International benchmarking of terrestrial laser scanning approaches for forest inventories

Liang, Xinlian; Hyyppä, Juha; Kaartinen, Harri; Lehtomäki, Matti; Pyörälä, Jiri; Pfeifer, Norbert; Holopainen, Markus; Wang, Di; Wang, Jinhu; More Authors

DOI
10.1016/j.isprsjprs.2018.06.021

Publication date
2018

Document Version
Final published version

Published in
ISPRS Journal of Photogrammetry and Remote Sensing

ISPRS Journal of Photogrammetry and Remote Sensing 144 (2018) 137–179

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

International benchmarking of terrestrial laser scanning approaches for forest inventories

Xinlian Liang[a], Juha Hyyppä[a], Harri Kaartinen[a,b], Matti Lehtomäki[a], Jiri Pyörälä[a,d], Norbert Pfeifer[c], Markus Holopainen[a,d], Gábor Brolly[e], Pirotti Francesco[f], Jan Hackenberg[g,h], Huabing Huang[i], Hyun-Woo Jo[j], Masato Katoh[k], Luxia Liu[l], Martin Mokroš[m,n], Jules Morel[o], Kenneth Olofsson[p], Jose Poveda-Lopez[q], Jan Trochta[r], Di Wang[c], Jinhu Wang[s], Zhouxi Xi[t], Bisheng Yang[u], Guang Zheng[v], Ville Kankare[a,d], Ville Luoma[a,d], Xiaowei Yu[a], Liang Chen[a], Mikko Vastaranta[a,d,w], Ninni Saarinen[a,d], Yunsheng Wang[a,*]

[a] Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, 02431 Masala, Finland
[b] Department of Geography and Geology, University of Turku, 20014 Turku, Finland
[c] Department of Geodesy and Geoinformation, Technische Universität Wien, 1040 Vienna, Austria
[d] Department of Forest Sciences, University of Helsinki, 00014 Helsinki, Finland
[e] Institute of Geomatics and Civil Engineering, Faculty of Forestry, University of Sopron (former University of West Hungary), H-9400 Sopron, Hungary
[f] CIRGEO—Interdepartment Research Center of Geomatics, University of Padova, 35020 Legnaro, PD, Italy
[g] Biogéochimie des Ecosystèmes Forestiers, INRA, 54280 Nancy, France
[h] Laboratoire d'Inventaire Forestier, Institut géographique national (IGN), 54250 Nancy, France
[i] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, 100101 Beijing, China
[j] Environmental GIS/RS Lab., Division of Environmental Science & Ecological Engineering, Korea University, 02855 Seoul, South Korea
[k] Forest Measurement and Planning Laboratory, Agriculture Faculty, Shinshu University, 399-4598 Nagano Pref., Japan
[l] Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, 100091 Beijing, China
[m] Department of Forest Management and Geodesy, Technical University in Zvolen, 96053 Zvolen, Slovakia
[n] Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, 16500 Praha, Czech Republic
[o] Institut Français de Pondichéry – Laboratoire des Sciences de l'Information et des Systèmes, India/France
[p] Department of Forest Resource Management, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden
[q] Treemetrics, Cork, T12 CCN3, Ireland
[r] 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
[s] Department of Geoscience and Remote Sensing, Delft University of Technology, 2628CN Delft, The Netherlands
[t] Department of Geography, University of Lethbridge, T1K3M4 Lethbridge, Canada
[u] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, University of Wuhan, 430079 Wuhan, China
[v] International Institute for Earth System Science, Nanjing University, 210023 Nanjing, China
[w] School of Forest Sciences, University of Eastern Finland, 80101 Joensuu, Finland

ARTICLE INFO

Keywords:
Benchmarking
State-of-the-art
Forest
Modeling
Point cloud
Terrestrial laser scanning
TLS

ABSTRACT

The last two decades have witnessed increasing awareness of the potential of terrestrial laser scanning (TLS) in forest applications in both public and commercial sectors, along with tremendous research efforts and progress. It is time to inspect the achievements of and the remaining barriers to TLS-based forest investigations, so further research and application are clearly orientated in operational uses of TLS. In such context, the international TLS benchmarking project was launched in 2014 by the European Spatial Data Research Organization and co-ordinated by the Finnish Geospatial Research Institute. The main objectives of this benchmarking study are to evaluate the potential of applying TLS in characterizing forests, to clarify the strengths and the weaknesses of TLS as a measure of forest digitization, and to reveal the capability of recent algorithms for tree-attribute extraction. The project is designed to benchmark the TLS algorithms by processing identical TLS datasets for a standardized set of forest attribute criteria and by evaluating the results through a common procedure respecting reliable references. Benchmarking results reflect large variances in estimating accuracies, which were unveiled through the 18 compared algorithms and through the evaluation framework, i.e., forest complexity categories, TLS data acquisition approaches, tree attributes and evaluation procedures. The evaluation framework includes

* Corresponding author.
E-mail address: yunsheng.wang@nls.fi (Y. Wang).

https://doi.org/10.1016/j.isprsjprs.2018.06.021
Received 9 December 2017; Received in revised form 28 June 2018; Accepted 28 June 2018

three new criteria proposed in this benchmarking and the algorithm performances are investigated through combining two or more criteria (e.g., the accuracy of the individual tree attributes are inspected in conjunction with plot-level completeness) in order to reveal algorithms' overall performance. The results also reveal some best available forest attribute estimates at this time, which clarify the status quo of TLS-based forest investigations. Some results are well expected, while some are new, e.g., the variances of estimating accuracies between single-/multi-scan, the principle of the algorithm designs and the possibility of a computer outperforming human operation. 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 show that the TLS-based approaches can provide the estimates of the DBH and the stem curve at a 1–2 cm accuracy 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 stems of 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 labeled in this paper can also serve as a reference for other terrestrial point clouds sources.

1. Introduction

Forest field inventory holds a central role in all forest research, monitoring and managements that rely on knowledge of forest structure, distribution and dynamics over time. Field inventories are conducted in sample plots, where tree information is usually collected through tree-by-tree measurements (i.e., plot-level inventory). Forest field inventories can be costly since the field measurements require many efforts and resources, consequently limiting the amount of field inventories that can be afforded. Attempts to improve the field inventory efficiency started ever since field inventory began. Countless techniques, instruments, and protocols have been introduced yet progress has been slow, until a laser-based measuring instrument called terrestrial laser scanner became practically available.

The first commercial terrestrial laser scanner was introduced to the market in 1998. It automatically measures the surrounding three-dimensional (3D) space using millions to billions of 3D points. 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 with maximum measurement distance of 100–300 m; the range precision is at a millimeter level, and the angular sampling capacity is less than 0.01° in both horizontal and vertical directions.

The major advantage of applying terrestrial laser scanning (TLS) in forest inventories lies in the digitization of the forest plots accurately, rapidly, automatically and in detail at millimeter-level. In addition to the regular tree attributes measured in practical field inventories, e.g., the diameter at breast height (1.3 m, DBH) and tree height, more detailed tree attributes, such as the stem curve or taper curve (stem diameter as a function of height) that reveals the wood productivity and quality yet difficult to acquire non-destructively in the field, can be derived from TLS with high degrees of accuracy and cost efficiency (Liang et al., 2014b).

Tremendous efforts have been put into research to investigate the automated interpretation of TLS data and to establish best practices for using TLS. In the past 20 years, significant progress has been made in deriving tree- and stand-level attributes from TLS data to depict forest productivity, evolution and ecological functions. 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. More recently, TLS has been shown to be capable of determining several high-quality tree attributes that are not directly measurable using conventional tools, such as the stem curve (Liang et al., 2014b). Tree-/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., 2014b).

However, the significant variance in the hardware properties, the scanning setups, the forest structures, and 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 detected trees from multi-scan TLS data ranged from 20% to 100% at the plot level as reported in previous research (Maas et al., 2008; Strahler et al., 2008; Brolly and Kiraly, 2009; Murphy et al., 2010; Lovell et al., 2011; Yao et al., 2011; Liang et al., 2012; Lindberg et al., 2012; Astrup et al., 2014; Olofsson et al., 2014). 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 by the European Spatial Data Research Organization (EuroSDR) and partly funded by the European Community's Seventh Framework Programme Project Advanced_SAR. The TLS benchmarking aims to clarify the potential and current status of the TLS application in field inventories by evaluating methodologies on the basis of a standard evaluation procedure and a

common dataset, thereby orienting further research and developments. As the project leader and coordinator, the Finnish Geospatial Research Institute (FGI) conducted the TLS and reference data acquisition, defined a series of plot- and tree-level attributes as standardized criteria, developed a standard and fully automated evaluation procedure, evaluated the performances of the algorithms, and benchmarked the results.

For the time being, this is the first international benchmarking of TLS-based forest inventories. The investigation 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.

The forest sample plots in the benchmarking project are selected by foresters to reflect different stand conditions in boreal forests. Considering the development stage, stem density, and density of the sub canopy vegetation, as well as the species composition in the forest stands, sample plots are classified into three complexity categories, i.e., “easy”, “medium”, and “difficult”, which also reflects the level of complexity in the TLS data processing. Both single- and multi-scan approaches are employed to acquire the sample plot TLS datasets. The TLS dataset was disseminated to all the benchmarking project partners, who processed the data utilizing their own algorithms and delivered the required products. All of the partners results were then evaluated using a standard evaluation procedure, so a comprehensive understanding can be achieved on the capacity of recent algorithms for extracting important forest attributes from TLS data. In particular, the influence of forest conditions and data acquisition methods on the algorithm performance can be investigated and interpreted from a practical perspective.

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 digital terrain model (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. 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 Appendix A 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; Yang et al., 2016; Wang et al., 2016a; Xi et al., 2016; Koreň et al., 2017; Trochta et al., 2017).

This paper summarizes the benchmarking project’s conceptual schema and reports the findings of the project. Section 2 describes the benchmarking project’s main objectives and fundamental concepts to support the main benchmarking objectives. Descriptions about the common TLS datasets of the forest sample plots representing different forest stand situations for the benchmarking are given in Sections 3.1 and 3.2, which explain the variety in the accuracy of tree attribute estimations across forest complexity categories, across algorithms and across the TLS-measurement approaches. The reference information and the evaluation procedures, which lay down the bases of the benchmarking, are detailed in Sections 3.3 and 4. New evaluation criteria for tree attribute estimates are established to analyze the TLS performances. Section 5 overviewed participants’ algorithms, which were shortly summarized in Appendix A. Section 6 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 7 reveals the achievements and remaining challenges of recent studies, providing recommendations and paving the way for further studies and applications in the field. Section 8 looks into the future of appling TLS in forest modelling and recommonds the road map of the algorithm design. Section 9 gives the conclusion.

Furthermore, since static TLS provides the plot-level point cloud with spatial precision and detailed richness that 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. The international TLS benchmarking project

Accurate forest inventories with strong degree of time and cost efficiencies have been long awaited by multiple forest-related applications and users. The TLS technology introduced 20 years ago was anticipated to have the potential to provide a high-quality solution that was highly automated for plot-level forest measurements. It is time to inspect the achievements of and the remaining barriers to the TLS-based forest investigations. This section summarizes the benchmarking project and the project’s conceptual schema.

2.1. The project

The TLS benchmarking project was launched in 2014 by EuroSDR and led by the Finnish Geospatial Research Institute (FGI). The FGI is responsible for the benchmarking processes’ architecture; the development of the evaluative criteria and procedures; the collection of participants; and the implementation, coordination and dissemination of the project. Forest researchers from the University of Helsinki (UH) selected the sample plots and measured tree attribute using calipers and clinometers in the field. The TLS data and field measurements were a joint effort of FGI and UH.

The benchmarking project’s targeted participants include national mapping agencies, companies, universities and research organizations, who develop their own processing methods or modify existing methods. Meanwhile, the project is open for techniques that are in the development phase. The project was actively advocated to potential participants, reached through research networks, during conferences and via various social media platforms.

Eighteen groups from three continents (Asia, Europe and North America) successfully processed the data and submitted their results for evaluation. Partners were encouraged to process both single- and multi-scan data, but had the option to process data according to their preference. 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. Of the 18 partners, 12 provided all requested parameters from both single- and multi-scan data, two provided all results from either single- or multi-scan data, and four provided part of the results. All the results were evaluated using the same reference data (Section 3) and evaluation methods (Section 4).

2.2. Conceptual schema

The benchmarking is carried out from two equally important perspectives: the capacity of the TLS data to digitize the forest plots and the performance of the data processing algorithms for attribute extractions. TLS digitization capacity in recording forest is influenced by the stand condition and scanning pattern, which determine what can best be achieved by a particular feature extraction method. In evaluating an

Table 1
List of the participants in the international TLS benchmarking.

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	UofL
University of Padova	Italy	UNIPD
University of Sopron	Hungary	NYME
Wuhan University	China	WHU

algorithm's performances, two major tasks are to establish a standardized criterion and to develop an evaluation procedure. This section details the benchmarking design.

2.2.1. To evaluate the capacity of the TLS to digitize a forest plot

The strength of TLS in forest field inventory lies in its capacity to record the forest environment automatically, accurately and rapidly. Two essential factors addressing the TLS-data quality are the spatial precision and completeness of tree information in the point-cloud data. The spatial precision is determined by the system calibration and by the registration of multi-scan data if applicable, which are typically sufficiently accurate for forest applications. The tree-information completeness is determined by the forest conditions and the field-inventory design but is however not guaranteed. The forest stands' complexity, the scanning patterns applied in the field and the distance/geometry between a tree and the scanning position(s) are the issues that determine the completeness of trees in the point-cloud data of a forest sample plot.

The accuracy of tree attributes' extraction can only be meaningfully discussed when the completeness of tree information in the data is clarified. Therefore, the impacts of different stand situations and scanning pattern to the quality of the collected TLS data and, as a consequence, to the results of attribute extraction are investigated in this benchmarking. The stand conditions of the sample plots, as well as the applied scanning pattern for the TLS data collection are described in Section 3.

2.2.2. To evaluate the algorithm performances

A foundation in evaluating the performance of attribute extraction is to establish a series of standardized criteria that suits most of the currently existing algorithms. Evaluation criteria are selected based on five main considerations: firstly, a criterion is of high interest and importance in forest inventories; secondly, the criterion's measurement is within the capacity of TLS equipment commonly available; thirdly, the criterion estimates can be evaluated against the corresponding references; fourthly, multiple algorithms for the criterion extraction from TLS data have been reported in previous studies; fifthly, for practical applications, such as forest inventories, the criterion is measurable with reasonable costs in practical inventories at a large scale, e.g., national forest inventories (NFIs).

Among various forest attributes, the most interesting tree-level attributes in conventional field inventory include the tree height, DBH and species, which are widely used in estimating tree volume and biomass. However, plausible results on species classification based on TLS merely exists until recently. Other highly interested tree-level attributes, e.g., stem-quality class, canopy layer and age, lack sufficient evidences to be measureable from TLS; therefore are not included in the benchmarking criteria.

A couple of other tree attributes that are highly important but not conventionally measurable due to the large amount of required resources but lie in the strengths of TLS should be considered to evaluate the added value of applying TLS in forest environments: the stem curve, a long awaited tree-level attribute that has been difficult to measure non-destructively; tree position, a parameter that reveals forest structure and bridges observations from different perspectives, e.g., terrain and airborne observations; and DTM that is essential for measuring tree height and DBH from the TLS point cloud.

Furthermore, two indirect attributes, i.e., the volume and the biomass, that are calculated using directly measured tree attributes should also be investigated. These indirect attributes not only reveal the overall performance of an algorithm since the calculation combines several estimates, but also reflect the error propagation, which help to understand the values and impacts of individual tree estimates. More importantly, the volume and biomass are among the most important tree attributes required by various applications, therefore, these estimates reveal the potential of applying TLS in forests.

Based on these factors, this benchmarking project's criteria consist of one plot-level attribute, i.e., the DTM; four direct tree-level attributes, i.e., tree location, tree height, DBH and stem curve; and two indirect attribute, i.e., the stem volume and total tree biomass. Fig. 1 illustrates five direct attributes that are taken as the criteria of this benchmarking.

In addition to standardized evaluation criteria, credible evaluation also requires robust evaluation procedures. To minimize the human-introduced influences, a series of fully automated procedures are developed to evaluate the attribute extraction, and a same set of parameter settings is applied for all the evaluated results. Thus, all the evaluations are solely based on the comparisons between the reference and the results delivered by the project partners. Details of the evaluation procedure for each criterion are given in Section 4.

In brief, all algorithms evaluated in this benchmarking project processed a unique set of TLS data and provided the attribute extraction results of a standardized set of criteria, which were projected to a common frame of reference and were independently evaluated by a series of automated evaluations.

3. Datasets

The data acquisition approaches are designed to support the main objectives of the benchmarking project. Thus, 24 sample plots were selected from varying forest-stand conditions representing different developing stages, stem densities and abundance of sub canopy growth in boreal forests and classified into three complexity categories. The amount of plots balanced the requirement of large amount of experimental data and the costs of data collection. The forest plots are scanned from 5 positions and the data for processing are delivered in both single- and multi-scan format. Reference datasets for the benchmarking were collected by integrating manual measurements from the TLS data and the conventional field measurements.

3.1. The sample plots and complexity categories

The 24 sample plots were selected and classified into three complexity categories by foresters to represent different stand situations, which vary in species, growth stages and management activities including both homogenous and heterogeneous forests. The sample plots are distributed in a southern boreal forest in Evo, Finland (61.19°N, 25.11°E), as displayed in Fig. 2. Each plot has a fixed size, 32-by-32 m.

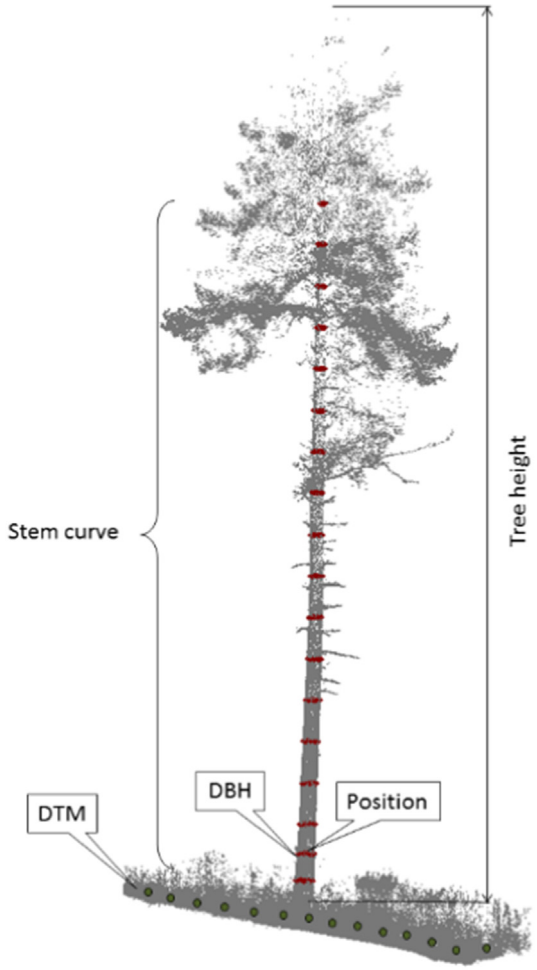


Fig. 1. The five directly measured criteria from TLS data at the plot- and tree-level: the diameter at breast height (1.3 m, DBH), tree height, tree position, stem curve (stem diameter as a function of height) and digital terrain model (DTM).

The main tree species are Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* L. Karst.) and silver (*Betula pendula* Roth) and downy (*Betula pubescens* Ehrh.) birches.

The sample plots were classified into three complexity categories: "Easy", "Medium" and "Difficult". The complexity categories were defined intuitively on stem visibility (the level of possible occlusion effects) at the ground level, stem density and DBH distribution in the sample plots. The category "Easy" represents a clear visibility with minimal understory vegetation and low stem density (~600 trees/ha); "Medium" represents sample plots with moderate stem densities (~1000 trees/ha) and sparse understory vegetation; and "Difficult" represents those plots having high stem densities (~2000 trees/ha) and dense understory vegetation. TLS data completeness in the three categories is expected to decrease as the complexity increases. Fig. 3 illustrates the examples of the three complexity categories.

Reference data of the sample plots were collected through an integration of field inventories and manual measurements from TLS data, which was conducted between May and August 2014 and detailed in Section 3.3. The sample plots' statistics are summarized in Table 2, where the plot attributes' mean and standard deviation values are presented by complexity categories. As the complexity of categories increase, the stem density increases sharply, the mean DBH and tree height decrease clearly and the basal area increase marginally, suggesting that, as the complexity level increases, the amount of young and small trees within a plot grows, the age of the forest stand decreases and human intervention in forest management also drops.

The differences between the three complexity categories are illustrated in more detail by the DBH distribution of each difficulty category in Fig. 4. For each difficulty category, the DBH is grouped at intervals of every 2 cm. The number of trees in each DBH group is separated and counted per difficulty category. In the category "Easy", most of the trees are mature with a DBH between 15 and 35 cm. The amount of small trees increases clearly in the "Medium" category, with most of trees having a DBH under 21 cm. Meanwhile, in the category "Difficult", the majority of trees have a DBH of approximately 10 cm, and the total population of trees in the plots increases significantly.

The species composition in each complexity categories is described using the tree-species-specific plot statistics, i.e., the mean and standard deviation values of DBH and tree height and the mean basal area of the main species, as presented in Fig. 5. Fig. 5 (a) and (b) indicate that DBH and tree height decreases for all species as the complexity category increases. The figure (c) shows that the plots in the categories "easy" and "medium" are pine and spruce dominated, respectively. The category "difficult" shows the heterogeneity of the species distribution in the plots, while the basal areas of pine, spruce and birch are close to each other.

3.2. The TLS data of the sample plots

The sample plots were scanned in April/May 2014, using a Leica HDS6100 (Leica Geosystems AG, Heerbrugg, Switzerland) terrestrial laser scanner. The scanner measure distances with a continuous wave of 650–690 nm. The field of view is 360° × 310° and the distance measurement accuracy is ± 2 mm at 25 m from the scanner. Data acquisition used a "High Density" mode. The angle increment is 0.036° in both horizontal and vertical directions, which gives a point spacing of 15.7 mm at 25 m from the scanning location in both horizontal and vertical directions. A full-field-of-view scan takes approximately 3 min.

The data acquisition speed is highly relevant to the forest structure. Per day, 3–7 sample plots were measured including scanning setup, 5 scans per plot and transportation between plots. In general, the field TLS measurement is pretty fast in the foresters' opinion.

The sample plots were scanned as is, i.e., without any pre-scan preparation, such as the removal of lower tree branches or the clearance of undergrowth. Five scans were made in each plot: one scan at the plot center and four scans at the four quadrant directions, as shown in Fig. 6. The theoretical position of the middle scan is at the plot's center and the distance between four quadrant scans to the center scan was 11.3 m. The exact scanning positions may be moved around the theoretical locations according to the forest structure, to find a suitable place for the scanner setup, e.g., away from tree stem next to the scanning position.

The influences of the scanning pattern on the tree attribute extraction of a sample plot are among the main targets of the benchmarking project. Therefore, TLS data were acquired using multi-scan approaches, and the project's partners had the option to process the registered multi-scan data and the single-scan data from the plot center. According to practical experience, the number of TLS-acquisition positions is a trade-off between the cost of field work (e.g., time and expense) and the data quality. In this project, five scanning positions, which is a typical setup in the multi-scan approach, was employed, because it normally leads to a merged TLS point cloud covering all trees within a forest plot and balances the completeness of tree information with cost and labor intensity.

