Agricultural and Forest Meteorology 216 (2016) 105–114

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Tree species classification based on explicit tree structure feature parameters derived from static terrestrial laser scanning data

Yi Lin [a,*], Martin Herold [b]

[a] Institute of Remote Sensing and Geographic Information Systems, School of Earth and Space Science, Peking University, 100871 Beijing, China
[b] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, 6708 PB Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 17 August 2015
Accepted 10 October 2015
Available online 11 November 2015

Keywords:
Tree species classification
Static terrestrial laser scanning (TLS)
Explicit tree structure feature parameters
Support vector machine (SVM) classifier
Leave-one-out-for-cross-validation (LOOCV)

ABSTRACT

Tree species information is essential for forest studies such as forest meteorology, botany and ecology, and across the relevant fields new techniques efficient for classifying tree species are desperately demanded. For this requirement, the state-of-the-art remote sensing technology of static terrestrial laser scanning (TLS) shows the potential of handling it, since TLS can represent tree structures in details. However, the literature review suggests that the endeavors of introducing TLS into tree species classification are still in shortage. Aimed at this technical gap, this study attempted TLS for distinguishing four typical boreal tree species, i.e. Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), European aspen (*Populus tremula*) and pedunculate oak (*Quercus robur*). After a theoretical comparison of the generic mechanisms of TLS- and airborne laser scanning (ALS)-based forest mapping, explicit tree structure (ETS) feature parameters, rather than conventional ALS-derived statistical-sense feature parameters, were proposed and derived from TLS point clouds. Then, based on a support-vector-machine (SVM) classifier, the specific classification was operated in a leave-one-out-for-cross-validation (LOOCV) mode. Tests indicated that the proposed TLS-based ETS feature parameters and the used classification algorithm can be validated for implementing the task of classifying the four tree species (the maximum total accuracy reaches 90.00% and the robust total accuracy reaches 77.5%). Overall, as a leading endeavor, this study has developed a procedural frame for TLS-based tree species classification.

1. Introduction

1.1. Background

Tree species information serves as the fundamental data of forest studies, such as forest meteorology, ecology, botany, climatology and management. This is evidenced by the fact that without this kind of information, almost no results of measuring or retrieving tree properties can be comprehensively mapped (Naidoo et al., 2012), and such ill situations may further lead to drawing biased conclusions about forest ecological processes (Mascaro et al., 2012) and making wrong decisions on forest managements (Pinard et al., 1999). Instead, obtaining this kind of information means a lot for both tree-associated scientific researches such as forest micrometeorology (Du et al., 2011) and practical applications such as forest precision harvesting (Roxby et al., 2015). Among the relevant communities, techniques for characterizing and classification of tree species are widely investigated due to popularity and growing demand.

Compared to the conventional approaches based on labor-intensive field identification, remote sensing (RS) has supplied a large variety of cutting-edge techniques to accomplish the task (Heinzel and Koch, 2012). A number of aerospace/airborne RS approaches have been proposed for tree species classification. The representative endeavors involve differentiating tree species based on satellite hyper- (George et al., 2014) and multi-spectral (Krahwinkler and Rossmann, 2013) images, aerial hyper- (Colgan et al., 2012) and multi-spectral (Franklin et al., 2000) images, as well as terrestrial hyper- (Puttonen et al., 2010) and multi-spectral (Cope et al., 2012) images. During the process of developing these RS solutions, it has also been realized that hyper- and multi-spectral images have their own limitations (Heinzel and Koch, 2012), e.g. two specimens of the same tree species may show different spectral features due to their varying shapes in details. For this problem, along with the continuous progress of aerospace/airborne RS imaging in terms of spatial resolution, crown structure features have been realized as a potential solution. The related feature parameters capable of characterizing crown structural profiles were derived

* Corresponding author.
E-mail address: yi.lin@pku.edu.cn (Y. Lin).

http://dx.doi.org/10.1016/j.agrformet.2015.10.008
0168-1923/

from the spectral images and attempted for improving the performance of tree species classification (Zhang and Hu, 2012).

As a state-of-the-art RS technology, light detection and ranging (LiDAR) can directly represent 3D structures of trees. In fact, LiDAR has already been applied for mapping forest species compositions. For instances, airborne LiDAR (also termed as airborne laser scanning, ALS) was applied for tree species classification in the leaf-off and -on cases (Brandtberg, 2007), and the performance-influencing factors such as feature parameter selection, different sizes of training sets and echo intensity normalization were discussed (Korpela et al., 2010). Meanwhile, it was found that usage of LiDAR cannot promise always giving sound results. Aimed at this problem, a number of endeavors tried the plan of integrating ALS data and hyper-/multi-spectral imagery in order to improve the classification results (Dalponte et al., 2008, 2012; Kantola et al., 2010; Puttonen et al., 2010; Colgan et al., 2012; Naidoo et al., 2012). Recently, the state-of-the-art technique of full-waveform (FWF) LiDAR capable of better characterizing tree lower parts was attempted for accomplishing the task (Heinzel and Koch, 2011; Yao et al., 2012), and high-density ALS was also introduced from the perspective of deriving more accurate structural features (Li et al., 2013). However, based on the current ALS systems, all of the above-listed proposals still cannot implement fine-scale characterizations of tree structures in a real sense (Alberti et al., 2013).

Potential solutions to the above-mentioned problems can be found by acquiring LiDAR data using static terrestrial laser scanning (TLS). Now, introducing TLS into forest inventory for investigating various tree properties at fine scales is a research topic of extensive interest (Hopkinson et al., 2004). This is rooted in the advantage of TLS on detailed representation of tree structures. Practices show that TLS is available for fulfilling the tasks such as prediction of diameter at breast height (DBH) (Henning and Radtke, 2006), retrieval of leaf area density (Hosoi and Omasa, 2006), estimation of tree growth (Lin et al., 2012), derivation of vertical crown profile (Calders et al., 2014), estimation of above-ground biomass (Calders et al., 2015), and detection and removal of vegetation to obtain a clear ground surface model (Pirotti et al., 2013). With the progress of LiDAR technology, the data attributes of laser backscatters such as intensity (Lovell et al., 2011) and FWF (Strahler et al., 2008) have also been applied in TLS-based mapping for better learning tree properties. In addition, from the perspective of measurement operation, the multi-scan mode (Thies et al., 2004), i.e. scanning an object from its multiple surrounding points, have been adopted for improving the completeness of tree structure characterization. In some cases, TLS, even, was assumed as the reliable surveying means for collecting the reference data (Hauglin et al., 2014).

