional (3D) space using millions to billions of 3D points. Attempts to achieve high-quality tree attributes in forests utilizing TLS started when the first commercial TLS system was introduced to the market in 1998. During the past two decades, the hardware has experienced rapid improvement, marked by its rapidly decreasing size, weight and price as well as its constantly increasing spatial resolution and measurement speed. The current systems measure up to million-level points per second at the range of 100-300 meters; the range precision is at a millimeter level, and the angular sampling capacity is less than 0.01° in both horizontal and vertical directions.

TLS has been in the spotlight in a forest-inventory context since its introduction. Early studies around the year 2000 (Erikson and Karin, 2003; Lovell et al., 2003; Simonse et al., 2003; Aschoff and Spiecker, 2004; Hopkinson et al., 2004; Pfeifer et al., 2004; Parker et al., 2004; Schütt et al., 2004; Thies et al., 2004; Watt and Donoghue, 2005) explored the potential of measuring tree attributes using TLS. The past two decades have witnessed steady progress in these types of studies (Henning and Radtke, 2006; Hosoi and Omasa, 2006; Lefsky and McHale, 2008; Maas et al., 2008; Strahler et al., 2008; Côté et al., 2009; Jupp et al., 2009; Tansey et al., 2009). More recently, TLS has been shown to be capable of determining several high-quality tree attributes that are not directly measurable using conventional tools. Plot-level stem volume and biomass components were also shown to be estimated at accuracy levels that are similar to those of the best national allometric models (Yu et al., 2013; Kankare et al., 2013; Astrup et al., 2014; Liang et al., 2014; Newnham et al., 2015).

However, the significant variance in the range of the hardware, the scanning setups, and the forest structures, as well as in the evaluation criteria and procedures, among the reported studies has made reliable assessment of the performances of TLS for forest inventory extremely difficult. For example, as a fundamental criterion of TLS-based forest in situ observation, the percentage of correctly detected trees from multi-scan TLS data ranged from 50 to 100% at the plot level as reported in previous research (Liang et al., 2016). Considering the diversity of the elementary components in the reported studies, such literature-based statistics do not reflect the capability and the overall performance of TLS due to the lack of a common frame of reference.

A proper understanding of the performance of TLS for forest in situ inventory can only be achieved when certain conditions are satisfied: that identical TLS data are processed; that common plot- and tree-level forest attribute are extracted; and that, the results from the algorithms are evaluated with reliable reference information utilizing standardized evaluation procedures. Under such conditions, all the algorithms are projected to a unique frame of reference, and an assessment of the status quo of the TLS-based forest inventory can be conducted by comparing the attribute extraction results of different algorithms.

As such, an international benchmarking study of TLS in forest inventories (TLS benchmarking) was launched in 2014, led by the European Spatial Data Research Organization (EuroSDR) and partly funded by the European Community’s Seventh Framework Programme Project Advanced_SAR. As the coordinator of the project, the Finnish Geospatial Research Institute (FGI) conducted the acquisition of the TLS data and the reference data in 24 sample plots at a boreal forest in southern Finland. The sample plots were selected by foresters from the perspective of forest conditions representing three difficulty classes, i.e., “easy”, “medium”, and “difficult”, including the development stage, the stem density, the richness of sub-canopy growth, as well as the species composition in the forest stands, which also reflects the level of complexity in the TLS data processing. For the evaluation of the performances of the algorithms, a series of plot- and tree-level attributes are defined as standardized criteria, and fully automated procedures are developed under the efforts of the FGI. For information about the conceptual and technical details, readers are referred to a separated paper about this benchmarking project (Part I).

Eighteen partners from Asia, Europe and North America delivered the required results after processing the single- and multi-scan TLS datasets of the 24 sample plots. The required attributes included the DTM of each sample plot, the location, the height, the DBH, and the stem curve of each tree in the sample plot; the stem volume and tree biomass were calculated based directly on the delivered attributes or through local allometric models. The evaluation of the results from all the partners was conducted by the FGI. Detailed information about the partners and about their algorithms are summarized in section 2. A brief description of each algorithm in this benchmarking project is available in the appendix of this paper. Some of the algorithms applied in this benchmarking study were new, while most have been published or are an updated version of previously reported algorithms. For the published algorithms, comprehensive method descriptions are found in (Liang et al., 2012; Olofsson and Holmgren, 2016; Pirotti et al., 2013; Hackenberg et al., 2015; Ma et al., 2016; Wang et al., 2016; Xi et al., 2016; Koreň et al., 2017; Trochta et al., 2017).

For the time being, this is the first international benchmarking of TLS-based forest inventories. The investigations on TLS performance is carried out from two different perspectives: first, from the TLS data point of view, i.e., the impact of the forest stand conditions and the data acquisition methods on the accuracy and completeness of the point cloud of a sample plot and, consequently, on the results of attribute extraction of an algorithm; and second, from the aspect of the algorithms, i.e., to what extent can the recent algorithms reach the best extraction of essential forest attributes from TLS data. Section 3 of this paper illustrates the evaluation results of the algorithms on the criteria utilizing the single- and multi-scan TLS data of the sample plots. In-depth analyses comparing the results in section 4 reveal the achievements and remaining challenges of recent studies, providing recommendations and paving the way for further studies and applications in the field.

Furthermore, since static TLS provides the plot-level point cloud with spatial precision and detailed richness that is surpasses all other contemporary terrestrial point cloud technologies, e.g., mobile laser scanning (MLS) and image-based structure from motion (SfM), the evaluation results reported in this benchmarking also indicate the best performance that can be achieved from terrestrial point clouds for forest inventory. The conclusions about the performance and the challenges of TLS from this benchmarking can also be generalized to other sources of terrestrial point clouds.

2 ALGORITHMS OF THE PARTNERS

To accumulate as many representative algorithms as possible for benchmarking, the project was promoted at various conferences, and the call for partners was disseminated through scientific networks, various websites, and social media platforms. Eighteen groups from around the world successfully processed the data and submitted their results for evaluation. In addition to universities and research institutions, there were also partners from the commercial sector. Table 1 lists the name and the country of the partners in alphabetical order; the abbreviations of the names of partners are used in reference to their processing algorithm in the following descriptions. Out of the eighteen partners, fourteen processed both single- and multi-scan datasets and provided the extraction results of all criteria. Two partners processed both datasets and provided part of the required criteria. Two partners processed the single- or multi-scan data only and provided all criteria.

Table 1: The partners and methods in the international TLS benchmarking for forest inventories

Full name	Country	Abbreviation
Chinese Academy of Forestry	China	CAF
Delft University of Technology	Netherlands	TUDelft
Finnish Geospatial Research Institute	Finland	FGI
Institut Français de Pondichéry – Laboratoire des Sciences de l'Information et des Systèmes	India/France	IFP-LSIS
INRA Biogéochimie des Ecosystèmes Forestiers – ING Laboratoire d'Inventaire Forestier	France	INRA-IGN
Institute of Remote Sensing and Digital Earth	China	RADI
Korea Univeristy	South Korea	KU
Nanjing University	China	NJU
Shinshu University	Japan	Shinshu
Swedish University of Agricultural Sciences	Sweden	SLU
Technical University in Zvolen	Slovakia	TUZVO
Technische Universität Wien	Austria	TUWien
The Silva Tarouca Research Institute for Landscape and Ornamental Gardening	Czech Republic	RILOG
Treemetrics	Ireland	TreeMetrics
University of Lethbridge	Canada	ULeth
University of Padova	Italy	UNIPD
University of Sopron	Hungary	NYME
Wuhan University	China	WHU

The eighteen algorithms in the benchmarking include a wide range of variation in terms of their methodological development. The variety of algorithms can be inspected based on the characteristics of their data structure, work flow and parameter settings for implementation. Despite the wide range of designs, the algorithms have a high level of automation; fifteen algorithms are fully automated, and the other three are semi-automated approaches. During data processing, twelve partners applied the same parameter settings for all the sample plots and single- and multi- scan data, which indicates the robustness of the algorithms towards different stand and data conditions.

Table 2 summarizes the main characteristics of the algorithms with an overview of the fundamental components of the algorithms in this benchmarking. Considering the length of this paper, more detailed descriptions of each algorithm are provided in the appendix.

Table 2: Brief summary of algorithms in the international TLS benchmarking for forest inventories

	Method	Data Processing			Data Structure		Methodological Concepts				
		data[1]	auto[2]	param[3]	stem detection	stem modelling	1. preprocessing thinning	filtering	2. DTM	3. individual tree detection	4. stem modelling
1	CAF	single	A	U	multiple 2D layers	raster	\\[*]	\	lastools	detecting circles in multi-layers	circles at different heights
2	TUDelft	both	A	D	voxel	2D plane	a random point in a voxel	point distance	morphological filtering + polynomial interpolation	clustering in voxel space	circles at different heights
3	FGI	both	A	U	point	point	the point closest to the center of mass in voxel	flatness + normal vectors	morphological filtering + the linear interpolation	point clustering and object modelling	cylinder along the trunk
4	IFP-LSIS	both	A	U	\	\	\	\	approximation in multi-scales + polygonization	\	\
5	INRA-IGN	both	A	U	a 2D layer	point	Center of mass in voxel	Statistical outlier filter	the lowest point in multi-scales + RANSAC plane fitting + inverse distance weighting (IDW) interpolation	clustering in 3D	circles along the trunk and cylinders for refinement for both stem and branches (LOD 4)
6	RADI	both	A	U	voxel + multiple 2D layers	point / raster	\	\	filtering based on distances to model in multi-scale	voxel distribution and point clustering	circles at different heights
7	KU	single; both	S-A; M	U; U	a 2D layer	\	\	\	minimum height + IDW interpolation	manually identifying (semi-) circular cluster	a circle at the DBH height

8	NJU	both	A	D	point	2D plane	one point in a neighbor	number of point + distance	surface class + IDW interpolation	classification based on models and training samples from data + point clustering in 2D plane	radius estimated at different heights
9	Shinshu	both	A	U	multiple 2D layers	raster	\	\	\	point number in voxel	\
10	SLU	both	A	U	voxel	point	\	flatness	minimum height in multi-scales	selecting curvature with same radius and originates + connected vertical cylinders	cylinder along the stem
11	TUZVO	both	S-A,	D	multiple 2D layers	2D plane	\	\	the lowest point + natural neighbor interpolation	segment in a 2D plane + fitting a circle	circles at different heights
12	TUWien	both	A	U	a 2D layer	point	\	normal vectors	hierarchical robust filtering + Robust Moving Plane / Delaunay TIN interpolation	-- project points onto a 2D horizontal plane. -- generate point density image and convert to a binary image	cylinders along the stem
13	RILOG	both	S-A	U	multiple 2D layers	2D plane	\	\	the lowest point + IDW interpolation	manual detection	circles at different heights
14	Treemetrics	both	A	U	a 2D layer	2D plane	\	curve smoothness	the lowest point + plan fitting	clustering in a 2D slice	circles at different heights
15	ULeth	both	A	U	voxel + 2D plane	2D plane	\	voxel distribution	the lowest point + IDW interpolation	finding the local extrema in 2D plane projected from voxels + filtering fine stem points by 3D region growing	circles at different heights
16	UNIPD	both	A	U	\	\	\	\	morphological filter + natural neighbour / Kriging interpolation	\	\

17	NYME	single	A	D	voxel + 2D plane	2D plane	\	voxel distribution+ penetration rate	hierarchical interpolation for the classified points	Finding voxels with high point density + segment in a 2D plane	circles at different heights
18	WHU	both	A	U	multiple 2D layers	raster/point	\	\		detecting cylinders in multi-layers and find linked cylinders	circles at different heights

[1.] Refers to the TLS dataset, which has been processed for the benchmarking; "both" means both single- and multi-scan data are processed, "single" means only single-scan data are processed

[2.] The level of automation of the algorithm: "A" is fully automated; "S-A" is semi-automated; "M" is manual.

[3.] The parameter settings for different sample plots and different TLS datasets: "U" means the universal parameter setting for all sample plots and all datasets; "D" means different parameters are applied for different sample plots; and single- and multi-scan datasets.

[*] \ indicates that no relevant processing is applied.

3 EVALUATION RESULTS

The evaluation of the algorithms is carried out using eight criteria, namely, 1) the DTM, 2) the overall stem detection accuracy at the plot level, 3) the tree location, 4) the DBH, 5) the tree height, 6) the stem curve, 7) the stem volume and 8) the tree biomass at the individual tree level. The first six criteria are directly extracted from the point cloud, and the volume is estimated from the extracted stem curve and tree height. The biomass is predicted using the extracted tree attributes and local biomass allometric model. It should be noted that this benchmarking has no intention of determining which algorithms surpass the others. One substantial challenge for algorithm development is that there are tradeoffs among different criteria, and the algorithm designs must assign priorities to the criteria respecting their own application requirements. Thus, each of the algorithms has its own strengths and weaknesses. This benchmarking only provides a spectra to describe the capability of recent TLS-based forest inventories, and the value of the evaluation results lies in the revealed status quo for the algorithms.

All the evaluations are separately conducted in each sample plot. To reveal the influences of the forest conditions, the results are summarized based on three stand complexity categories, namely, an average is calculated for the evaluation results over all the sample plots in the same stand complexity category. Therefore, the performance of the algorithms is linked with the stand conditions of the forest. For more detailed information about the stand conditions in the sample plots and the definition of the complexity categories, readers are referred to a separated paper about this benchmarking.

3.1 Digital Terrain Model (DTM)

The DTM influences and is needed for the estimation of tree attributes, e.g., the tree height, the DBH and the stem curve. The more accurate the DTM is, the higher the chance to derive accurate parameter estimations of individual trees. All the algorithms in this benchmarking filtered and removed ground points before the stem detection step leaving an impression that this is a standard step in the processing chain. However, it is worth noting that the removal of ground points decreases the data volume but is not necessary for feature extraction.

Ground point filtering and terrain surface modelling have been among the most focused topics ever since laser scanning (or LiDAR) point clouds became available. Most of the DTM generation methods involve two main steps, i.e., the extraction of ground points and the interpolation of the terrain surface.

The major challenge for TLS-based DTM generation comes from 1) complex terrains; 2) the occlusion effects caused by the shadows brought by objects, e.g., bushes, low vegetation and tree stems; and 3) the TLS point distribution that becomes sparser with increasing distance from the scanning position, especially in the single-scan approach. Therefore, a new factor called DTM coverage is introduced as an additional indicator for DTM evaluation. This factor indicates the ratio between the areas of the extracted and the reference DTMs. The reference DTMs were built from multi-scan TLS data and cover the entire plot area. The closer the ratio is to 100%, the larger the plot area that is covered by the DTM built from the point cloud data. In general, a low RMSE and almost 100% DTM coverage are expected.

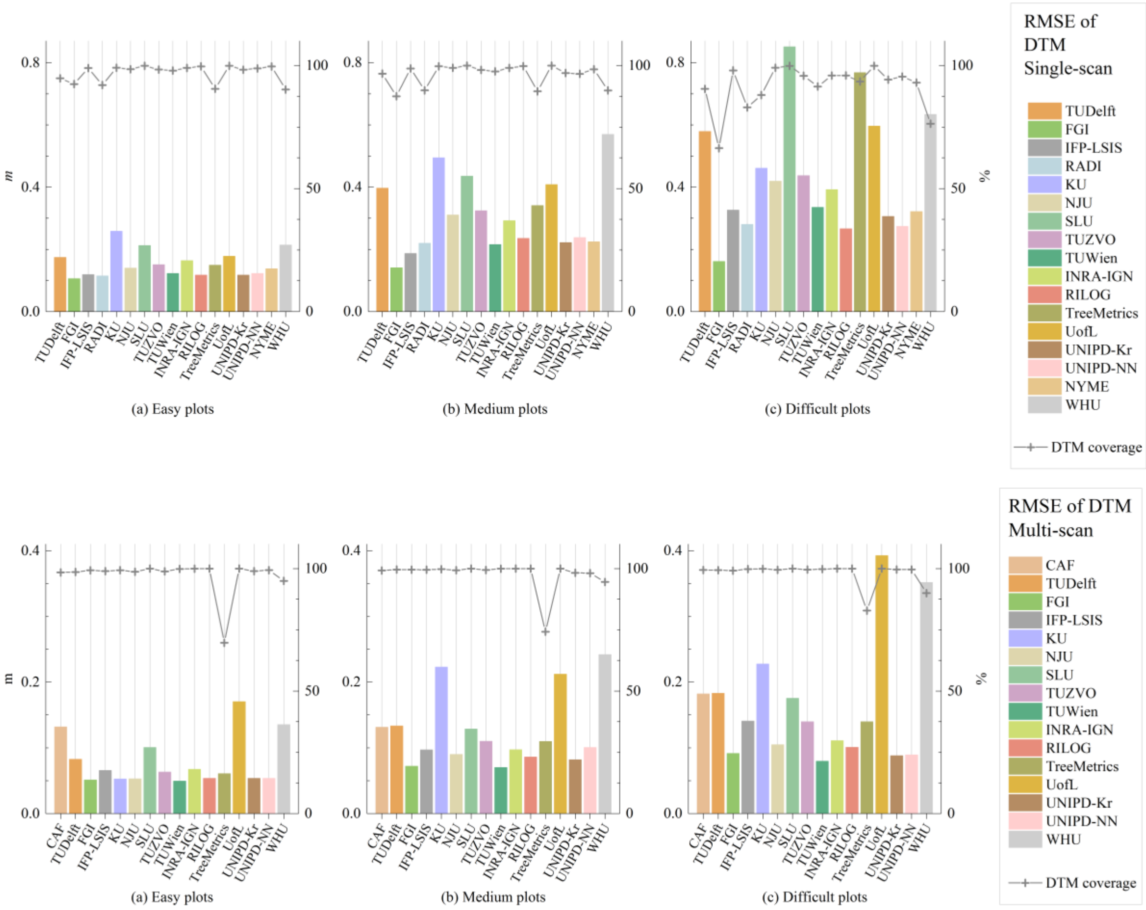


Figure 1: RMSE of DTM built from the single-scan (upper) and multi-scan (lower) data. The left vertical axis corresponds to the RMSE, and the unit is meter; the right vertical axis corresponds to the DTM coverage, and the unit is percentage.