Artificial spheres, i.e., approximately six in each plot, with a constant radius of 198 mm were set up as reference targets throughout the plot for data registration. For each sample plot, all five scans were registered using targets and merged as multi-scan TLS data with an average registration accuracy of 2.1 mm; the center scan was employed as single-scan TLS data. Examples of test data in the single- and multi-scan TLS in the three complexity categories are presented in Fig. 7. In

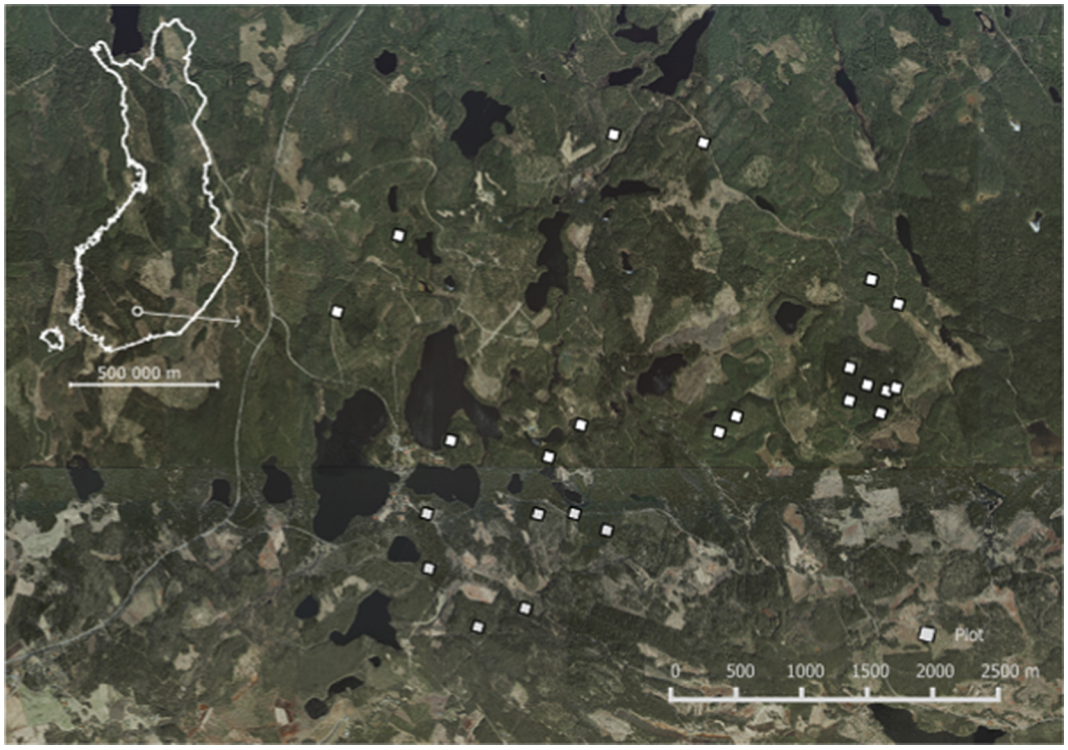


Fig. 2. The location of the study area in Finland and the distribution of the sample plots.

the TLS point cloud, the complexity category 'Easy' typically has good visibility for both single- and multi-scan TLS data. The visibility or completeness of trees in the point cloud of the complexity category 'Difficult' can be low, even in multi-scan TLS, due to the heavy occlusion effects created by the dense stands.

TLS and reference data from six test plots, i.e., two from each complexity categories in both single- and multi-scan, are open for non-profit research purposes. (The link to the data can be found at http://laserscanning.fi/tls-benchmarking-results/.)

Table 2
The statistics of the forest plots in three complexity categories, i.e., mean and standard deviation values of the stem density (stems/ha), basal areas (m^2/ha), diameter at the breast height (cm) and tree height (m).

Complexity categories	Stem density (stems/ha)	DBH (cm)	Tree height (m)	Basal area (m^2/ha)
Easy	592 ± 189	20.7 ± 8.5	18.4 ± 6.4	23.2 ± 5.9
Medium	968 ± 370	17.2 ± 10.7	16.2 ± 7.3	31.2 ± 8.6
Difficult	2021 ± 553	12.3 ± 7.2	13.2 ± 5.9	32.3 ± 7.1

3.3. Acquisition of reference datasets

Reference information was collected through a design integrating field inventories and manual measurements from TLS data to evaluate it credibly; thus, the ground truth of the sample plots can be presented as accurately as possible in the reference datasets. This section details the reference data collection.

3.3.1. Tree map and basic tree-level attributes

A detailed map of trees whose DBHs are greater than 5 cm for each sample plot was generated by integrating manual measurements from

(a) Easy (b) Medium (c) Difficult

Fig. 3. Three complexity categories of the sample plots in the TLS benchmarking. The category "Easy" represents lower stem densities and little understory vegetation, "Medium" represents moderate stem densities and sparse understory vegetation, and "Difficult" represents high stem densities with dense understory vegetation.

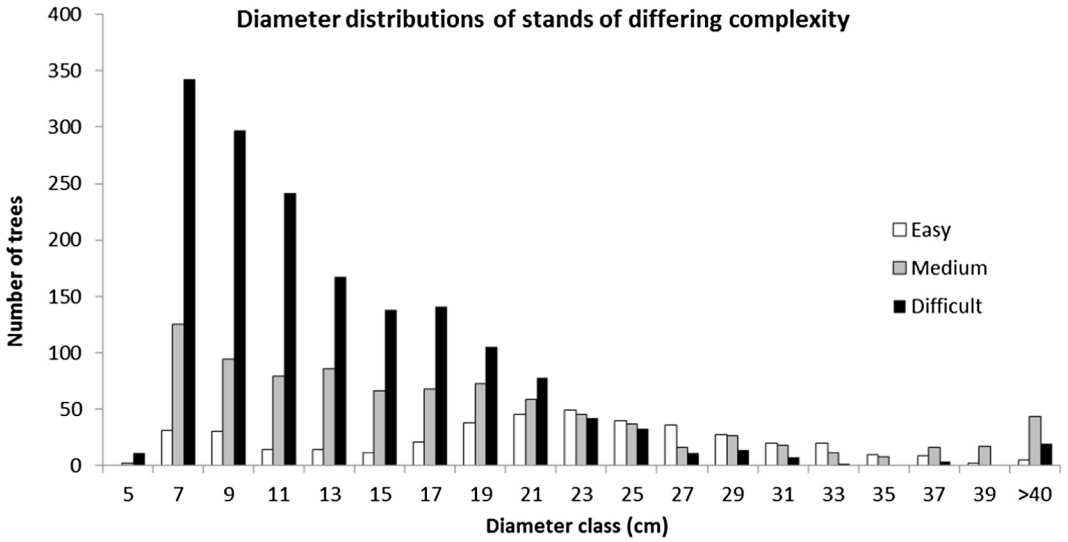
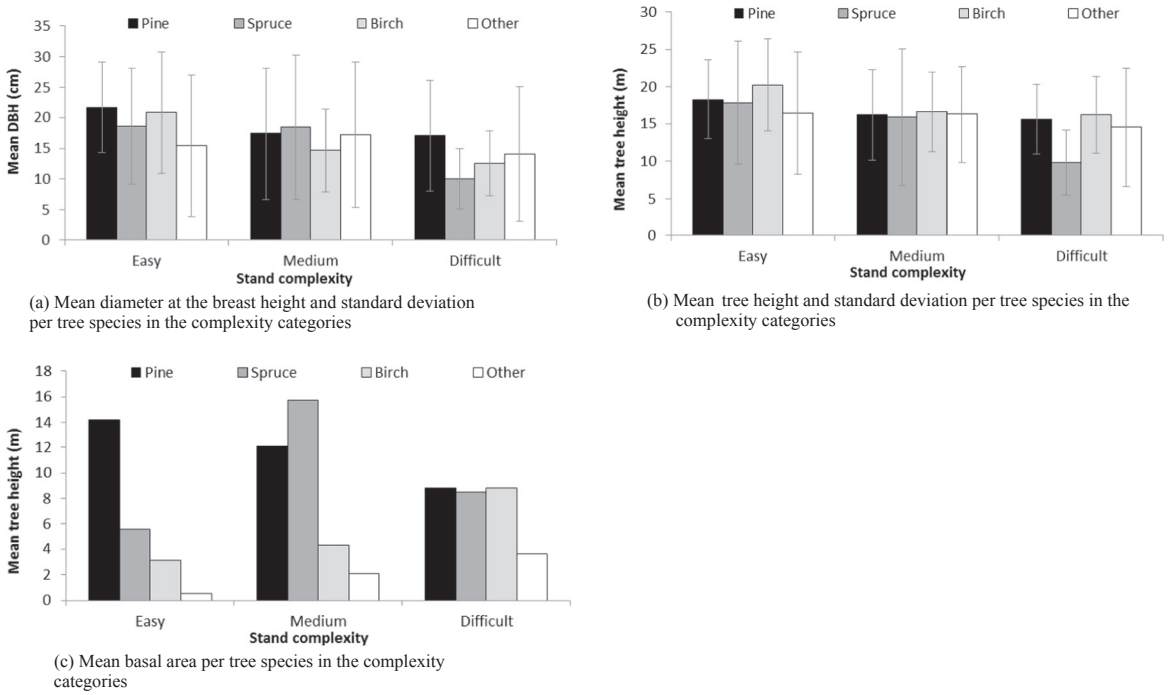


Fig. 4. DBH distribution in three complexity categories in DBH classes in 2 cm interval.



(a) Mean diameter at the breast height and standard deviation per tree species in the complexity categories

(b) Mean tree height and standard deviation per tree species in the complexity categories

(c) Mean basal area per tree species in the complexity categories

Fig. 5. Statistics of the test plots per tree species in the complexity categories. (a) Mean diameter at breast height and standard deviation. (b) Mean tree height and standard deviation. (c) Mean basal area.

the multi-scan TLS data and in the field. A preliminary tree map was manually measured from multi-scan TLS data for trees having high-quality 3D points in TLS data. The tree location was defined as the stem's center point at the breast height. This preliminary map was verified in situ during a revisit to the field, and the location of omitted trees in the preliminary tree map was determined by the distances and directions of the omitted tree to its four neighboring known trees on the preliminary tree map. A full tree map was created after the field verification, and the full tree map was double-checked again with reference to multi-scan TLS data, ensuring that the locations of field-measured trees were consistent with the TLS-recorded tree locations.

Tree-level attributes such as tree height and DBH were measured for each tree using conventional field measurement methods. For DBH, stem diameter was measured at the breast height from two perpendicular directions utilizing steel calipers to the nearest millimeter, the average value of these two diameters is recorded as DBH of a tree. Tree height was measured with Vertex 3.0 (Haglöfs, Sweden) to a resolution of 0.1 m. Vertex 3.0 utilizes a tangent method to calculate tree height. The manufacturer promises 1% accuracy in distance measurement and 0.1 degree accuracy in angle measurement. The expected accuracy of tree-height measurement was 0.5 m. Tree-height was measured from a location where the whole tree was clearly visible, normally from a distance equal to the tree length.

3.3.2. Digital terrain model

The digital terrain model can be retrieved through either point cloud data or field mensuration. In general, the point cloud from the multi-scan TLS records the terrain information in great detail. However, both terrain and dense ground vegetation may block the laser pulses, consequently creating large shadows on the terrain surface where no 3D

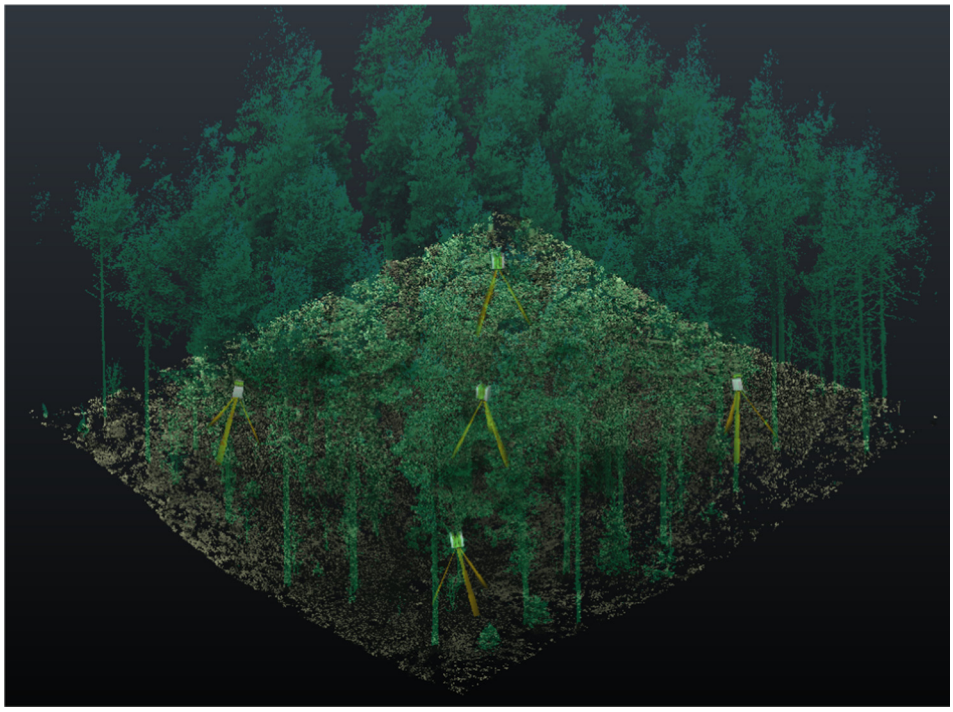


Fig. 6. The scanning positions in the sample plots.

points are recorded. Accurate ground-point classification from the point cloud is another challenge, which is hard to accomplish with fully automated algorithms. Alternatively, the DTM can be measured in field inventories, e.g., using total station. The field inventory has the potential to be the most accurate measurement since the operator can find the best observational perspectives, measure the true ground surface and have full coverage of the whole plot. But the associated cost is high since the manual measurement takes a long time.

To balance the requirement for the high accuracy with the time and labor costs, the multi-scan TLS point cloud was selected as the data source, and the reference DTMs of the sample plots were retrieved through a semi-automated approach that combines the automated data processing and manual editing. The ground points were first identified utilizing the ground classification algorithm in TerraScan software (Terrasolid, Finland). The algorithm is based on a triangulated irregular network (TIN) densification algorithm that uses local low points as initial points and starts to densify the TIN by adding more ground points according to the given parameters (Axelsson, 2000). In the automated phase, the same parameter setting was applied for all sample plots. Remaining non-ground objects, such as stones and stumps whose diameters are larger than a predefined threshold, were visually checked and manually removed. The threshold was defined as 63 cm according to a manual estimation of the average stones and stumps size in the sample plots.

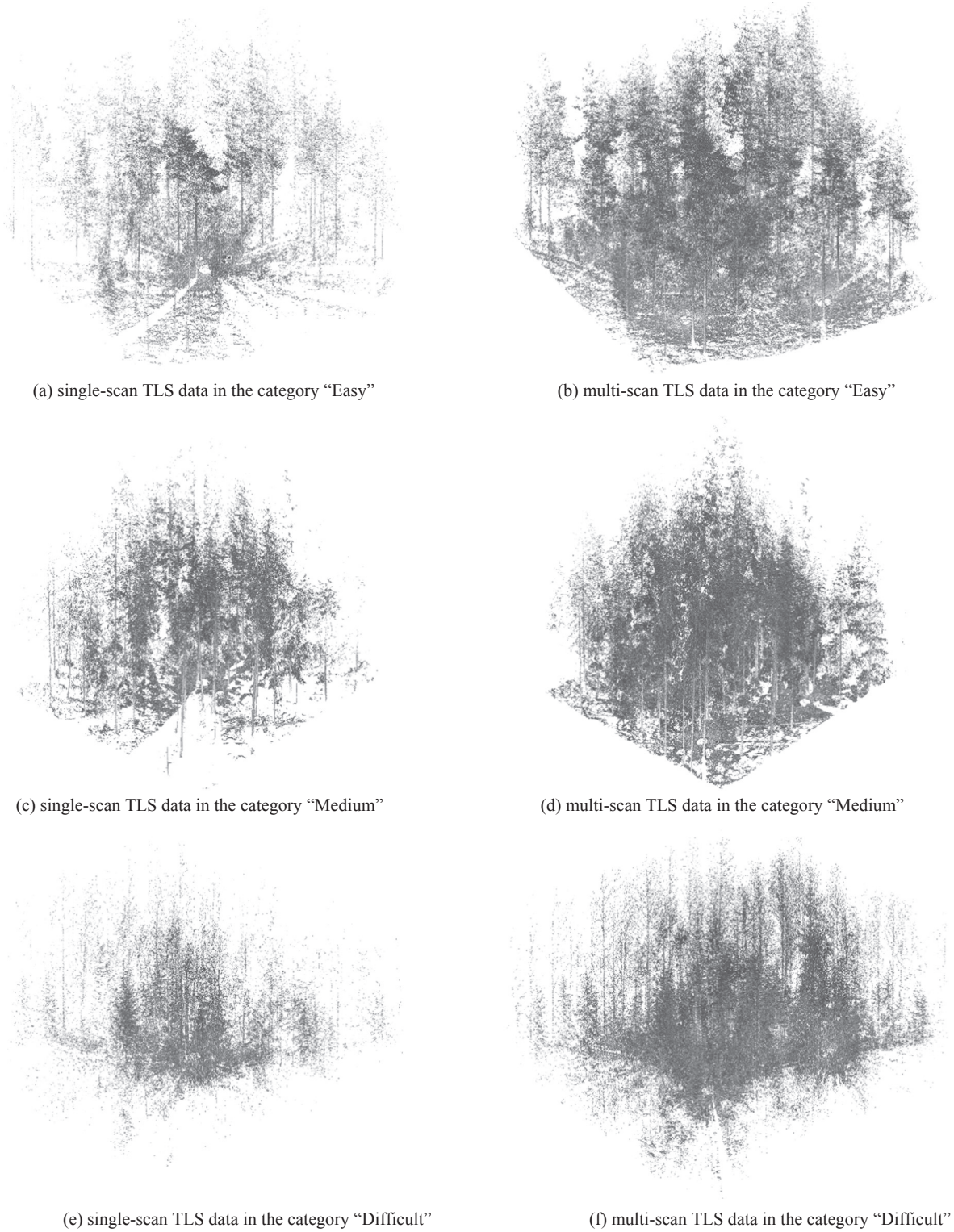
The reference DTMs were generated through rasterizations of the classified ground points. The resolution of the reference DTMs was 20 cm, considering the plot size, the details in final DTMs and the amount of interpolations required at the shadowed areas on the terrain surfaces. In the TLS-based forest inventories, the structure of the DTM is not that important; the elevation accuracy is the prime concern since the DTM gives a reference surface for the estimation of tree attributes, such as the DBH and tree height. In the rasterization approach, a grid of 20 cm resolution was overlaid on the ground points. For a cell where multiple ground points exist, an average of the z values of the points was calculated and taken as the cell's z value. For a cell in the shadowed area where no ground point exists, the z value was interpolated as an average of its neighboring cells. Fig. 8 illustrates an example of the ground points and the DTM reference of a sample plot.

3.3.3. Stem curve

The most precise method of determining the stem curve in the field is to measure the stem from the stump level to the tree top using a logging machine, which, however, exposes the stem to damage during the measurement and makes the measurement possible only after the tree has been felled (Liang et al., 2014b). Manual digitization from the precise point cloud data is so far the most accurate and practical method for non-destructive stem curve measurement of a large amount of trees. In this study, each tree stem was manually digitized through multi-scan TLS point cloud to measure the stem curves. The stem curve of an individual tree consisted of stem diameters starting at the height of 0.65 m above the ground, followed by diameters at the DBH height and at every meter above the DBH height, i.e., 0.65 m, 1.3 m, 2 m and 3 m, till the maximum measurable heights from the point cloud data.

For each sample plot, the multi-scan TLS point cloud was first cut for each individual tree. The points of each tree were then sliced on specific heights above local ground-height level. Points in each cross-section were inspected from a top view, and a circle was manually fitted on the stem points using the TerraScan software. In many cases, the stems did not present exact circular shapes on the cross-sections. Each circle was thus fitted to minimize the least square error between the stem points and the arc of the circle. Stem curves started from the lowest height and continued up the stem so long as a sufficient amount of points could be recognized as a stem cross section. At each measurement position, the central coordinates and diameter of the fitted circle were recorded. An example of the stem-curve measurement is presented in Fig. 9.

Even though the multi-scan TLS data provides a large amount of details in forests, in most cases, regardless of the stand conditions in the sample plots, the stems are blurred close to the treetops, due to occlusion effects and the distances to the scanning positions. The severity of the occlusion effect varies from plot to plot and from tree to tree, depending on the stand density, species and the tree's position in the plot. If an insufficient amount of points was found at a particular height, the diameter of the previous measurement at a lower height was used to estimate the diameter at that height. In some very special cases, the tree stems were divided into smaller sub-stems from the root and no clear main stem could be identified. In such cases, multiple stems were



(a) single-scan TLS data in the category “Easy”

(b) multi-scan TLS data in the category “Easy”

(c) single-scan TLS data in the category “Medium”

(d) multi-scan TLS data in the category “Medium”

(e) single-scan TLS data in the category “Difficult”

(f) multi-scan TLS data in the category “Difficult”

Fig. 7. Examples of forest sample plots in the single- and multi-scan terrestrial laser scanning data in the three complexity categories.

recorded for a single tree if the separation of the main stem occured below the breast height and if the divided sub-stems satisfied the 5 cm DBH threshold. Examples of the reference stem curve measurement are shown in Fig. 10, which compares the stem curve measurements in the sample plots of “Easy” and “Difficult” complexity categories, where the stem curve was measured for visible parts in the point clouds.

3.3.4. Stem volume and total aboveground biomass

Stem volume was estimated based on the stem curve. The stem was divided into sections based on the retrieved diameters. In addition, the total stem volume was calculated as the sum of the sections. The stem section between two adjacent diameters was modeled as a cylinder whose radius was the mean of the radii of the top and bottom of the block. The upper most tree stem was modeled as a cone using the highest stem diameter and tree height. The base of the tree, i.e.,

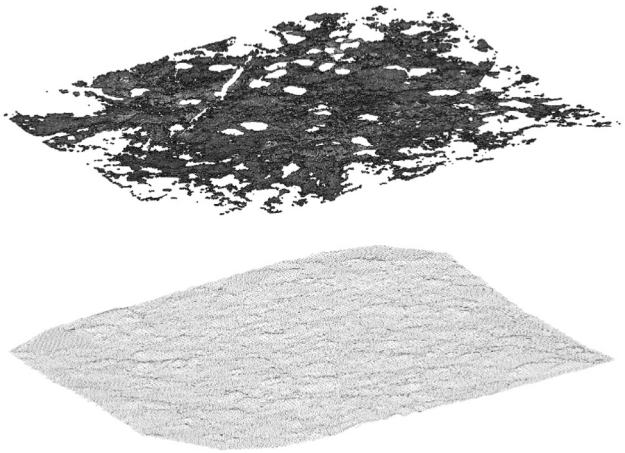


Fig. 8. An example of the ground points (upper) classified by a semi-automated approach and the rasterized DTM (lower). Holes in the ground points are created by shadows from big rocks and stumps or the low point density towards the plot borders.

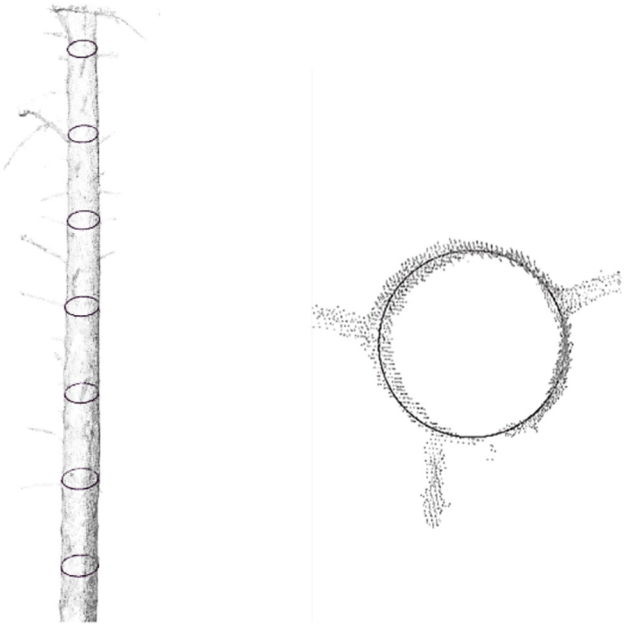


Fig. 9. An example of a set of stem diameter measurements on a Scots pine tree (left) and of a single measurement circle fitted on stem points (right).

between the ground and the lowest diameter was estimated using a cylinder whose height was 65 cm and diameter equaled to the lowest retrieved diameter.

The total aboveground biomass of a tree was estimated using the multivariate statistical models presented in (Repola, 2008, 2009). The models use DBH and tree height as explanatory variables to predict the biomass. Repola's models were developed for birch, Scots pine and Norway spruce trees in Finland. For other tree species, the birch model was applied.

4. Methods of evaluation

The partners of the benchmarking project are required to deliver their extraction results for the criteria, i.e., the DTM, tree locations, tree height, DBH, and stem curve, of each sample plot utilizing the TLS datasets of the 24 sample plots. Partners' results are evaluated respecting the relevant reference data described in Section 3.3 and utilizing standardized evaluation procedures defined in this section.

4.1. The accuracy of the DTM

Partner DTMs were evaluated using the 'Output control report' tool of the TerraScan software, which is designed for elevation comparison between laser ground points and the known ground control points. For each sample plot, the reference DTM was employed as the ground control data, and the DTM from partners was compared against the reference DTM.