However, the related literature review indicates that the endeavors of applying TLS for tree species classification have not been extensively reported. The only work was still based on the texture attribute of tree stems reconstructed from TLS point clouds (Othmani et al., 2013), whereas the strength of TLS capable of representing the structures of individual trees well has not been made use of. In other words, the implications of launching the studies of TLS-based tree species classification have not been fully realized, though which indeed exist. First, practices require efficient methods. After all, it cannot be guaranteed that TLS-collected point clouds are distributed together with the data attribute of tree species to data analysts, since it is hard to ensure each of the field operators of TLS also being an expert knowing tree species well; on the other hand, it is also hard to promise that every data analyst possibly from different disciplines can accurately identify tree species from the 3D scattered laser points. Second, from the perspective of following application of its results, TLS-based tree species classification is of high potential for supporting fine-scale forest studies at the region or even plot levels. For instances, grasping tree species is essential for learning tree diversity which may decide tree transpiration (Kunert et al., 2012) or for investigating the influences of mixed forest compositions as the eddy-covariance (EC) footprints on measuring the fluxes of CO_2 exchange (Nagy et al., 2006). These all suggest that it is time to develop new TLS-based methods for tree species classification.

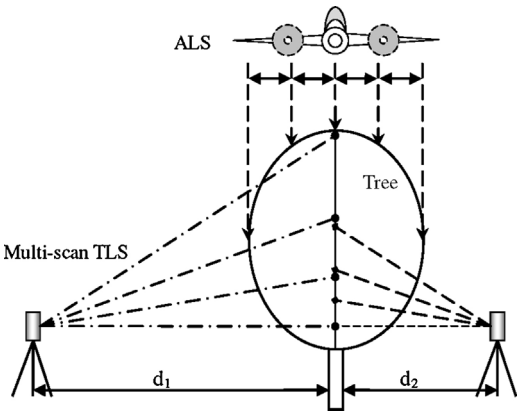

Fig. 1. Schematic comparison of ALS- and TLS-based individual tree sampling mechanisms.

1.2. Technical review

As regards to the laser-based methods of tree species classification, most of the existing algorithms were developed aiming at ALS. The representative ALS-based studies are listed in Table 1, in terms of different categories of feature parameters. Specifically, all of the feature parameters involved in these endeavors can be classified into three dominant categories, i.e. laser return intensity, laser point distribution and tree shape index. These categories of feature parameters may be referred to in the process of attempting TLS for tree species classification. In other words, the question "Can the feature parameters listed in Table 1 be directly adopted into TLS-based tree species classification?" can be asked. If the answer is positive, the task of developing TLS-based tree species classification algorithms will become relatively simpler. After all, proposing feature parameters capable of reflecting the peculiarities of different tree species constitutes the basis of implementing accurate tree species classification, and hence, a positive answer means that the existing ALS-based algorithms can be directly, or after a little adaption, applied for TLS-based tree species classification.

The possibility concerned in the question was first examined by analyzing the schematic paradigms of ALS- and TLS-based individual tree mapping, as shown in Fig. 1. In the case of ALS-based tree sampling, flight heights are usually tens of times of individual tree crown diameters, and thus, it can be deemed that ALS laser beams are parallel and have equal spatial intervals. Actually, this is a common presumption for simplification of data pre-processing in practical ALS-based forest mapping (Pfeifer and Briese, 2007), and this also constitutes the theoretical foundation of proposing such as the quantile-typed feature parameters (Dalponte et al., 2012). The reasoning is as follows. As shown in Fig. 1, the upper part of a crown that is segmented into equal-interval horizontal layers generally shows more laser echoes than the lower part, and thus, different crown horizontal layers can be characterized by the percentages of laser echoes at different height spans.

However, the ALS-oriented presumption is unavailable in the case of TLS-based tree sampling. In TLS-based tree measurements, the scanner-tree distances tend to be only several times of tree heights, and it is typically deemed that TLS laser beams are in a center-scattering mode. For a same tree, TLS samplings from different distances show different modes of point spatial distributions.

Table 1
List of the representative feature parameters used for tree species classification in the existing literature.

Type	Parameter	Definition	Ref.
Laser return intensity	I	Laser echo intensity (I)	Dalponte et al. (2008) Heinzel and Koch (2011)
	I_{mean}, I_{std}, I_{skew}, I_{kurt}	Intensity moments in the whole crown, including such as mean of intensity (I_{mean}), standard deviation of intensity (I_{std}), skewness of intensity (I_{skew}) and kurtosis of intensity (I_{kurt})	Korpela et al. (2010) Yao et al. (2012)
	SW	Mean pulse width of single and first reflections (SW) in the entire tree segment	Yao et al. (2012)
Laser point distribution	nth hq,	nth height quantile (hq) in percents (from tree top)	Puttonen et al. (2010) Kantola et al. (2010) Dalponte et al. (2012) Yao et al. (2012)
	PD	Point density (PD) at normalized height	Puttonen et al. (2010)
	dd_n	Height density deciles (dd_n)	Korpela et al. (2010)
	E	Elevation of digital terrain or surface models (E)	Dalponte et al. (2008) Heinzel and Koch (2011)
	max$D$	Maximum crown diameters (max$D$) when crown is considered as an ellipse	Kantola et al. (2010)
	RCD	Relative clustering degree (RCD)	Li et al. (2013)
Tree shape index	H_{max}	Maximum height of points indicating tree height (H_{max})	Naidoo et al. (2012) Puttonen et al. (2010) Kantola et al. (2010) Dalponte et al. (2012)
	CH	Crown height (CH)	Kantola et al. (2010)
	CV	Crown volume (CV) as a convex hull in 3D	Kantola et al. (2010)
	TEX	3-D texture (TEX)	Li et al. (2013)
	DT	Delaunay triangulation (DT)	Li et al. (2013)

As illustrated in Fig. 1, the spacing interval of any two adjacent laser echoes in the vertical direction for the close TLS location, theoretically, is smaller than the opposites for the distant TLS location. Thereby, the ALS-related feature parameters such as the nth height quantiles in percents (Dalponte et al., 2012) cannot be directly applied into TLS-based tree species classification.

1.3. Objectives

The technical review indicates that development of effective methods for tree species classification by playing the full role of TLS in tree structure representation, substantially, starts from scratch. Meanwhile, the background analysis suggests that the methods efficient for TLS-based tree species classification indeed are demanded. Aimed at this technical gap, this study attempted to validate TLS for tree species classification, via deriving appropriate feature parameters and introducing high-performance classification methods. The specific tasks involve:

1) Proposing and then calculating new feature parameters appropriate for reflecting the characteristics of TLS-based tree sampling for tree species classification;
2) Playing the full role of the introduced classification method to implement tree species classification with the optimal accuracies acquired in both of the maximum and robust ways.