The RMSE of the DTM increases as the stand complexity increases in both single- and multi-scan point cloud data. The more complex the stand is, the more shadows exist on the ground, and the more difficult it is to reconstruct terrain surface.

As shown in Figure 1, in many cases, a high DTM coverage requires not only interpolation but also the extrapolation based on the extracted ground points, and the amount of applied extrapolation significantly influences DTM accuracy. One strategy to build accurate DTMs is to focus on areas where the signal penetrates ground vegetation well and where point cloud data are reflected from the ground, which may sacrifice DTM coverage, e.g., giving up the extrapolation at the plot border leads to a smaller size of the DTM, especially in the single-scan scenario. In such cases, the best achievable RMSE values of the DTM (FGI) are 0.10 m (92.5% coverage), 0.14 m (87.5% coverage), and 0.16 m (66.4% coverage) in easy, medium and difficult plots, respectively, with the single-scan data. In contrast, when high DTM coverage is pursued, the best achievable RMSE (RILOG) values are 0.12 m (99.7% coverage), 0.24 m (99.8% coverage), and 0.27 m (95.9% coverage) in easy, medium and difficult plots, respectively.

The application of the multi-scan approach can reduce shadows on the ground; therefore, high accuracy can be expected without losing the coverage of the DTM. Seven out of sixteen algorithms, e.g., TUWien, provide similar DTM results in terms of accuracy and coverage when the multi-scan data are applied. The average RMSE and coverage of DTM across the seven algorithms with similar accuracies are 0.05 m (99.7% coverage), 0.08 m (99.6% coverage), and 0.10 m (99.7% coverage) in easy, medium and difficult plots, respectively. For the seven methods, the differences in the DTM accuracies between different stand-complexity categories are moderate, indicating that the algorithms for DTM generation are well designed.

Extrapolation introduces errors in DTM generation, as revealed by the results from the FGI and TUWien. Algorithms by the FGI and TUWien have similar performances in all stand-complexity categories using the multi-scan data. The differences were visible in single-scan data. The FGI gave smaller RMSE and coverage values, while TUWien gave a lightly larger RMSE, but it covers the plots area more completely, indicating that extrapolation is the main error source.

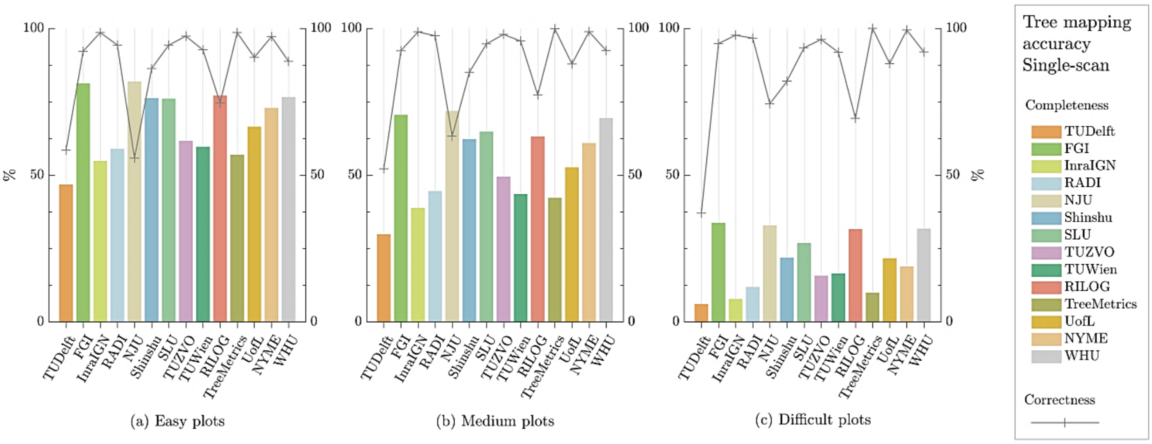
3.2 Stem Detection Accuracy

Stem detection accuracy is evaluated by the completeness, the correctness and the mean accuracy of the detected trees in each sample plot. The completeness measures how many reference trees have been found by an algorithm. The correctness measures how many detected trees from an algorithm correspond to the reference trees. The mean accuracy provides an indication of how an algorithm is balanced between the omission (missing reference trees) and the commission (finding redundant trees) errors.

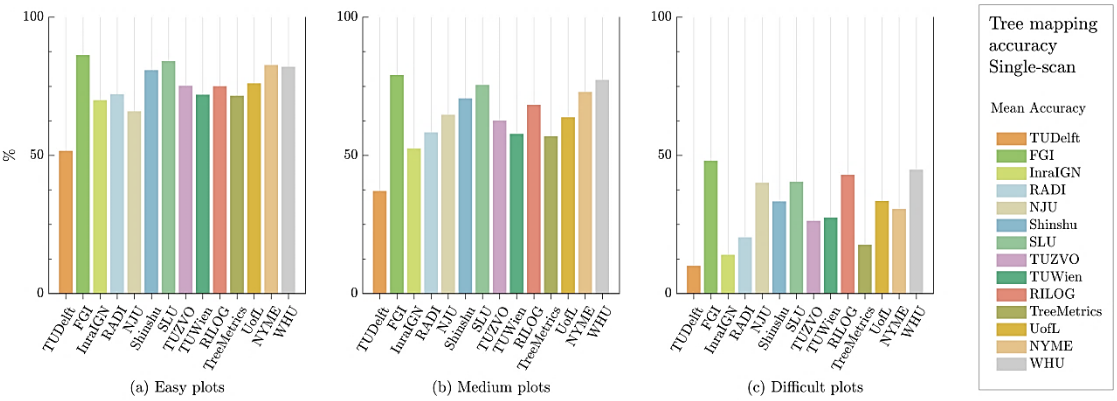
Evaluation results for tree detection accuracy utilizing the single- and multi-scan TLS data are presented in Figures 2 and 3, respectively. The completeness and correctness are illustrated in the same figure to intuitively demonstrate the trade-off between these two characteristics and how different algorithms choose their priorities. In general, the efforts placed on detecting more trees, especially the small trees, lead to higher commission errors, namely, when pursing higher completeness of tree detection, the risk of obtaining a lower correctness increases.

In an ideal scenario, an algorithm should be capable of providing high level of both completeness and correctness, which remain as a great challenge in reality. For most of the cases, the cost of higher completeness is a lower correctness and vice versa, which can be seen based on the relationship between the crossed lines and the bars in Figures 2 (a) and 3 (a). For example, a tall bar, i.e., high completeness, is usually accompanied by a low corresponding cross, i.e., low correctness, and a high cross tends to be paired with a low bar. In extreme cases, an algorithm can achieve over 80% completeness in easy forest stand conditions from single-scan data, with the price of obtaining correctness that is below 60%, i.e., approximately 40% of the detected trees are commission errors. At the other end of the spectrum, an algorithm can achieve 100% correctness, i.e., all the detected trees are correct in terms of corresponding to reference trees while having a relative low completeness at approximately 60%.

The mean accuracy illustrated in Figures 2 (b) and 3 (b) presents a balanced evaluation between the completeness and correctness. High correctness will compensate for low completeness of an algorithm and vice versa. The comparison of the results indicates that when the completeness and correctness are averaged, most of the algorithms perform at a similar level. The variations in the mean accuracies amongst the algorithms in each complexity category are much less than the completeness and the correctness, regardless of the stand condition and the applied TLS data.



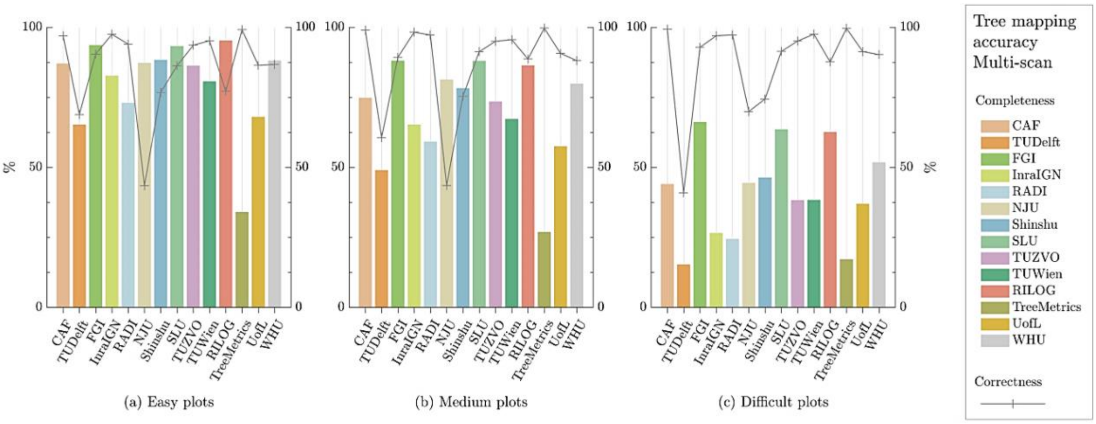
(a) Completeness and correctness



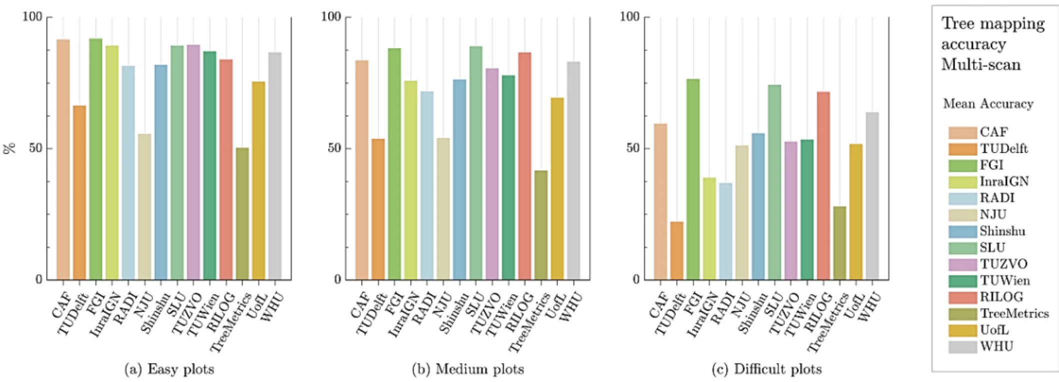
(b) Mean accuracy

Figure 2: The accuracy of tree mapping from single -scan TLS data. (a) The completeness and the correctness: the left vertical axis corresponds to completeness (bars), and the right vertical axis corresponds to the correctness (crossed line). (b) The mean accuracy. Units in both (a) and (b) are percentages.

Considering the tendency of the algorithms toward detecting more trees or safeguarding the credibility of the detected trees, three types of algorithms can be distinguished, namely, “Aggressive”, “Conservative” and “Robust”. An “Aggressive” algorithm emphasizes the completeness by trying to detect as many trees as possible and has a relatively high tolerance for false detection, which might lead to low correctness. In contrast, a “Conservative” algorithm cares more about the correctness of the detected trees by focusing on trees that are comprehensively recorded in the point cloud, which would reduce the total amount of detected trees. A “Robust” algorithm tends to keep a balance between the completeness and the correctness by pursuing the highest possible accuracy for both factors. According to the evaluation results, most of the algorithms in this benchmarking were designed in a "Conservative" (InraIGN, RADI, TUZVO, TUWien, TreeMetrics, and UofL) or "Robust" (CAF, TUDelft, FGI, Shinshu, SLU, NYME, and WHU) manner, and two algorithms were following an "Aggressive" (NJU and RILOG) principle.



(a) Completeness and correctness



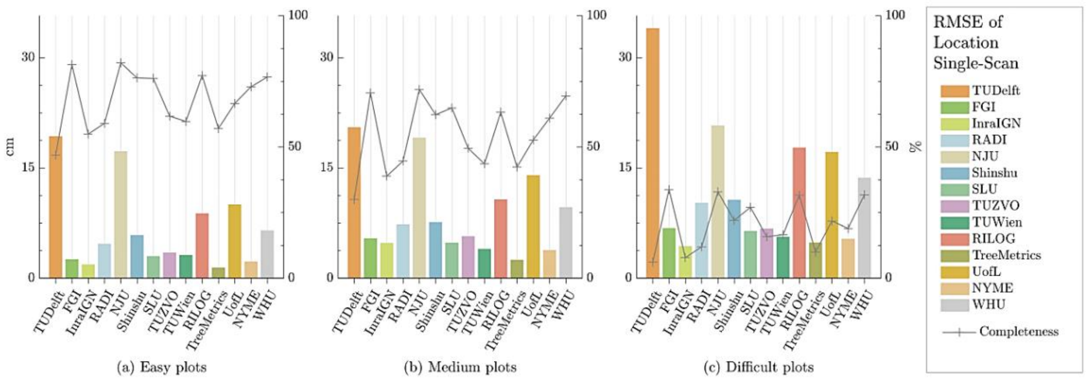
(b) Mean accuracy

Figure 3: The accuracy of tree mapping from multi-scan TLS data. (a) The completeness and the correctness, the left vertical axis corresponds to completeness (bars), and the right vertical axis corresponds to the correctness (crossed line). (b) The mean accuracy. Units in both (a) and (b) are percentages.

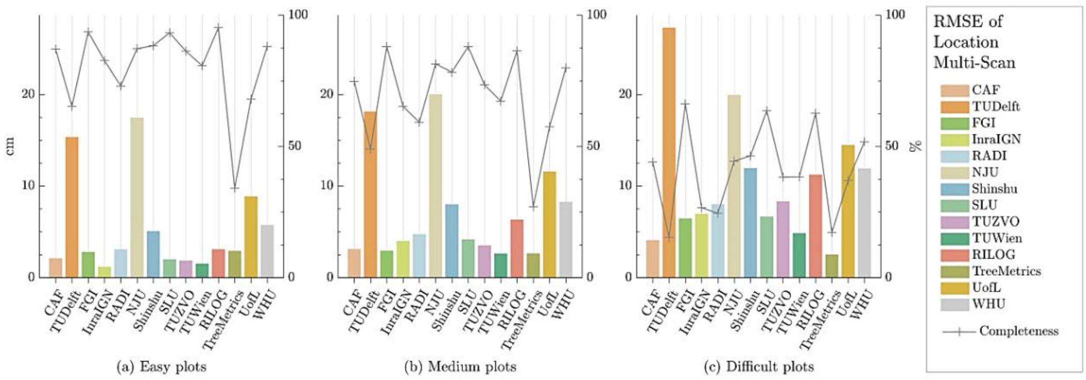
From the detection results, the status quo of the algorithms for stem detection from the TLS data can be summarized as follows: 1) In a simple forest stand condition, it is normal to achieve a mean accuracy that is approximately 75% with single-scan data and 80% with multi-scan data. With the best efforts focusing on balancing between the omission and the commission errors, the completeness can reach 81.3% while having a correctness that is 92.2% in a single-scan scenario, and 90.4 % completeness with 93.6% correctness utilizing multi-scan data. 2) With an increase in forest stand complexity, the performance of stem detection decreases significantly. In a single-scan condition, the average mean accuracy of all the algorithms for medium difficulty stands is approximately 64% and is even lower, i.e., ca. 31% for difficult stands. The best achievable pairs of completeness and correctness are 70.6% and 92.4%, respectively, for medium plots and 33.8% and 94.8%, respectively, for difficult plots. 3) The application of multi-scan strategy will improve detection accuracy for the medium and difficult stands by raising the average mean accuracy to approximately 74% for medium plots and to approximately 53% for difficult plots. The best achievable completeness pairing with correctness is 88.0% with 89.2% for medium plots and 66.2% with 92.8% for difficulty plots. 4) While the completeness decreases sharply in complex stand conditions, the correctness of the algorithms appears to be stable in different forest stands. The correctness is commonly above 90% in all three complexity categories, indicating that the detecting algorithms are mostly reliable. 5) The application of multi-scan strategy has a greater impact on the completeness than on the correctness. Algorithms, as well as associated TLS data, seem to be reliable when the detections are mostly correct and when commission errors are low in both single- and multi-scan scenarios.

3.3 Stem Location

The location of detected stems has a high accuracy level. Using the single-scan data, most of the algorithms can provide the stem location at RMSE levels of below 5 cm in easy plots, below 8 cm in medium plots, and below 10 cm in difficult plots, as shown in Figure 4. With multi-scan data, the RMSE of stem location can commonly be controlled to 2-3 cm in easy plots, 2-5 cm in medium plots, and 4-9 cm in difficult plots.



(a) Stem location accuracy from single scan data



(b) Stem location accuracy from multi-scan data

Figure 4: The accuracy of stem location (a) from single-scan data and (b) from multi-scan data. In both (a) and (b), the left vertical axis corresponds to the RMSE of the stem location (bars), and the unit is cm; the right vertical axis corresponds to the correctness (crossed line), and the unit is percentage.

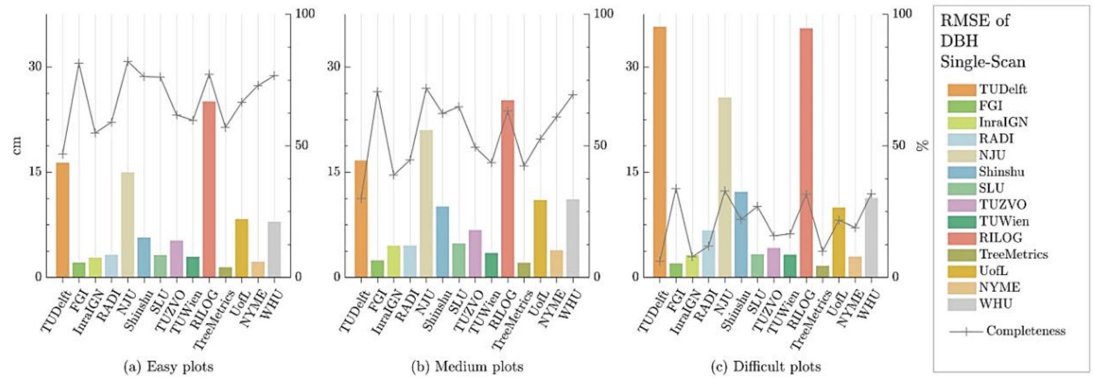
It is worth noting that the estimation accuracy of an individual parameter itself cannot represent the overall performance of an algorithm. The plot-level feature-extraction results require inspection in the context of the stem detection rate. For example, a method may achieve the best parameter estimation by focusing on only the trees that are creditably recorded in the data while omitting those are inadequately recorded, which gives high correctness and accurate parameter estimates, but sacrifices the completeness of stem detection. In contrast, a method may provide high completeness but sacrifices the parameter estimation accuracy. In between these two cases, a method may manage to provide decent estimation results while achieving a high level of stem detection completeness. Such results demonstrate typical cases for the selection of different algorithm development principles, i.e., Conservative by TreeMetrics, TUWien and InraIGN, Aggressive by NJU and RILOG, and Robust by FGI and SLU. Similar phenomena can also be observed for other parameters in the following sections. In this context, the completeness of stem detection is always illustrated as background information in the figures on parameter accuracy.