For each given XY location of the reference DTM, three nearest points from the DTM to be evaluated were selected, and a small 3D triangulate plane was created using the selected points. An elevation value z_e derived from the triangulate plane was compared with the reference z value at the location XY. The root mean square error (RMSE) of the built DTM was calculated based on the elevation difference between z and z_e.

In addition to the RMSE, the percentage of the reference DTM covered by partners' results was taken into the evaluation, since the combination of the RMSE and the area covered gives a more comprehensive evaluation than the single factor. For example, a small RMSE can be achieved by limiting the extracted DTM to areas where TLS data coverage is good, because errors in the DTM estimation tend to increase in areas where TLS data coverage is inadequate.

4.2. Tree matching

The first step in evaluating the tree-level attributes is to verify the correctness of the detected individual trees using various algorithms. An automated tree-matching approach is developed to evaluate whether a tree in a plot is correctly detected or not.

The detected and reference trees were matched according to both tree locations XY and DBHs. For each detected tree, all reference trees within a neighborhood of 50 cm radius were retrieved. The detected tree was linked to the neighboring reference tree whose DBH was closest to that of the detected tree to form a preliminary match.

In the preliminary matching, more than one detected tree may correspond to the same reference tree. To remove such duplicate matches, the following four steps were repeated until unique links between the detected and reference trees are found: (1) a non-unique match was sought; (2) a match was established if the DBH of a detected tree was closest to the reference's DBH; (3) other links to the reference were removed, and the reference tree was also removed from reference map; and (4) a new matching iteration started from step 1 using the remained detected and reference trees. The iteration continued until all reference or detected trees found a match in the other list, if possible. If no reference tree can be found for a detected tree, the detected tree was considered a commission error. If no detected tree can be found for a reference tree, an omission error was counted.

With the matching approach, tree detection, location, DBH, and the tree height were all evaluated simultaneously.

4.2.1. Tree detection accuracy

Tree detection accuracy was evaluated using three measures, i.e., the completeness, the correctness and the mean accuracy.

Completeness measures how large a part of the reference trees is found using an algorithm. Correctness measures how large a part of the trees extracted using an algorithm is correct. They are defined as

$$\text{Completeness} = \frac{n_{\text{match}}}{n_{\text{ref}}} \tag{1}$$

$$\text{Correctness} = \frac{n_{\text{match}}}{n_{\text{extr}}} \tag{2}$$

where n_{match} is the number of found reference trees, n_{ref} is the number of reference trees and n_{extr} is the number of trees detected.

The mean accuracy of detection was defined as the joint probability that a detected tree randomly chosen was a correct detection and that a

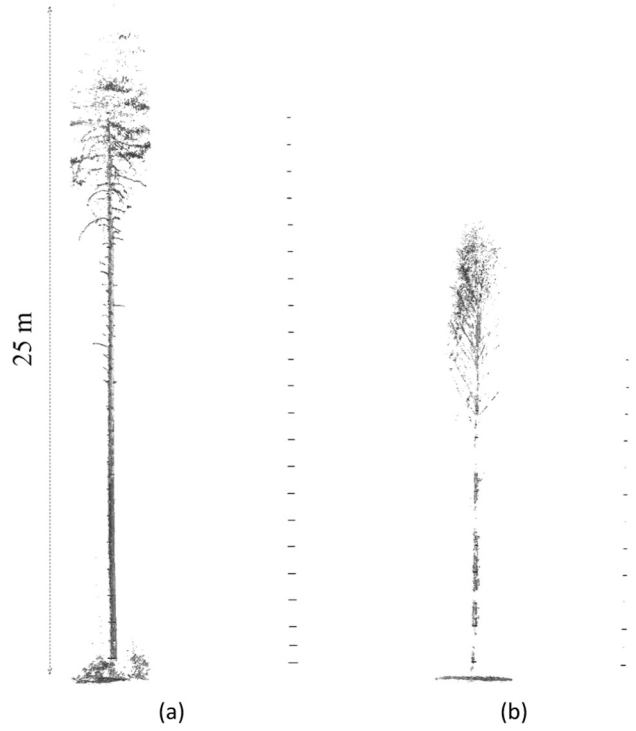


Fig. 10. Examples of the reference stem-curve measurement. (a) A Scots pine tree on an easy plot. Most of the tree trunk is visible in the TLS data and the top part the tree stem is occluded by the tree crown. Stem curve was measured to the tree top but not to the apex. (b) A birch tree on a plot in the complexity category difficult. The stem curve is measurable only for the visible parts. The dashed line on the left is 25 m and gives the scale for both sub-figures.

reference tree randomly chosen is detected by an algorithm. It is defined as

$$\text{Mean accuracy of detection} = \frac{2n_{\text{match}}}{(n_{\text{ref}} + n_{\text{extr}})} \tag{3}$$

4.2.2. The accuracy of tree location, DBH, and height

The accuracy of the extracted tree location, tree height and DBH is evaluated using the RMSE and bias, except for the tree location where only RMSE is calculated. In addition, relative RMSE and relative bias, denoted by RMSE% and Bias%, respectively, were calculated for DBH and tree height. The accuracy measures were calculated by comparing the extracted values $\hat{y}_i$ to the reference values y_i, i.e., tree parameter such as tree location, height or DBH, where i is the index of the match. RMSE is defined with the equation

$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{match}}} \sum_{i=1}^{n_{\text{match}}} (\hat{y}_i - y_i)^2} \tag{4}$$

Bias is defined with the equation

$$\text{Bias} = \frac{1}{n_{\text{match}}} \sum_{i=1}^{n_{\text{match}}} (\hat{y}_i - y_i) \tag{5}$$

The relative RMSE and bias, in percentages, were calculated by comparing the RMSE and bias to the mean reference value $\bar{y}$ defined as

$$\bar{y} = \frac{1}{n_{\text{match}}} \sum_{i=1}^{n_{\text{match}}} y_i \tag{6}$$

The RMSE% was calculated with the equation

$$\text{RMSE\%} = \frac{\text{RMSE}}{\bar{y}} \times 100\% \tag{7}$$

and the bias% with the equation

$$\text{Bias\%} = \frac{\text{Bias}}{\bar{y}} \times 100\% \tag{8}$$

4.3. The stem-curve accuracy

At an individual tree level, the accuracy of the stem-curve estimates was evaluated using RMSE and bias of the extracted stem curve which were calculated using Eqs. (9) and (10), respectively.

The extracted stem curves consisted of diameters $\hat{d}_i(\hat{z}_{i,j})$ at heights $\hat{z}_{i,j}$, where i is the index of the match and j is the index of the extracted diameter. The corresponding reference measurements are denoted by $d_i(z_{i,k})$, where k is the index of the measured diameter. Since the heights $\hat{z}_{i,j}$ at which the diameters were extracted may vary between participants, i.e., not equal to the defined reference heights $z_{i,k}$, the accuracy of the extracted curve was evaluated by comparing the diameters $\hat{d}_i(\hat{z}_{i,j})$ to the linearly interpolated reference values at the same heights $d_i^{\text{interp}}(\hat{z}_{i,j})$.

$$\text{RMSE}_i = \sqrt{\frac{1}{m_i} \sum_{j=1}^{m_i} (\hat{d}_i(\hat{z}_{i,j}) - d_i^{\text{interp}}(\hat{z}_{i,j}))^2} \tag{9}$$

$$\text{Bias}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} (\hat{d}_i(\hat{z}_{i,j}) - d_i^{\text{interp}}(\hat{z}_{i,j})), \tag{10}$$

where i is the index of the match and m_i the number of extracted diameters in the ith match. The extracted diameters outside the range of the reference diameters were ignored in the accuracy evaluation.

At a plot level, the accuracy of the extracted stem curves is evaluated using averages of the tree-wise RMSEs and biases using equations

$$\overline{\text{RMSE}} = \frac{1}{n_{\text{match}}} \sum_{i=1}^{n_{\text{match}}} \text{RMSE}_i, \tag{11}$$

$$\overline{\text{Bias}} = \frac{1}{n_{\text{match}}} \sum_{i=1}^{n_{\text{match}}} \text{Bias}_i \tag{12}$$

In addition to the measurement accuracy, the efficiency of each algorithm is evaluated by the proportion of the stem covered by the extracted diameters. For this purpose two measures are defined, i.e., curve length ratio (CLR) and the percentage of the tree height covered (PHC). Both measures are calculated using histograms whose bins correspond to height intervals along the stem. The bin edges are designed such that the default heights of the retrieved diameters (see Section 3.3.3) are in the middle of the bins. The bin is occupied if at least one diameter is retrieved between the bin edges; otherwise, it is empty. The lengths of the occupied bins are summed to determine the stem length that is covered by the retrieved stem curve. CLR is the ratio of the stem length covered by the extracted curve to the stem length covered by the reference curve in percentage. PHC is otherwise the same as CLR, but the denominator is replaced by the measured reference tree height. The units of both the CLR and PHC are in percentages.

CLR measures how large a part of the manually measured reference stem curve is retrieved with an algorithm-extracted curve, which also reveals how well the (semi-) automatic stem-curve extraction methods perform compared to manual measurements by laser scanning experts, i.e., the best a human being can achieve. CLR may have a value larger than 100%, meaning the method extracts more curve than the manually measured reference data, or the computer over-performs human beings if the method is fully automated. However, the accuracy of the extracted diameter outside the reference curve is unknown. PHC reveals the degree of the whole tree retrieved by the extraction methods, 100% being the ultimate goal where an algorithm fully depicts the object. PHC indicates the capacity of the TLS point cloud and an algorithm to depict the object in the field.

The average CLR and PHC over all matched trees in a plot were also

calculated in the same way as the average of tree-wise RMSEs and biases of stem curves as mentioned above, to gain an overall measure of how large a part of the trees were covered by the extracted curves using different algorithms.

Since some extracted matches do not contain an extracted stem curve, a modified completeness is used in the curve-related evaluations, which considers only those matches with an extracted stem curve. It is defined as

$$\text{Completeness}_{\text{RMNSC}} = \frac{(n_{\text{match}} - n_{\text{match, no stem curve}})}{n_{\text{ref}}} \quad (13)$$

where $n_{\text{match, no stem curve}}$ is the number of matches that do not have an extracted stem curve. $\text{Completeness}_{\text{RMNSC}}$ is a modified completeness, where the subscript 'RMNSC' comes from the words 'removed matches with no stem curve'.

4.4. The accuracy of stem volume and total biomass

Although same mathematical models are applied to all participants for the stem volume and the total biomass estimation, the evaluation actually reveals the combined impacts of extracted tree height, DBH and stem curves to the volume and biomass estimates. Because each algorithm has its own strength and weakness with respect to different tree attributes, e.g., improving estimate accuracy by sacrificed completeness, the volume and biomass evaluation provides an overview of the overall performance of all the extracted tree attributes of an algorithm.

The absolute and relative RMSE and bias of volume and biomass over the trees in each plot are calculated using Eqs. (4)–(8). In the biomass estimates, tree species information was from the reference of the linked tree. In addition, a volume ratio and a biomass ratio are used to evaluate the performance of volume and biomass estimates at a plot-level. Volume ratio is the ratio of the total volume of all extracted trees to all reference trees in the plot. It evaluates the overall volume estimations. Biomass ratio is the ratio of the total biomass of the matched extracted trees to the total biomass of all reference trees in the plot. It compares the biomass of the correct tree detections to the reference biomass at a plot level.

5. Algorithms of participants

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 3 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 Appendix A.

6. 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.

6.1. Digital Terrain Model (DTM)

The DTM is used as reference surface 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 modeling 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.

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.

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. Without extrapolation, 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, as shown in Fig. 11.

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,

Table 3
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 modeling	1. Preprocessing		2. DTM	3. Individual tree detection	4. Stem modeling
							Thinning	Filtering			
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 modeling	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 neighborhood	Number of points + 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 count in voxel	\
10	SLU	Both	A	U	voxel	Point	\	Flatness	Minimum height in multi-scales	Selecting curvature with same radius and origin + 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 + plane fitting	Clustering in a 2D slice	Circles at different heights
15	UofL	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.

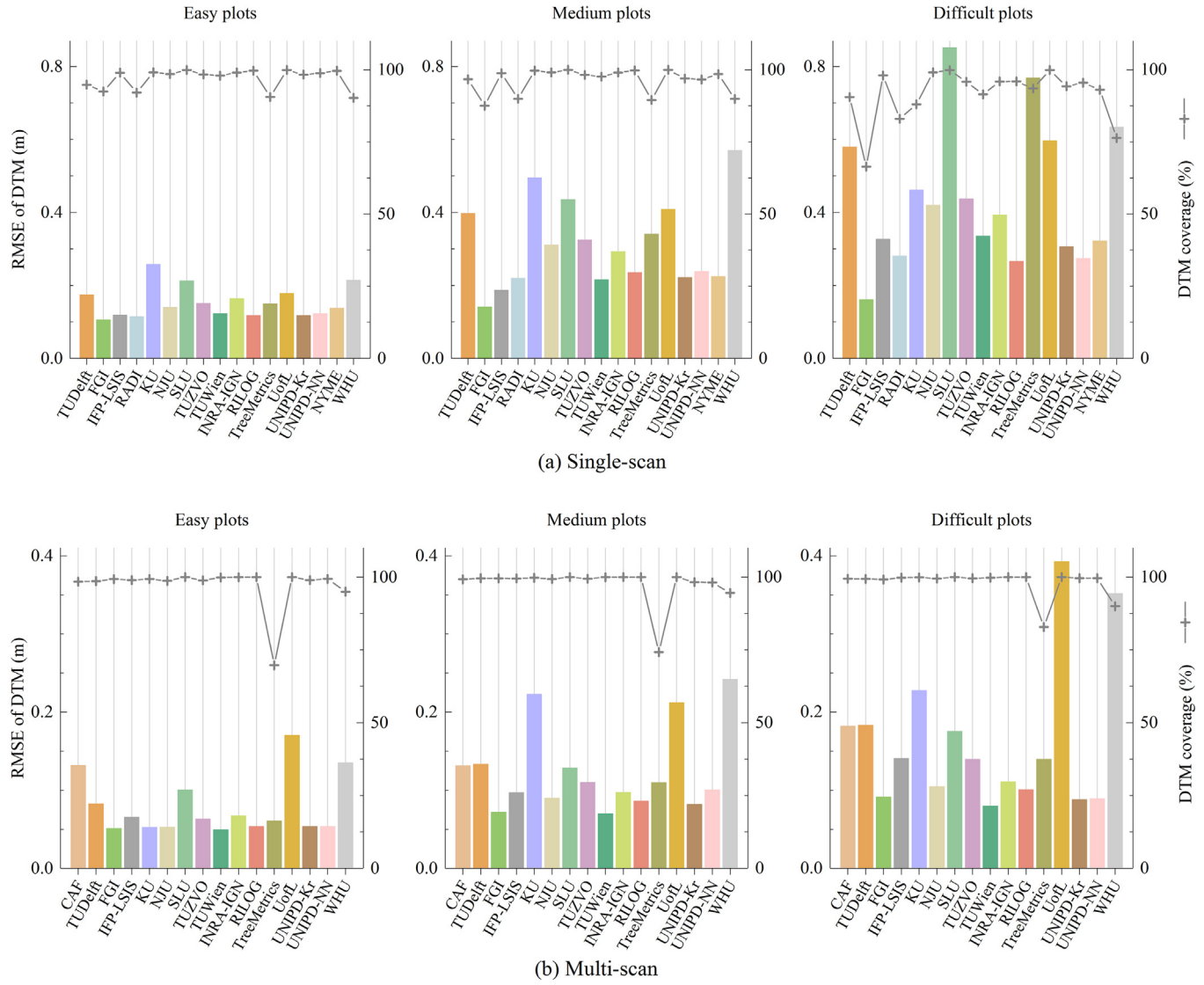


Fig. 11. RMSE of DTM from the single- (a) and multi-scan (b) TLS data. The left vertical axes correspond to the RMSE (bars), and the units are in meters. The right vertical axes correspond to the DTM coverage (solid line with '+' markers), and the units are in percentages.

provide similar DTM results in terms of accuracy and coverage from the multi-scan data. 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 did not have extrapolation operation and gave smaller RMSE and coverage values, while TUWien extrapolated DTM and gave a slightly larger RMSE but covered the plots area more completely, indicating that extrapolation is the main error source.

6.2. 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 Figs. 12 and 13, 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 Figs. 12(a) and 13(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

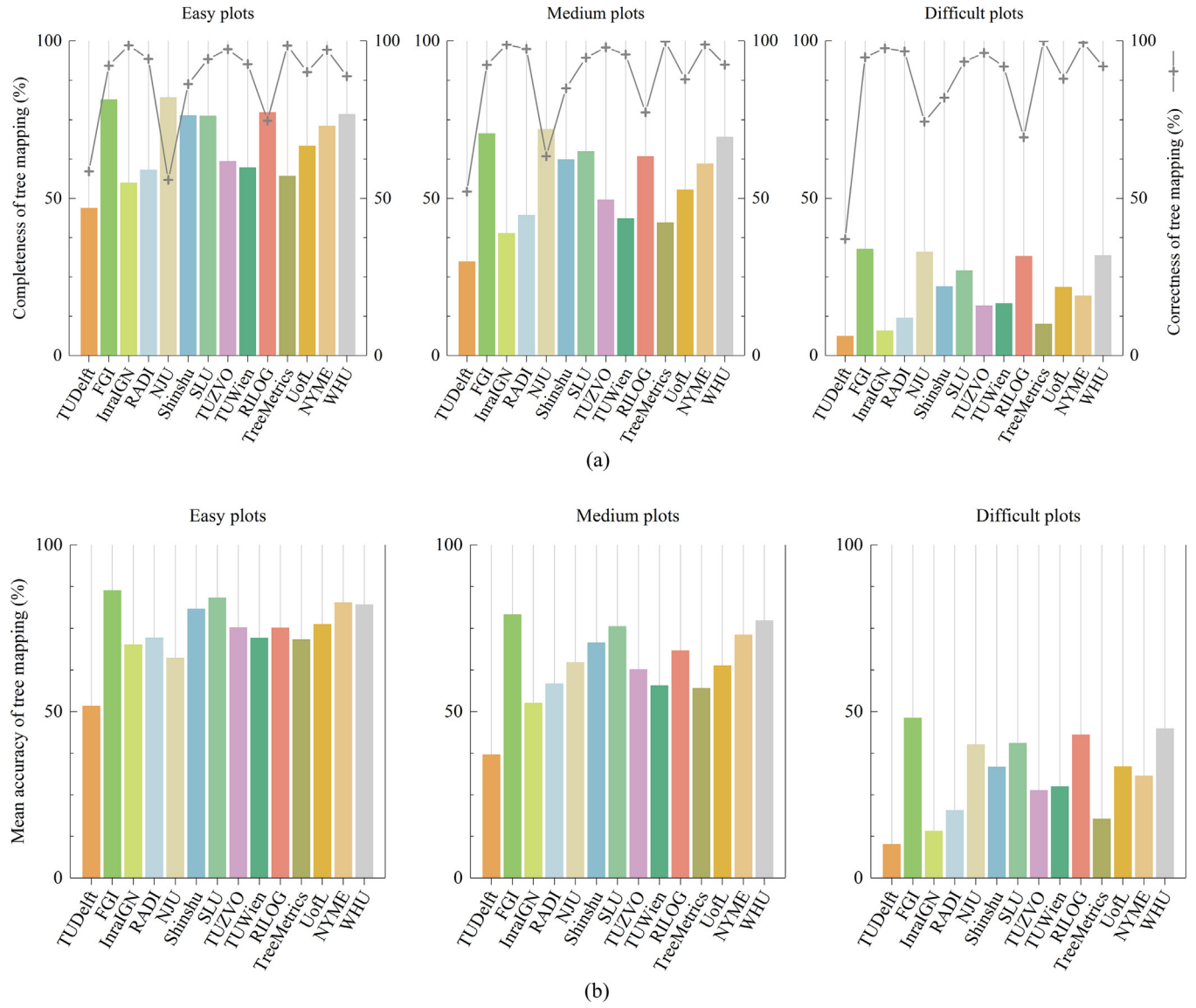


Fig. 12. The accuracy of tree mapping from the single-scan TLS data. (a) The completeness and the correctness: the left vertical axes correspond to the completeness (bars), and the right vertical axes correspond to the correctness (solid line with '+' markers). (b) The mean accuracy. Units in both (a) and (b) are in percentages.

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 Figs. 12(b) and 13(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 indicate that when the completeness and correctness are averaged, most of the algorithms perform at a similar level. The variations in the mean accuracies among the algorithms in each complexity category are much lower than the completeness and the correctness, regardless of the stand condition and the applied TLS data.

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.

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

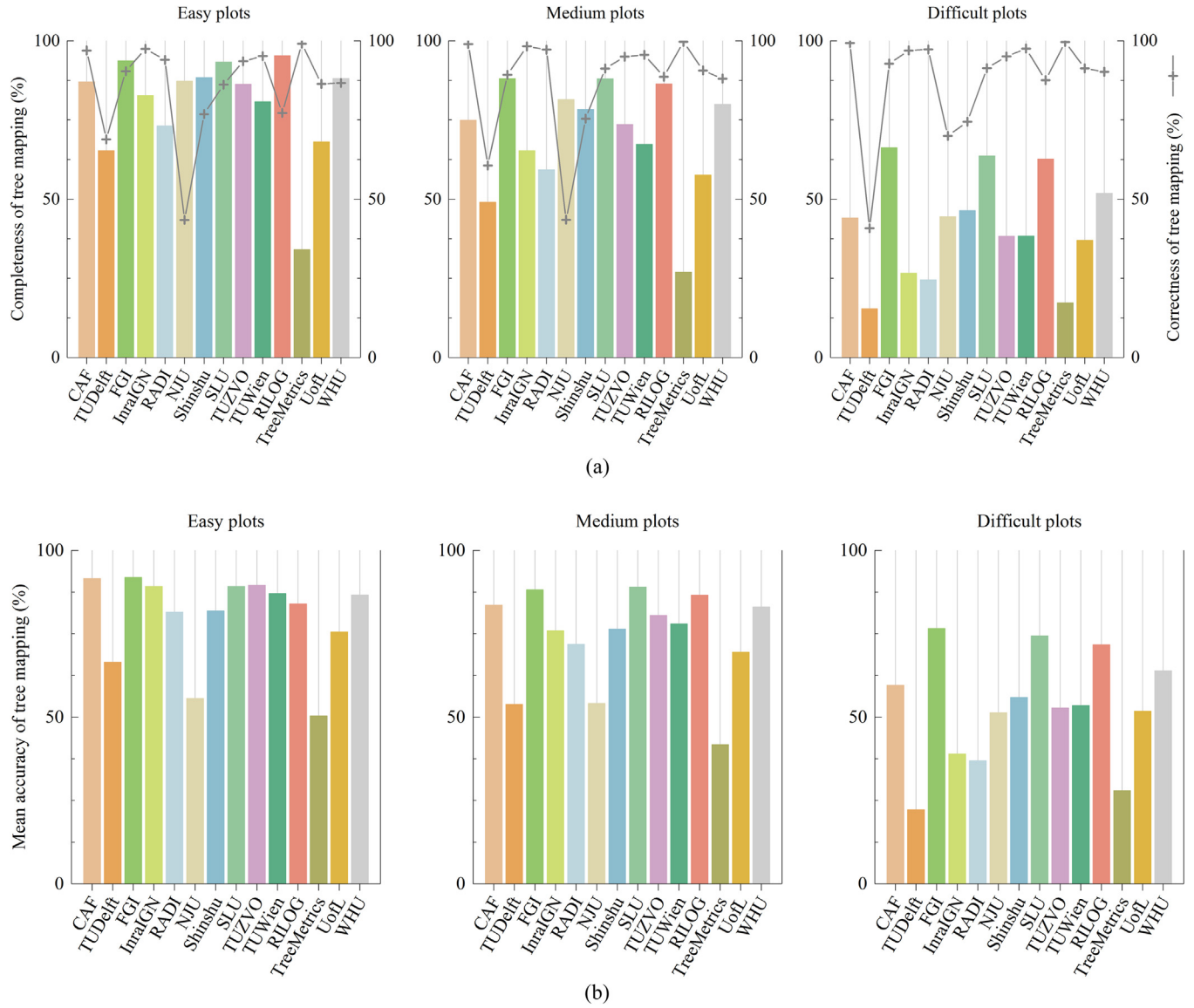


Fig. 13. The accuracy of tree mapping from the multi-scan TLS data. (a) The completeness and the correctness: the left vertical axes correspond to the completeness (bars), and the right vertical axes correspond to the correctness (solid line with '+' markers). (b) The mean accuracy. Units in both (a) and (b) are in percentages.