2. Materials and methods

2.1. Study site and data collection

The study site is located in the Seurasaari Island, Helsinki, southern Finland (24°10′52″ E, 60°53′3″ N). The island covering ~46 ha is a wooded area, with rocks, hills, wetlands and herb-rich forests. The northern part is a well-managed urban park briefly with old oaks, spruces and pines, and the southern part is a natural unmanaged park forest mainly with oaks, spruces, poplars and pines.

The data were collected with a Leica HDS6100 TLS system (Leica Geosystem AG, Heerbrugg, Switzerland) in September 2010. HDS6100 is a 690-nm phase-based scanner with a 360° × 310° field-of-view upward and its laser rate is 508 000 points per second. The range measurement accuracy is ±2 mm at a distance of 25 m. The circular laser beam diameter and the beam divergence at the scanner exit are 3 mm and 0.22 mrad respectively. The TLS-based measurement here was conducted in the multi-scan mode (Thies et al., 2004), and this setting resulted in the delineations of tree structures in a relatively more complete way. More details about the experiment specifications such as how many scan positions used for data recording and data pre-processing such as the methods used for individual tree segmentation can refer to (Holopainen et al., 2013). The species and the diameters at breast height (DBHs) of the trees selected for test were manually identified, measured and recorded during the in-field investigation.

Four typical boreal tree species were finally selected for the test. There are 9 Norway spruces (*Picea abies*, PA), 14 Scots pines (*Pinus sylvestris*, PS), 7 European aspens (*Populus tremula*, PT) and 10 pedunculate oaks (*Quercus robur*, QR). Their structures represented by the multi-scan TLS point clouds are separately illustrated in Fig. 2. Their structure features are representatively listed in Table 2, in terms of tree height and DBH in a statistical sense. The heights of the trees are in the range of 16.14 m and 28.61 m, and the DBH values lie in the range of 14.30 cm and 73.30 cm. A comparison indicates that there are no immoderate differences between the sample trees of the four tree species, and there are also no extreme cases occurring in the chosen trees. Overall, it is appropriate to validate the feature parameters and the species-classification method proposed later based on the selected trees.

2.2. Data preprocessing

The premise of fulfilling TLS-based tree species classification is a sound performance of TLS-based tree structure characterization, rather than a high-density representation. Thus, the performance of the TLS-based tree structure characterization was first assessed

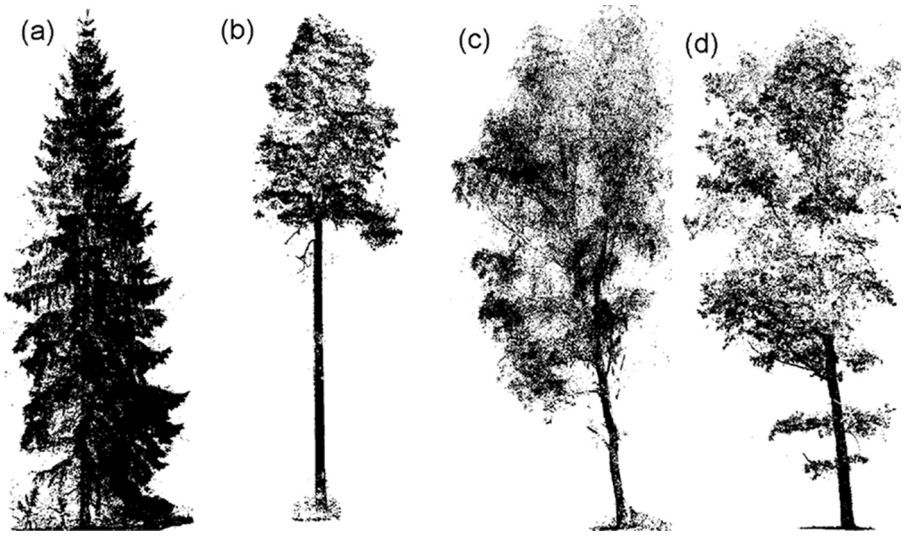

Fig. 2. Illustration of the four tree species represented by the TLS-collected point clouds for the test: (a) *Picea abies* (PA), (b) *Pinus sylvestris* (PS), (c) *Populus tremula* (PT), and (d) *Quercus robur* (QR).

in the step of data preprocessing. Here, DBH was assumed as a representative feature variable. Compared to other variables, the reference data about DBHs can be more readily acquired via manual measurements; more importantly, DBH is among the features at the finest scales, which can demonstrate the capacity of TLS on characterizing tree structural details. The DBHs of the sample trees were estimated using the algorithm of cone fitting, i.e. least-square fitting of the laser points within a horizontal layer (with a length of 1 m) across the DBH-related height (~1.37 m) to a cone model in order to acquire its diameter. More detailed operations of DBH estimation can refer to (Liang and Hyyppä, 2013). After correlation analysis of the estimated and manually measured DBH values, the performance of the characterization was quantified using the coefficient of determination (R^2) and root mean squared error (RMSE), each with the definition as

$$R^2 = 1 - \frac{\sum (d_i - \hat{d}_i)^2}{\sum (d_i - \bar{d})^2}, \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum (d_i - d_i^{\mathrm{R}})^2}{n}}, \quad (2)$$

where d_i is the estimated DBH value for the ith tree, d_i^{R} is its reference DBH value, their difference $(d_i - d_i^{\mathrm{R}})$ is termed as DBH deviation, $\hat{d}_i$ is the DBH value estimated by following the estimation-reference DBH regression relationship revealed by correlation analysis, and $\bar{d} = \sum d_i/n$.

Table 2
Statistics of the basic structural features of the sample trees for test.