3.4 Diameter at Breast Height

The accuracy of DBH estimations is evaluated using the statistical factors RMSE, RMSE%, bias, and bias%, respecting the field measurement results in the reference data. The evaluation results are illustrated in Figures 5 (RMSE) and 6 (bias) for single-scan data, and in Figures 7 (RMSE) and 8 (bias) for multi-scan data. The bars in the figures represent the corresponding values of a statistical factor. In addition, the completeness of stem detection is illustrated with plus signs and a crossed line, through which a more comprehensive understanding of the accuracy of the DBH estimations can be derived.

3.4.1 RMSE of DBH estimations

Among the fourteen algorithms that provided DBH estimations, three algorithms (FGI, SLU, and NYME), which are in the "Robust" stem detection category, delivered an RMSE ranging from 2-4 cm and RMSE% ranging from 8-20% for all three stand difficulty categories from single-scan data, with above average completeness of stem detection. On the other hand, the Conservative algorithms (e.g. TreeMetrics) are capable of providing DBH estimates with even lower RMSE and RMSE% values, at 1-3 cm and below 10%, respectively, given that the completeness of stem detection is below the average level. For many algorithms, the accuracy of DBH estimations remained stable across the three stand complexity categories.



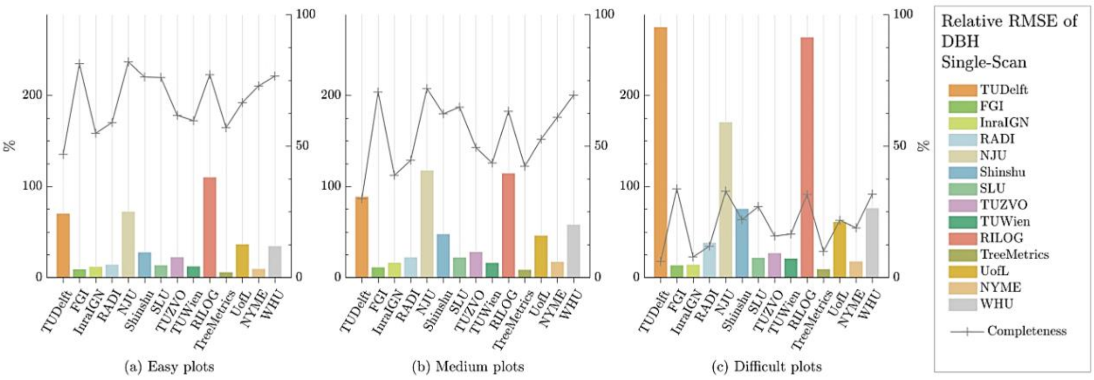


Figure 5: RMSE and RMSE% of the DBH estimation from single-scan TLS data. The left vertical axis corresponds to the RMSE (bars in upper subfigure, the unit is centimeter) and RMSE% (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage

Multiple scans clearly improve the accuracy of the DBH estimations. The RMSE can be reduced to less than 2 cm for all three complexity categories with the “Robust” and “Conservative” algorithms based on the multi-scan data. The RMSE% can be reduced to a range of 5-10% for easy and medium stands and 10-15% for the difficult forest stands.

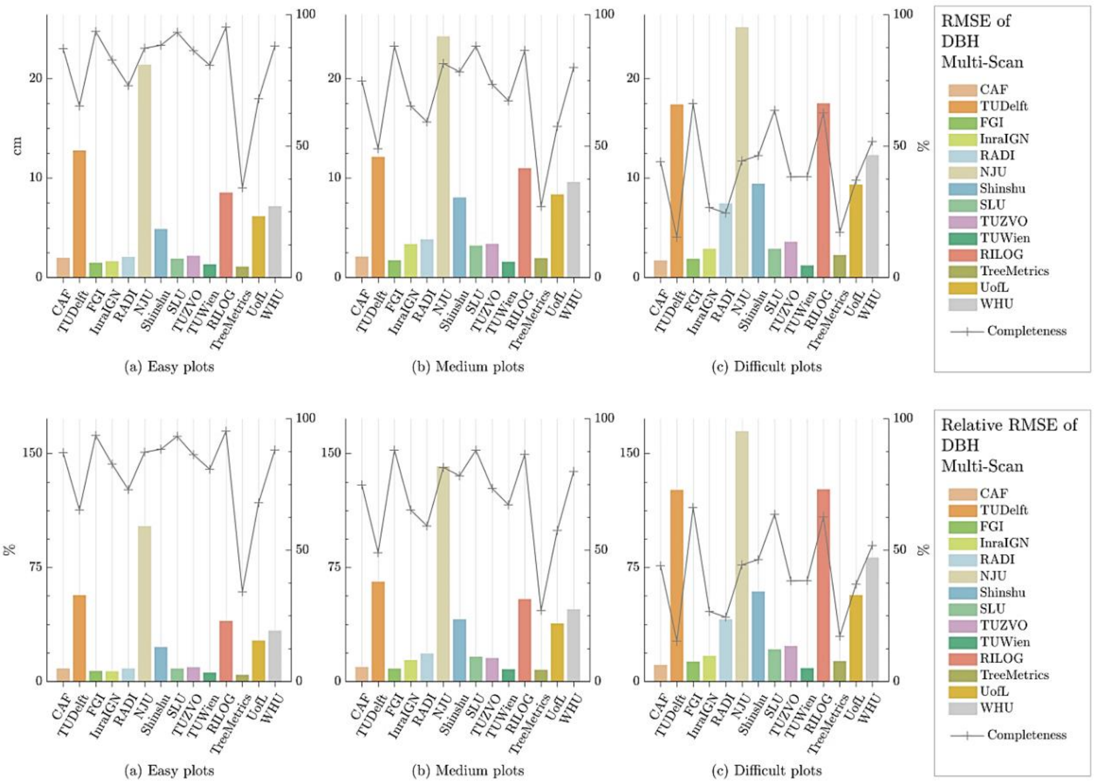


Figure 6: RMSE and RMSE% of the DBH estimation from multi-scan TLS data. The left vertical axis corresponds to the RMSE (bars in upper subfigure; the unit is centimeter) and

RMSE % (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage

3.4.2 Bias of the DBH estimation

The bias of the DBH estimate is expected to be as close to zero as possible. A small bias accompanied by high stem detection completeness indicates that an algorithm is capable of carrying out unbiased DBH estimation while detecting more trees.

When utilizing the single-scan data, the biases of DBH estimation of different algorithms diverse. Almost half of the algorithms tend to overestimate, while the other half of the algorithms tend to underestimate the DBH (see Figure 7). The results in this benchmarking reveal that current algorithms are capable of estimating DBH at a level close to zero for bias and bias% for sample plots in simple and medium complexity categories. At least four algorithms in this benchmarking achieved this goal. It is easier for the “Conservative” and “Robust” algorithms to achieve low bias on DBH estimations. For some algorithms, the bias of the DBH estimation can be extremely high, e.g., over ±50% for bias%, which is most likely because the DBH estimations of those algorithm (e.g., circle fitting) are largely impacted by a few outliers or gross errors.

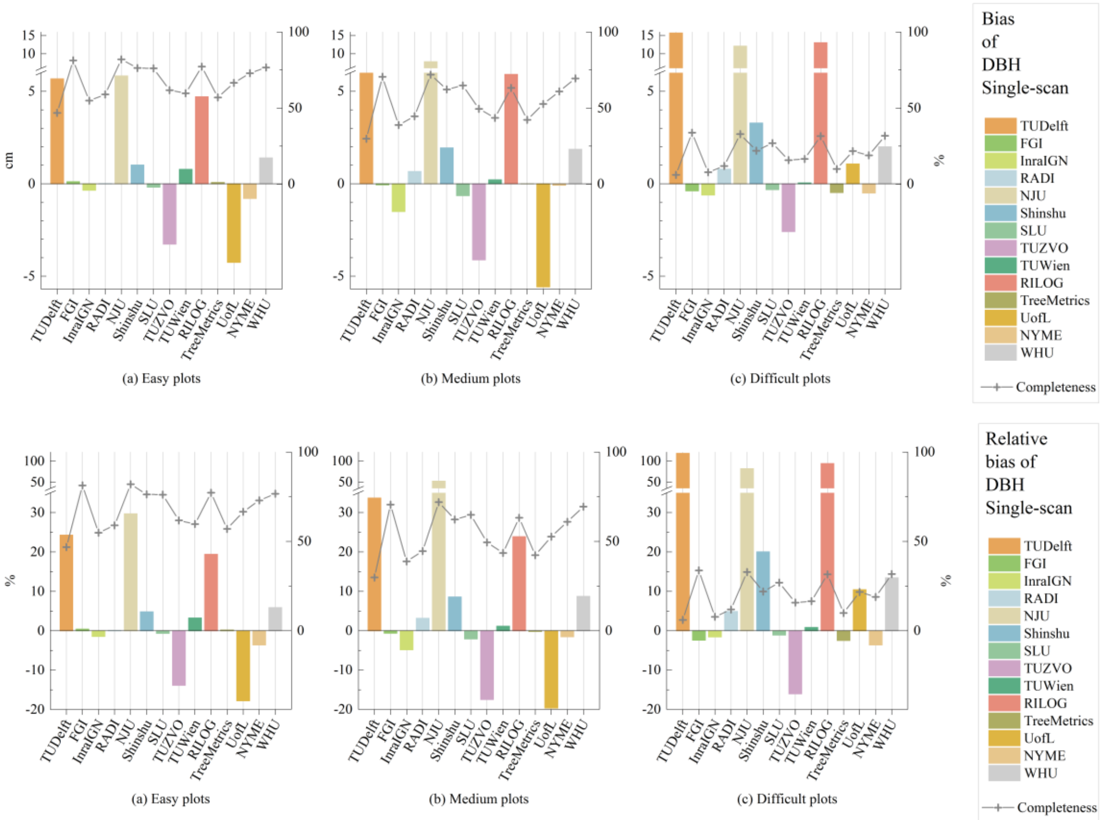


Figure 7: Bias of the DBH estimation from single-scan TLS data. The left vertical axis corresponds to bias (bars in upper subfigure; the unit is centimeter) and bias% (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage

The application of multi-scan data can largely reduce the underestimation of the DBH; only three algorithms produced negative bias (see Figure 8). The bias of DBH estimations also confirms that the multi-scan approach is helpful for improving DBH estimation accuracy. The more complex the stand condition is, the more significant the advantage of applying the multi-scan strategy is. For the “Conservative” and “Robust” algorithms, the bias and bias% can be kept close to zero in all three stand complexity categories.

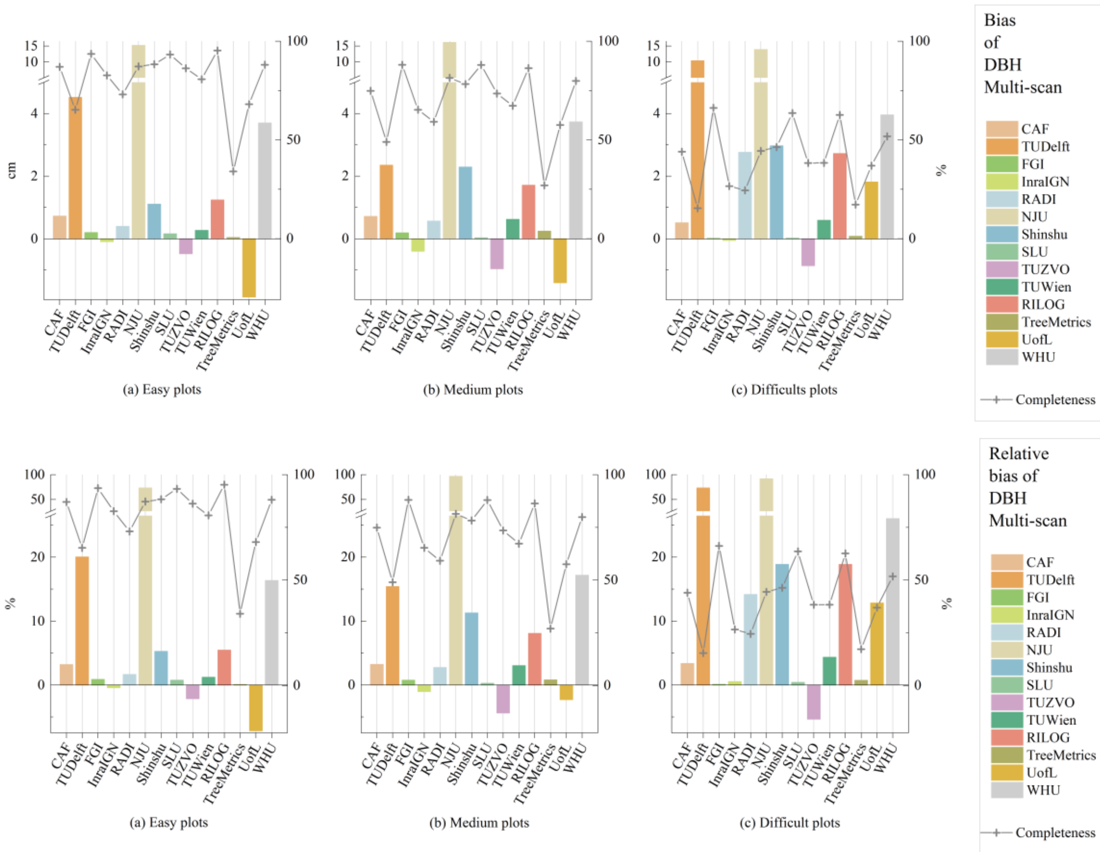


Figure 8: Bias of the DBH estimation from multi-scan TLS data. The left vertical axis corresponds to bias (bars in upper subfigure; the unit is centimeter) and bias% (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage

3.5 Tree Height

The tree height estimation is evaluated using the statistic factors RMSE, RMSE%, bias, and bias%, respecting the reference data, as with the DBH. The evaluation results are illustrated in Figures 9 (RMSE) and 10 (bias) for single-scan data, and in Figure 11 (RMSE) and 12 (bias) for multi-scan data. For the convenience of interpreting the benchmark results, the figures have layouts that are similar to those used for the DBH. Ideally, an algorithm is expected to provide a low RMSE and an almost zero bias with high stem detection completeness.

3.5.1 RMSE of the tree height estimation

A similar performance for tree height estimation is observed for different algorithms. With the single-scan data, the RMSE and RMSE% of most algorithms in the easy plots ranged from 2.4-4.5 m and 12-23%, respectively. Tree height estimations become more difficult when the stand complexity increases because the determination of the treetops of sub canopy trees and of small trees in dense forest stands is much more demanding. The RMSE and RMSE% of tree height estimation decrease to ranges of 3.5-7.8 m and 18–41%, respectively, in medium sample plots and of 4.0 – 7.7 m and 28% - 57%, respectively, in difficult sample plots. In medium and difficult plots, although the absolute values of the RMSE were similar, the RMSE% in the difficult plots was clearly larger than that in the medium plots because the difficult plots have many small trees.

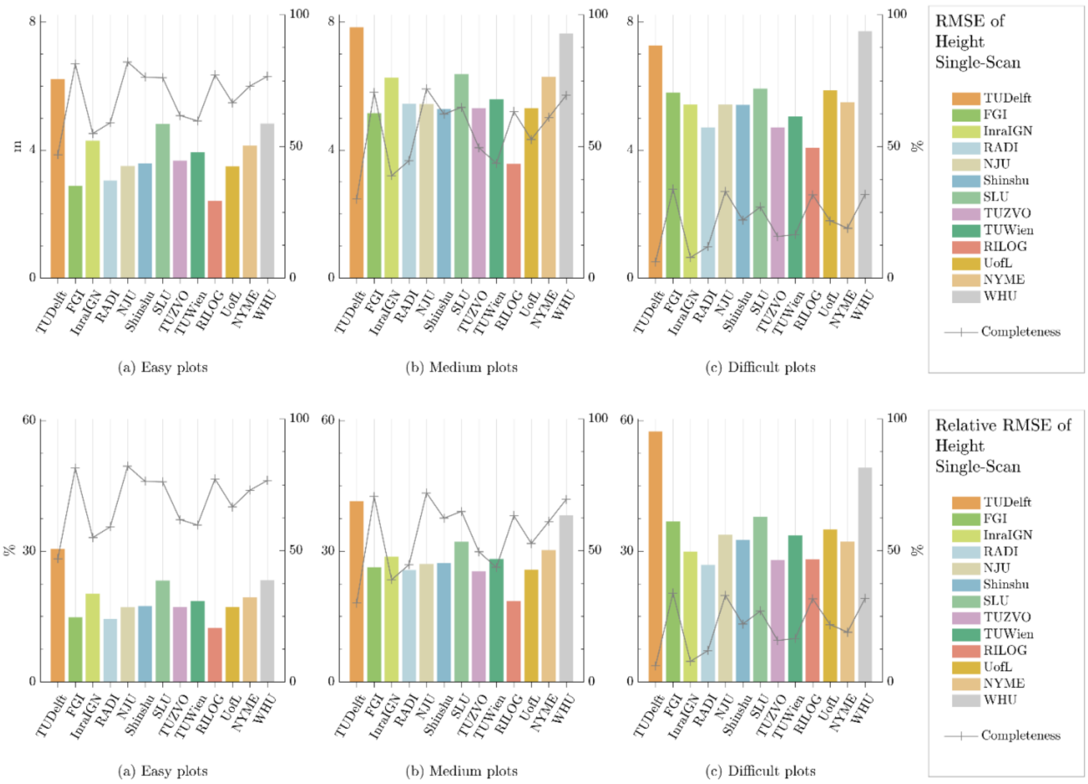


Figure 9: RMSE of the tree height estimation from single-scan TLS data. The left vertical axis corresponds to RMSE (bars in upper subfigure; the unit is meter) and RMSE % (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage

The improvement brought by the multi-scan approach for the tree height estimations is not as significant as for DBH estimation. For all the algorithms, the RMSE and RMSE% are approximately 2.8 m and 13% on average, respectively, under simple stand conditions; 4.4 m and 23%, respectively, for medium stand conditions; and 4.7 m and 30%, respectively, for complex stand conditions.