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 detection 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.

6.3. 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 Fig. 14. 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.

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

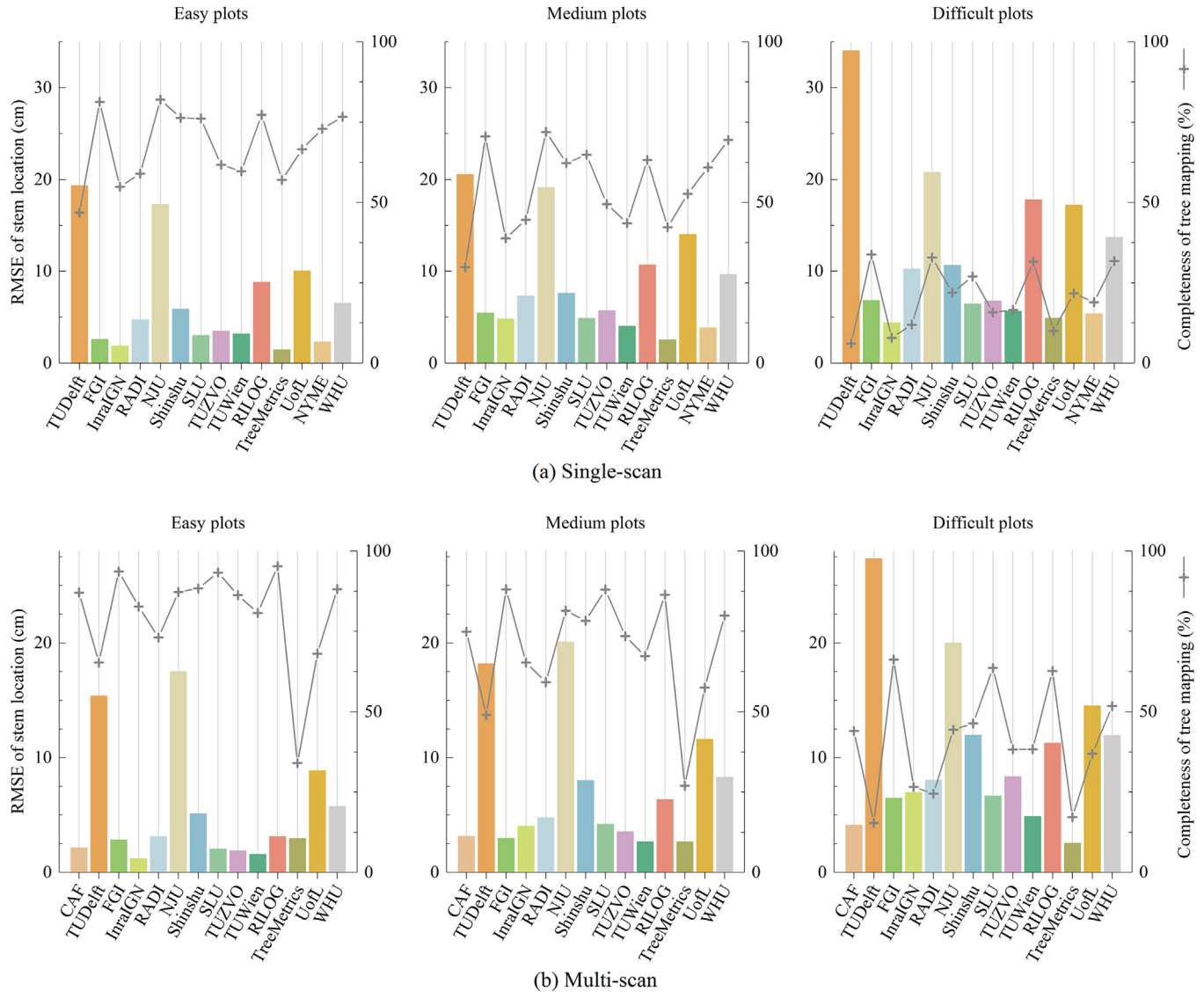


Fig. 14. RMSE of the stem location estimation from the single- (a) and multi-scan (b) TLS data. The left vertical axes correspond to the RMSE (bars), and the units are in centimeters. The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

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.

6.4. Diameter at breast height

The accuracy of DBH estimations is evaluated using the statistical metrics RMSE, RMSE%, bias, and bias%, with respect to the field measurement results in the reference data. The evaluation results are illustrated in Figs. 15 (RMSE and RMSE%) and 16 (bias and bias%) of the DBH estimates from single-scan data, and in Figs. 17 (RMSE and RMSE%) and 18 (bias and bias%) of the DBH estimates from 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.

6.4.1. RMSE of the DBH estimation

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 to 4 cm and RMSE% ranging from 8 to 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.

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.

6.4.2. 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

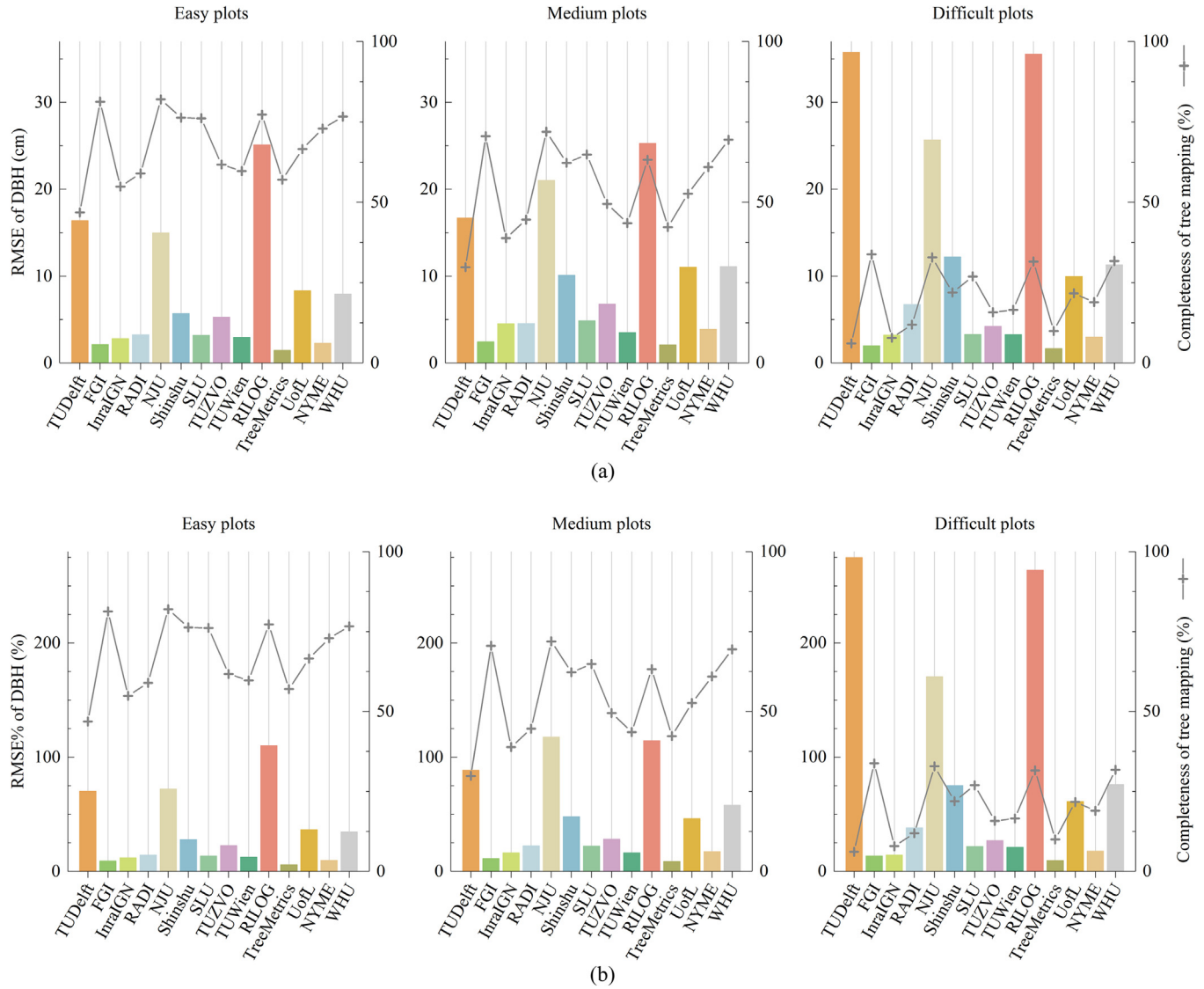


Fig. 15. RMSE (a) and RMSE% (b) of the DBH estimation from single-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in centimeters) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

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 Fig. 17). 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.

The application of multi-scan data can largely reduce the underestimation of the DBH; only three algorithms produced negative bias (see Fig. 18). The bias of DBH estimations also confirms that the multi-scan approach improves DBH estimation accuracy. The more complex the stand condition is, the more significant the advantage of applying the multi-scan strategy. For the "Conservative" and "Robust" algorithms, the bias and bias% can be kept close to zero in all three stand complexity categories.

6.5. Tree height

The tree height estimation is evaluated using the statistic metrics RMSE, RMSE%, bias, and bias%, with respect to the reference data, as with the DBH. The evaluation results are illustrated in Figs. 19 (RMSE) and 21 (bias) for single-scan data, and in Figs. 20 (RMSE) and 22 (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.

6.5.1. 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 to 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

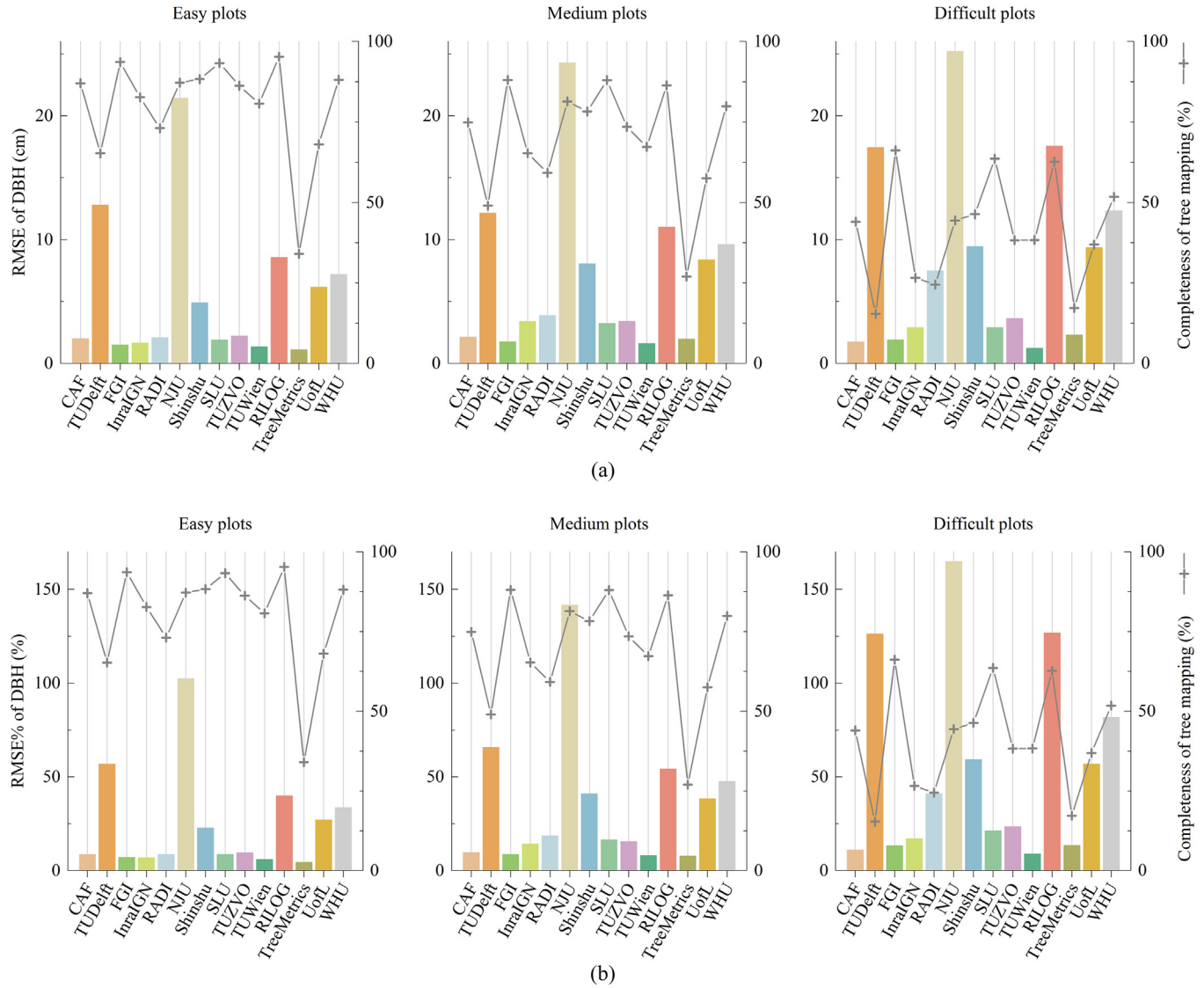


Fig. 16. RMSE (a) and RMSE% (b) of the DBH estimation from multi-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in centimeters) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

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. Although the absolute values of the RMSE were similar in medium and difficult plots, the RMSE% in the difficult plots was clearly larger than that in the medium plots because the difficult plots have many small trees.

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.

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.

6.5.2. *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 underestimations 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

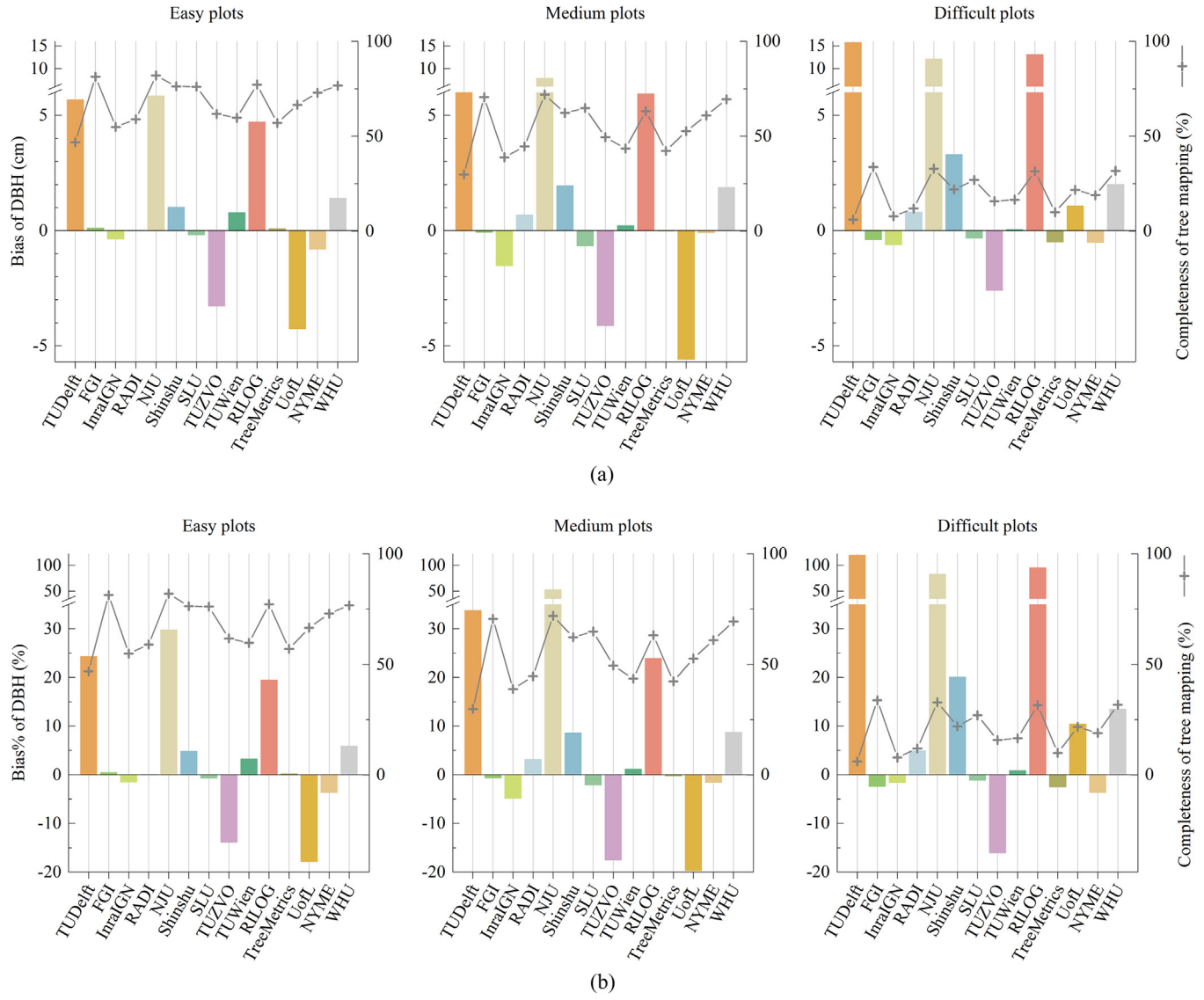


Fig. 17. Bias (a) and bias% (b) of the DBH estimation from single-scan TLS data. The left vertical axes correspond to the bias (bars in (a), the units are in centimeters) and bias% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

heights, especially in complex forest stands where the sub canopy growth is rich, which explains the positive bias and bias% in Fig. 21.

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 Fig. 22 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.

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.

6.6. 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) defined in Section 4.3, are investigated. The CLR is 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 indicates 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 and the estimates are accurate. The PHC reveals the degree of the whole tree that is retrieved by the

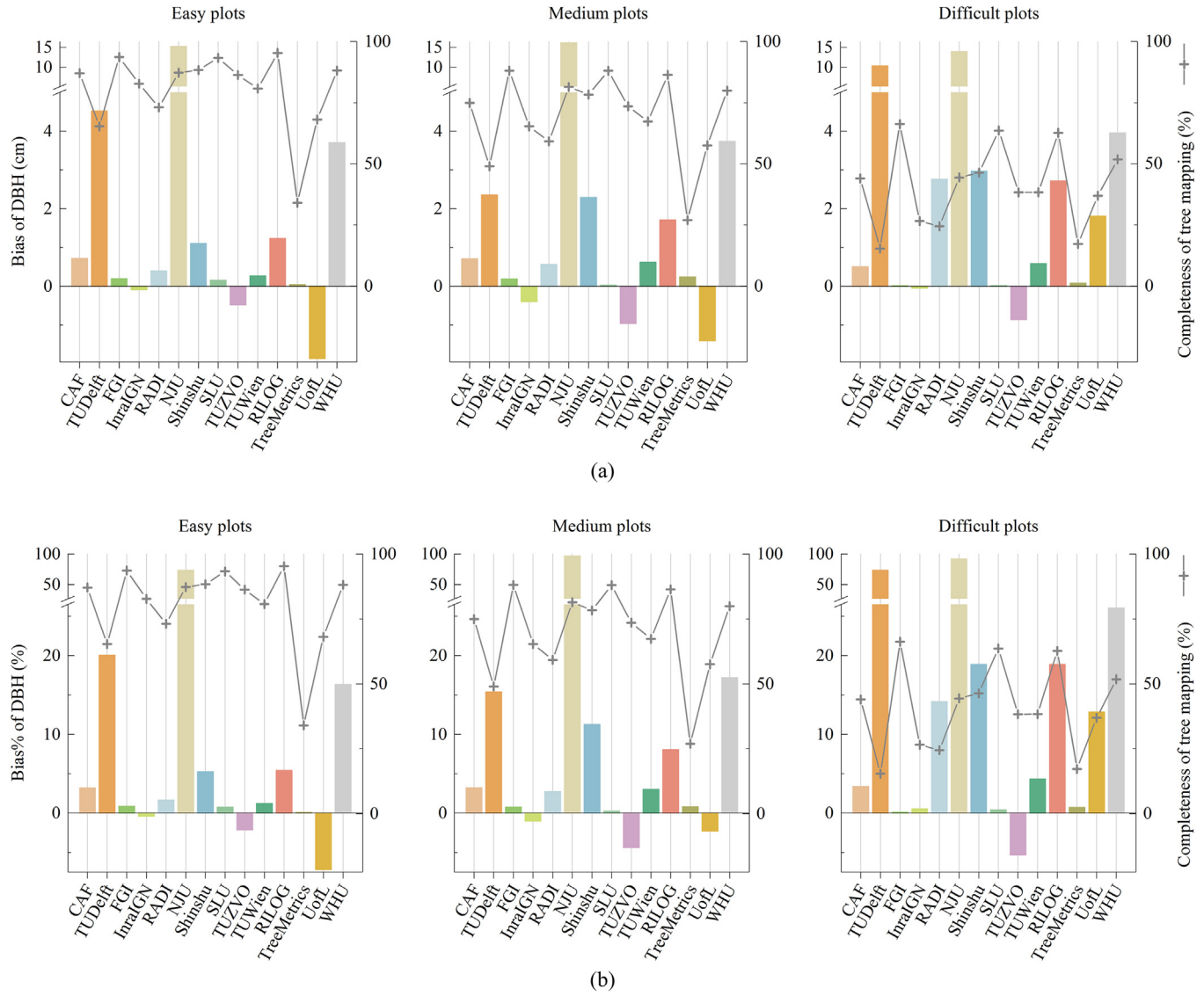


Fig. 18. Bias (a) and bias% (b) of the DBH estimation from multi-scan TLS data. The left vertical axes correspond to the bias (bars in (a), the units are in centimeters) and bias% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

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.

6.6.1. RMSE of the stem curve estimation

The RMSE of the tree-wise stem curve estimation from the TLS data is illustrated in Fig. 23. 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 modeling, 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 modeled. 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.

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

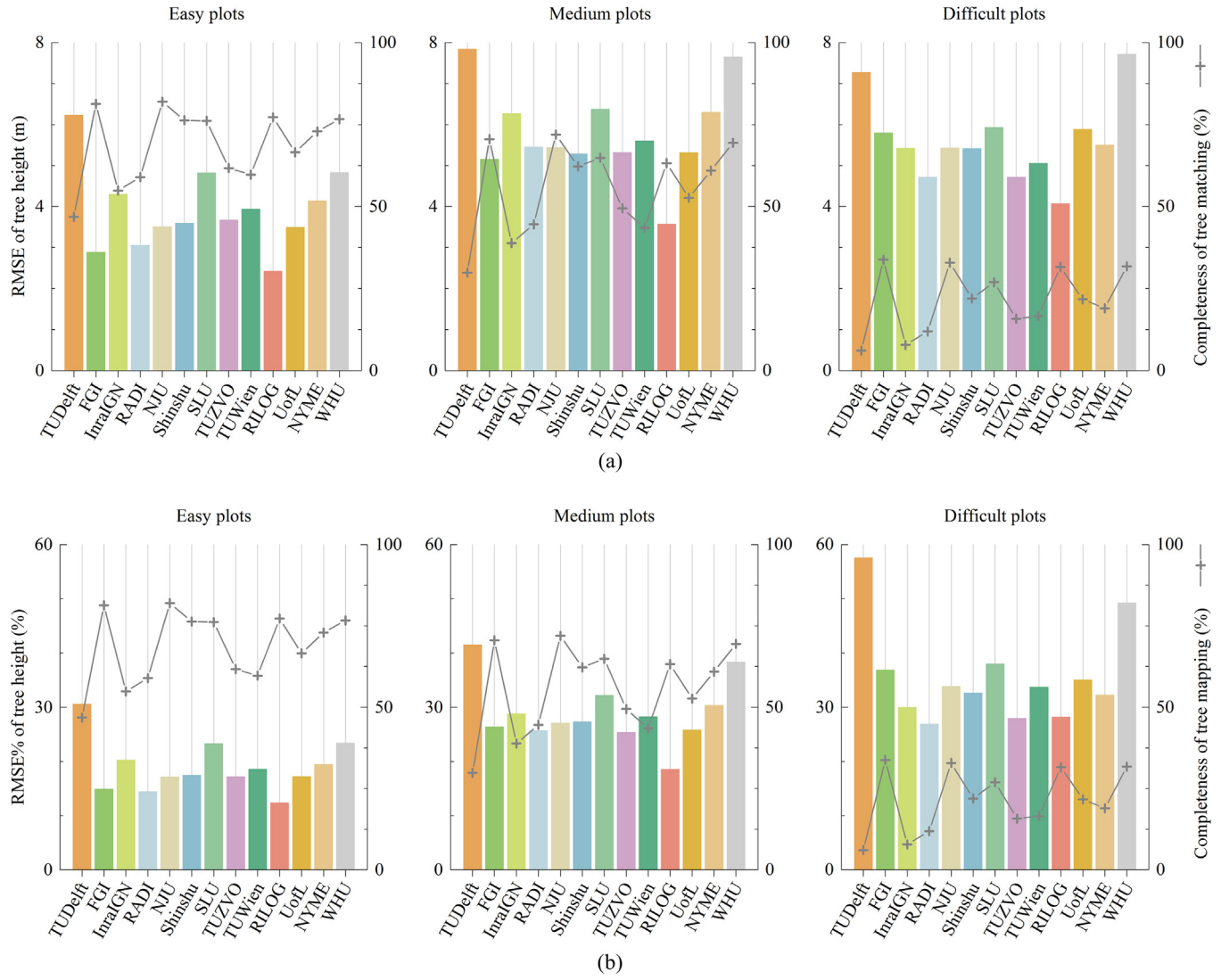


Fig. 19. RMSE (a) and RMSE% (b) of the tree height estimation from single-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in meters) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

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 easy, 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 leads 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.