Species	Number		Tree height (m)	Manual DBH (cm)
PA	9	Min	18.88	14.30
		Max	28.61	55.70
		Mean	24.30	38.52
PS	14	Min	17.35	30.10
		Max	23.96	73.30
		Mean	21.25	43.19
PT	7	Min	20.58	25.50
		Max	25.40	56.70
		Mean	23.68	42.60
QR	10	Min	16.14	24.80
		Max	26.46	52.10
		Mean	20.27	42.31

2.3. Derivation of explicit tree structure feature parameters

2.3.1. Proposal of the feature parameters

Compared to the case of ALS laser beams often being unable of reaching the lower parts of crowns due to laser obscuration, TLS can represent the details of tree twigs in many cases. Accordingly, it can be easily conceived that explicit tree structure (ETS) feature parameters can be derived from TLS point clouds and used as the variables in tree species classification. This strategy can play the full role of TLS in tree structure characterization. The common feature parameters used for characterizing tree structures include tree height, stem DBH, crown height (or termed as height of the first live branch, FLBH), branch angle, crown length, height of the highest branch within crown lower surface (HLS), height of the lowest branch within crown lower surface (LLS), the longest spread of crown cover (LS) and the longest cross-spread of crown cover (LCS), as illustrated in Fig. 3. There have been many works validating TLS for derivation of these parameters (e.g., Henning and Radtke, 2006; Hosoi and Omasa, 2006; Lin et al., 2012; Calders et al., 2014, 2015), and consequently, it is convenient to choose them as the ETS feature parameters for tree species classification.

In addition to these common tree structure feature parameters, some other ETS feature parameters can be proposed by seeking their corresponding robust forms or by combining two common feature parameters. For the first situation, the diameter of the circle with area equal to the crown cover (D_{EA}) as shown in Fig. 3b was put forward, which can overcome the influence of crown structure variability to some extent; for the second situation, the ratios between the common structure feature parameters and tree height were proposed, which can reduce the impacts of different tree ages on tree species classification. Based on the rules as mentioned above, all of the proposed specific ETS feature parameters for this study are listed in Table 3.

2.3.2. Definition of the feature parameters

The proposed ETS feature parameters can be classified into five categories, individually reflecting the structural characteristics of an entire tree as well as its stem, branches, crown and leaves. Two

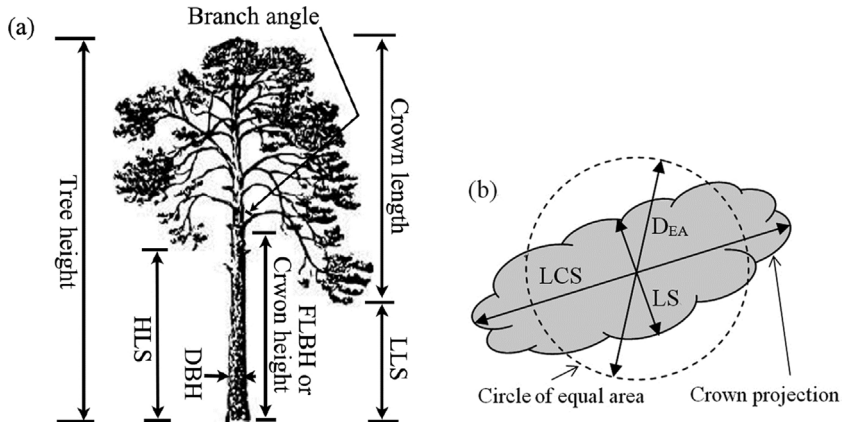

Fig. 3. The common structure parameters indicated in (a) a whole tree from side view and (b) a crown from top view.

ETS parameters are of stem type and five are of crown type. As listed in Table 3, the first seven ETS feature parameters are defined as

$$P_{\mathrm{LcHt}} = \frac{L_c}{H_t}, \tag{3}$$

$$P_{\mathrm{DsHt}} = \frac{D_s}{H_t}, \tag{4}$$

$$P_{\mathrm{HcHt}} = \frac{H_c}{H_t}, \tag{5}$$

$$P_{\mathrm{Ab}} = a\cos\left(\frac{L_{\mathrm{HS}}}{L_{\mathrm{HB}}}\right), \tag{6}$$

$$P_{\mathrm{LcD_{EA}}} = \frac{L_c}{D_{\mathrm{EA}}}, \tag{7}$$

$$P_{\mathrm{LlsLhls}} = \frac{H_t - H_{\mathrm{LLS}}}{H_t - H_{\mathrm{HLS}}}, \tag{8}$$

$$P_{\mathrm{LsLcs}} = \frac{L_s}{L_{cs}}, \tag{9}$$

where H_t indicates tree height, L_c means crown length, D_s is DBH, H_c relates to crown height, L_{HB} means the length of branch within a certain height span, L_{HS} means the length of stem within the same height span, H_{HLS} is HLS, H_{LLS} is LLS, L_s is LS and L_{cs} means LCS.

As regards to the last three ETS feature parameters, they need to be derived based on the rastered data after 3D gridding of the point clouds, rather than directly from the raw scattered points. The gridding operation is as follows: The space occupied by the points is divided into an integration of 3D voxels with the same length unit, and the points lying within each voxel are marked by the same 3D coordinates. This operation can reduce the influences such as the asymmetric effects in multi-scan TLS mappings. Based on the rastered data for each tree, the parameter of GcHt was derived by calculating the ratio between the mean height for all of the voxels and tree height, and it can suggest the relative height of the "gravity" of each crown. R_{SG} was derived as the average ratio between the total area of voxel one-side covers and the area of the 2D convex hull of the crown from two perpendicular side views, and it can show the leaf and branch gap fraction for each tree from the horizontal perspective. The definition of LAI is similar with R_{SG}, just from a top view. These three ETS feature parameters can be defined as

$$P_{\mathrm{Gc}} = \frac{\sum_{i=1}^{n} H_{G_i}/n}{H_t}, \tag{10}$$

$$P_{R_{\mathrm{SG}}} = \frac{\sum(n_{G_{XZ}} \times A_G)/A_{CV}^{XZ} + \sum(n_{G_{YZ}} \times A_G)/A_{CV}^{YZ}}{2}, \tag{11}$$

$$P_{\mathrm{LAI}} = \frac{\sum(n_{G_{XY}} \times A_G)}{A_{CV}^{XY}}, \tag{12}$$

where H_{G_i} means the height of the ith grid, A_G indicates the side area of each grid, $n_{G_{XZ}}$, $n_{G_{YZ}}$ and $n_{G_{XY}}$ each is the number of grids in the XZ, YZ and XY projection plane, A_{CV}^{XZ}, A_{CV}^{YZ} and A_{CV}^{XY} each is the area of the convex hull of the grids in the XZ, YZ and XY projection plane respectively, and n relates to the total number of the voxels generated for each tree.

Table 3
List of the ETS feature parameters proposed in this study.