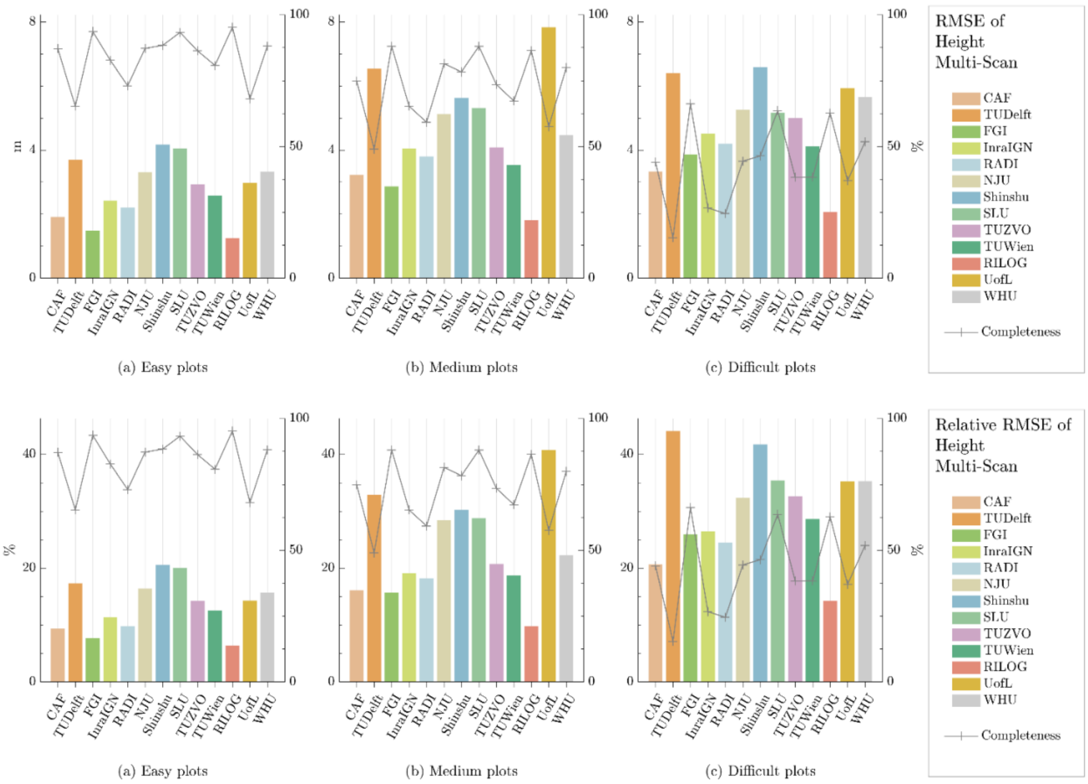


Figure 10: RMSE of the tree height estimation from multi-scan TLS data. The left vertical axis corresponds to RMSE (bars in upper subfigure; the unit is meter) and RMSE % (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage.

These results indicate that tree height estimation from the TLS data is still quite challenging. Due to the limitation of the terrestrial perspective for data acquisition, the treetops for most of the trees in a sample plot can hardly be recorded, even with the multi-scan approach. The algorithm RILOG presents the best results for tree height estimation by giving RMSE values of 2.4 m, 3.6 m and 4.1 m for the easy, medium and difficult plots, respectively, using the single-scan data and 1.2 m, 1.8 m and 2.1 m, respectively, using the multi-scan data, based on the condition that the individual trees were manually extracted from the point cloud. Therefore, the evaluation results of the RILOG algorithm can be interpreted as milestones of tree height estimation based on the TLS data, where tree detection errors were minimized by using manual segmentation.

The hardware can also influence the accuracy of tree height estimates. In this study, the scanner is phase-based, which is prone to noise points. A pulse-based scanner may capture point cloud data that are less noisy and may have a better chance of recording treetops from multi-returns.

3.5.2 Bias of the tree height estimation

It is approved in this benchmarking that the TLS-based approaches underestimate tree heights. The tree height estimations present negative biases for almost all the algorithms, with only a few exceptions. The average underestimation for the tree heights is approximately 2.2 m (bias) and 10% (bias %) across all the sample plots and algorithms when utilizing the single-scan data.

Since it is commonly assumed that the TLS will record lower tree heights, some algorithms act aggressively and risk taking the tree heights from the upper layer tree crowns for the secondary layer trees. Under such circumstances, it is possible to overestimate the tree heights, especially in complex forest stands where the sub canopy growth is rich, which explains the positive bias and bias% in Figure 11.

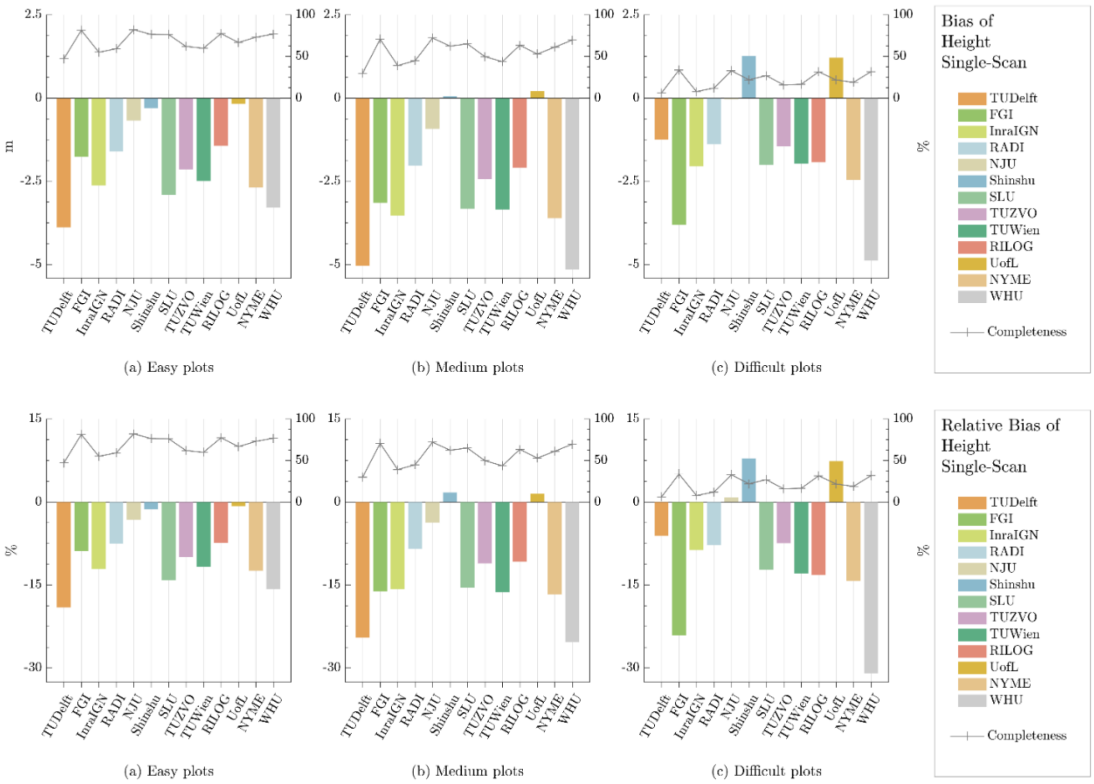


Figure 11: Bias of the tree height estimation from single-scan TLS data. The left vertical axis corresponds to bias (bars in upper subfigure; the unit is meter) and bias % (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage.

The effect of being “Aggressive” in estimating the tree heights from the TLS data becomes more obvious when applying the multi-scan data, where more algorithms overestimate the tree heights. The results in Figure 11 suggest the degree of aggressiveness of the algorithms. When more TLS points are provided in the multi-scan data, the risk of taking the wrong treetop locations for small and sub-canopy trees becomes higher by extracting the tree height from the TLS points that are close to the stem area. The overestimation of tree heights worsens when the forest stand condition becomes complex and when the amount of small and sub-canopy trees increases. On the other hand, the algorithms that continue to underestimate the tree heights in the difficult sample plots when utilizing the multi-scan data can be considered as “Conservative” for tree height estimations, and approximately 2/3 of the algorithms in this benchmarking belong to this category.

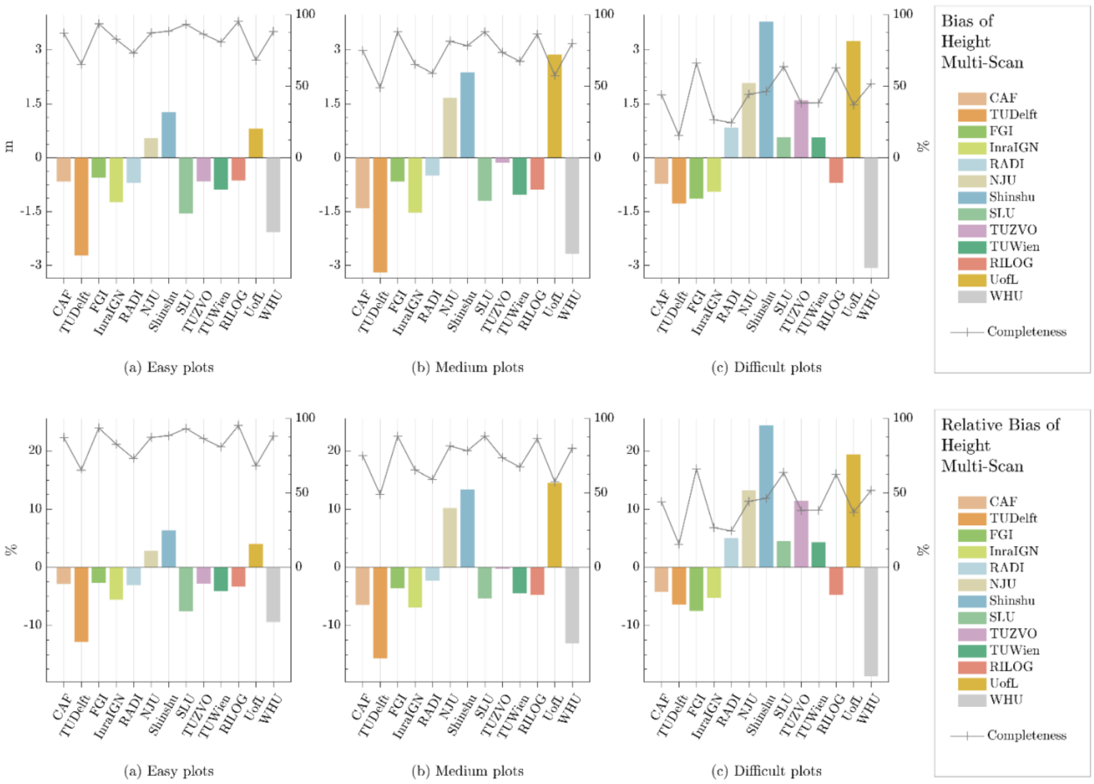


Figure 12: Bias of the tree height estimation from multi-scan TLS data. The left vertical axis corresponds to Bias (bars in upper subfigure; the unit is meter) and bias % (bars in lower subfigure; the unit is percentage). The right vertical axis corresponds to the completeness of stem detection (crossed line), and the unit is percentage.

By using multi-scan data, both the RMSE and bias decrease, but the bias is substantially reduced, which is an advantage of using multi-scan data.

3.6 Stem Curve

The accuracy of the stem curve estimation is evaluated with the mean RMSE and mean bias of the stem diameters at different height levels of each matched stem, and these mean values are further averaged over the plots in the same complexity category. Additionally, to measure the efficiency of the algorithm in the stem curve estimation, two new evaluation factors, i.e., the curve length ratio (CLR) and the percentage of the tree height covered (PHC), are investigated. The CLR is defined as the ratio (in percentage) of the stem length covered by the extracted curve to that covered by the reference curve. The PHC is defined as the ratio (in percentage) of the stem length covered by the extracted curve to the reference tree height. The CLR reveals how well the stem curve extraction methods perform compared to reference measurements, e.g., manual measurements by laser scanning experts which indicate the best that a human operator can achieve from a point cloud. The CLR may be larger than 100%, which means the method extracts more diameters than the manually measured reference data or that the computer over-performs human beings if the method is fully automated. The PHC reveals the degree of the whole tree that is retrieved by the extraction methods, with 100% being the goal where an algorithm fully depicts the object.

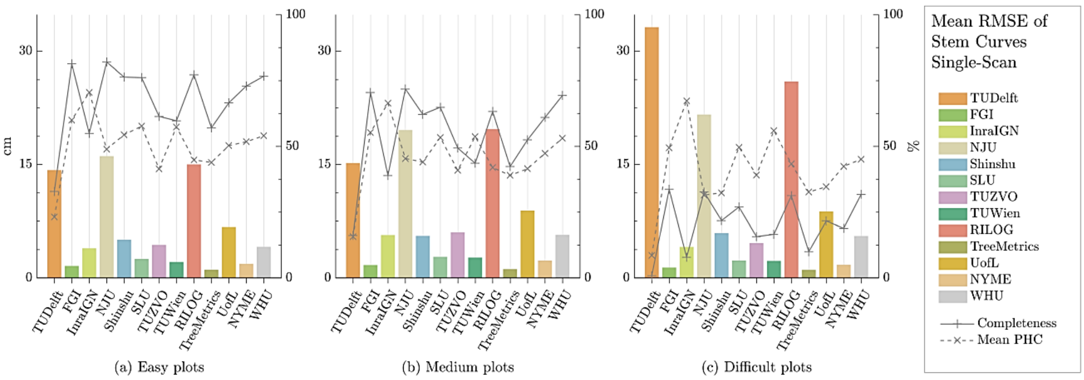
The PHC indicates the capability of the TLS point cloud and an algorithm to depict the object in the field.

3.6.1 RMSE of the stem curve estimation

The RMSE of the tree-wise stem curve estimation from the TLS data is illustrated in Figure 13. To remain consistent with the DBH and tree height, the basic layouts of the figure remain the same as in previous figures, except for an extra dashed line representing the PHC of each algorithm.

A faithful understanding of the performance of stem curve extraction can only be derived through a consideration that integrates of the RMSE, the PHC, and the stem detection completeness. While the completeness indicates the number of trees detected in a plot, the RMSE and PHC indicate the capability of an algorithm in stem modelling, with the RMSE referring to the accuracy of the estimated diameters at different heights of a stem and the PHC measuring the proportion of the stem that is modelled. A “Robust” algorithm should present a small RMSE and large PHC with high completeness of stem detection. With an “Aggressive” algorithm, a large RMSE can be expected for large a PHC with high completeness. When an algorithm presents a small RMSE with low completeness, it is considered “Conservative” regardless of its PHC value.

It is important to simultaneously take the three factors into account, and missing any of them will lead to a biased evaluation. For instance, if only the RMSE and PHC are considered, the “Conservative” and the “Robust” algorithms may perform similarly, where both designs give small RMSE and large PHC values. However, there is a clear difference in how these two strategies achieve these results. The “Conservative” algorithms achieve a small RMSE and a large PHC by accurately reconstructing stems that were completely and clearly recorded in the point cloud, which may constitute only a small portion of the stems in the plot. In contrast, the “Robust” algorithms provide plausible stem curve estimations for a large number of stems, part of which are not recorded by the high-quality TLS points, which is the strength of the algorithm. The perception of the performance of the algorithm can only be justified when the completeness of stem detection is properly referenced.



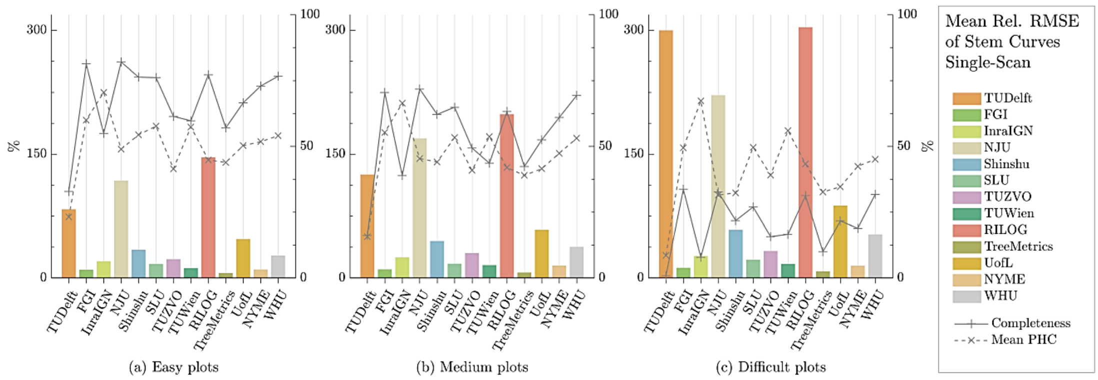


Figure 13: RMSE of the tree-wise stem curve estimation from the single- (upper) and multi-scan (lower) TLS data. The left vertical axes correspond to the RMSE value (bars), and the unit is centimeter. The right vertical axes correspond to the completeness of stem detection (crossed line) and the mean PHC of the stem curve (crossed dash-line), and the unit is in percentage.

The benchmarking reveals that the mean RMSE of stem curve estimation is relatively stable in terms of the stand conditions. Except for minor cases, most of the algorithms can achieve a mean RMSE that is between 1.3 and 6.0 cm from single-scan data, and between 0.9 and 5.0 cm from multi-scan data for all three stand complexity categories. The impact of the stand complexity on the PHC is greater than that on the RMSE. A decrease in the PHC can be observed when the complexity of the forest stand increases. As expected, the PHC from the multi-scan data is higher than that from the single-scan data. The average PHCs across all the algorithms are 52%, 48% and 43% for the simple, medium and difficult plots, respectively, from the single-scan data and 57%, 54% and 50%, respectively, from the multi-scan data.