6.6.2. 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, e.g., Fig. 24. 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 10 cm 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 become positive when the multi-scan data is applied, regardless of the stand situation. One reason is that multi-scan data provides a more complete tree structure and there is more noise surrounding the stems in the multi-scan data, especially inside the crown, due to the branches or the registration errors of multiple scans; therefore, the stem diameters at different stem heights may be over estimated. Another possible reason is the wind that blows through canopies and pushes stems bent. When bent stems are recorded by the multi-scan TLS data, they appear bigger in merged TLS data as shown in examples in (Vaajaet al., 2016; Pyorala et al., 2018).

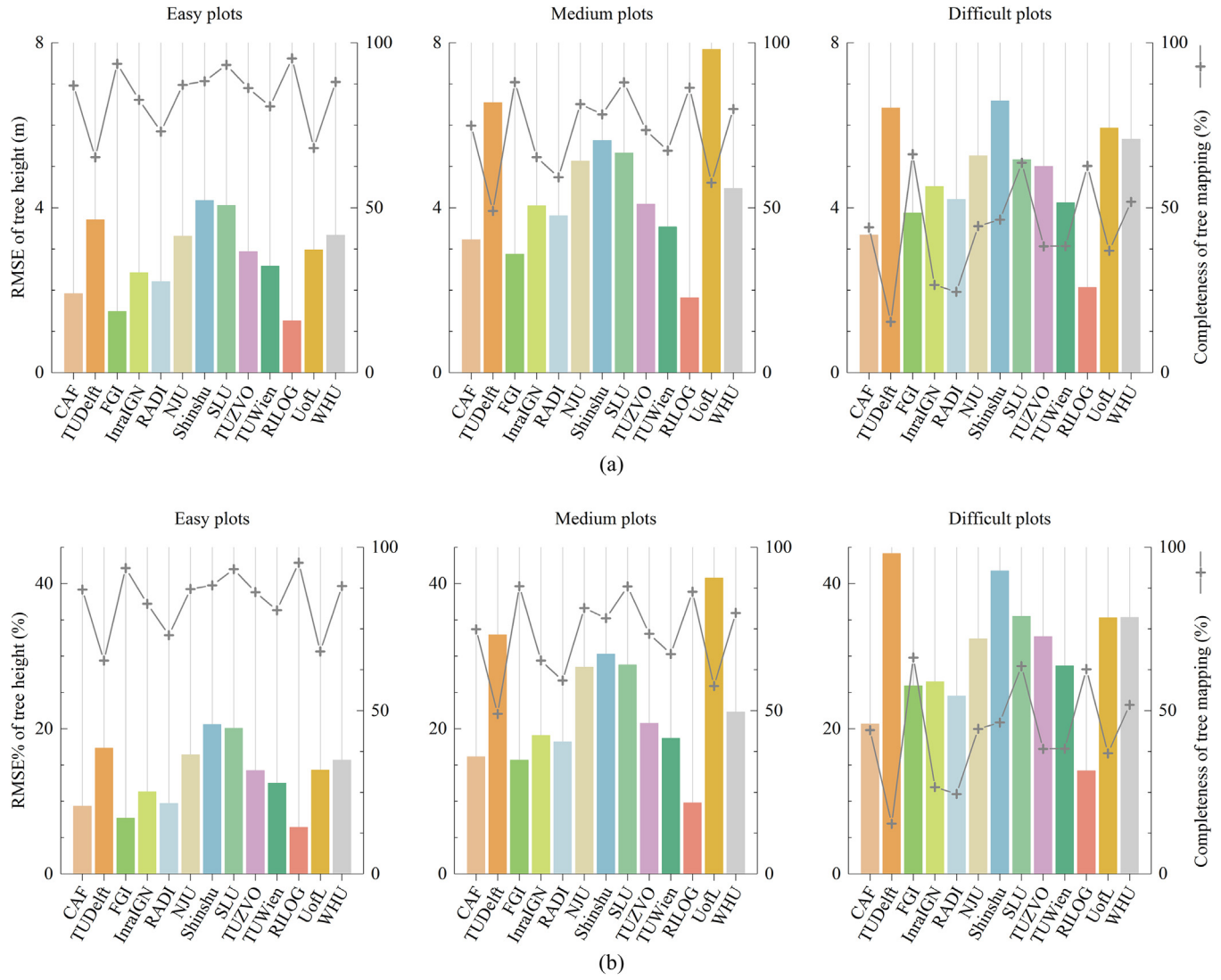


Fig. 20. RMSE (a) and RMSE% (b) of the tree height estimation from multi-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in meters) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

6.6.3. Curve length ratio and the percentage of the tree height covered 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, as shown in Fig. 25. An almost 100% CLR is expected, which indicates that the automated results reach at least 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.

The results in this benchmarking also show that (Fig. 24) 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.

6.7. 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

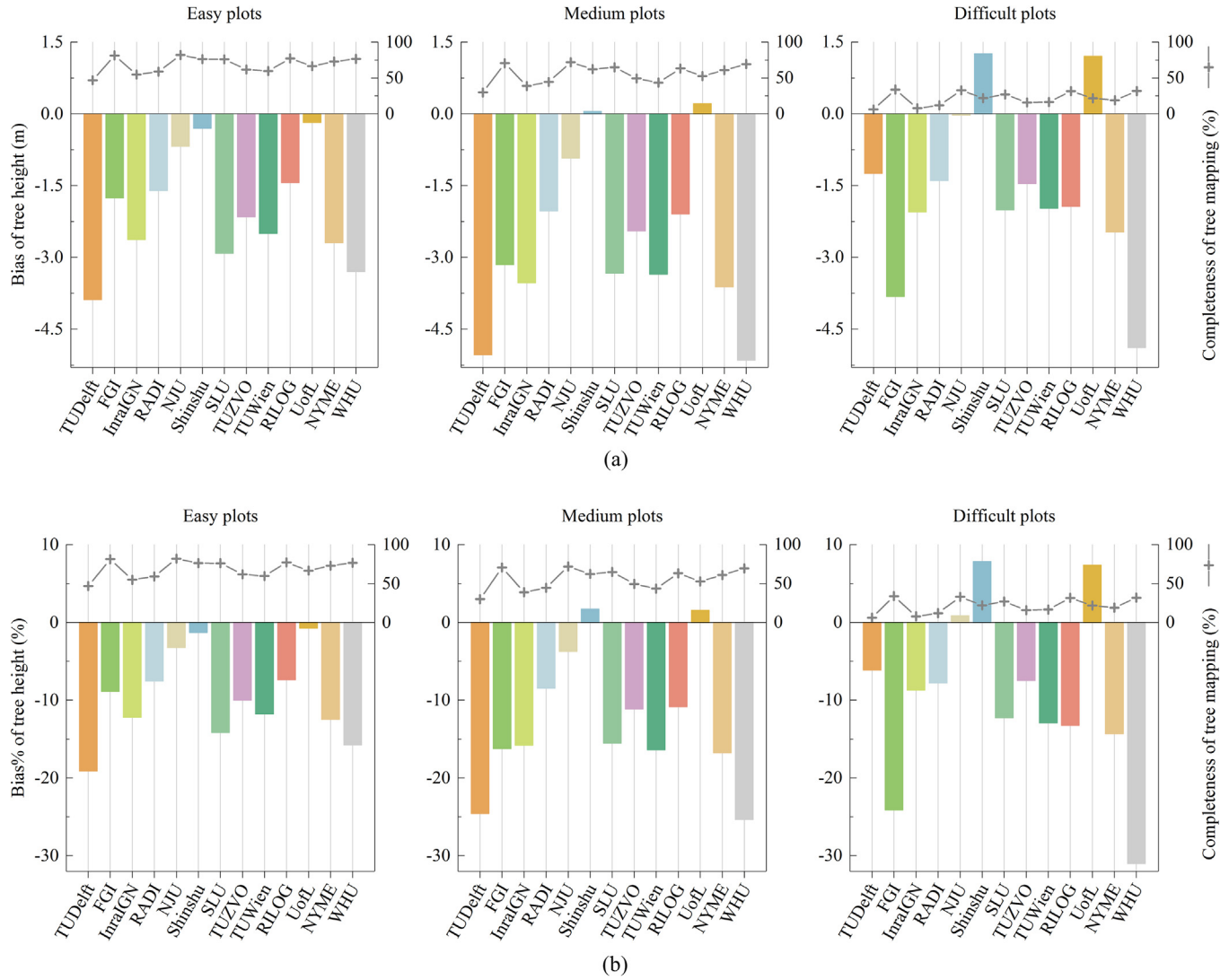


Fig. 21. Bias (a) and bias% (b) of the tree height estimation from single-scan TLS data. The left vertical axes correspond to the bias (bars in (a), the units are in meters) and bias% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

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.

6.7.1. Tree-level stem volume

The tree-level stem volume is calculated utilizing the stem curve and the tree height as described in Section 3.3.4. 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. Figs. 26 and 27 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.

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

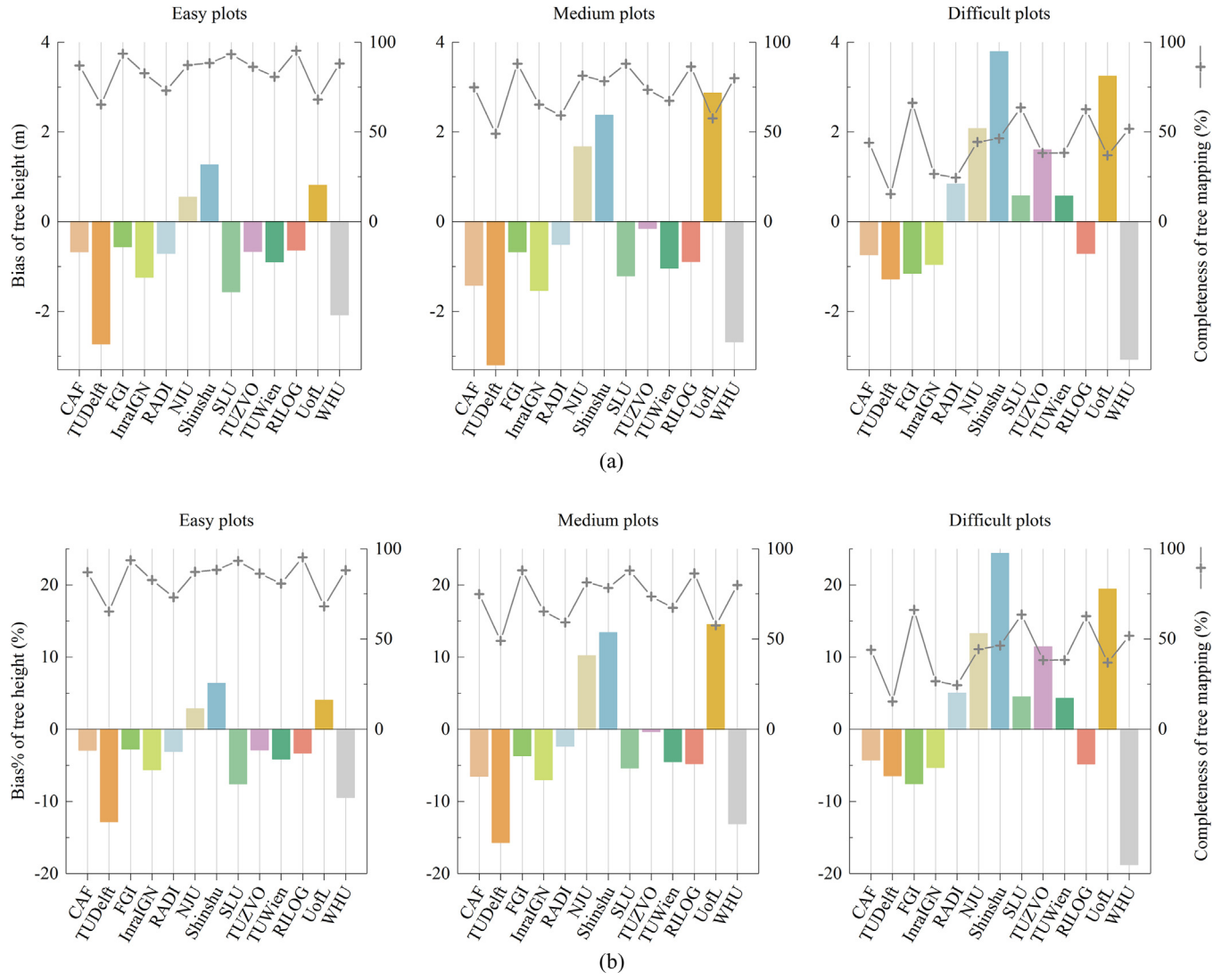


Fig. 22. Bias (a) and bias% (b) of the tree height estimation from multi-scan TLS data. The left vertical axes correspond to the bias (bars in (a), the units are in meters) and bias% (bars in (b); the units are in percentages). The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

difficult when the stand conditions becomes 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.

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 from the RMSEs of the stem volume estimation from sincle- and multi-scan TLS data in Figs. 26 and 27. Although the stem volume is a dependent variable of both 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 is 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.

6.7.2. 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 the plot-level trunk volume ratio from the single- and multi-scan TLS data in Fig. 28, 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

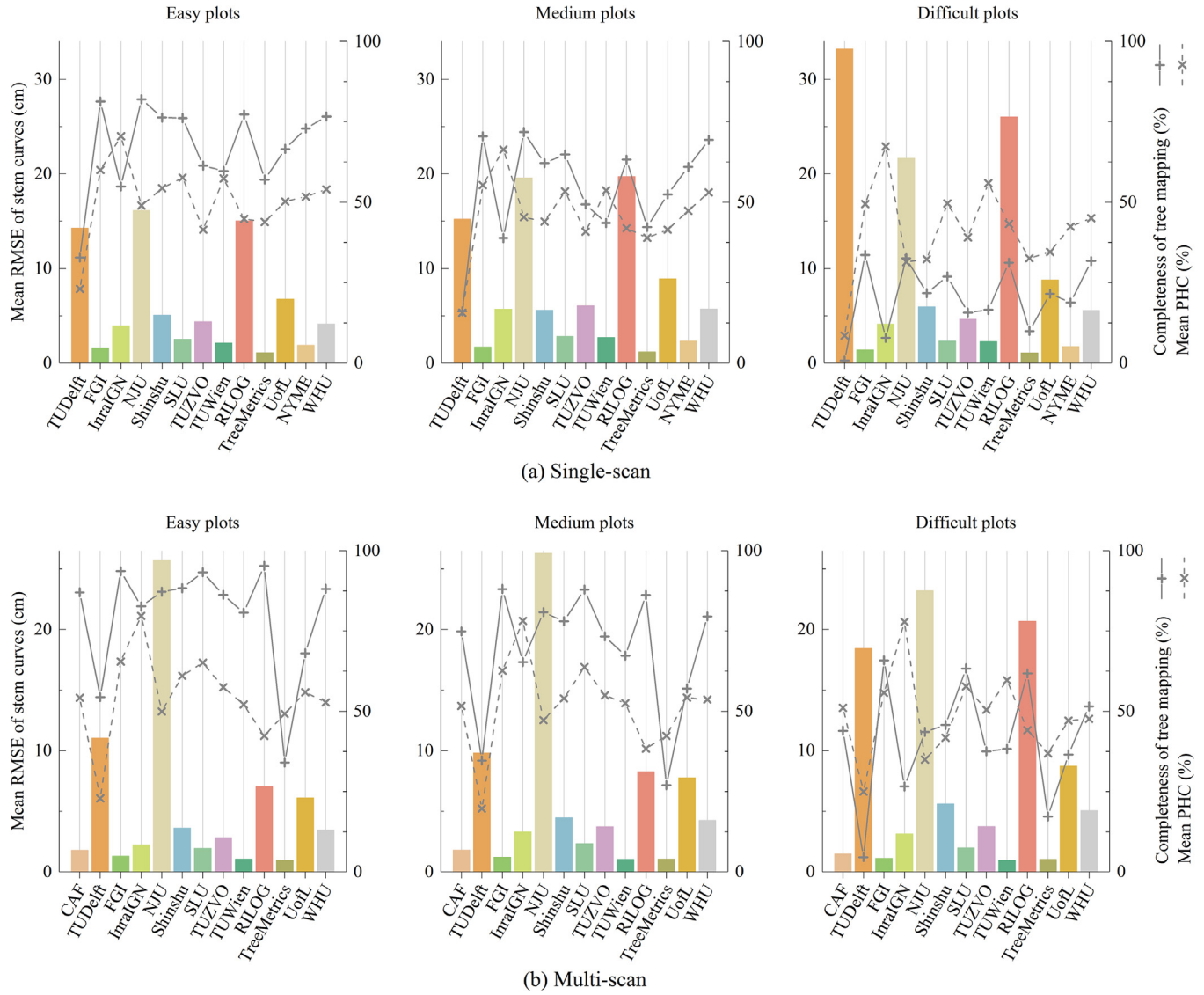


Fig. 23. RMSE of the stem curve estimation from the single- (a) and multi-scan (b) TLS data. The left vertical axes correspond to the RMSE (bars), and the units are in centimeters. The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers) and the mean PHC of stem curve (dashed line with 'x' markers), and the units are in percentages.

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 6.6.2.

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.

6.8. 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.

6.8.1. 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 the RMSE of the biomass estimation from single- an multi-scan TLS data in Figs. 29 and 30; 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.

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

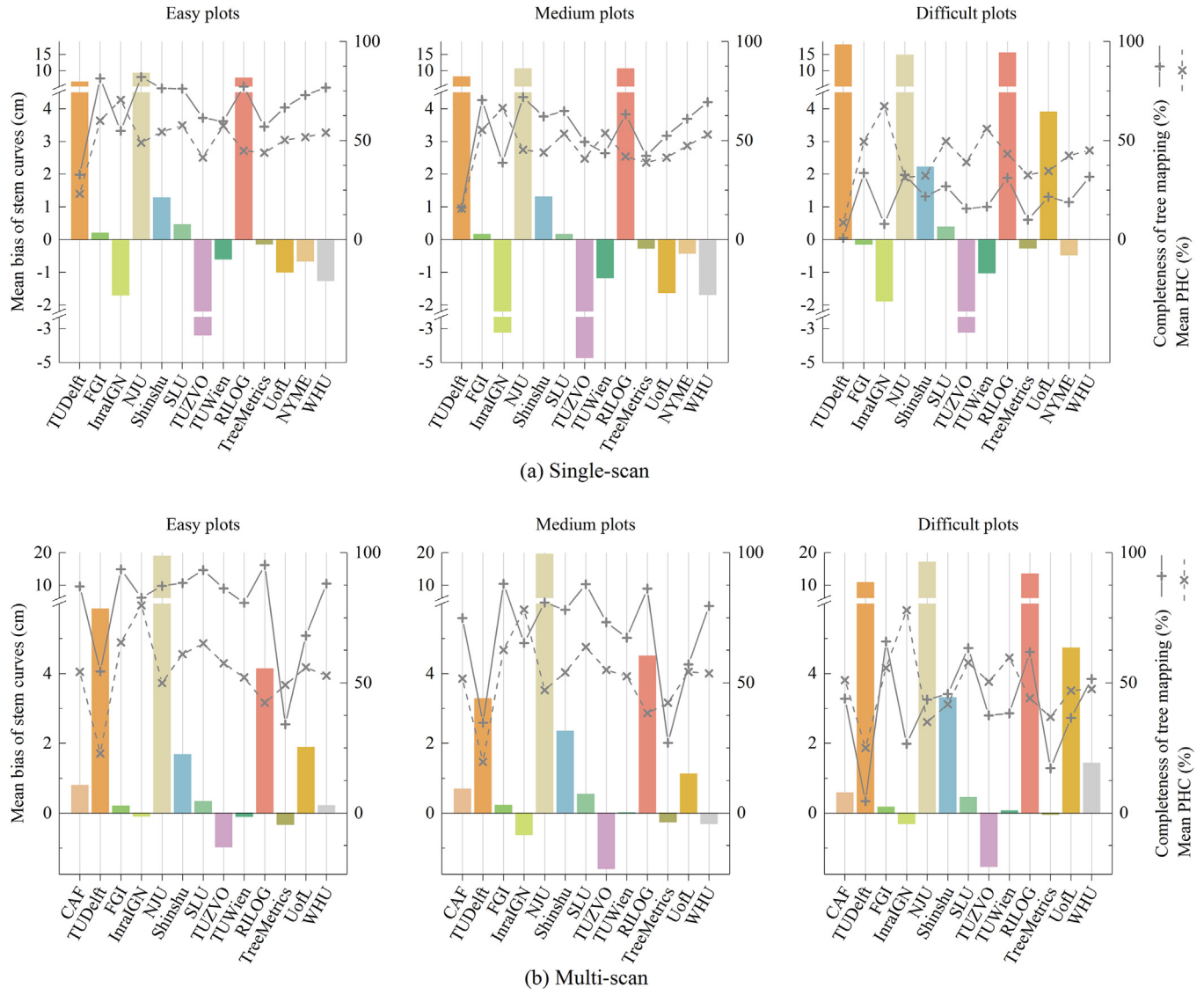


Fig. 24. Bias of the stem curve estimation from the single- (a) and multi-scan (b) TLS data. The left vertical axes correspond to the bias (bars), and the units are in centimeters. The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers) and the mean PHC of stem curve (dashed line with 'x' markers), and the units are in percentages.

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.

6.8.2. *Plot-level biomass estimation*

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 the extracted and all 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. The closer the biomass ratio to 100%, the better the performance of an algorithm is at estimating the plot-level biomass estimation is. In Fig. 31, the biomass ratios are illustrated in combination with the stem detection completeness and correctness. The influence of the stem detection accuracy on the plot-level biomass estimation is clearer when comparing Figs. 30 and 31. 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.

7. 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 methods, 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.