Type	Definition	Abbr.	No.
Tree	(Crown length)/(tree height)	LcHt	1
Stem	DBH/(tree height)	DsHt	2
	FLBH/(tree height)	HcHt	3
Branch	Branch angle	Ab	4
Crown	(Crown length)/(crown diameter)	$\mathrm{LcD_{EA}}$	5
	(Tree height-LLS)/(tree height-HLS)	LlsLhls	6
	LS/LCS	LsLcs	7
	(Mean height for all of the grids)/(tree height)	GcHt	8
	Side gap fraction for all of the grids	R_{SG}	9
Leaf	Leaf area index for all of the grids	LAI	10

2.3.3. Calculation of the feature parameters

The method for calculating DBH D_s has been introduced in Section 2.2. Tree height H_t can be calculated by extracting the laser point with the maximum Z value from the isolated laser points corresponding to that tree, although this simplified operation may introduce errors. The calculations of the other proposed feature parameters actually rely on the derivations of the related specific structural variables. Specifically, derivations of LS, LCS and D_{EA} involve dealing with the 2D scattered points resulting from crown projection onto ground, as shown in Fig. 3b. Principal component analysis (PCA) (Monnier et al., 2012) was deployed on the spatial coordinates (X, Y, Z) of these points, and the first and second principal components relate to LS and LCS respectively. The area of the minimum circumscribed circle can give the corresponding D_{EA} value.

For the remaining variables of crown length L_c, crown height H_c, branch angle P_{Ab}, H_{HLS} and H_{LLS}, they were derived by segmenting the laser points into different spaces. As illustrated in Fig. 4, the spaces were drawn by setting three different-size cylinders

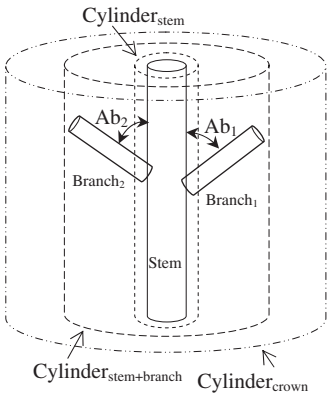

Fig. 4. Schematic diagram of restricting different cylinder-typed spaces for calculating different tree structural variables.

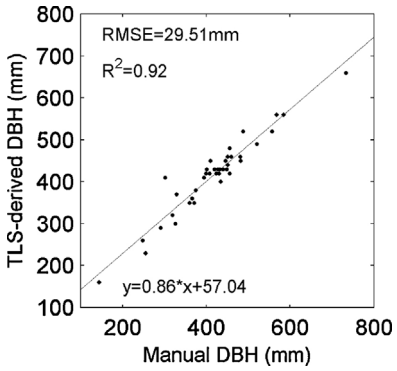

Fig. 5. Scatterplots of the TLS-derived and the reference DBH values of the trees for test and their linear relationship acquired by using linear regression analysis.

around the stem. The cylinders are defined as Cylinder$_{stem}$, Cylinder$_{stem+branch}$ and Cylinder$_{crown}$, and their radiuses were set to be 1.5 times and 4 times of the stem radius and 2 times of D_{EA}, respectively. Crown height H_c and branch angle P_{Ab} were calculated based on the laser points lying between Cylinder$_{stem}$ and Cylinder$_{stem+branch}$, and their specific derivations were accomplished by interactively segmenting the points individually corresponding to stems and branches and comparing the directions of their axes that are acquired as the first principal components after the PCA operation (Monnier et al., 2012). Crown length L_c, H_{HLS} and H_{LLS} were calculated based on the laser points between Cylinder$_{stem+branch}$ and Cylinder$_{crown}$, and their derivations were implemented by automatically searching the lowest points in different vertical segments of the restricted space.

2.4. SVM-based classification

The implementation of tree species classification was based on the LIBSVM, i.e. a support vector machine (SVM) package (Chang and Lin, 2014). As a branch of the generalized linear classifiers, the SVM-schematic classifiers generally are distribution-free. The specific algorithms are to carry out the optimization of finding a separating hyperplane between the feature vectors of any two classes (Melgani and Bruzzone, 2004). The separating hyperplanes identified by the SVM maximize the margin between any two classes and can robustly identify the classes of the class-unknown samples. In fact, this SVM-based classification plan has once been used in the previous works on tree species classifications (Melgani and Bruzzone, 2004; Dalponte et al., 2008; Puttonen et al., 2010; Heikkinen et al., 2010).

The sample trees for test in this study were classified using the SVM classifier in a leave-one-out-cross-validation (LOOCV) way. Specifically, in LOOCV, each sample is classified using the associated SVM classifier trained by the rest of the sample set (LOOCV-SVM in short hereafter). For this study, each tree was classified using the associated SVM classifier trained by the other trees. The evaluation was based on recall and precision, i.e. for each tree species the fraction of the number of the right-classified trees to the number of the original trees and the fraction of the number of the right-classified trees to the number of the trees classified into that tree species, respectively.

The operation of LOOCV-SVM classification was deployed on, one by one, all of the cases of combinations of the feature parameters. The numbers of the cases of combining from 1 to 10 ETS feature parameters are 10, 45, 120, 210, 252, 210, 120, 45, 10 and 1, respectively. Then, for each case, the corresponding recall and precision, i.e. the ratio between the number of the right-classified trees and the number of the original trees for the concerned tree species and the ratio between the number of the right-classified trees and the number of the trees all classified into the concerned tree species, respectively, were calculated. Finally, evaluation was conducted in both of maximum and robust scenarios. The maximum scenario just seeks the total accuracy for all of the specimens to be as high as possible, while the robust scenario asks the minimum accuracy for the specimens of each tree species to be as high as possible, which may result in the total accuracy lower than the maximum-scenario-related value.

3. Results

The results of TLS-based DBH estimation were first acquired and evaluated to verify the premise of TLS-based tree species classification. Then, the proposed ETS feature parameters were derived from the TLS-collected point clouds of the sample trees. Based on these feature parameters, the trees were classified and the performance was assessed.

3.1. Validation of TLS-based tree characterization

The derived and reference pairs of DBH are reported in the scatterplot in Fig. 5. After correlation analysis of the two data, the resulting indices of TLS performance, namely, R^2 and RMSE, are 0.92 and 29.51 mm, respectively. After the linear regression analysis, the relation between the derived and reference DBHs (mm) can be expressed as $y = 0.86 \times x + 57.04$. For the four PA, PS, PT and QR species, their individual R^2 values are 0.99, 0.88, 0.96 and 0.92, and their RMSE values are 15.25 mm, 42.50 mm, 21.58 mm and 22.77 mm, respectively. All of these results have primarily verified the premise of carrying out the study of TLS-based tree species classification.