The RMSE% of the DBH and stem curve estimates are quite similar for some algorithms (e.g., FGI, SLU, TUWien and TreeMetrics) from both single- and multi-scan and across the three forest complexity categories. The stem curve estimates may even be more accurate because more data are used, which lead to a better average. This similar accuracy, leaving aside the mean PHC, indicates that algorithms can have the same capacity in estimating the DBH and stem curve. In the future, when TLS-based forest inventories are applied, stem curves can be used as a tree attribute similar to DBH.

3.6.2 Bias of the stem curve estimation

The mean bias of the stem curve estimation also needs to be investigated in combination with the PHC and detection completeness. An almost zero mean bias with a high PHC and high completeness is expected for the stem curve estimation. In this benchmarking, diverse reactions are observed from the algorithms. Some algorithms have difficulty extracting the stem curve above the first branch height, leading to an over 10cm mean bias per stem from the single-scan data or even multi-scan data, indicating that those algorithms are intolerant of noisy stem data, especially in the tree crown, or give biased estimates. Such performance indicates that the estimation of the stem curve is a challenging task for most of the algorithms.

Stem curves extracted from the multi-scan data tend to have a higher positive bias than those extracted from the single-scan data. In general, the mean bias of the stem curve estimations becomes positive when the multi-scan data is applied, regardless of the stand situation. One reason is that multi-scan data provides a more complete stem structure; therefore, the estimated stem diameters at different stem heights increase. Another possible reason is that there is more noise surrounding the stems in the multi-scan data, especially at the crown, due to the branches or the registration errors of multiple scans.

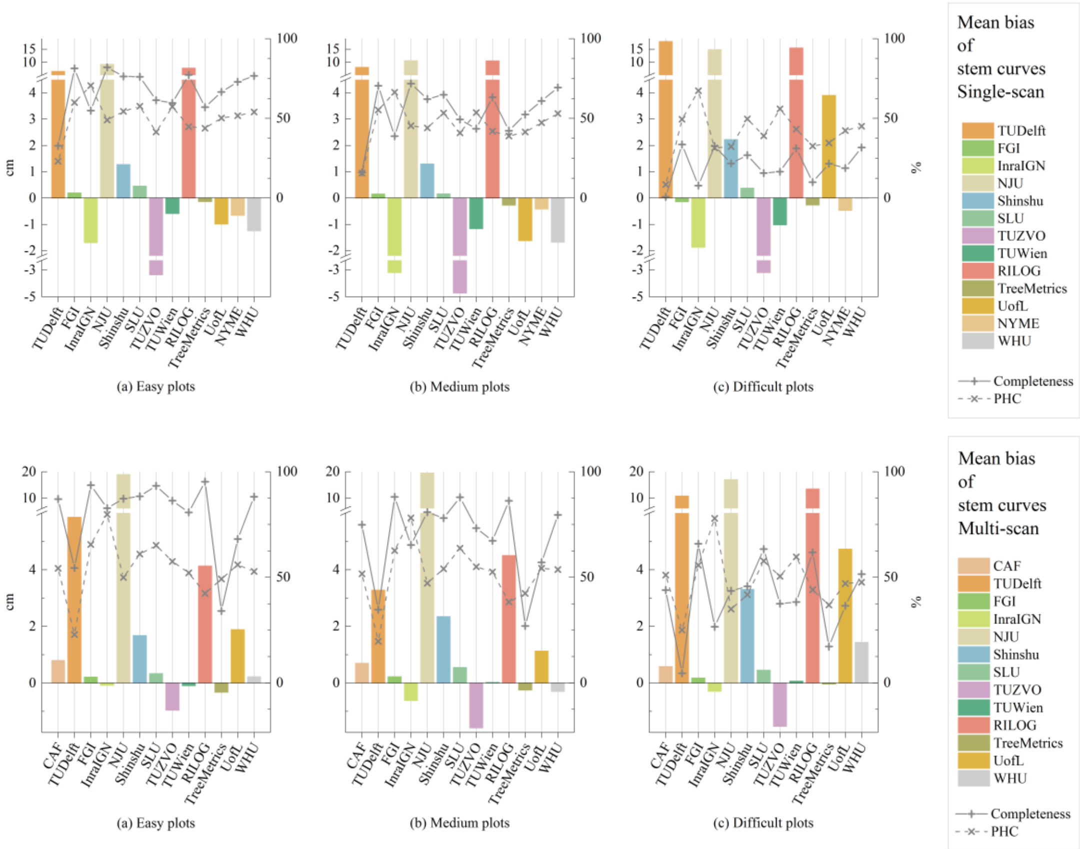


Figure 14: Bias of the tree-wise stem curve estimation from the single- (upper) and multi-scan (lower) TLS data. The left vertical axes correspond to the RMSE value (bars), and the unit is centimeter. The right vertical axes correspond to the completeness of stem detection (crossed line) and the mean PHC of stem curve (starred dashed line), and the unit is in percentage

3.6.3 Curve length ratio of the stem curve estimation

The CLR, which is the ratio between the lengths of the automatically and manually extracted stem curves, is a more effective indicator of the capacity of the automated algorithms in extracting the stem curve. An almost 100% CLR is expected, which indicates that the automated results at least reach a similar level of coverage as human interpretation. However, the current automated approaches lag behind the manual process of stem-curve extraction except for time efficiency, which indicates that the current algorithms are significantly affected by the incomplete stem structure in the point cloud and by the noise.

On the other hand, when only large trees that are recorded with high quality points are considered, i.e., with low completeness in stem detection, it is possible to extract more diameters automatically than manually, as shown by the results from the InraIGN method, i.e., CLR > 100%. In a more balanced scenario where both high stem detection completeness and high-quality stem curve extraction are expected, the best achievable CLR values with average completeness are 87%, 81% and 74% in the easy, medium and difficult plots, respectively, with single-scan data and 97%, 92%, 88%, respectively, with multi-scan data. The application of the multi-scan approach clearly improves the length of the extracted curve from the automated methods, adding approximately an extra 10% of the tree height of the extracted stem curve.

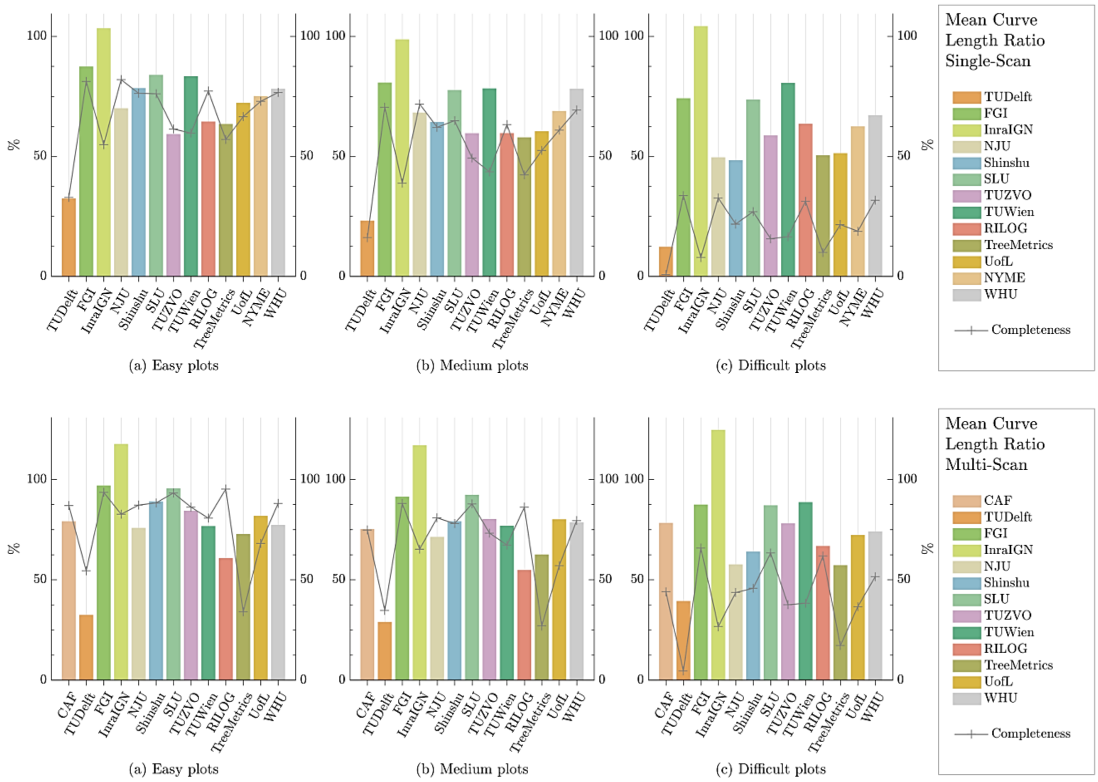


Figure 15: Mean CLR from the single- (upper) and multi-scan (lower) TLS data. The left vertical axes correspond to the CLR of stem curve (bars), and the unit is in percentage. The right vertical axes correspond to the completeness of stem detection (crossed line), and the unit is in percentage.

The results in this benchmarking also show that (Figure 14) the best achievable accuracy for the stem curve estimation from the single-scan data is approximately 0.2 cm mean bias, with 60% PHC and 81% completeness in simple plots; 0.2 cm mean bias, with 55% PHC and 70% completeness in medium plots; and -0.1 cm mean bias, with 49% PHC and 34% completeness in difficult plots. For the multi-scan data, the accuracy of the stem curve estimation can reach to 0.2 cm mean bias, with 65% PHC and 94% completeness in simple plots; 0.2 cm mean bias, with 63% PHC and 88% completeness in medium plots; and 0.2 cm mean bias, with 56% PHC and 66% completeness in difficult plots.

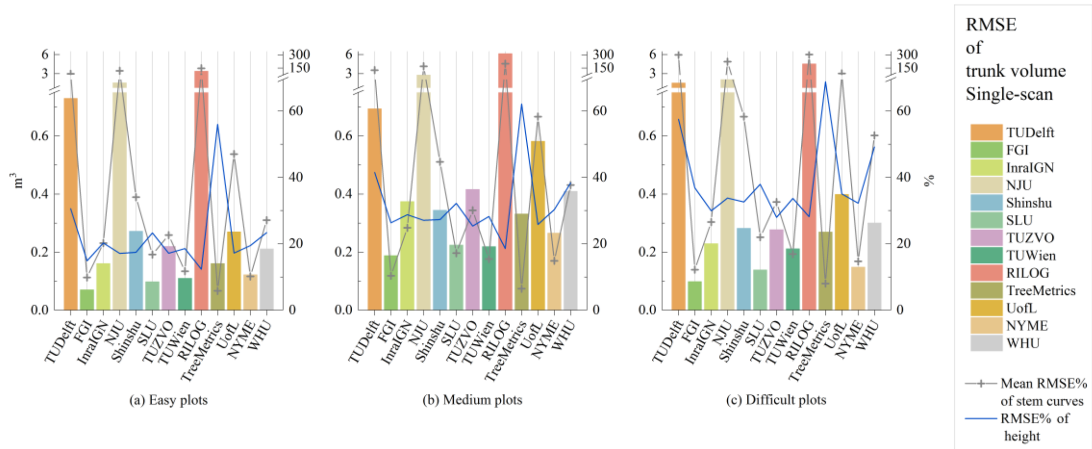
3.7 Stem volume estimation

As a function of the tree height and stem curve, the stem volume estimation reveals the overall performance of the extracted tree height and stem curve in an algorithm and also reveals the potential of applying TLS in forest inventories since stem volume is one of the most important tree attributes required by various applications. The stem volume estimation is evaluated on two different levels, i.e., the tree level and the plot level. While the tree-level evaluation explains the joint impacts of the tree height and stem curve estimations on the stem volume, the plot-level evaluation inspects the integrated impact of the tree-level attribute estimations and the stem detection accuracy.

3.7.1 Tree-level stem volume

The tree-level stem volume is calculated utilizing the stem curve and the tree height. More detailed information about the mathematical model can be found in another paper relevant to this benchmarking. The reference stem volume is calculated based on the manually extracted reference stem curve and the field-measured reference tree height using the same mathematical model as for the participant results. The estimated stem volume of each algorithm is calculated based on the stem curve and tree height results of the algorithm itself. While most methods estimate tree height from the TLS data, the TreeMetrics algorithm predicts the tree height from silviculture models for known tree species. In this comparison, the TreeMetrics algorithm did not estimate tree height. The stem volume is therefore calculated based on only the stem curve, so that the volume estimates of the treetop is missing in the calculation. It is worth noting that even though the volume tended to be underestimated; the volume estimates from TreeMetrics are still good, which indicates that the treetop volume makes up a relatively small part of the total volume.

The performance of the algorithms is evaluated by the RMSE and RMSE% between the estimated and the reference stem volumes. Figures 16 and 17 illustrate the evaluation results, and the basic layout of the figures is consistent with the other figures. The differences in these figures are that the completeness of the stem detection is left out and that two more indicators that are closely related to the stem volume estimates, i.e., the RMSE% of the stem curve and the tree height, are integrated to conduct more insightful analyses.



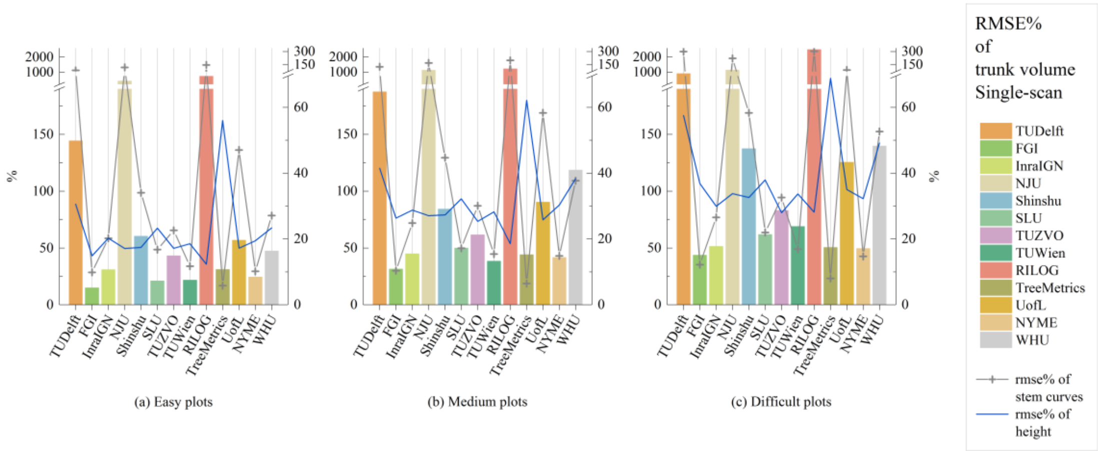


Figure 16: RMSE (upper) and RMSE% (lower) of the stem volume estimation from the single-scan TLS data. The left vertical axis corresponds to the RMSE value (bars), and the unit is m^3. The right vertical axis corresponds to the two relative indicators, i.e., the mean tree-level RMSE% of the stem curves (crossed line) and the RMSE% of the tree height (blue line), and the unit is percentage.

Both the RMSE and RMSE% should be considered when evaluating tree attribute estimates, which can clearly be seen from the stem volume evaluation. Excluding the three extreme cases (TUDelft, NJU and RILOG), the average absolute RMSE of the tree-level stem volume across the compared algorithms are 0.17 m^3, 0.33 m^3, and 0.24 m^3 in easy, medium and difficult plots, respectively, with the single-scan data, and 0.12 m^3, 0.21 m^3 and 0.18 m^3, respectively, with the multi-scan data. The absolute RMSE of the stem-volume estimation in medium plots is higher than that of the difficult plots, seemingly hard to explain. The situation can, however, be clarified when the relative RMSE is referenced, e.g., the stem volume RMSE% values in easy, medium and difficult plots are 35.1%, 60.4% and 81.0%, respectively, with the single-scan data, and 28.3%, 47.3% and 77.1%, respectively, with the multi-scan data. Obviously, the stem volume estimation becomes more difficult when the stand conditions become more complicated. The reason for a smaller absolute RMSE of the tree-level stem volume in the difficult plots is because of the sizes/ages of the trees in the stand. With a much smaller overall tree size in the difficult forest stands, the absolute RMSE of the estimate of the tree-level stem volume is clearly smaller, but the accuracy of the stem volume estimation is still worse than that in the medium forest stands, given that the RMSE% in the difficult plots is higher.

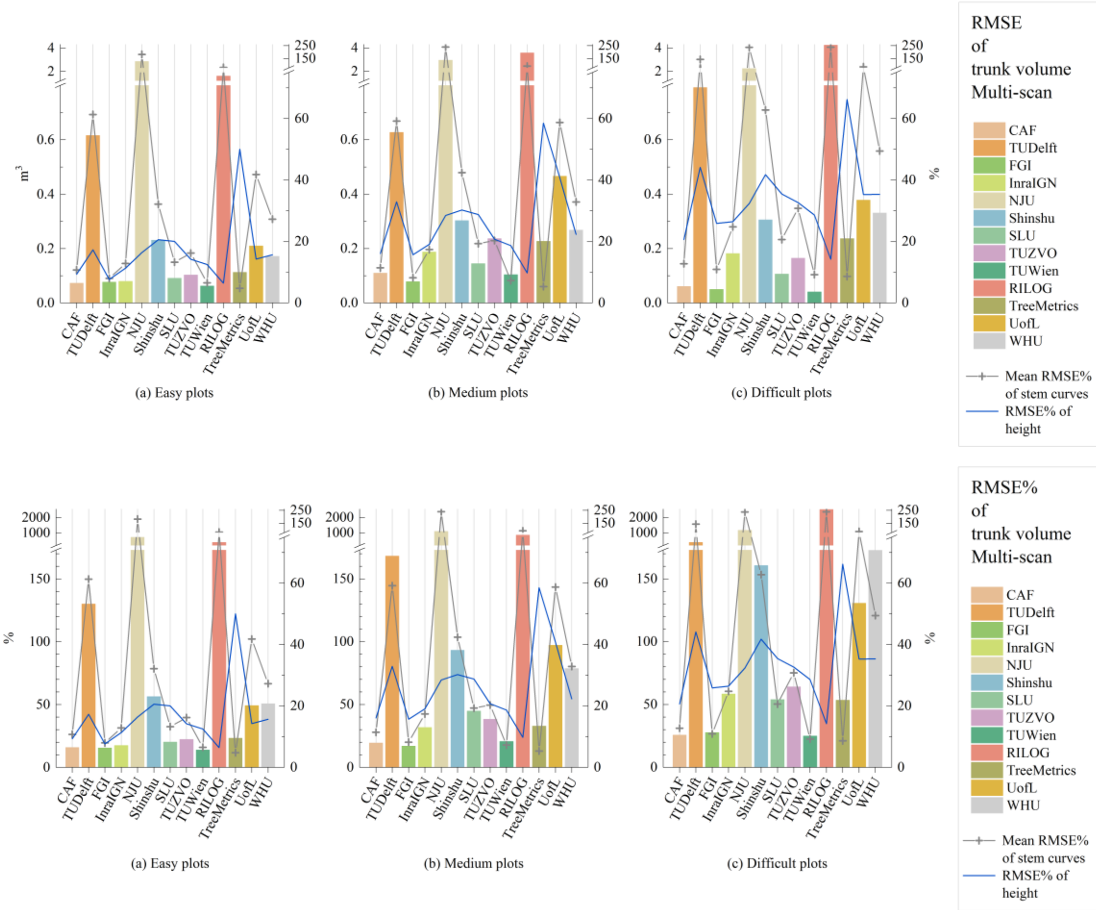


Figure 17: RMSE (upper) and RMSE% (lower) of the stem volume estimation from the multi-scan TLS data. The left vertical axis corresponds to the RMSE value (bars), and the unit is m^3. The right vertical axis corresponds to the two relative indicators, i.e., the mean tree-level RMSE% of the stem curves (crossed line), and the RMSE% of the tree height (blue line), and the unit is percentage.