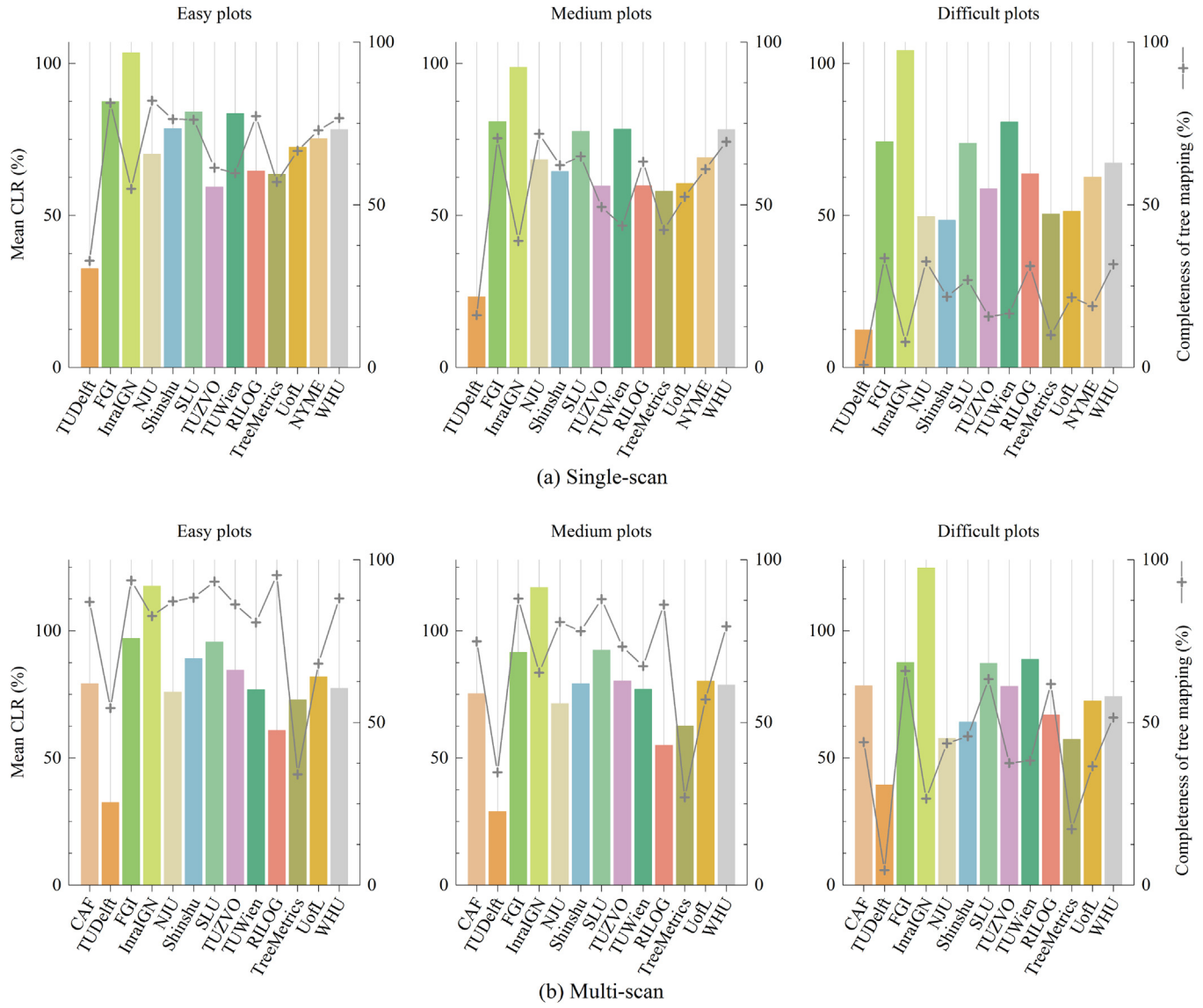


Fig. 25. CLR from the single- (a) and multi-scan (b) TLS data. The left vertical axes correspond to the CLR (bars), and the units are in percentages. The right vertical axes correspond to the completeness of stem detection (solid line with '+' markers), and the units are in percentages.

7.1. 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 methods in this benchmarking. 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.

7.1.1. 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 preprocessing 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 modeling in the following steps.

The art of the method design 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 modeling 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

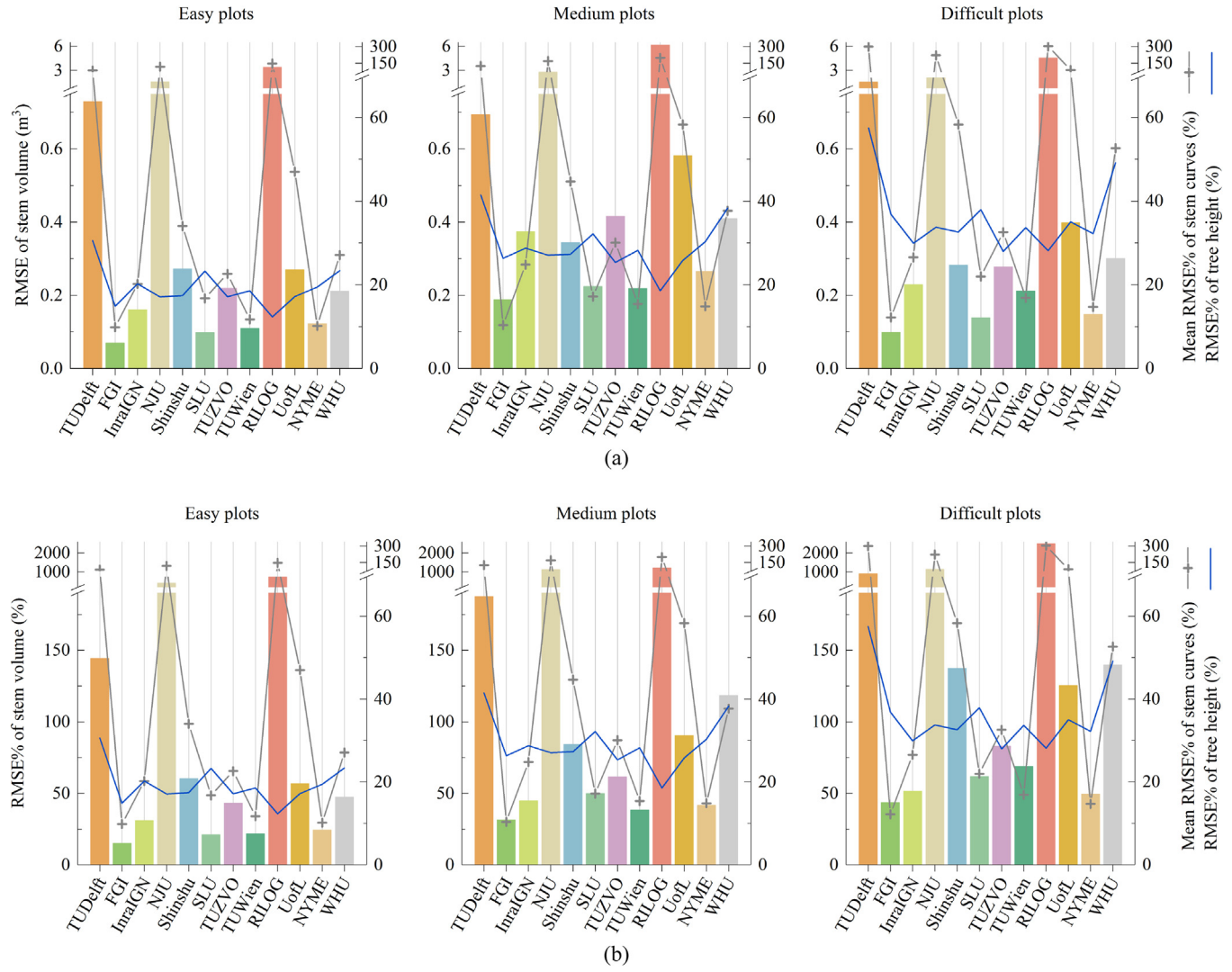


Fig. 26. RMSE (a) and RMSE% (b) of the stem volume estimation from single-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in cubic meters) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the two relative indicators, that is, the RMSE% of stem curve estimation (gray line with '+' markers), and the RMSE% of tree height estimation (blue line), and the units are in percentages.

on these extracted points. In another group of the methods, e.g., CAF and TUZVO, stem detection is achieved by the feature fitting/modeling, 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 modeling results.

A large variety of method designs is observed in the first step, i.e., stem detection, among the algorithms. 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 modeling are similar. A tree stem was modeled 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., stem modeling 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.

7.1.2. Data structure

The most commonly applied data structures for stem detection and modeling 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 detections 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

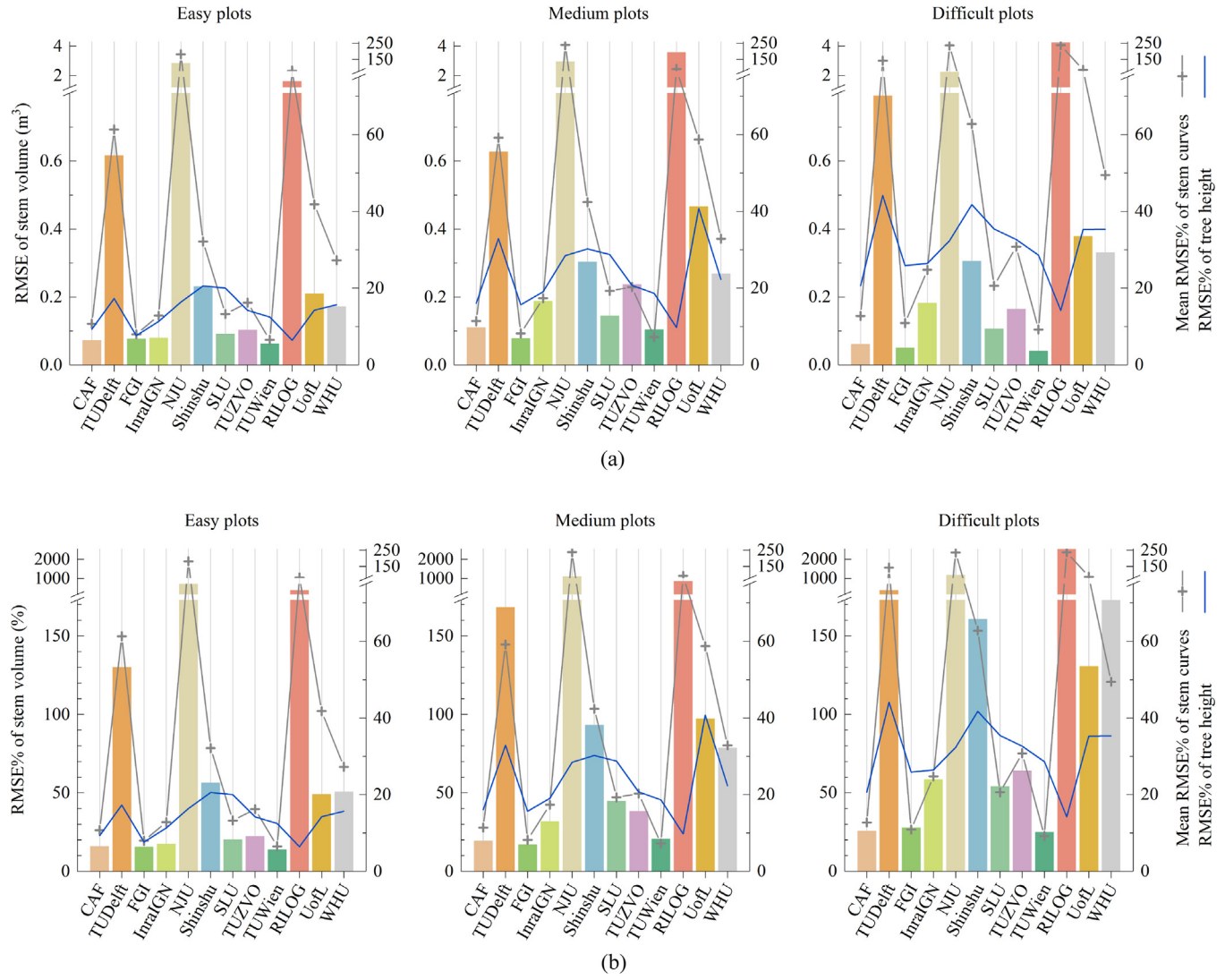


Fig. 27. RMSE (a) and RMSE% (b) of the stem volume estimation from multi-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in cubic meters) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the two relative indicators, that is, the RMSE% of stem curve estimation (gray line with '+' markers), and the RMSE% of tree height estimation (blue line), and the units are in percentages.

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 from 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 preprocessing. The advantage of the point-based data structure is the completeness of the available information, which improves the overall performance of tree detection and modeling. 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.

7.1.3. *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.

Prioritization 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 recorded 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.

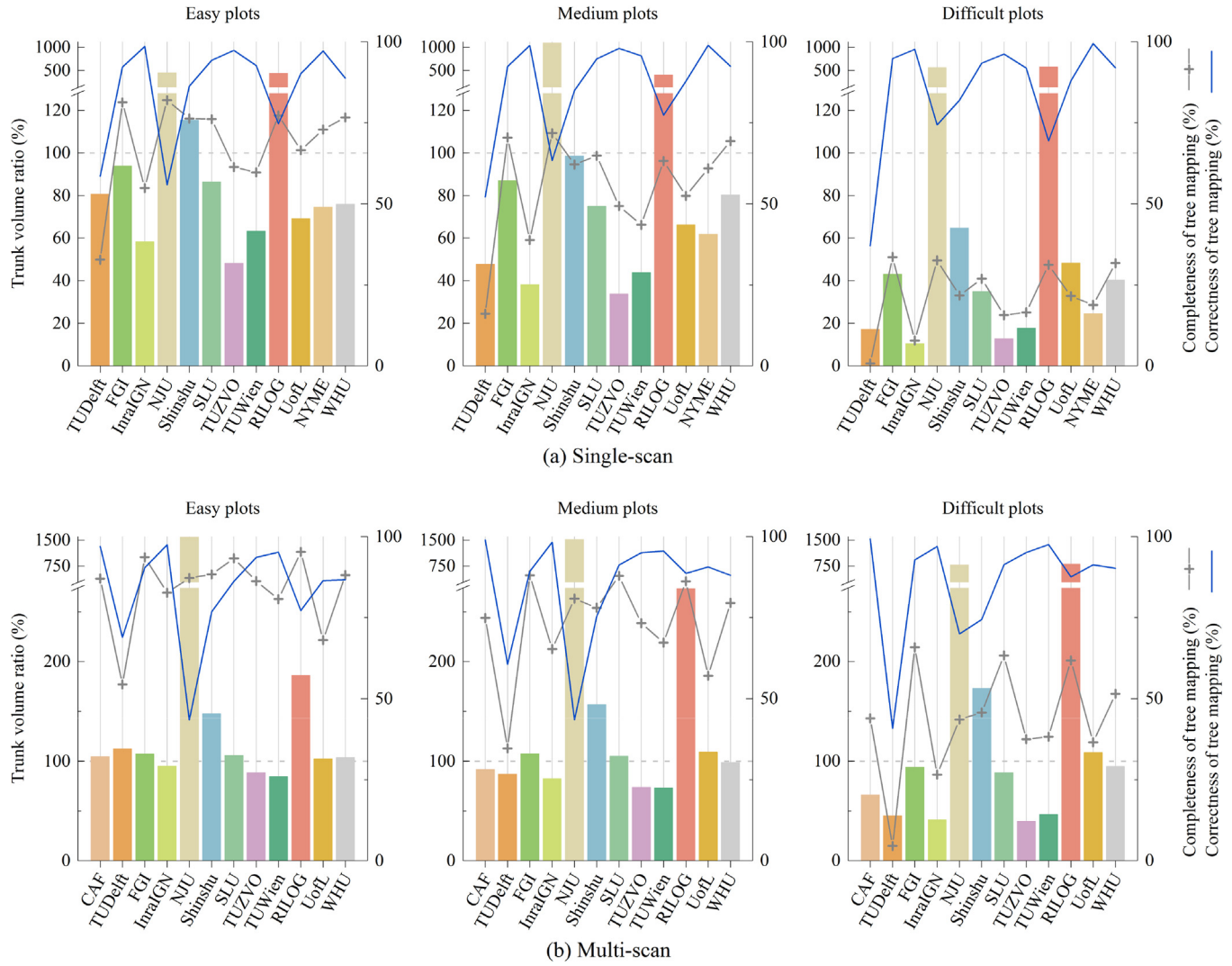


Fig. 28. Plot-level trunk volume ratio from the single- (a) and multi-scan (b) TLS data. The left vertical axes correspond to the trunk volume ratio (bars), and the units are in percentages. The right vertical axes correspond to the two relative indicators, that is, the completeness (gray line with '+' markers) and the correctness (blue line) of stem detection, and the units are in percentages. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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.

7.1.4. 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 modeling 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 modeling 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 modeled segments are checked.

7.2. The milestones of tree attribute extraction

The results of the plot/tree attribute estimation were introduced in the Section 6. 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 are 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

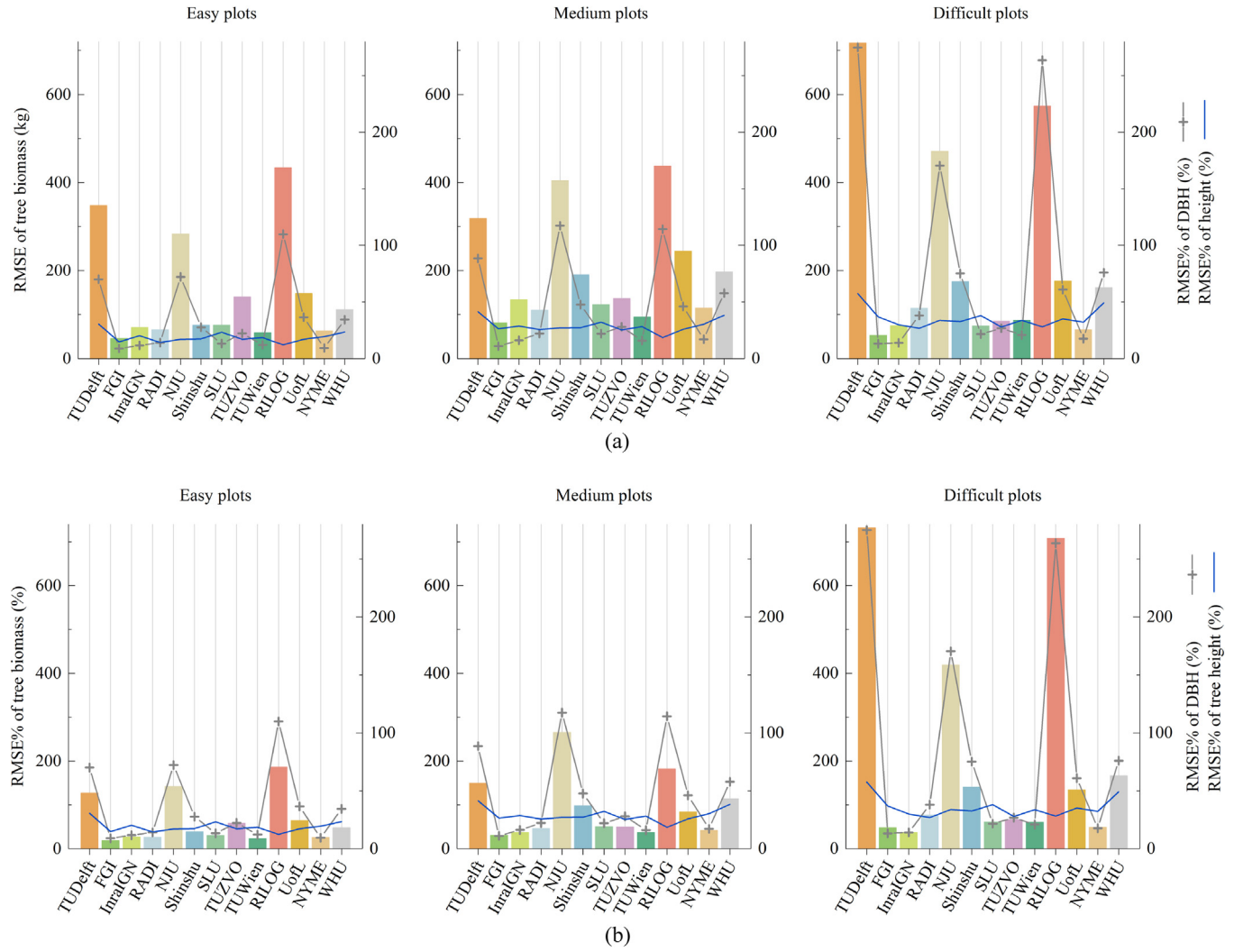


Fig. 29. RMSE (a) and RMSE% (b) of the biomass estimation from single-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in kilograms) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the two relative indicators, that is, the RMSE% of DBH estimation (gray line with '+' markers), and the RMSE% of tree height estimation (blue line), and the units are in percentages. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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.

7.3. 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.

7.3.1. 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

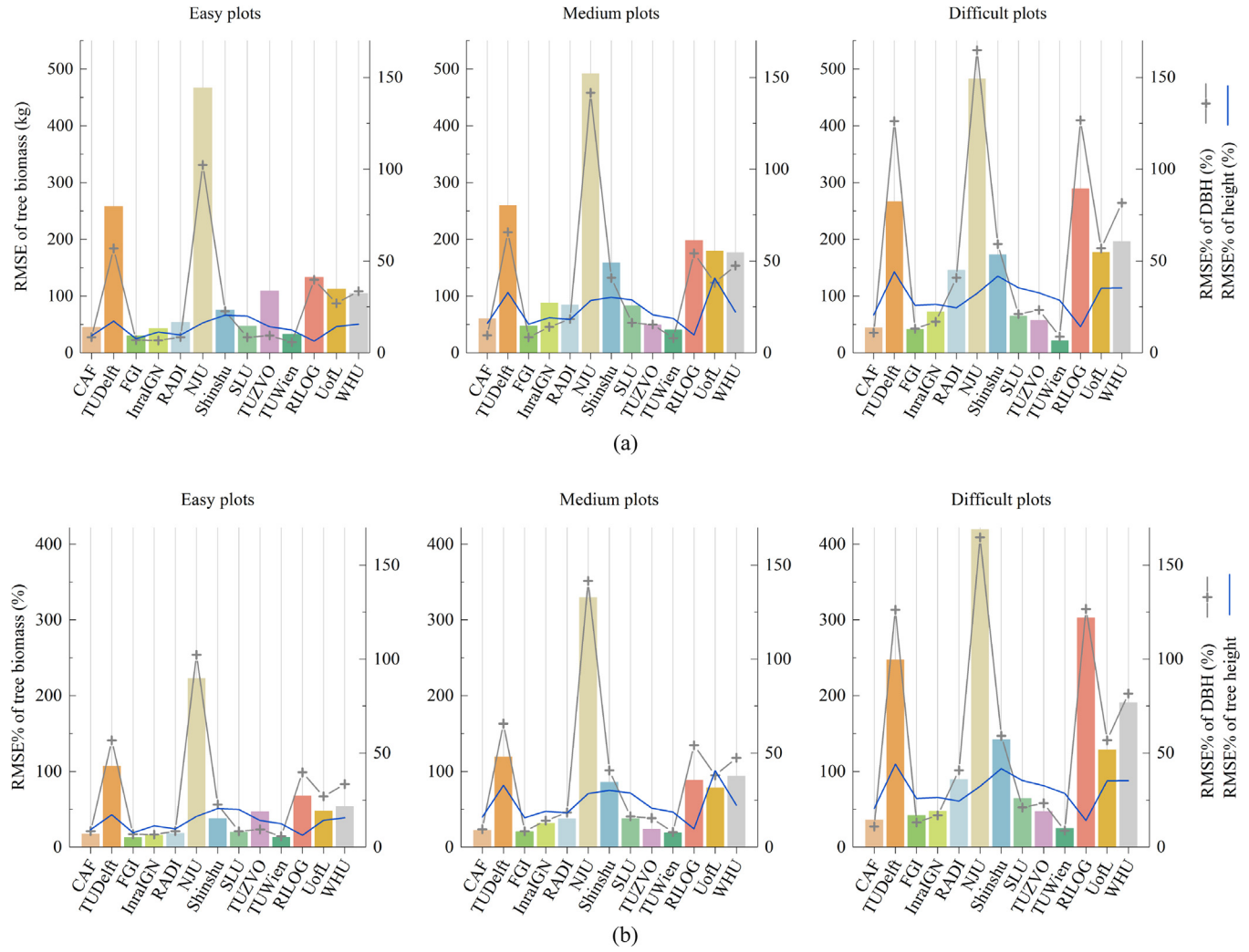


Fig. 30. RMSE (a) and RMSE% (b) of the biomass estimation from multi-scan TLS data. The left vertical axes correspond to the RMSE (bars in (a), the units are in kilograms) and RMSE% (bars in (b); the units are in percentages). The right vertical axes correspond to the two relative indicators, that is, the RMSE% of DBH estimation (gray line with '+' markers), and the RMSE% of tree height estimation (blue line), and the units are in percentages. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 gained by 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 modeling. 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 modeling. 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.