3.2. Derivation of the feature parameters

For the three ETS feature parameters of GcHt, R_{SG} and LAI, their derivations were implemented based on the rastered data. The operations on the rastered TLS datasets were conducted by setting the edge width of 3D voxels at 1 cm and 1 dm, and the resulting tree representations are illustrated in Fig. 6a and b, respectively. Fig. 6b shows that the 1dm-gridded data still can characterize tree structures in a relatively complete way.

The derived values of the proposed ten ETS feature parameters are separately shown in Fig. 7, in which each of the four tree species relates to a boxplot. Specifically, each boxplot indicates the range of the derived ETS feature parameter values for all of the specimens of the corresponding tree species. Based on all of the sub-figures in Fig. 7, it can be realized that there are no cases of merely one ETS feature parameter capable of distinguishing all of

Fig. 6. Illustration of the (a) 1 cm-gridded and (b) 1 dm-gridded TLS data that represent the tree same as shown in Fig. 1(d).

the four tree species. But fortunately, there are some cases of one feature parameter capable of distinguishing one tree species from the other three. For instances, LcD_{EA} may separate PA from PS, PT and QR, while R_{SG} may separate QR from PA, PS and PT in a large ratio. Such information proved to play its role in the following step of SVM-based classification, which has the capacity of dealing with multiple feature parameters.

3.3. LOOCV-SVM classification

The LOOCV-SVM classification was first conducted based on the proposed ETS feature parameters one by one. Actually, this configuration is equivalent to the scenario of classifying the trees in terms of the boxplots, as shown in Fig. 7. Based on each feature parameter, the resulting recalls and precisions for different tree species are individually listed in Fig. 8a and b, respectively. There are no cases that the recalls are all larger than zero for the four tree species, and these phenomena occur to the precisions as well. Namely, in the case of using only one parameter, there is always one species missed. Although this is not a common rule available for other trees, it is at least revealed that the proposed feature parameters each is not enough for finishing the task of tree species classification.

Fig. 8 indicated that the task of tree species classification needed to be accomplished by using the combinations of the ETS feature parameters. Then, the total accuracies for all of the cases of combining from 1 to 10 ETS feature parameters were individually calculated, and the results are shown in Fig. 9. The best performance corresponding to the maximum total accuracy occurred in the case of combination of five parameters, and the robust performance keeps improving in a whole sense. Their optimal values relate to the case of combing the five R_{SG}, LcD_{EA}, DsHt, HcHt and Ab parameters and the case of combining all of the parameters, respectively. Their individual results are listed in Table 4. The total accuracy in the maximum mode is 90.00% and the total accuracy

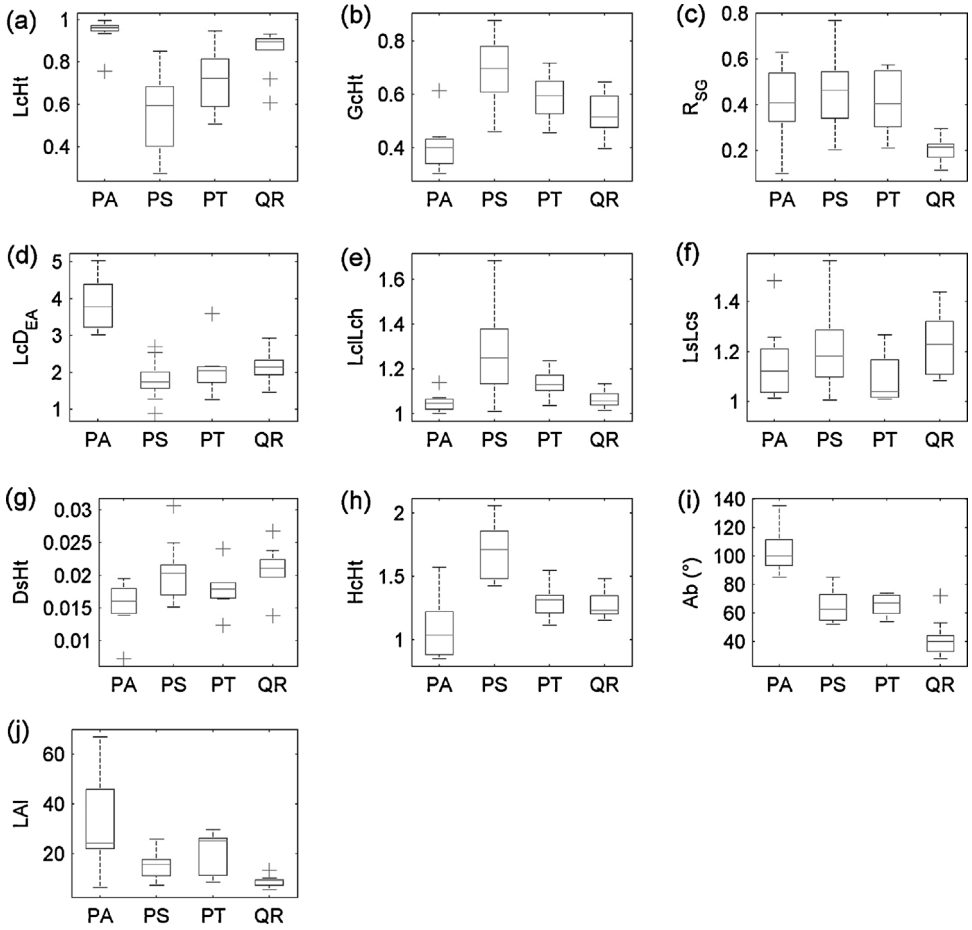

Fig. 7. Boxplots of the derived values of the ETS feature parameters: (a) LcHt, (b) GcHt, (c) R_{SG}, (d) LcDEA, (e) LllsLhls, (f) LsLcs, (g) DsHt, (h) HcHt, (i) Ab, and (j) LAI.