Another interesting finding is the strong correlation between the RMSEs of the stem curve and stem volume estimations from the compared methods, which can be clearly observed in Figures 16 and 17. Although the stem volume is a dependent variable of the stem curve and tree height, no similar coherence is observed between the stem volume and the tree height. In other words, an algorithm that provides better stem curve estimations always gives better results for stem volume estimations, but if an algorithm gives better tree height estimations, it cannot guarantee an accurate estimation of stem volume. Such phenomena are possibly related to the method of the stem-volume estimation, where the stem volume beyond the highest diameter measured is estimated by a cone-shape-model. More importantly, considering the relatively accurate volume estimates from the method TreeMetrics that missed the volume from the treetop totally and the low coherence between the stem volume and the tree height estimate accuracies, it turned out that the stem curve plays a more determining role than the tree height.

3.7.2 Plot-level stem volume estimation

To further investigate the performance of TLS-based plot-level estimations of stem volume, a new factor called trunk-volume-ratio, i.e., the ratio between the total volume of all extracted trees and all reference trees in a plot, is introduced as an evaluating indicator. A value close to 100% is expected for the automated algorithms. As shown in Figure 18, the ratio can be below and above 100% for different algorithms. Underestimation of the stem volume at the plot level can be explained by at least four factors: 1) the omission errors of the stem detection, 2) the limited total length of the extracted stem curves, 3) the underestimation of the stem diameters at different height of the stem curves (negative bias of stem curve estimation), and 4) the underestimation of tree heights.

Overestimation, i.e., a larger than 100% ratio, can be attributed to 1) the commission errors of the stem detection, 2) the exaggerated estimation of the stem curves (positive bias of the stem curve estimation), and 3) the exaggerated tree height estimations. It is easier to derive larger trunk-volume-ratio from the multi-scan data due to the fact that the stem curve estimation tends to get positive bias, i.e., the stem diameter tends to be overestimated from the multi-scan data, as explained in section 3.5.2.

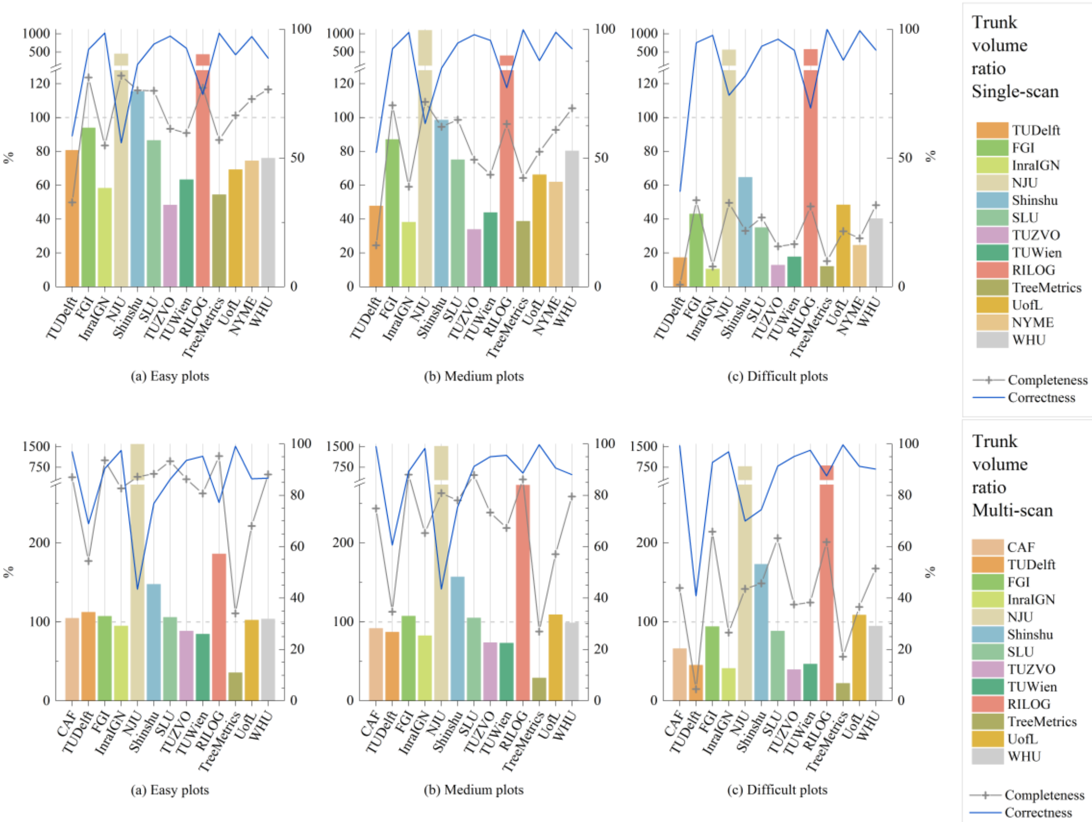


Figure 18: Plot-level trunk volume ratio, single-scan (upper) and multi-scan (lower). The left vertical axis corresponds to the plot-level trunk ratio, and the unit is percentage. The right vertical axis corresponds to the two relative indicators, i.e., the completeness (crossed line) and the correctness (blue line) of stem detection, and the unit is percentage

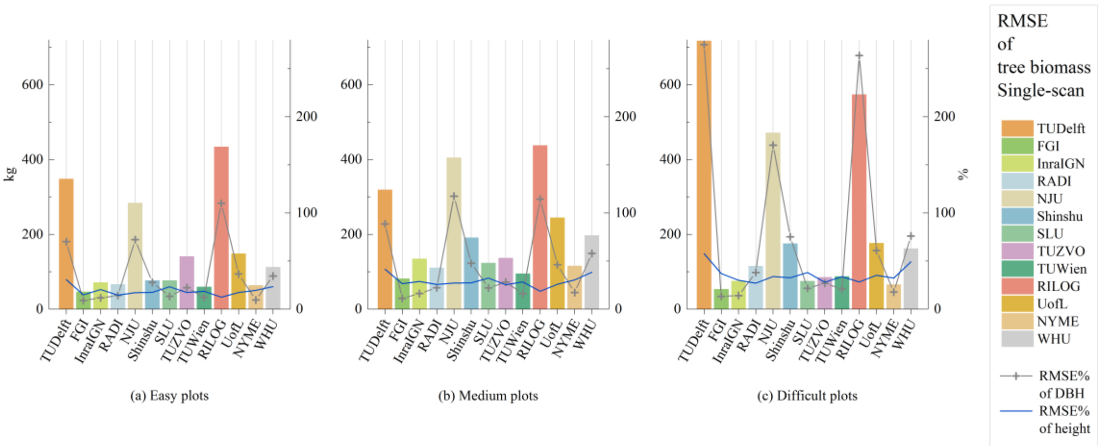
The most important information discovered for the stem volume estimation is that with the best performances, the automated algorithms are capable of carrying out plot-level stem volume estimation at a similar accuracy level as the reference data from multi-scan data (94%, 87%, and 43% trunk-volume-ratio for easy, medium, and difficult plots, respectively, with the single-scan data and 107%, 107%, and 94% for easy, medium, and difficult plots, respectively, with the multi-scan data). Despite the high level of omission errors in the medium and difficult forest stands, the estimated total stem volumes in the plots are close to the reference value (i.e., 100%) using the multi-scan TLS data, indicating that the omitted trees by the stem detection are mainly small trees, and the total volume of those small tree plays a minor role in the plot-level stem volume.

3.8 Biomass estimation

The biomass was predicted using an allometric model as a function of DBH and tree height, and evaluated on both the tree and plot levels. The influence of the DBH and the tree height on the biomass calculation is investigated.

3.8.1 Tree-level tree biomass estimation

Approximately half of the algorithms (CAF, FGI, InraIGN, RADI, Shinshu, SLU, TUWien, and NYME) perform quite similarly in the biomass estimation as shown in Figures 19 and 20; thus, the average of the RMSE value of these algorithms provided a general RMSE level, e.g., 64.9 kg (23.9%), 109.3 kg(43.2%), and 78.8 kg(53.2%) for easy, medium and difficult plots, respectively, with the single-scan data, and 46.4 kg(15.9%), 64.4 kg(27.2%), 49.9 kg(39.3%), respectively, with the multi-scan data.



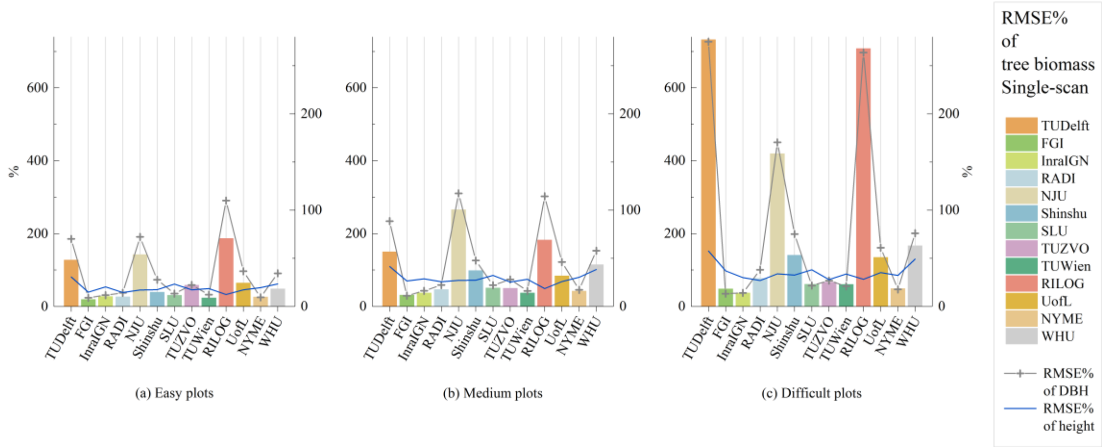
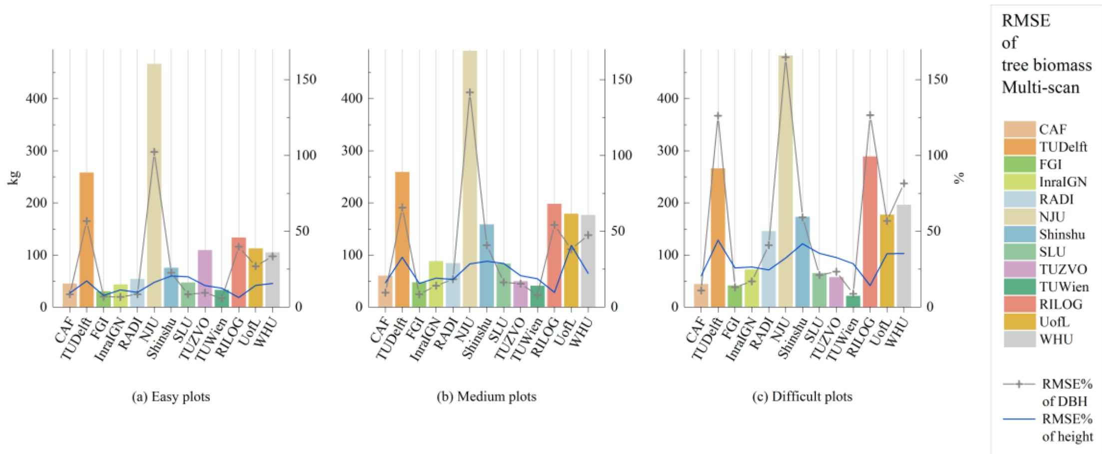


Figure 19: RMSE (upper) and RMSE% (lower) of the biomass estimation from the single-scan TLS data. The left vertical axis corresponds to the RMSE value (bars), and the unit is kg. The right vertical axis corresponds to two relative indicators, i.e., the mean tree-level RMSE% of the DBH (crossed line) and the RMSE% of the tree height (blue line), and the unit is percentage.

A stronger correlation is observed between the biomass accuracy and the accuracy of the DBH than with the tree height. However, the DBH is not the determining factor in the biomass estimation. Therefore, to support a reliable estimation of biomass, a robust algorithm should be able to provide plausible results for both the DBH and tree height. On the other hand, the benefit of multi-scan approach is more obvious in complex forest stands where the RMSE of tree-level biomass is reduced by approximately 15% in medium and difficult plots.



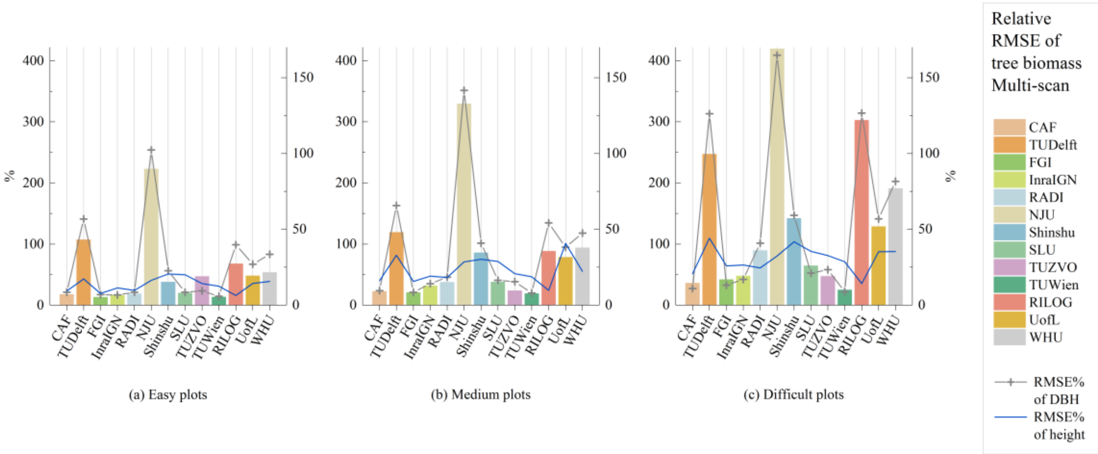


Figure 20: RMSE (upper) and RMSE% (lower) of the biomass estimation from the multi-scan TLS data. The left vertical axis corresponds to the RMSE value (bars), and the unit is kg. The right vertical axis corresponds to two relative indicators, i.e., the mean tree-level RMSE% of the DBH (crossed line) and the RMSE% of the tree height (blue line), and the unit is percentage.

3.8.2 Plot-level biomass estimation

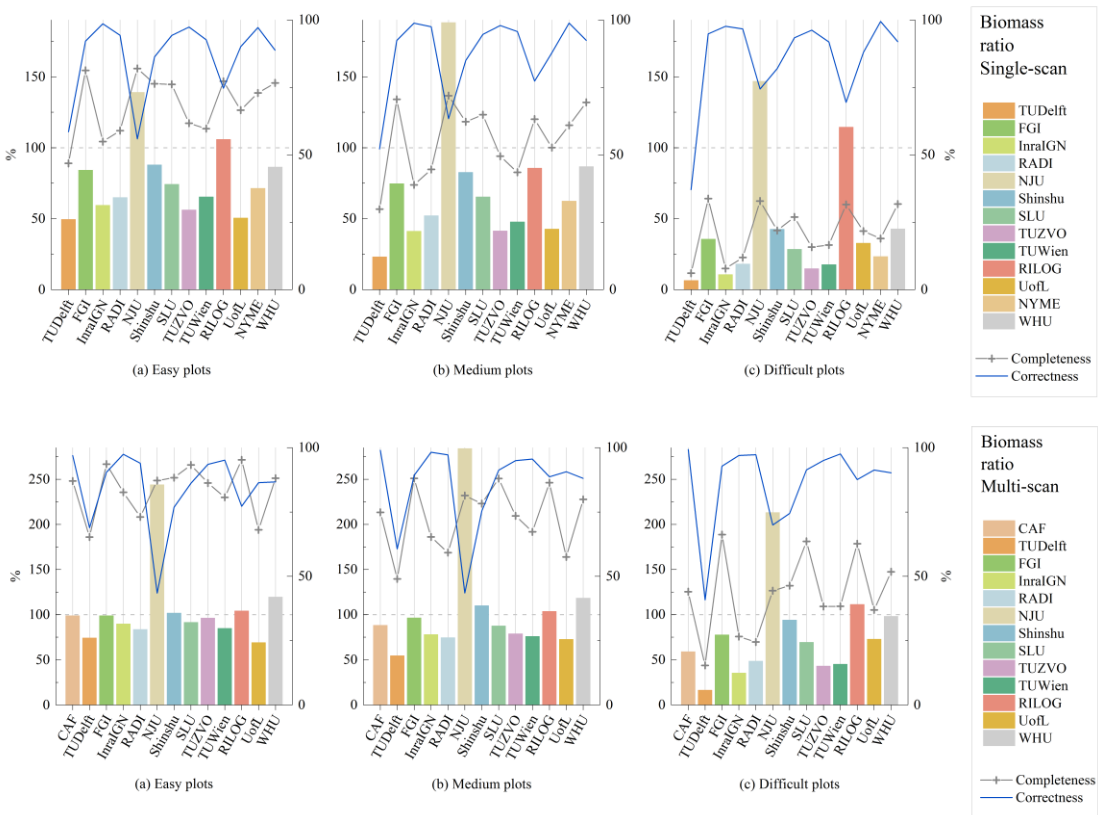


Figure 21: Plot-level biomass ratio, single-scan (upper) and multi-scan (lower). The left vertical axis corresponds to the ratio, and the unit is percentage. The right vertical axis

corresponds to two relative indicators, i.e., i.e., the completeness (crossed line) and the correctness (blue line) of stem detection, and the unit is percentage

The accuracy of the biomass estimation at the plot level is evaluated through the biomass-ratio, which is the ratio between the sums of the tree-level above ground biomass of all extracted and all the reference trees in a plot. The ratio reflects the combined influence of four factors, i.e., the completeness and correctness of stem detection, as well as the DBH and tree height accuracy. In Figure 21, the biomass ratios are illustrated in combination with the stem detection completeness and the correctness. The influence of the stem detection accuracy on the plot-level biomass estimation is clearer when comparing Figures 20 and 21. The closer the biomass ratio to 100%, the better the performance of an algorithm is at estimating the plot-level biomass estimation is. Due to the limitation of the overall stem detection accuracy, an algorithm that is capable of providing good results for tree-level biomass does not necessarily provide satisfying results for plot-level biomass.