7.3.2. 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 modeling, 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 modeling 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

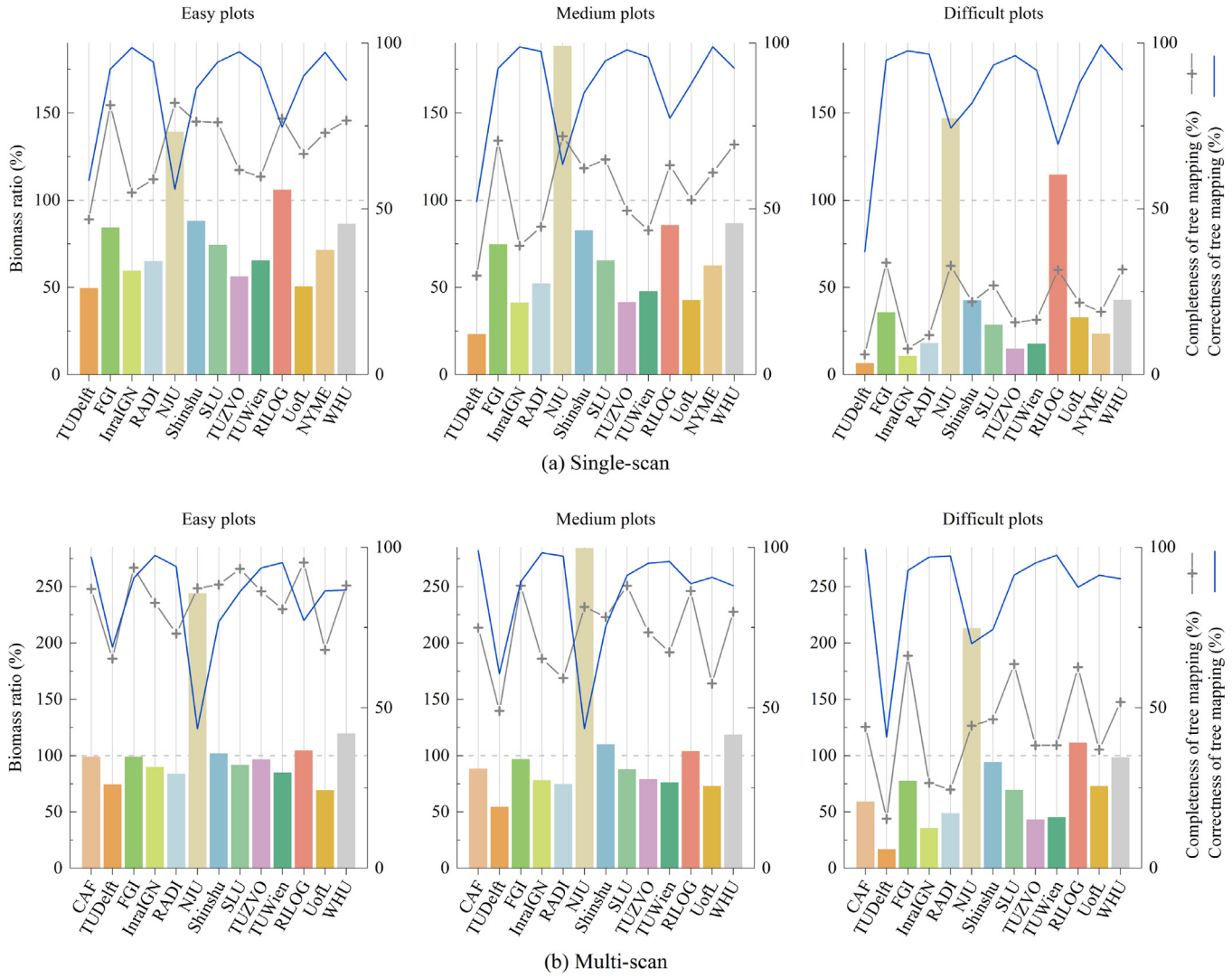


Fig. 31. Plot-level biomass ratio from the single- (a) and multi-scan (b) TLS data. The left vertical axes correspond to the biomass ratio (bars), and the units are in percentages. The right vertical axes correspond to the two relative indicators, that is, the completeness (gray line with '+' markers) and the correctness (blue line) of stem detection, and the units are in percentages. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hand, in terms of the stem modeling 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 are 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 in 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.

8. 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, modeling tree stem as a series of 3D primitives (e.g., overlapped cylinders), and verifying the tree model by the parameters of the modeling 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 (7.1.3) 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 issues, such as 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 a 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.

The benchmarking results indicated that some tree attributes that were previously not measurable using conventional tools, e.g., too costly to measure, have become measurable in practice using TLS, such as the tree position and stem curve. Tree position reveals the forest structure and bridges observations from different perspectives, e.g., terrain and airborne observations (Kaartinen et al., 2012; Wang et al., 2016b; Liu et al., 2017). Stem curve defines the merchantable volume, e.g., (Murphy et al., 2010), and can calibrate the local allometric model, e.g., (Sun et al., 2016). Such results indicate the impacts of TLS in forests not only stay in the plot-level inventory in NFIs, but also lies in many other aspects to be discovered.

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 that, instead of the inaccurate attribute estimates, it is the number of recorded trees that may hinder the practical use of TLS in complex forest conditions. 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 was the only practical tool to collect terrestrial point cloud data 10–15 years ago. Recently, similar point clouds have been available from the structure-from-motion, structured-light and mobile laser scanning, e.g., (Liang et al., 2014a, 2015; Forsman et al., 2016a, 2016b; Hyyppä et al., 2017; Tomaštík et al., 2017; Mokroš et al., 2018). However, the quality of TLS data remains as the best of all terrestrial point clouds due to the fact that the stationary laser ranging is typically very accurate and that the registration errors in TLS are minimized by utilizing the artificial registration targets (Liang et al., 2015). In addition, TLS also has better capability to record the structure inside tree crowns. Therefore, the results of this benchmarking also label a standard for all the terrestrial point clouds.

9. Conclusion

The awareness of the potential impacts of terrestrial laser scanning in forest applications is substantially increasing during the past 20 years in both public and commercial sectors. Even though a great leap forward in the hardware evolution can be witnessed, brought about by the enormously decreased size and weight of the equipment, the continuously advanced data quality, and the considerably dropped hardware costs, the data-processing sophistication has not kept pace with the hardware surge, leading to a very limited application of TLS in practical operations. After 20 years tremendous efforts invested in the relevant studies, it is now time to inspect the achievements and the remaining barriers of the TLS-based forest investigations, so further research and applications receive a clear orientation on their path toward practical TLS operations in forest inventories.

In such a context, the international TLS benchmarking project was launched in 2014. 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 benchmarking data include 24 sample plots selected by foresters and classified into three complexity categories, i.e., easy, medium, and difficult, considering the complexity of forest conditions. The TLS data were collected utilizing both single-scan and multi-scan approaches on each plot, representing the highest field measurement efficiency and the best data quality of the terrestrial point cloud.

The bechmarking project assembled eighteen partners worldwide to participate. Algorithms for TLS-based forest modeling from the partners, including two commercial software, 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 the influences of the TLS data quality and the forest stand complexity on the algorithms were clarified, the impact of algorithm development concepts and principles were more revealed, and a comprehensive understanding on the status quo of the TLS-based forest investigation algorithms was achieved.

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 decreases 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 trees 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 modeling. A higher tolerance for a fragmented and noisy stem structure is required when a higher stem detection rate is pursued. Consequently, stem modeling 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 above-mentioned 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.

A high level of automation is a commonly shared standard among the approaches in this benchmarking, i.e., approximately 80% of the methods are fully automated. The rasterized data formats, i.e., the 2D raster layers and the voxels are the most popular data structures for stem detection and modeling 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.

Three principles are observed from the algorithms in this benchmarking confronting the conflicts between the higher detection rate and more accurate parameter estimation, 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 weaknesses, and the selection of a principle/algorithm depends on the final objective of the applications.

In terms of further algorithm development, a new method is likely to achieve accurate tree attribute estimation by taking a couple of components into the method design, based on the similarities observed in the benchmarked algorithms and the results: a filtering step to reduce noise, a 3D-feature-based individual tree detection, modeling tree stem as a series overlapped cylinders, and verification of the tree model by the parameters of the basic modeling elements. A step to reduce the data volume seems necessary at this moment, but it 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., national forest inventories (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 modeled, 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, structured-light, and the image-based structure from motion, which means that all the milestones labeled 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 bechmarking project. Juha Hyyppä, Harri Kaartinen, Matti Lehtomäki, Jiri Pyörälä, Xiaowei Yu, Liang Chen and Yunsheng Wang are the team in FGI to carry out the project. Norbert Pfeifer is the advisor from EuroSDR. 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, Jinhu Wang, Zhouxi Xi, Bisheng Yang, Guang Zheng from partner groups processed the TLS data and the names are listed in alphabetical order according to the last name. The FGI and University of Helsinki teams prepared the field data. Yunsheng Wang writes the paper jointly with Xinlian Liang.

Acknowledgements

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.

Appendix A

This section summarizes the main components of the method involved in the international TLS benchmarking.

A.1. Method CAS

The processing by the Chinese Academy of Forestry, China, is a multi-layer circle detection algorithm.

The digital terrain model (DTM) was derived using the lasground function in the lastools (Rapidlasso GmbH, Gilching, Germany). Point cloud data were normalized using the derived DTM. The tree detection was based on the 2D Hough transform. The point cloud was sliced to multiple layers where the layer thickness was 10 cm, and the points were projected to the 2D layers. The projected point cloud data were rasterized to a grid of 5 mm resolution for each layer and the raster image is then converted to binary image. The Hough transform method was used to detect circles from the raster images and the circles that fitted outside stem were removed. The DBH was the detected diameter of the stem circle at the DBH height (1.3 m) and the tree location was the circle center. The stem curve was extracted according to the tree location, grow direction and DBH and then smoothed by a mean shift method. After the stem curve extraction, a line was fitted to the circle centers. The tree height was considered as the highest value of return points that lay within buffer zone around the fitted line. The buffer zone was defined as a space with 3 times distance of DBH along the grow direction of the stem.

The method is fully automated. All multi-scan sets were processed with one parameter set.

A.2. Method TUDelft

The processing by the Delft University of Technology, Netherlands, useed a voxel-based algorithm.

The original point clouds were first uniformly down-sampled through an octree with a minimum cell size of 1.0 centimeter. One randomly picked point was selected in each cell.

The ground points were identified by a progressive morphological filtering method (Zhang et al., 2003). Based on the obtained ground points, polynomial interpolation was conducted to determine the grid points of a DTM of a predefined resolution.

Individual tree detection was based on a voxel based approach. First the lower front left and the upper back right points of all the imported tree points were obtained. Then all tree points were traversed and each point was assigned to its corresponding cell in voxel space. The voxel space was then clustered to find connected cells to form multiple component clusters, where more than one object may be present. Consecutively, trees were individualized from those multiple component clusters based on an adjacency analysis of the voxel cells. The individualization started either from the bottom layer upwards or from the top layer downwards. In each layer, region growing was performed based on the seed cells that were inherited from the previous layer. The procedure continued until all the layers were traversed and trees were individualized. Points of voxel cells from the final clusters were output as individual trees. This segmented point cloud, corresponding to an individual tree, was then filtered to remove outliers. In each point's neighborhood defined by a radius of 10 cm, points were assumed to have a Gaussian distribution. Points whose mean distance to the centroid of the neighborhood was larger than an internal defined threshold were considered as outliers and removed from the dataset. The threshold was derived from the global mean distance and standard deviation.

After noise removal, the tree height was calculated from the lowest to the highest point. DBH was estimated from the points of an individualized tree by circle fitting in a 20 cm height bin. Points that deviated most from a fitted circle were considered noise and removed for DBH estimations through RANSAC. The tree location was the center of the circle at breast height. A stem curve was calculated by fitting a circle to the projected trunk points at 20 cm height bins at different heights again using RANSAC. The stem curve was estimated until the end of the trunk.

The method is fully automatic. The parameters were different for different plots. For one specific plot, the parameters were the same for the single- and multi-scan.

A.3. Method FGI

The method from the Finnish Geospatial Research Institute (FGI) is a point-based tree-modelling method (Liang et al., 2018), i.e., an improved version of the one in (Liang et al., 2012).

The original point cloud was thinned through an equivalent sampling, where the original point was thinned while the point distribution was preserved. The 3D point data was digitized into a voxel space and one point in each voxel, who was closest to the center of gravity, was selected to represent point's distribution within the voxel. A digital terrain model (DTM) was reconstructed using the morphological filter and linear interpolation. The 3D points were digitized into a 2D raster space. The lowest point in each pixel was selected as the seed point. And the seed points were clustered in 3D and the largest connected group was understood as part of the ground. Detached groups were accepted as ground if they were smoothly connected with the accepted ground, i.e., the slope between a detached group and the ground was gentle. The DTM was then built through the linear interpolation of the identified ground points

The stem detection and modeling follows the method proposed in (Liang et al., 2012). Points on vertical planar surfaces were first identified by analyzing the structure in its immediate neighborhood using the principal component analysis. Tree stem models were built from the recognized stem points as a series of 3D cylinders representing stem's changing growth directions. The DBH and location of the stem were then estimated from the cylinder element at the breast height (1.3 m above the ground) and the stem curve was estimated from the cylinder element at pre-defined heights.

The tree height was estimated separately for big and small trees, where the tree groups are separated based on the DBH, i.e., 15 cm. Tall trees were assumed to be dominant or co-dominant trees (Wang Y. et al., 2016), where no other trees crown above. The tree-height estimate was the elevation difference between the highest and lowest points around the stem location. Small trees are typically shadowed by other trees in the near vicinity. In order to find the treetop of a small tree, points around the stem model were clustered and the largest group was assumed to belong to the tree. The tree height estimate was the elevation difference between the highest point in the group and the DTM beneath.

The method is fully automated. All single- and multi-scans were processed with one parameter set.

A.4. Method RADI

The method from the Institute of Remote Sensing and Digital Earth, China, combines voxel- and point-based processing and the DTM was built iteratively based on triangulated irregular network. In each iteration, a grid was created to record the absolute minimum height within a cell, and the TIN was created with those selected minimum heights. The distances between the remaining point clouds with the TIN were calculated. A laser point was determined as terrain when the distance was smaller than a threshold according to scale, and the threshold used here was half of scale. The resolutions of the grid in iteration was 4 m, 2 m, 1 m and 0.5 m.

The non-ground points were used for the tree detection. The point cloud was transferred into a voxel space with the resolution 0.2 m. The vertical column voxels were from tree stem if all voxel elements had at least 1 point. Individual tree stems were segmented from the selected points using a clustering algorithm based on Euclidean distance. The stem curve was estimated in horizontal slices using a circle fitting algorithm combined with the Hough transform to remove noise laser points in the horizontal slice. The resolution of rasterized point cloud was 0.01 m. The DBH and tree position were the diameter and center of the circle at the DBH height. The tree height was the difference between the minimum and maximum z values in a buffer (1 m) around the tree position. For more details, please refer (Huang et al., 2011).

The method is fully automatic. The parameters were the same for all plots and all scans.

A.5. Method KU

The processing by the Korea University, South Korea, consists of a semi-automated DTM generation and manual tree attributes estimation.

The DTM was built using ArcGIS software (ESRI, Redlands, USA). The point cloud data were converted to a raster file with a 0.05 m resolution, where the pixel value was the minimum z value. Data from trees were semi-automatically eliminated, and the rest points were grouped. Pixels who had very high z-coordinate and curvature values were identified and deleted, where the curvature was a difference value between neighboring pixels. The results were interpolated using inversedistance weighting (IDW) and pixels with big differences between predicted and observed values were removed.

To measure tree locations and DBHs, a slice with 10 cm thickness at the DBH height was cut from the original data. Curvature were built to emphasize the drastic increase of the point density. Trees were identified in the density map in the slice as a circular or semicircular cluster. Three points of the circular shape were selected to build a complete circle and to estimate the tree position and DBH. To minimize the detection errors, trees that were identified in each plot by all 4 persons were selected as a tree detection.

The method is semi-automated for the DTM generation and manual for the tree detection. The parameters are the same for the single-scan of all plots.

A.6. Method NJU

The processing by the Nanjing University, China, is a classification based tree detection method.

The original point clouds were sampled to improve processing efficiency. A random point was picked as a seed point and all other points falling within a 5 cm searching sphere were deleted. After repeating this procedure for all the rest points, the distance between two neighboring points in the sampled point cloud was at least 5 cm.

Geometrical-based Automatic Forest Points Classification (GAFPC) algorithm (Ma et al., 2016) was used to classify the forest point cloud, which was based on the saliency feature and Gaussian mixture models (GMM). Training samples of different classes were selected before classification, and training samples were different between the single-scan plots and multi-scan plots due to the different patterns of local geometrical features. The point data were classified into scatter class (i.e., photosynthetic canopy components), linear class (i.e., non-photosynthetic canopy components) and surface class (i.e., ground). Saliency features were calculated using points in a searching radius defined in each plot from visual inspection, which corresponded to the maximum DBH and 1.5 times of that value in the single- and multi-scan, respectively.

The surface class points were used to produce the DTM using the IDW interpolation method. The points in the linear class were normalized based on the DTM to remove the topographic effects, and filtered to remove the noisy and branch points. The filter has two searching cones (i.e. up- and down-facing) with the searching radius and cone apex angle as 0.5 m and 10°, and the point will be kept if the number of points within either of the two searching cones is larger than 8 and at least one point from this cone whose distance to the point is larger than 0.45 m. To detect individual tree stems, a random point was first picked as a seed point to build a searching circle with the radius of 1 m in XY plane for finding the linear class points from a tree. The stem curve was calculated in slices with 1 m thickness ($\pm$ 0.5 m). Points in the slice were projected onto a horizontal plane and the diameter was estimated as the maximum distance between any two points. The DBH of the tree was the average of all diameters. The tree height was the local maximium height within a cylinder volume with the stem location as center point and 1 m as radius.

The process is fully automatic. The parameters were different for different plots and scans. Small adjustments were made according to factors such as topographic variations, stem density.

A.7. Method Shinshu

The processing by the Shinshu University, Japan, was carried out using a commercial software, DigiForet, developed by WoodInfo (Tokyo, Japana).

Functionality of the software is to automatically extract the trees from TLS measurements based on predefined parameters which determine the criteria for tree extraction. Details of the algorithm have not been published. Major steps of the algorithm include: (1) DTM creation; (2) tree stem extraction and (3) estimation of DBH, tree height and stem curve. In this study, the parameter settings in each step were as follows: in the DTM creation, surveyed area was divided into 2 m $\times$ 2 m grid and the minimum elevation value within the grid was taken. If grid normal was greater than 30 degrees, the point was considered a branche point and was removed from DTM, and then interpolated from neighboring points. In the tree stem detection, 2 m $\times$ 2 m grid was further divided into 5 cm $\times$ 5 cm grid. In each sub-grid, the point numbers within 10 cm range at 2 m, 2.5 m, 3 m, 3.5 m, 4.5 m and 6 m height were retrieved. If the point number was greater than 4, the sub-grid was supposed to belong to a tree stem. In the DBH, tree height and stem curve estimation, the maximum diameter of the tree was set to 1 m; the maximum tree height to 40 m; beast height to 1.3 m; the minimum distance between trees to 0.5 m.

The process is fully automatic. All single- and multi-scans were processed with one parameter set.

A.8. Method SLU

The processing by the Swedish University of Agricultural Sciences is a voxel-based method to search, model and link cylindrical objects to detect tree stems and to build stem models.

The DTM was extracted based on the idea of using minimum height rasters of different resolutions to find the overall shape of the ground and then to refine the detailed shape of the ground. The raster cell sizes in this investigation was chosen to be 2 m and 0.5 m respectively. For a detailed description of the algorithm see (Olofsson et al., 2014; Lindberg et al., 2012).

The tree extraction algorithm is a voxel-based solution. The point cloud data were transferred into a voxel space with a resolution of 3 cm. In each voxel element, a surface patch was identified if the flatness value was larger than 0.8. If several patches had the same radius of curvature from the same origin they were assumed to be part of the same curved object and nearly cylindrical objects were assumed to be tree stems or large branches. For details of the algorithm see (Olofsson and Holmgren, 2016).

The detected curve-objects were segmented in sections, i.e., half meter, and each section was modelled as a cylinder, and was accepted as part of a stem if it was almost vertical (between 0-45 degrees zenith angle). Single cylinders were assumed to be noise and removed from further calculations. Cylinders positioned above each other with a direction difference less than 20 degrees were assumed to be part of the same tree stem and were connected as stem segments. Stems were supposed to be nearly vertical passing the height span of 1 to 2 meters or with a length of more than 3 meters. The connected vertical cylinders were used to calculate the stem curve and DBH of each stem using interpolation and a ground model.

To estimate the tree height, the canopy height model (CHM) was created with a resolution of 0.5 m and with the cell value of the 99.9 percentile height value. The CHM values for each cylinder's *XY*-position of a stem were searched and the highest canopy height was chosen as the tree top. The tree height was estimated as the difference between the ground height at the cylinder positioned at the root and the top value of the tree.

The algorithm used a curvature weight and a cluster weight described in (Olofsson and Holmgren, 2016). The curvature weight was set to 0.8 and the cluster weight was set to 10 in this study. All single- and multi-scan sets were processed with one parameter set.

The method is fully automatic. To be able to run several field plots simultaneously, the processing for this study was conducted using the resources of High Performance Computing Center North (HPC2N). However, the processing can be also run on a standard office computer.

A.9. Method IFP-LSIS

The processing by the Institut Français de Pondichéry – Laboratoire des Sciences de l'Information et des Systèmes, India/France, is an automated DTM generation algorithm, combining simultaneously ground point classification, DTM reconstruction and hole filling.

A quadtree construction guided by the local density of local minima was used to provide a dynamic multi-scale approximation of the ground surface, using compactly supported Wendland's radial basis functions (CSRBF). The reconstruction was further refined by identifying and correcting approximations in each cell of the quadtree. A global implicit model was then constructed by polygonising the local approximations by means of CRBF, from which a DTM was constructed using a triangular irregular network approach. At the end, the local implicit patches were merged into a global implicit function using Wendland's CSRBF, and a DTM surface is computed through a polygonization of its zero level set.

The algorithm is fully automatic. The algorithm was initiated using local minima selected according to a user defined grid resolution. For 33 out of the 48 TLS data, the grid size was the same (20 cm); and for the rest 15, because of the low point densities in single-scans or complex topographical figures, the grid sizes varied from 40 to 100 cm.

A.10. Method INRA-IGN

The processing by the INRA Biogéochimie des Ecosystèmes Forestiers – IGN Laboratoire d'Inventaire Forestier (INRA-IGN), France, was performed using the open source SimpleTree plugin to the Computree software developed by INRA-IGN (Office national des forêts, ONF).

The DTM generation is a multi-scale approach inspired by (Kraus and Pfeifer, 1998), i.e., top-down scales at 3.2 m, 1.6m and 0.8m, respectively. In each scale, point cloud was rasterized. The lowest points in pixels, i.e., in a 50 cm buffer above the lowest point in the coarsest scale, and the points in the range of ± 1 m of the previous plane in other scales, were used to fit a plane using RANSAC. The DTM values were computed from the plane equations and pixel coordinates. The final DTM with a resolution of 20 cm was derived by IDW interpolation of the finest scale.

To detect individual trees, a horizontal cross-section with a 60 cm thickness was sliced at the DBH height from the normalized point cloud. Points were spatially clustered and the detected clusters served as seeds to segment the point cloud into tree and undergrowth clusters. The path lengths of points to all seeds were computed by applying Dijkstra's algorithm to all vegetation points above the DTM (Dijkstra, 1959) and the point was associated to the seed with the minimum path length.

From visual inspection on the segment results on the training plot (No. 1), we decided to only model clouds with a height larger 8 m. Less harsh thresholds would presumingly resulted in a higher tree detection rate but with worse diameter modelling results. The remaining individual clouds were de-noised using filters described in (Hackenberg et al., 2015), e.g., the radius outlier filter, the voxel-grid filter and the stem/major branch point detection. A cylindrical buffer was built to remove needle hits since fully automated filtering of needle or leaf hits cannot achieve satisfying results yet, where the buffer was located at the tree position and along the *z*-axis with an 80 cm radius. The preliminary tree position was estimated by the center of mass of the according seed cluster.

In each tree cluster, the sphere following approach was applied to model the complete trees and later extracting the stems. The method fits circles first on search spheres located on the tree skeleton. Two neighboring circles are connected to a preliminary cylinder which is later refined by a 3d cylinder fitting routine. Tree position, DBH and stem curve are extracted from the resulting tree model. The tree height was computed as the maximum height in the normalized point cloud of each segmented tree, from the point data before the described filtering operations have been applied. For more details of the method, readers are referred to (Hackenberg et al., 2014, 2015).

The processing is fully automatic. All single scan and multi scan sets were processed with one parameter set.

A. 11. Method TUZVO

The processing by the Technical University in Zvolen, Slovakia, was done by the software DendroCloud (gis.tuzvo.sk/dendrocloud/), developed by Milan Koreň at the Technical University in Zvolen.

To build the DTM, the grid with a 0.2 m resolution was first created, and the point with minimal *z*-value within each cell was extracted. DTM construction method is described in (Koreň et al., 2015). The extracted DTM points were interpolated by natural neighbor interpolation after removing isolated points within ArcGIS for Desktop 10.2 (www.esri.com).