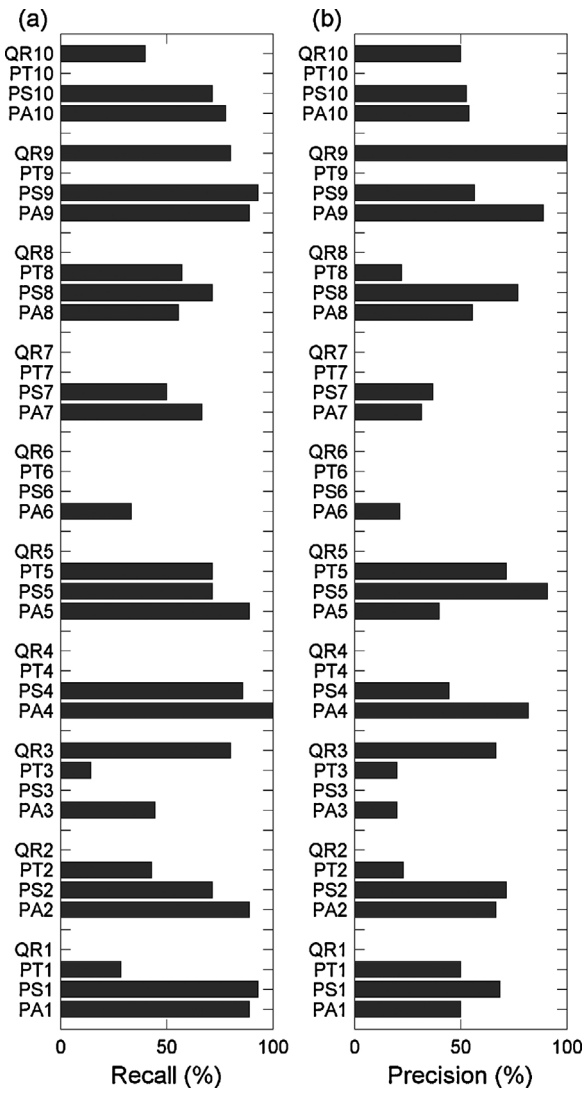

Fig. 8. The (a) recall- and (b) precision-indicated accuracies of tree species classifications based on each ETS feature parameter.

in the robust mode is 77.50%, which are larger than the maximum-value result of 70.25% in Fig. 8 (note that in this case the specimens of one tree species were completely missed in the result of the classification). These results have briefly validated the proposed plan of TLS-based classification of the specimens for the four boreal tree species of interest.

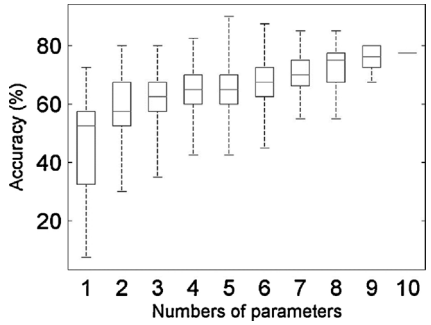

Fig. 9. Boxplots of the classification accuracies for the combinations of the derived ETS feature parameters with the numbers increasing.

Table 4
Results of the tree species classification based on the proposed ETS feature parameters and the used classifier in the two cases.

Case	Species	Precision (%)	Recall (%)	Total accuracy (%)
Maximum	PA	100.00	77.78	90.00
	PS	86.67	92.86	
	PT	100.00	85.71	
	QR	83.33	100.00	
Robust	PA	87.50	77.78	77.50
	PS	83.33	71.43	
	PT	71.43	71.43	
	QR	69.23	90.00	

4. Discussions

4.1. Performance influences

The final classification results are positive but far from perfect. Even in the case showing the maximum classification accuracy, there is still at least one misclassified specimen existing for each of the PA, PS and PT species. The specific samples are illustrated in Fig. 10. From the figure, it can be realized that even if we make visual interpretation, it is still very hard to discern these trees. This is due to that their morphologies are similar and also differ from the common patterns of their corresponding tree species. The causes of their structural deviations may include environmental influences such as tree competition (Seidel et al., 2015) and pest hurt (Han et al., 2012). The impacts of limited growth space may also alternate the shapes of tree crowns (Seidel et al., 2015).

The illustration in Fig. 10 also demonstrates an embarrassing situation existing in the proposals of some feature parameters. The embarrassment involves the definitions of these parameters with tree height as their denominators. This measure was used for the purpose of avoiding the influences caused by different tree ages on tree species classification, and thus, the method developed based on the limited number of sample trees can work for the trees in a more extensive range. Although the introduction of the "ratio" manner into the definitions of the ETS feature parameters proved to play its full role in avoiding the influence of tree ages, a little ratio of misclassified cases were also generated. This is shown by Fig. 10c, in which the TH of the tree actually is smaller than the other two trees in Fig. 10.

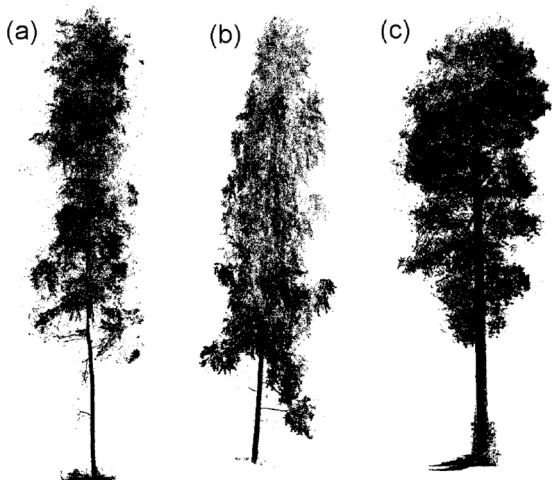

Fig. 10. Illustrations of the cases of tree species misclassification: (a) the PA tree (TH = 19.69 m) was misclassified as a PT tree, (b) the PT tree (TH = 20.72 m) was misclassified as a PA tree, and (c) the PS tree (TH = 17.51 m) was misclassified as a QR tree.

In addition, the best performance of classifications tends also to be up to the optimal combination of the feature parameters, and how to find the optimal combination is an issue (Puttonen et al., 2010). This is due to that the optimal combinations turned out to vary with different choices of sample trees (Puttonen et al., 2010). Consequently, given that the number of the trees in this study is quite small, this study operated the classifications in terms of total accuracy in both of the maximum and robust manners. After all, the rule about parameters combination corresponding to the maximum total accuracy pursuit here may fail for different tree samples in other regions. The rules acquired in the robust mode can be more referred to in the real applications.

4.2. Applicability analysis

The size of the chosen sample-tree set and the used LOOCV-SVM classifier both have some reference significance for forest species composition mapping. The reasoning is that in practice, tree species classification for large areas tends to first undergo training-test in order to establish appropriate classification models (Dalponte et al., 2012). Specifically, a limited number of trees are manually classified and used to train the classifier, through which the remaining trees can be classified; the performance of the trained classifier is evaluated by checking another set of limited-number sample trees. In this framework, about ten sample trees for each tree species are a common but effective choice. The number of the sample trees concerned in this study is approximate to the size of the training set in real forest inventory. In this sense, the used LOOCV-SVM classifier can play well, which each time only check one sample. Consequently, this has also verified the proposed TLS-based method being available for classifying tree species in practical applications.