When an algorithm is capable of providing accurate estimates on the DBH and the tree height while maintaining a high completeness and correctness of stem detection, the biomass-ratio can reach to 86.1%, 81.2%, and 40.2% for easy, medium and difficult plots, respectively, with the single-scan data, and 98.9%, 95.8% and 80.0%, respectively, with the multi-scan data. These results indicate the value of applying the multi-scan approach for plot-level biomass estimations, which significantly improves the biomass ratio, especially in complex forest stands.

4 DISCUSSION

Supported by the international community, the benchmarking project was capable of covering eighteen different methods that were originally developed for different forest conditions on three continents. Considering the amount and the diversity of the evaluated method, the results and findings of this benchmarking project mark the milestones of TLS performance in forest investigations. The status quo of the methodology development and the accuracy of attribute extraction as well as the influences of data quality and forest conditions can be drawn from the analyses.

4.1 The State-of-the-art of algorithm development

The processes of forest measurements from the TLS point cloud data have reached a high level of automation. The majority of the methods in this benchmarking are fully automated, i.e., approximately 80% of the total methods. In addition, twelve of the fifteen fully automated methods use the same parameter setups for data with different forest conditions and different scanning setups, indicating a plausible flexibility and adaptability of these methods. From the perspective of the methodology development, the methods in the benchmarking demonstrated a wide range of variations. This section summarizes the major findings about the algorithms.

4.1.1 General method design

A general challenge of the automated methods for TLS-based forest investigations is the quality of the point cloud. On the one hand, TLS provides a dense point cloud that can be a heavy computational burden for processing. On the other hand, valid information on trees in the TLS data is almost always insufficient and noisy. In the single-scan data, all the trees are incompletely recorded due to the single view point and occlusion effects. The multi-scan approach can compensate for stem completeness to a certain degree; however, it also increases the noise level due to the registration errors and the mixed scan-to-object distances in the merged data.

To reduce the computational load and to improve the quality of the input data, two typical pre-processing operations are point thinning and noise filtering. For the algorithms that use a rasterized data structure, i.e., 2D raster layers and voxels, the application of the data structure is a sampling approach that reduces the data volume, which means that a point thinning approach is implicitly embedded in the procedure. For the algorithms that directly process the 3D points, an extra point thinning step is explicitly attached. Moreover, more than one-third of the methods in this benchmarking carried out a point filtering process to denoise the input data, expecting to improve the accuracy of the stem modelling in the following steps.

The art of the design method lies in the efforts to produce accurate tree/stem models from limited information recorded in the TLS data. With no exception, all compared methods consist of three essential steps, i.e., the detection of the stem points, the modelling of the detected stems and the validation of the preliminary results. For the majority of the compared methods, there is a clear separation between the first two steps, i.e., to extract the stem points first and to model the stem based on these extracted points. In another group of the methods, e.g., CAF and TUZVO, stem detection is achieved by the feature fitting/modelling, so the first two steps are accomplished simultaneously. A validation step is implicitly or explicitly involved in all the methods to select the “correct” stem models from the preliminary modelling results.

A large variety within the algorithms is observed for the first step, i.e., stem detection. It is where manual processing is applied in the two semi-automated methods, i.e., KU and RILOG, which indicates the difficulty of algorithm development. In contrast to stem detection, algorithms for stem modelling are similar. A tree stem was modelled either with a series of 2D circles or with a series of cylinders. This means that tree/stem detection plays a determining role for the whole processing chain and most of the efforts were put into this specific step.

It summary, most efforts to develop the algorithms serve a clear task, namely, to effectively and accurately extract the stem points from the point cloud. The quality of the remaining efforts, e.g., stems modelling and parameter extraction, is largely determined by the quality of the extracted stem points. This task has been and will remain the most fundamental step for the algorithm development.

4.1.2 Data structure

The most commonly applied data structures for stem detection and modelling are the 2D raster layer, the voxel and the 3D point. Among the eighteen methods in this benchmarking, ten are based on a 2D raster layer, six use the voxels, and two are point based.

The 2D raster layer is widely accepted because of its simplicity and convenience. The algorithm development is relatively easy due to the richness and the capacity of the existing processing tools. When the 2D-raster-layer is applied, satisfactory results can be derived for mature and sparse forest stands with less effort. The main drawback of this data structure is that the accuracy of the results is limited by the signal-to-noise ratio in a 2D slice; tree detection becomes challenging in a slice that has a low information-noise ratio since noise, e.g., from the tree crown, may have patterns similar to those of the targets. It is possible to reduce the amount of false detection by adopting a series of 2D layers. Another major drawback is that the accuracy is restricted by the resolution of the rasterization. Details might be lost due to the space partition during rasterization, which further hinders the detection of small stems and the estimation of stem parameters. The methods using 2D layer(s) in this benchmarking seem to have lower completeness of tree mapping than other algorithms.

The voxel is another popular data structure beside the raster layers, which digitizes the 3D space into cubes of the same size. The main advantages of voxel are the data volume reduction, and the intuitive link between the 2D images and the 3D space, where voxel elements with the same height equal to a 2D raster layer, which raises the flexibility of the voxel structure when dealing with complex forest stands,; therefore, better results can be expected from the voxel structure than the 2D raster layers in the young and dense forests. Similar to the 2D raster layers, the major drawback of the voxel is its sensitivity toward the resolution of rasterization.

Compared to the two different rasterized data structures, the point-based structure is much less applied. The heavy computational load of the dense TLS point cloud is the greatest challenge; therefore, point thinning is usually required in pre-processing. The advantage of the point-based data structure is the completeness of the available information, which improves the overall performance of tree detection and modelling. Another challenge comes from limited available data processing tools, which explains the small number of point-based algorithms in this benchmarking.

All the data structures have their own benefits and weaknesses. There are no strict rules for the selection of a data structure. Developers may choose a data structure according to their own preferences for the ease algorithm development, the capacity of processing large amounts of data, or the quality of the final outcomes.

4.1.3 Implementation principles

The point cloud data in forest conditions are characterized by incomplete and fragmented trees due to the limited view point(s) and occlusion effects. Consequently, conflicts among priorities occur in the processing, where the most commonly seen conflict lies between the completeness and the correctness of stem detection, and trade-offs are observed throughout the processing chain.

Prioritizing between the completeness and the correctness reveals the fundamental implementation principle of an algorithm. Three principle categories, named Aggressive, Conservative and Robust, were defined to describe the benchmarked algorithms. A method that follows an Aggressive principle gives a high priority to the detection rate, namely, it takes the risk of accumulating high commission errors and tries to delineate as many targets, e.g., trees, as possible. In contrast, a Conservative method allocates the highest priority to the correctness; it tries to focus on trees that are highly visible and record completely in the point cloud. When a method applies the Robust principle, it balances the options of detecting more targets and detecting the correct ones. The forest field inventory typically requires an ideal scenario to achieve high completeness along with high correctness, i.e., to follow a Robust principle; however, the task is challenging, and it usually involves higher methodological complexity and computational costs.

Naturally, a high detection rate, i.e., a high completeness of stem detection, is expected when an algorithm is developed. However, the performance evaluation cannot simply rely on any single factor. A higher completeness usually implies a higher tolerance to the fragmented and noisy targets in the data, which potentially leads to higher commission errors and possibly further reduces parameter-estimation accuracy when the accepted stem points are fragmented and noisy. The balance point between completeness and correctness should be determined by the final objective of the application. When a reliable parameter estimation is expected, the Conservative principle can also be a good option.

4.1.4 Best practices

Despite being significantly diverse, the benchmarked algorithms showed some common features, which suggest a road map for best practices.

Tree-attribute estimation seems to follow a series of general steps. Data-volume reduction is practically used by all the methods, but it works in totally different ways, through either direct sampling or space partitioning. Steps such as noise filtering, individual tree detection, tree modelling and validation are commonly adopted. Among them, individual tree detection holds the most significant position in that it directly decides the quality of attribute estimations. The 2D detection method is adequate for locating standing alone trees. Regarding trees in close proximity, the 3D detection method works more efficiently. Tree modelling shows few variations. Two methods model the tree as a series of either cylinders or circles at different heights. The cylinder model seems to be more competitive than the circle model as shown by the superior stem curve accuracy. This is most likely because the cylinder model considers both vertical and horizontal information simultaneously. The validation step also is highly similar among the methods where the diameter and positions of the modelled segments are checked.

4.2 The milestones of tree attribute extraction

The results of the plot/tree attribute estimation were introduced in the section 3. An overview reveals the milestones.

In general, the completeness levels of tree detection are at 70%, 60% and 30% for easy, medium and difficult plots, respectively, using the single-scan data. With the multi-scan data, the completeness level clearly improves and is at 90%, 80% and 50% for easy, medium and difficult plots, respectively. Meanwhile, 90% correctness, i.e., very low commission errors, can be expected from most of the methods, regardless of plot complexity and the scanning approaches. These results indicate that the TLS-based approaches are capable of mapping the trees accurately and that the TLS-based forest inventories can serve as a reliable information source, especially in less complex forest stands. The completeness levels also suggest that the TLS-based approaches are hindered by the visibility of stems. The multi-scan approach can improve stem visibility, but its effectiveness is highly related to the stand complexity.

The DBH estimation accuracy is at a 1-3 cm RMSE level for the best results. Forest stand complexity slightly influences the DBH estimation, but no significant difference in the DBH accuracy is observed between the forest stand complexity categories. The multi-scan TLS data improves the DBH accuracy, resulting from the improvement in the data coverage on the tree stems. The bias of the DBH estimation is close to zero for the best results, which satisfies the requirement of forest inventories. The small impact of stand complexity reveals that the determining factor for DBH accuracy is the quality of the stem points. Once a tree is correctly detected, which implies that the stem is recorded with a satisfactory TLS data (related to point coverage, distribution, number), the estimation of the DBH is reliable.

For the tree height estimation, the results are at the 3-5 m RMSE level, and there is no clear difference between the methods. The results of this benchmarking also confirm that being limited by the view-point, TLS has limited capacity for measuring tree heights in forested conditions, except in simple cases. The accuracy of tree-wise stem curve estimations is at the 1-3 cm RMSE level for the best results, and the level is the same for both single-scan and multi-scan datasets. The bias of the tree-wise stem curve estimation is stable in different stand complexity categories and data acquisition approaches, which are close to zero for the best results.

At a plot level, the estimations of stem volume from the multi-scan TLS data can be quite close to the reference data, e.g., close to the 100% trunk volume ratio in all stand conditions. For the biomass, the main challenge comes from the difficult plots, where the best ratio between the estimated and the reference biomass is 78% in a multi-scan scenario, but the ratio can be above 95% for easy and medium plots.

4.3 Scanning approach and forest stand condition

In addition to the overall performances of different stem mapping methods, this benchmarking also investigates the impacts of the different scanning approaches and the different stand conditions in forests.

4.3.1 Single-scan vs. multi-scan approaches

The single-scan approach has the simplest data acquisition setting and the fastest speed. The major problem is that single scans are limited in terms of the view angle; therefore, the point cloud is highly influenced by occlusion effects. In most of the cases, not all trees in a plot are recorded by a single-scan, and only the side facing the scanner is recorded. The multi-scan approach has the potential to map all the trees and to provide full coverage of the stem surface. However, the multi-scan approach requires more time to acquire the field data and more efforts in processing the data, e.g., the registration of multiple scans.

The result of this benchmarking shows that the improvements of the multi-scan are remarkable for the overall detection rate, the stem curve estimation height percentage (PHC), and the tree height estimation. When utilizing the multi-scan data, averaged for all stand complexity categories, the completeness is improved by approximately 20%; the PHC, by approximately 10%; and the tree height, by approximately 1 m. It is also clearly demonstrated that the more complex the stand condition in the plot is, the more important it is to apply the multi-scan approach.

However, the benefits associated with the multi-scan approach are mainly related to individual tree mapping, rather than tree modelling. This can be seen by the fact that no significant improvement in the parameter estimations, e.g., the DBH, the stem curve and the volume, can be observed when applying the multi-scan data. These results indicate that once a tree is recorded at a satisfactory level, e.g., the tree is visible and can be correctly detected, the information captured in the single-scan data is also sufficient for stem modelling. Considering the marginal effects between costs and benefits, single-scan is quite competitive when the objective is not to map all the trees but to achieve accurate parameter estimation for the visible trees.

On the other hand, the results in this benchmarking also reflect a clear improvement in the algorithm development during the past two decades, given that the recent algorithms are capable to build the stem models from the single-scan data with a quality similar to those from the multi-scan data.

4.3.2 Forest stand condition

The test plots in this project were selected by foresters. The complexity categories cover a wide range of forest conditions, considering the species composition and the development stage. When interpreted in the context of stem detection and modelling, the most influential factor is the visibility, which is closely related to the forest stand factors, including the stem density, the mean DBH and species. It is intuitive that with an increasing stem density and a decreasing mean DBH in the forest stands, the occlusion effects are strengthened and the stem detection and modelling become more difficult.

The stand condition significantly influences stem detection. The results in the benchmarking showed that the higher the complexity of the stand, the lower the completeness of stem detection. On the other hand, in terms of the stem modelling and parameter estimation, the impact of the stand condition mainly lies in the parameters that are relevant to tree height. The accuracy of the DBH estimations is stable among different stand complexity categories, but the tree height accuracy decreases when the stand complexity increases. Consequently, the accuracy of other height relevant parameters, such as the PHC of the stem curve, the volume and the biomass, changes according to the stand conditions.

These findings indicate that the occlusion effects in the forest stands mainly increase the difficulty of stem detection, and reduce the effective tree heights recorded in the TLS point cloud. The application of multiple scans can improve the stem detection rate, but the improvement on the in parameter estimations is limited. In summary, the complex forest stands that are difficult for foresters remain a challenge for new technologies, such as TLS.

5 OUTLOOK

The initial motivation for applying TLS in forest inventories was to automatically derive tree attributes (e.g., tree positions, DBH, stem curve, tree height, and stem volume) to replace manual field tree measurements. The results of the benchmarking indicated that this has been mostly achieved in easy forest plots; from the multi-scan TLS data, tree mapping accuracy at the plot level is close to 100% and the tree-attribute estimates, i.e., tree position, DBH, stem curve, volume and biomass, from the best solutions are at or very close to the acceptable level. In this sense, the multi-scan TLS is technically applicable in practice under easy and homogeneous forest conditions. The best solutions can also provide accurate tree attribute estimates in complex forest conditions or using single-scan TLS, for the trees successfully recorded in point clouds. The accuracy of tree detection remains a main challenge, and the lower the stem visibility is, the lower the point-cloud data quality is; consequently, the detection rate of the algorithms is lower. For tree height, the recent algorithms are still not capable of providing the expected accuracy, i.e., 0-0.5 m accuracy, mainly because of the limited visibility of treetops from a single or several terrestrial viewpoints.

In general, the turning point of accepting any new technique in practice is that the added value from the new technique surpasses previously available techniques. For forest field inventories, the influential factors include the estimation accuracy, cost (hardware and software) and usability (the hardware weight, software readiness, training, etc.). Considering the accuracy reported in this benchmarking, cost and usability are the main factors that limit the added value of applying TLS in forest investigations.

Meanwhile, these recent algorithms still need improvements. Many variances were observed among the algorithms in this benchmarking. Many algorithms perform similarly in easy and homogeneous forest conditions, but the results in complex forest conditions, at both the forest and tree levels, are what differentiate the performance of algorithms, which highlights the necessity of evaluating an algorithm on a wide spectrum of forest conditions to properly interpret the performance of a particular algorithm.

Considering the variances and similarities observed in the algorithms, as well as the results shown in the benchmarking, a new method is likely to achieve an improved tree attribute estimation by taking several components into the new method design: a filtering step to reduce noise, a 3D-feature-based individual tree detection, modelling tree stem as a series of 3D primitives (e.g., overlapped cylinders), and verifying the tree model by the parameters of the modelling elements. Having or not having the step to reduce the data volume depends on computational power. This step seems to be necessary, but it may become a selective option when the computing power is sufficient to handle the large amount of point cloud data. It is also possible to combine some steps into one procedure. In addition, the implementation principles (4.1.3) are fulfilled in the processing steps. In general, a Robust principle should be given a higher priority.

This TLS benchmarking focuses on the tree attributes and data acquisition methods that have been widely reported in previous literature. Several important tree attributes, e.g., tree species; acquisition approaches, e.g., multi-single-scan; processing approaches, e.g., automated multi-scan registration; and applications, e.g., multi-temporal analyses, are not included in this project because few studies have been reported on these topics. Future studies should focus more on these topics.

As shown in the benchmarking, forest conditions have significant influences on the quality of TLS data. Reconstructing individual tree models at different level of details (LoD) (e.g., 1 and 2) can already be very difficult. For example, even in easy forests and with multi-scan data, no algorithms in the benchmarking gave unbiased tree height estimates, and no algorithms gave a stem curve that covers more than 70% of the total tree height while keeping above 60% stem detection completeness. As the stand complexity increases, the quality of the point cloud data sharply decreases and the multi-scan approach is barely capable of recording all the trees in a plot with a generally accepted number of scans, indicating that the reconstruction of individual tree models at higher LoDs (e.g., LoD 3 and 4) is either extremely difficult or costly, especially in complex forest conditions.