For the tree parameter estimation, a DTM was created with a 0.5 m grid size, where the grid value was the minimum *z*-value of points within the cell. The grid size was set to be greater than the largest tree diameter on the plot. Horizontal cross-sections with a 5 cm thickness were extracted for all given heights above the ground. For all datasets, points within cross-sections that do not represent stems were manually filtered out, and then the points in cross-sections were grouped. Tree positions and diameters were estimated using a least-square circle-fitting algorithm (Koreň et al., 2017). Tree diameters were calculated in horizontal cross-sections at all heights above the ground. The stem curve was constructed by connecting cross-sections of each single tree; the cross-sections at other heights were assigned to the nearest cross-section at the breast height. Position of a tree was defined by the circle center in the cross-section at the breast height. Tree height was calculated as the difference between DSM and DTM at the tree position.

The processing is semi-automatic. Each scan was processed using different parameters.

A.12. Method TUWien

The processing by the Technische Universität Wien, Austria, combines 2D/3D tree detections and 3D stem reconstruction.

The DTM was derived from a hierarchical approach (Pfeifer and Mandlburger, 2008). The robust filtering approach was originally developed for DTM extraction from airborne laser scanning data (Kraus and Pfeifer, 1998). One main assumption is that terrain points are located at the lowest. Starting from a thinned point cloud, the lowest points within 4×4 m^2 raster cells were selected and a robust moving plane interpolation was applied. The gaps were filled using a triangulated model. The interpolation is applied hierarchically for 1.5 m, 0.5 m and 0.2 m raster resolutions whereas the points are only selected around a buffer zone ($\pm$ 0.3 m) of the former elevation model. The point cloud was normalized using the DTM generated.

To detect tree stems, the point cloud was firstly filtered by normal vectors (Liang et al., 2012). Then, the point cloud data were rasterized with a resolution of 2 cm on a horizontal plane and converted to a binary image. A fuzzy classification method was applied to find pixels having large point density that corresponded to tree stems. We empirically set the class number to 4 for all plots. However, sensitivity tests with a different number of classes in fuzzy classification seem in general advisable. The pixel with the largest point density was selected as the seed, and a region growing method was applied to group the pixels. Then, all groups whose projected centers of gravity were close enough were merged as tree stems.

Stem model was built by inscribing a series of cylinders along the tree stem, as described in (Wang D. et al., 2016a). A stem was divided into 0.5 m sections. The section containing the largest amount of points was selected as the starting section and was inscribed with a cylinder by the RANSAC method. Afterwards, the cylinder started to stretch upwards and downwards. In the next stretched section, various candidate cylinders were tested by adjusting the cylinder radius and axis. The one containing most inlier points was then chosen. We further applied a moving window approach to smooth the connection parts of successive cylinders for each stem. Position, DBH and stem curve were retrieved from individual fitted stems. The tree height was determined simply based on the extracted location of each tree. A square centered by the stem location was formed, and the highest point within such a square was regarded as the top of the tree.

The deployed processing chain is fully automated. All single- and multi-scan sets of all plots were processed with one parameter set.

A.13. Method RILOG

The processing by the Silva Tarouca Research Institute for Landscape and Ornamental Gardening, Czech Republic was done by the software 3D Forest (www.3dforest.eu), developed in the same institute.

The terrain and vegetation were firstly segmented automatically. This step used octree search with smallest leaf size 5 cm. The lowest leaves and their upper neighbors (in maximal distance 5 cm), if present, were selected as terrain points. The rest points were classified as vegetation points. Remaining vegetation points from the classified terrain points were visually inspected and merged with the classified vegetation points to build the vegetation point cloud. Terrain points were transformed into voxels of 20 cm size, where the lowest points were selected as terrain and used as a seed points in an inverse distance weighted (IDW) interpolation to generate DTM.

Individual trees were manually segmented from the vegetation cloud. For each individual, tree parameters were automatically calculated. Tree position was initially estimated by the median *XY* coordinates of all points from the tree bottom, i.e., from the lowest tree point up to 50 cm above it, and then refined by the mean of 5 terrain points closest to the initial tree position. DBH was estimated in a 2D slice at the breast height, i.e., 1.25–1.35 m above tree position. Points were projected onto a horizontal plane to estimate circle parameters (position and radius) by the Randomized Hough transform, where the minimum number of points was 5 in order to compute a circle successfully. The method calculated tree parameters iteratively (maximum 200 iterations in this study). In each iteration, three points were randomly selected to fit a circle and results were saved in an accumulator. The circle parameter with the most votes were selected as the final result. Stem curve was estimated as circles at given heights in the same way as the DBH, with the exception that the slice spanned $\pm$ 3.5 cm around the estimated heights. The computation stopped when estimated circle was greater than any two previous computed circles. Tree height was the vertical distance between the tree position and highest point of tree.

The processing is semiautomatic. The process uses automated method to segment vegetation and terrain with manual adjustment, manual method to segment vegetation cloud into individual trees and automatically estimated tree parameters. All single scan and multi scan sets were processed with one parameter set.

A.14. Method treemetrics

The processing by the Treemetrics, Ireland, was done by the software AutoStem developed by the Treemetrics. AutoStem analyzes the TLS data and additional field data to identify individual trees from the raw scanner point cloud, creating 3D tree stem models and extracting parameters. The analysis is optimized for the single scans. To improve the accuracy in multi-scan TLS data, further research will be required.

The processing were focused on a plot area with an 18m radius, which corresponds to the maximum distance typically used in the AutoStem analysis. All other data in the point cloud was omitted in the processing. By default, Autostem omits isolated points and removes obvious noise and enhancing solids objects.

To build the DTM, Autostem first removed ghost points below the ground level using a density allocation along the *z*-axis in each 20cm DTM grid cell. The lowest remaining point inside each grid define the DTM center points and these points were used to calculate an adjusted plane DTM (Kampmann and Renner, 2004).

An advanced filter was used to remove foliage and branches. Objects were detected in the point cloud by isolating high point density regions, and the shape and direction of these objects were analyzed. Smooth curved surfaces distinguished stems and branches from foliage and detritus which were made up of amorphous point clusters. Size and axial direction then distinguished branches from main stems.

AutoStem detected stems in a slice with a thickness of 0.1 m at the breast height. An initial tree object was assumed where point clusters with sufficiently high density were detected. These objects were discarded if the point cluster did not have an approximate 160 degrees circular arc. Analyses were then performed on the sub-section of the point cloud in a cylinder surrounding this initial tree position. The stem curve was estimated at 10cm intervals from the ground to the stem top. At each height interval, points were projected onto a horizontal plane and a structure element moved over the slice to search point clusters with more than *n* points. *n* is the density threshold calculated from the distance from the scan origin and the scan resolution. A circle was fitted to the point cloud. The DBH was the diameter at 1.3 m height.

Several rules were applied to the stem model in Autostem: (1) The diameter of the stem must never increase as the stem ascend in the *Z* direction. The model in one disc was deleted if the fitted diameter was bigger than the lower one. (2) Each disc must be formed by a cluster of points with approximate 160 degrees circular arc. (3) Based on these parameters and the density of points in the arc, a reliability value is assigned to each disc (based on analysis of arc length, point density *v* expected resolution, circularity etc.) Discs with a reliability below the required threshold were excluded from the stem profile. A 70%

threshold was used in this analysis. If a disc was removed or if there was an insufficient number of points to assign a disc at a specific height, the diameter of the disc was interpolated using discs above and below.

Treemetrics system usually implements a tree height and taper modelling system to complete the upper sections of the stems. The tree height modelling is based on dbh/height regression analysis using manual data collected supplementary to the scan. In this test no taper was applied and no tree height was calculated.

The processing is fully automated. All single- and multi-scan sets were processed with one parameter set.

A.15. Method UofL

The processing by the University of Lethbridge, Canada, was done by a 3-D directional filter, region growing, and circle fitting. An improved version of this method can be found in (Xi et al., 2016).

DTM was generated using a piecewise inverse-distance-weighted algorithm. The above-ground points were extracted after the DTM buffer (0.1 m) was removed from the whole point clouds.

To extract stem locations, the 25% top and 25% bottom of the point clouds were discarded first. The remaining point clouds were converted to voxel space with 0.1 m resolution and point counts as the voxel value. The voxel space was convoluted with a 3-D directional filter to enhance the vertical pattern. The size of the filter was $3 \times 3 \times 3$ voxels and the center and edge column was 1 and -1, respectively. The convoluted values in each voxel were then projected to a 2D horizontal plane, where the local extrema gave the potential stem locations. Stem range was defined by Thiessen polygon around the stem and the point clouds for individual stems were thus coarsely extracted from the above-ground point clouds.

Based on each stem point clouds, the tree height was obtained as the vertical distance between the highest and lowest points. To extract the curve, the point cloud of an individual tree was segmented into cylinder-shaped point regions using region growing method. The difference from Xi et al. (2016) was that region growing step became 5 cm and that region growing terminated when region length exceeded 0.3 m for acceleration purposes. Then segment points were projected to a 2-D plane perpendicular to the stem direction, which was estimated by the PCA. In this stem-cross plane, a circle was fitted by the simple estimation method. The circle diameter were considered diameter of stem or branch segment. The centers of circles were then connected based on a joint Gaussian kernel of affinity parameterized by direction and diameter differences. The connected nodes with the longest length formed the required stem curve.

The processing is fully automated. All single scan and multi scan sets in all plots were processed with same setup of parameters.

A.16. Method UNIPD

The DTM generation by the University of Padova (UNIPD), Italy, was done using a processing chain that applies a morphological filter and an interpolation method. Two different interpolation methods were tested to provide two resulting DTMs to test for accuracy.

To deal with the data-size, the points were organized in a disk-based octree, using Point Cloud Library. An empirical resolution of 10 m was used. The resolution parameter described the length of the smallest voxels at lowest octree level. A regular grid of 0.2 m resolution was then used to subset the point cloud keeping the lowest point falling in each cell. This produced a set of points with true positives (real ground points) and false positives, a common initial filtering step (Kobal et al., 2015). To remove false positives, a progressive morphological filter (Zhang et al., 2003) was applied which uses opening operation (erosion followed by dilation). Interpolation was then used to derive a regular grid. Two results were provided by applying two different interpolation algorithms: Natural Neighbour and Universal Kriging with automatic variogram.

The processing is fully automated. All plots and both scans (single- and multi-scan) in a single plot were processed with one parameter set.

A.17. Method NYME

The processing by the University of Sopron, Hungary uses voxels for stem detection and tree height estimation, while stem cross-sections are reconstructed from the original points within stem voxels.

The DTMs were generated using Robust Filtering (Kraus and Pfeifer, 1998) implemented in SCOP++ software package. Since this program has been developed for processing airborne data, the original points were reduced in advance. A coarse surface was generated from the lowest points within a grid of 2 m. Points within 0.5 m vertical distance of this surface were used as input for Robust Filtering

Laser points above the terrain were converted into a voxel space of 5 cm resolution and the point count for each voxel was registered. Knowing the sensor positions (for single scans only) and the angular resolution of the scanning, the relative point density for each voxel was obtained by normalizing the actual point count with the number of beams emitted toward the corresponding voxel. Voxels exceeding a given threshold on relative point density were considered to represent larger surfaces, i.e., stems, and the rest were supposed to be branches. The appropriate threshold on of relative point density for filtering stem voxels was set upon visual inspection. Vertical structures were enhanced using an anisotropic 3D filtering kernel to remove voxels of branches (Brolly et al., 2013). Stem voxels were grouped into regions applying Connected Component Labelling on the horizontal projection of the voxel-space. Regions with vertical extent exceeding 1 meter were regarded as stem.

In each layer of the voxel space, stem cross-sections were estimated by circle-fitting using the points in voxels. Circles fitted on too few points or with high RMSE were removed. Center coordinates and stem diameters at the required heights were computed from linear interpolation of the preceding and succeeding circles. Linear regression of diameters versus height obtained from the fitted circles for each stem was used when the required height was below the lowest detected circle.

Branch voxels were assigned to the identified stem regions to get tree tops and estimate tree height. The stem voxels were used as seeds, and tree regions were expanded simultaneously. Tree heights were estimated from the height difference between the top and bottom voxels.

The processing is fully automated. All single-scan data sets were processed by setting one parameter (threshold on relative point density) for different plots.

A.18. Method WU

The processing by the University of Wuhan, China, was based on detecting cylinder in multi-layers and extended Min-Cut approach.

Ground points were identified by finding the points with lowest height in the local neighborhood and were removed using the method of (Hernández and Marcotegui, 2009) before point cloud segmentation. DTM was built by constructing a Delaunay triangulation using the extracted ground points.

The remaining points were sliced into a series of horizontal cross-sections with a thickness predefined according to the point densities of the point clouds. In each slice, points were projected to *XY* plane and a Delaunay triangulation was built. Neighboring points were grouped if the distance between them was less than the threshold, which was defined by the mean and standard deviation of the edge lengths in the triangulation. A cylinder was fitted to each group. A tree stem was identified if there existed a series linked cylinders along the vertical direction.

Trees were further isolated in individual tree point clouds by an extended Min-Cut approach. Points around each stem model were selected, where the radius was defined by the horizontal distance between

the current and its nearest neighbor tree and an enlarge factor. Points of an individual tree were segmented in the graph by cutting edge, formed between every pair of 3D points, with small weight defined as a function of distance. Most of the tall trees can be segmented correctly after this step. For small trees, the result may contain disconnected components which come from adjacent trees or bushes. To remove the disconnected components, a region growing was performed from stem points as seeds, and points unreachable were filter out.

The intersection points between trunks and ground points were regarded as the tree locations. DBH was determined by cylinder fitting in the slice at the 1.2–1.4 m height. Stem curves were estimated by cylinder fitting on slices at specific heights. If the cylinder fitting failed and the height was less than the stem top, the cylinder was fitted to a section above the current height. The tree height was calculated by the difference of the maximum elevation in the single tree point cloud and the elevation of the ground close to the location of the tree.

The processing is fully automated. All single- and multi-scan data sets in all plots were processed with same setup of parameters.

References

Aschoff, T., Spiecker, H., 2004. Algorithms for the automatic detection of trees in laser scanner data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 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. Can. J. For. Res. 44, 666–676. https://doi.org/10.1139/cjfr-2013-0535.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. Int. Arch. Photogramm. Remote Sens. 33, 111–118.

Brolly, G., Kiraly, G., 2009. Algorithms for stem mapping by means of terrestrial laser scanning. Acta Silv. Lignaria Hung. 5, 119–130.

Brolly, G., Király, G., Czimber, K., 2013. Mapping forest regeneration from terrestrial laser scans. Acta Silvatica et Lignaria Hungarica 9. https://doi.org/10.2478/aslh-2013-0011.

Dijkstra, E.W., 1959. Numerische Mathematik 1, 269–271.

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.

Forsman, M., Börlin, N., Holmgren, J., 2016a. Estimation of tree stem attributes using terrestrial photogrammetry with a camera rig. Forests 7, 61. https://doi.org/10.3390/f7030061.

Forsman, M., Holmgren, J., Olofsson, K., 2016b. Tree stem diameter estimation from mobile laser scanning using line-wise intensity-based clustering. Forests 7, 206. https://doi.org/10.3390/f7090206.

Hackenberg, J., Morhart, C., Sheppard, J., Spiecker, H., Disney, M., 2014. Highly accurate tree models derived from terrestrial laser scan data: a method description. Forests 5, 1069–1105. https://doi.org/10.3390/f5051069.

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.

Hernández, J., Marcotegui, B., 2009. Point cloud segmentation towards urban ground modeling. In: Urban Remote Sensing Event, 2009 Joint. IEEE, pp. 1–5.

Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004. Assessing forest metrics with a ground-based scanning lidar. Can. J. For. Res. 34, 573–583. https://doi.org/10.1139/x03-225.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W., Wang, L., 2011. Automated methods for measuring DBH and tree heights with a commercial scanning lidar. Photogram. Eng. Remote Sens. 77, 219–227.

Hyyppä, J., Virtanen, J.-P., Jaakkola, A., Yu, X., Hyyppä, H., Liang, X., 2017. Feasibility of google tango and kinect for crowdsourcing forestry information. Forests 9, 6. https://doi.org/10.3390/f9010006.

Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B.M., Wu, J.-C., 2012. An international comparison of individual tree detection and extraction using airborne laser scanning. Remote Sens. 4, 950–974. https://doi.org/10.3390/rs4040950.

Kampmann, G., Renner, B., 2004. Vergleich verschiedener methoden zur bestimmung ausgleichender ebenen und geraden. AVN 2, 56.

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 J. Photogramm. Remote Sens. 75, 64–75. https://doi.org/10.1016/j.isprsjprs.2012.10.003.

Kobal, M., Bertoncelj, I., Pirotti, F., Dakskobler, I., Kutnar, L., 2015. Using lidar data to analyse sinkhole characteristics relevant for understory vegetation under forest cover—case study of a high karst area in the dinaric mountains. PLOS one 10, e0122070. https://doi.org/10.1371/journal.pone.0122070.

Koreň, M., Slančík, M., Suchomel, J., Dubina, J., 2015. Use of terrestrial laser scanning to evaluate the spatial distribution of soil disturbance by skidding operations. iForest – Biogeosciences and Forestry 8, 386–393. https://doi.org/10.3832/ifor1165-007.

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

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS J. Photogram. Remote Sens. 53, 193–203. https://doi.org/10.1016/S0924-2716(98)00009-4.

Liang, X., Litkey, P., Hyyppa, J., Kaartinen, H., Vastaranta, M., Holopainen, M., 2012. Automatic stem mapping using single-scan terrestrial laser scanning. IEEE Trans. Geosci. Remote Sens. 50, 661–670.

Liang, X., Kankare, V., Yu, X., Hyyppa, J., Holopainen, M., 2014b. Automated stem curve measurement using terrestrial laser scanning. IEEE Trans. Geosci. Remote Sens. 52, 1739–1748. https://doi.org/10.1109/TGRS.2013.2253783.

Liang, X., Hyyppa, J., Kukko, A., Kaartinen, H., Jaakkola, A., Yu, X., 2014a. The use of a mobile laser scanning system for mapping large forest plots. IEEE Geosci. Remote Sens. Lett. 11, 1504–1508. https://doi.org/10.1109/LGRS.2013.2297418.

Liang, X., Wang, Y., Jaakkola, A., Kukko, A., Kaartinen, H., Hyyppa, J., Honkavaara, E., Liu, J., 2015. Forest data collection using terrestrial image-based point clouds from a handheld camera compared to terrestrial and personal laser scanning. IEEE Trans. Geosci. Remote Sens. 53, 5117–5132. https://doi.org/10.1109/TGRS.2015.2417316.

Liang, X., Kukko, A., Hyyppä, J., Lehtomäki, M., Pyörälä, J., Yu, X., Kaartinen, H., Jaakkola, A., Wang, Y., 2018. In-situ measurements from mobile platforms: An emerging approach to address the old challenges associated with forest inventories. ISPRS J. Photogram. Remote Sens. https://doi.org/10.1016/j.isprsjprs.2018.04.019.

Lindberg, E., Holmgren, J., Olofsson, K., Olsson, H., 2012. Estimation of stem attributes using a combination of terrestrial and airborne laser scanning. Eur. J. For. Res. 131, 1917–1931.

Liu, J., Liang, X., Hyyppä, J., Yu, X., Lehtomäki, M., Pyörälä, J., Zhu, L., Wang, Y., Chen, R., 2017. Automated matching of multiple terrestrial laser scans for stem mapping without the use of artificial references. Int. J. Appl. Earth Obs. Geoinformation 56, 13–23. https://doi.org/10.1016/j.jag.2016.11.003.

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. Can. J. Remote Sens. 29, 607–622. https://doi.org/10.5589/m03-026.

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Culvenor, D.S., 2011. Measuring tree stem diameters using intensity profiles from ground-based scanning lidar from a fixed viewpoint. ISPRS J. Photogramm. Remote Sens. 66, 46–55.

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 non-photosynthetic components within terrestrial lidar point cloud data of forest canopies. IEEE Trans. Geosci. Remote Sens. 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. Int. J. Remote Sens. 29, 1579–1593. https://doi.org/10.1080/01431160701736406.

Mokroš, M., Liang, X., Surový, P., Valent, P., Čerňava, J., Chudý, F., Tunák, D., Saloň, Š., Merganič, J., 2018. Evaluation of close-range photogrammetry image collection methods for estimating tree diameters. ISPRS Int. J. Geo-Inf. 7, 93. https://doi.org/10.3390/ijgi7030093.

Murphy, G.E., Acuna, M.A., Dumbrell, I., 2010. Tree value and log product yield determination in radiata pine (*Pinus radiata*) plantations in Australia: comparisons of terrestrial laser scanning with a forest inventory system and manual measurements. Can. J. For. Res. 40, 2223–2233. https://doi.org/10.1139/X10-171.

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.

Olofsson, K., Holmgren, J., Olsson, H., 2014. Tree stem and height measurements using terrestrial laser scanning and the RANSAC algorithm. Remote Sens. 6, 4323–4344. https://doi.org/10.3390/rs6054323.

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. J. Appl. Ecol. 41, 755–767. https://doi.org/10.1111/j.0021-8901.2004.00925.x.

Pfeifer, N., Mandlburger, G., 2008. LiDAR data filtering and DTM generation. In: Topographic laser ranging and scanning: principles and processing. CRC Press, Boca Raton, FL, USA, pp. 307–334.

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. Appl. Geomat. 5, 311–322. https://doi.org/10.1007/s12518-013-0119-3.

Pyorala, J., Liang, X., Vastaranta, M., Saarinen, N., Kankare, V., Wang, Y., Holopainen, M., Hyyppa, J., 2018. Quantitative Assessment of Scots Pine (Pinus Sylvestris L.) Whorl Structure in a Forest Environment Using Terrestrial Laser Scanning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 1–10. https://doi.org/10.1109/JSTARS.2018.2819598.

Repola, J., 2008. Biomass equations for birch in Finland. Silva Fenn. 42, 605–624.

Repola, J., 2009. Biomass equations for Scots pine and Norway spruce in Finland. Silva Fenn. 43, 625–647.

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. Int Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 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®). Can. J. Remote Sens. 34, 426–440.

Sun, Y., Liang, X., Liang, Z., Welham, C., Li, W., 2016. Deriving merchantable volume in poplar through a localized tapering function from non-destructive terrestrial laser

scanning. Forests 7, 87. https://doi.org/10.3390/f7040087.
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. Scand. J. For. Res. 19, 571–581.
Tomaštík, J., Saloň, Š., Tunák, D., Chudý, F., Kardoš, M., 2017. Tango in forests – an initial experience of the use of the new Google technology in connection with forest inventory tasks. Comput. Electron. Agric. 141, 109–117. https://doi.org/10.1016/j.compag.2017.07.015.
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.
Vaaja, M.T., Virtanen, J.-P., Kurkela, M., Lehtola, V., Hyyppä, J., Hyyppä, H., 2016. The Effect of Wind on Tree Stem Parameter Estimation Using Terrestrial Laser Scanning. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences III–8 117–122. https://doi.org/10.5194/isprsannals-III-8-117-2016.
Wang, D., Hollaus, M., Puttonen, E., Pfeifer, N., 2016a. Automatic and Self-Adaptive Stem Reconstruction in Landslide-Affected Forests. Remote Sensing 8, 974. https://doi.org/10.3390/rs8120974.
Wang, Y., Hyyppa, J., Liang, X., Kaartinen, H., Yu, X., Lindberg, E., Holmgren, J., Qin, Y., Mallet, C., Ferraz, A., Torabzadeh, H., Morsdorf, F., Zhu, L., Liu, J., Alho, P., 2016b. International benchmarking of the individual tree detection methods for modeling 3-D canopy structure for silviculture and forest ecology using airborne laser scanning. IEEE Trans. Geosci. Remote Sens. 54, 5011–5027. https://doi.org/10.1109/TGRS.2016.2543225.
Watt, P.J., Donoghue, D.N.M., 2005. Measuring forest structure with terrestrial laser scanning. Int. J. Remote Sens. 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.
Yang, B., Dai, W., Dong, Z., Liu, Y., 2016. Automatic forest mapping at individual tree levels from terrestrial laser scanning point clouds with a hierarchical minimum cut method. Remote Sens. 8, 372. https://doi.org/10.3390/rs8050372.
Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G., Ni-Meister, W., Schaaf, C., Woodcock, C., Wang, J., Li, X., Strahler, A., 2011. Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. Remote Sens. Environ. 115, 2965–2974. https://doi.org/10.1016/j.rse.2010.03.019.
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 Sens. Lett. 4, 344–353. https://doi.org/10.1080/2150704X.2012.734931.
Zhang, K., Chen, S.-C., Whitman, D., Shyu, M.-L., Yan, J., Zhang, C., 2003. A progressive morphological filter for removing nonground measurements from airborne LIDAR data. IEEE Transactions on Geoscience and Remote Sensing 41, 872–882. https://doi.org/10.1109/TGRS.2003.810682.