In addition, the role that TLS plays in forest field inventories is worth reconsidering. As mentioned previously, the initial motivation of applying TLS in forest inventories was to replace manual field tree measurements. From that perspective, it is challenging to use TLS in medium and difficult plots where a significant number of trees are not recorded in the point clouds. However, the benchmarking revealed that the accuracy of tree attribute estimates of detected trees, whether from single-scan or multi-scan data, are relatively good. This means what may hinder the practical use of TLS in complex forest conditions is the number of recorded trees instead of the inaccurate attribute estimates. Therefore, the emerging question is challenging to the initial motivation for the use of TLS, namely, whether it is necessary to record all the trees in small areas, e.g., sample plots, to achieve accurate quantitative evaluations of large forests. It is time to rethink modern forest investigations beyond conventional forest inventories that rely on circular or rectangular sample plots.

It is worth noting that the source of terrestrial point cloud data is constantly increasing. TLS had been the only practical tool to collect terrestrial point cloud data ten to fifteen years ago. In the past five years, more techniques, such as the structure-from-motion, structure-light-type and mobile laser scanning, have become available to produce similar point clouds. However, the quality of TLS data remains as the best, or at least among the best, of all terrestrial point clouds due to the fact that stationary laser ranging is typically very accurate and that the registration errors in TLS are minimized by utilizing the artificial registration targets. Therefore, the results of this benchmarking label a standard for all the terrestrial point clouds.

6 CONCLUSIONS

To reveal the state-of-the-art of TLS-based forest investigations, an international benchmarking project was launched in 2014 by EuroSDR and assembled eighteen partners worldwide to participate. Algorithms for TLS-based forest mapping and modelling from the partners, including two commercial software products, were required to process identical TLS datasets and to deliver a common set of results, including the DTM, the tree map, the height, and the DBH as well as the stem curve of each individual tree at the plot level. The outcomes from the partners were evaluated with a standard evaluation procedure; thus, the performances of different algorithms were projected to a unique evaluation system so that a comprehensive understanding on the status quo of the TLS-based forest investigation algorithms could be achieved.

The benchmarking project was designed to inspect TLS performance from the perspectives of the data acquisition, the algorithm development, and the forest stand conditions. The TLS data were collected from 24 sample plots utilizing both single-scan and multi-scan approaches. The sample plots were selected by foresters and classified into three complexity categories, i.e., easy, medium, and difficult, considering the complexity of forest conditions. When comparing the results from the different TLS data acquisition approaches, as well as the results of different stand categories, the influences of the TLS data quality and the forest stand complexity on the algorithms were clarified; hence, the impact of different algorithm development concepts and principles for the stem detection and modelling were more precisely and effectively studied.

Considering the amount and the diversity of the evaluated methods in this benchmarking, the performance evaluations provide milestones for TLS-based forest investigations. With the single-scan data, most of the recent algorithms are capable of achieving approximately 75% completeness with 90% correctness for stem detection in the easy forest stands, where the stem density is approximately 600 stems/ha and the mean DBH is at 20 cm level. For the most competitive methods, the completeness can reach to over 80% with 90% correctness. The detection rate decreased when the stand conditions become more complex. For the medium forest stands, i.e. approximately 1000 stem/ha and 15 cm mean DBH, the completeness is at a 60% level with 90% correctness. In a difficult dense, young and multilayer stand, i.e., approximately 2000 stems/ha and 10 cm mean DBH, the detection rate decreases to 30% completeness with 90% correctness. The improvement with the multi-scan approach is substantial, which increases the detection rate by approximately 20% in all forest stand types, i.e., the completeness increases to 90%, 80%, and 50% levels in easy, medium and difficult stands, respectively, with a correctness that is close to 100%. Despite the high level of omission errors in the medium and difficult forest stands, the estimated total stem volumes in the plots are close to 100% of the reference value using the multi-scan approach, which indicates that the omitted trees by the stem detection are mainly small trees, and the total volume of those small tree takes a small proportion of the stem volume at the plot level.

The influences of the scanning approach are insignificant in terms of the accuracy of parameter estimations, except for tree height. Similarly, the impact of the stand condition is less substantial for parameter estimations. Once a stem is successfully detected, the estimation of its DBH and stem curve remained relatively robust for each algorithm, regardless of the scanning approach. The accuracy of parameter estimation is determined by the completeness and the clearness of the stem points; therefore, a trade-off exists between the stem detection and the stem modelling. A higher tolerance for a fragmented and noisy stem structure is required when a higher stem detection rate is pursued. Consequently, stem modelling becomes more difficult when dealing with such fragmental and noisy stem points. With the precondition of a plausible stem detection rate, i.e., the completeness is at the abovementioned average level, the RMSE% of the DBH estimation can be kept at the 10% level with both the single-scan and multi-scan data in easy plots, at the 15% level in medium plots, and at the 20% level in difficult plots. For the stem curve, the most promising results provide approximately 10% level of the RMSE% with both the single-scan and multi-scan data and in all three stand conditions.

Hindered by the limited variation in viewing directions, it is challenging for TLS to capture the treetops; therefore, tree height measurements from TLS are commonly underestimated, and the situation becomes worse when the stand condition becomes more complex. With single-scan data, the RMSE% of the tree height estimation is at approximately 15% level in easy plots, at the 25% level in medium plots and at the 40% level in difficult plots. The multi-scan approach improves the RMSE% of tree height to approximately 10% level in easy plots, the 15% level in medium plots and the 25% level in difficult plots.

The tree-level volume and biomass are also estimated based on the height, the DBH and the stem curve delivered from the partners. The results in this benchmarking demonstrate that the accuracy of the volume estimation has a strong correlation with the accuracy of the stem curve, while the biomass estimate correlates more strongly to DBH than to the tree height. An algorithm that gives a better estimation of DBH and stem curve always provides better results for the volume and the biomass. On the other hand, no such correlation is observed for the tree height estimations. When the single-scan approach is used, the RMSE% of the tree-level volume estimation can be expected to be at approximately 25% level for easy plots, the 40% level for medium plots, and the 50% level for difficult plots. The tree-level RMSE% of the biomass estimation can be expected to be at approximately 25%, 40%, and 60% levels, for easy, medium and difficult plots, respectively. For the multi-scan approach, the RMSE% of the tree-level volume is at approximately 20%, 30%, and 40% levels for easy, medium and difficult plots, respectively, and the RMSE% of the tree-level biomass is at approximately 15%, 30%, and 45% levels for easy, medium, and difficult plots, respectively.

In terms of algorithm development, a high level of automation is a commonly shared standard. Approximately 80% of the approaches in this benchmarking are fully automated, the rest are semi-automated. The rasterized data formats, i.e., the 2D raster layers and the voxels are the most popular data structures for stem detection and modelling from TLS point clouds. The 3D points as a data structure may provide more details and benefits with higher accuracies, but it is also hindered by high computational cost and heavy programming load.

Confronting the conflicts between the higher detection rate and better parameter estimation, three main principles are observed from the algorithms in this benchmarking, i.e., Aggressive, Conservative and Robust. The Aggressive principle allocates the highest priority to the stem detection rate, sacrificing the correctness of stem detection and the accuracy of parameter estimation. In contrast, the Conservative principle focuses on the correctness and the accuracy of stem models, resulting in lower completeness. The Robust principle pursues high stem detection correctness and accurate parameter estimations while maintaining a high detection rate, with the cost of highly complex algorithms. Each principle/algorithm has its own advantages and weakness, and the selection of a principle/algorithm depends on the final objective of the applications.

Based on the similarities observed in the benchmarked algorithms and the results, a new method is likely to achieve good tree attribute estimation by taking a couple of components into the new method design: a filtering step to reduce noise, a 3D-feature-based individual tree detection, modelling tree stem as a series overlapped cylinders, and verification of the tree model by the parameters of the basic modelling elements. A step to reduce the data volume seems necessary at this moment, but it totally depends on computational power.

The results from this benchmarking showed that the TLS-based approaches have the capability to provide DBH and the stem curve estimations that are close to what is required in practical applications, e.g., NFIs. Stem detection achieves high correctness regardless of the data acquisition approaches and the stand conditions. However, the bottleneck is at the completeness of stem detection and the accuracy of tree height estimation, especially in young and dense forest stands. These findings indicate that more research is needed to optimize TLS application in forest investigations. Improving algorithms to further explore the potential of TLS point clouds is one direction that is worth investigating. Another viable starting point would be to change to the concept of conventional field inventories in which all the trees in a sample plot must be measured.

If the areal forest parameters can be achieved based on a sufficiently large sample of randomly located individual trees that are accurately modelled, the application of TLS or other terrestrial point clouds would be much more meaningful. It is therefore worth noting that, TLS currently provides the best quality terrestrial point cloud compared to all the other technologies, such as the mobile laser scanning, personal laser scanning, structure-light, and the image-based structure from motion, which means that all the milestones labelled in this benchmarking mark achievable targets for all types of terrestrial point clouds. Thus, the results in this benchmarking also provide information on the selection of terrestrial systems for point cloud acquisition.

Author Contributions: Xinlian Liang is the PI of the benchmarking project. Juha Hyyppä, Harri Kaartinen, Matti Lehtomäki, Jiri Pyörälä, Xiaowei Yu are the team in FGI to carry out the project. Norbert Pfeifer is the advisor from EuroSDR. Gábor Brolly, Hopkinson Cristopher, Pirotti Francesco, Jan Hackenberg, Huabing Huang, Hyun-Woo Jo, Masato Katoh, Luxia Liu, Martin Mokroš, Jules Morel, Kenneth Olofsson, Jose Poveda-Lopez, Jan Trochta, Di Wang, Jinbu Wang, Bisheng Yang, Guang Zheng from partner groups processed the TLS data and the names are listed in alphabetical order according to the last name. Yunsheng Wang writes the paper jointly with Xinlian Liang. All coauthors comments the paper.

7 ACKNOWLEDGEMENT

The authors would like to thank financial aid from the Finnish Academy projects “Centre of Excellence in Laser Scanning Research (CoE-LaSR) (272195)” and from European Community’s Seventh Framework Programme ([FP7/2007–2013]) under grant agreement No. 606971.

Jan Hackenberg would like to thank the financial support from the DIABOLO project from the European Union’s Horizon 2020 research and innovation program under grant agreement No 633464.

Group University of Lethbridge would like to thank the contribution of Prof. Cristopher Hopkinson.

8 REFERENCE

Aschoff, T., Spiecker, H., 2004. Algorithms for the automatic detection of trees in laser scanner data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36, 71–74.

Astrup, R., Ducey, M.J., Granhus, A., Ritter, T., von Lüpke, N., 2014. Approaches for estimating stand-level volume using terrestrial laser scanning in a single-scan mode. Canadian Journal of Forest Research 44, 666–676. https://doi.org/10.1139/cjfr-2013-0535

Côté, J.F., Widlowski, J.L., Fournier, R.A., Verstraete, M.M., 2009. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. Remote Sensing of Environment 113, 1067–1081.

Erikson, M., Karin, V., 2003. Finding tree-stems in laser range images of young mixed stands to perform selective cleaning, in: Proceedings of the ScandLaser Scientific Workshop on Airborne Laser Scanning of Forest. pp. 244–250.

Hackenberg, J., Spiecker, H., Calders, K., Disney, M., Raumonen, P., 2015. SimpleTree — An Efficient Open Source Tool to Build Tree Models from TLS Clouds. Forests 6, 4245–4294. https://doi.org/10.3390/f6114245

Henning, J.G., Radtke, P.J., 2006. Detailed stem measurements of standing trees from ground-based scanning lidar. Forest Science 52, 67–80.

Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004. Assessing forest metrics with a ground-based scanning lidar. Canadian Journal of Forest Research 34, 573–583. https://doi.org/10.1139/x03-225

Hosoi, F., Omasa, K., 2006. Voxel-based 3-D modelilng of individual trees for estimating leaf area density using high-resolution portable scanning lidar. IEEE Transactions on Geoscience and Remote Sensing 44, 3610–3618.

Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., Strahler, A.H., Woodcock, C.E., 2009. Estimating forest LAI profiles and structural parameters using a ground-based laser called ‘Echidna®’’.’ Tree Physiol 29, 171–181. https://doi.org/10.1093/treephys/tpn022

Kankare, V., Holopainen, M., Vastaranta, M., Puttonen, E., Yu, X., Hyyppä, J., Vaaja, M., Hyyppä, H., Alho, P., 2013. Individual tree biomass estimation using terrestrial laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing 75, 64–75. https://doi.org/10.1016/j.isprsjprs.2012.10.003

Koreň, M., Mokroš, M., Bucha, T., 2017. Accuracy of tree diameter estimation from terrestrial laser scanning by circle-fitting methods. International Journal of Applied Earth Observation and Geoinformation 63, 122–128. https://doi.org/10.1016/j.jag.2017.07.015

Lefsky, M., McHale, M.R., 2008. Volume estimates of trees with complex architecture from terrestrial laser scanning. Journal of Applied Remote Sensing 2, 023521-023521-19.

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F., Holopainen, M., Vastaranta, M., 2016. Terrestrial laser scanning in forest inventories. ISPRS Journal of Photogrammetry and Remote Sensing 115, 63–77. https://doi.org/10.1016/j.isprsjprs.2016.01.006

Liang, X., Kankare, V., Yu, X., Hyyppa, J., Holopainen, M., 2014. Automated Stem Curve Measurement Using Terrestrial Laser Scanning. IEEE Transactions on Geoscience and Remote Sensing 52, 1739–1748. https://doi.org/10.1109/TGRS.2013.2253783

Liang, X., Litkey, P., Hyyppa, J., Kaartinen, H., Vastaranta, M., Holopainen, M., 2012. Automatic stem mapping using single-scan terrestrial laser scanning. IEEE Transactions on Geoscience and Remote Sensing 50, 661–670.

Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., Coops, N.C., 2003. Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. Canadian Journal of Remote Sensing 29, 607–622. https://doi.org/10.5589/m03-026

Ma, L., Zheng, G., Eitel, J.U.H., Moskal, L.M., He, W., Huang, H., 2016. Improved Salient Feature-Based Approach for Automatically Separating Photosynthetic and Nonphotosynthetic Components Within Terrestrial Lidar Point Cloud Data of Forest Canopies. IEEE Transactions on Geoscience and Remote Sensing 54, 679–696. https://doi.org/10.1109/TGRS.2015.2459716

Maas, H.G., Bienert, A., Scheller, S., Keane, E., 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data. International Journal of Remote Sensing 29, 1579–1593. https://doi.org/10.1080/01431160701736406

Newnham, G.J., Armston, J.D., Calders, K., Disney, M.I., Lovell, J.L., Schaaf, C.B., Strahler, A.H., Danson, F.M., 2015. Terrestrial Laser Scanning for Plot-Scale Forest Measurement. Current Forestry Reports 1, 239–251. https://doi.org/10.1007/s40725-015-0025-5

Olofsson, K., Holmgren, J., 2016. Single Tree Stem Profile Detection Using Terrestrial Laser Scanner Data, Flatness Saliency Features and Curvature Properties. Forests 7, 207. https://doi.org/10.3390/f7090207

Parker, G.G., Harding, D.J., Berger, M.L., 2004. A portable LIDAR system for rapid determination of forest canopy structure: Portable LIDAR for canopy structure. Journal of Applied Ecology 41, 755–767. https://doi.org/10.1111/j.0021-8901.2004.00925.x

Pfeifer, N., Gorte, B., Winterhalder, D., 2004. Automatic reconstruction of single trees from terrestrial laser scanner data. pp. 114–119.

Pirotti, F., Guarnieri, A., Vettore, A., 2013. Vegetation filtering of waveform terrestrial laser scanner data for DTM production. Applied Geomatics 5, 311–322. https://doi.org/10.1007/s12518-013-0119-3

Schütt, C., Aschoff, T., Winterhalder, D., Thies, M., Kretschmer, U., Spiecker, H., 2004. Approaches for Recognition of Wood Quality of Standing Trees b Based on Terrestrial Laser Scanner Data. International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences 36, 179–182.

Simonse, M., Aschoff, T., Spiecker, H., Thies, M., 2003. Automatic determination of forest inventory parameters using terrestrial laserscanning. pp. 252–258.

Strahler, A.H., Jupp, D.L.B., Woodcock, C.E., Schaaf, C.B., Yao, T., Zhao, F., Yang, X., Lovell, J., Culvenor, D., Newnham, G., 2008. Retrieval of forest structural parameters using a ground-based lidar instrument (Echidna®). Canadian Journal of Remote Sensing 34, 426–440.

Tansey, K., Selmes, N., Anstee, A., Tate, N.J., Denniss, A., 2009. Estimating tree and stand variables in a Corsican Pine woodland from terrestrial laser scanner data. International Journal of Remote Sensing 30, 5195–5209.

Thies, M., Pfeifer, N., Winterhalder, D., Gorte, B.G.H., 2004. Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees. Scandinavian Journal of Forest Research 19, 571–581.

Trochta, J., Krůček, M., Vrška, T., Král, K., 2017. 3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR. PLOS ONE 12, e0176871. https://doi.org/10.1371/journal.pone.0176871

Wang, D., Hollaus, M., Puttonen, E., Pfeifer, N., 2016. Automatic and Self-Adaptive Stem Reconstruction in Landslide-Affected Forests. Remote Sensing 8, 974. https://doi.org/10.3390/rs8120974

Watt, P.J., Donoghue, D.N.M., 2005. Measuring forest structure with terrestrial laser scanning. International Journal of Remote Sensing 26, 1437–1446. https://doi.org/10.1080/01431160512331337961

Xi, Z., Hopkinson, C., Chasmer, L., 2016. Automating Plot-Level Stem Analysis from Terrestrial Laser Scanning. Forests 7, 252. https://doi.org/10.3390/f7110252

Yu, X., Liang, X., Hyyppä, J., Kankare, V., Vastaranta, M., Holopainen, M., 2013. Stem biomass estimation based on stem reconstruction from terrestrial laser scanning point clouds. Remote Sensing Letters 4, 344–353. https://doi.org/10.1080/2150704X.2012.734931