More than for tree species classification, the proposed ETS feature parameters also have some reference significance for grasping the brief characteristics of individual tree structures for different tree species. This is beneficial for various studies on forest ecology, meteorology and management, in which the mode of combining aerospace/aerial imagery or LiDAR data with TLS reference data is often assumed (Hopkinson et al., 2004; Heinzel and Koch, 2012). Specifically, for different forest properties of interest their correlations with different tree structure feature parameters are explored, and the optimal results can determine the retrieval formulas. In this kind of processes, the performance of retrieval of large-area forest attributes is generally up to the accuracy of TLS-based tree structure characterization, which serves as the basis in the up-scaling operation. The methods of deriving ETS feature parameters used in this study can help to quickly grasp the structural characteristics of the targeted trees, and this is beneficial for selecting the readily-acquirable structure feature parameters appropriate for deriving forest properties of interest.

4.3. Potential improvements

The first improvement plan is about how to add the automatic degree of deriving tree structural variables. In fact, automatic extraction of structural parameters directly from TLS point clouds of large volumes or even from the gridded data is not a trivial task. For example, although TLS data can give relatively complete representations of tree shapes, it was found that dead branches often disturb the extraction of the FLBH values (Zhou and Lin, 2013). As regards to the variable of crown length, dense understory layers sometimes may make it hard to automatically identify the boundaries of crowns. This issue also occurs in the process of finding the highest and lowest branches within each crown lower surface. Development of automatic methods for more efficient parameter extraction is an important task, which can help to bring TLS into fields for practical applications. This will also help to handle the issue of using limited samples for classifier training, as indicated in the last sub-section.

The final robust accuracy of 77.5% indicates that the TLS-based solution has a lot of space for improving its performance. The parameters capable of characterizing the shapes, structures and morphologies of crown-insides need to be proposed in the next step. By this means, the influences such as tree competition and tree age can be somehow reduced in the process of tree species classification. More than those extracted structural parameters, even more than the crown-inside structural parameters as proposed above, the feature parameters capable of characterizing the whole or local shapes, geometries, structures and morphologies of trees of different species need to be explored. With all of such factors taken into account, the feature parameters inherent to trees can be derived for fulfilling more reliable tree species classification. In addition, the disturbances caused by exterior factors such as tree competition can also be directly considered in the process of classification, e.g. by introducing the related compensation models.

In addition to deriving more inherent tree structure parameters from TLS point clouds, improvements can also be implemented by adding other kinds of RS means. For example, multi- and hyper-spectral imaging devices can be integrated with the TLS systems, and the spectral signatures can be used for information enhancement. The spectral features can reflect the biophysical and biochemical tree attributes, which serve as the substantial markers of tree species. Such integrated systems have emerged in market and laboratories, and the associated approaches have been put forward for object classification (González-Aguilera et al., 2011), even already for plot-level forest inventory (Vastaranta et al., 2009). Overall, adding other RS instrumental attachments to the common TLS systems may mean a lot for finishing the task of tree species classification.

5. Conclusion

This study has validated TLS for tree species classification, in the case of four typical boreal tree species. The strength of TLS on tree structure characterization at fine scales was fully made use of, and explicit tree structure feature parameters were proposed and derived. Based on these feature parameters, this study as a leading endeavor has further established a procedural frame for TLS-based tree species classification. These both are evidenced by the results that based on the derived feature parameters and the used LOOCV-SVM algorithm, the total accuracy of the classification proved to reach 90.00% in the maximum mode and 77.5% in the robust mode. Overall, this work has achieved the goal of filling the technical gap, i.e. on the one hand, TLS has been increasingly applied for forest inventory; on the other hand, TLS has been almost absent from accomplishing the task of tree species classification.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Grant No. 41471281), in part by the Beijing Natural Science Foundation (Grant No. 4154074), in part by the Research Fund for Doctoral Program of Higher Education of China (Grant No. 20130001120016), and in part by the SRF for ROCS, SEM, China. Thank the Mobile Mapping group in the Finnish Geodetic Institute for distributing the TLS data. The work of Martin Herold for terrestrial LIDAR has been supported by Silvacarbon research program. Sincere thanks to the two anonymous reviewers for their instructive comments.

References

Alberti, G., Boscutti, F., Pirotti, F., Bertacco, C., De Simon, G., Sigura, M., Pirotti, F., Bonfani, P., 2013. A LiDAR-based approach for a multi-purpose characterization of Alpine forests: an Italian case study. iForest Biogeosci. For. 6, 156–168.

Brandtberg, T., 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS J. Photogramm. Remote Sens. 61, 325–340.

Calders, K., Armston, J., Newnham, G., Herold, M., Goodwin, N., 2014. Implications of sensor configuration and topography on vertical plant profiles derived from terrestrial LiDAR. Agric. For. Meteorol. 194, 104–117.

Calders, K., Newnham, G., Burt, A., Raumonen, P., Herold, M., Culvenor, D., Avitabile, V., Disney, M., Armston, J., Kaasalainen, M., 2015. Nondestructive estimates of above-ground biomass using terrestrial laser scanning. Methods Ecol. Evol. 6, 198–208.

Chang, C.-C., Lin, C.-J., 2014. LIBSVM: A Library for Support Vector Machines, Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm (accessed 12.06.14).

Colgan, M., Baldeck, C., Feret, J., Asner, G., 2012. Mapping Savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. Remote Sens. 4, 3462–3480.

Cope, J., Corney, D., Clark, J., Remagnino, P., Wilkin, P., 2012. Plant species identification using digital morphometrics: a review. Expert Syst. Appl. 39, 7562–7573.

Dalponte, M., Bruzzone, L., Gianelle, D., 2008. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. IEEE Trans. Geosci. Remote Sens. 46, 1416–1427.

Dalponte, M., Bruzzone, L., Gianelle, D., 2012. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. Remote Sens. Environ. 123, 258–270.

Du, S., Wang, Y.L., Kume, T., Zhang, J.G., Otsuki, K., Yamanaka, N., Liu, G.B., 2011. Sapflow characteristics and climatic responses in three forest species in the semiarid Loess Plateau region of China. Agric. For. Meteorol. 151, 1–10.

Franklin, S., Hall, R., Moskal, L., Maudie, A., Lavigne, M., 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. Int. J. Remote Sens. 21, 61–79.

George, R., Padalia, H., Kushwaha, S., 2014. Forest tree species discrimination in western Himalaya using EO-1 Hyperion. Int. J. Appl. Earth Obs. Geoinf. 28, 140–149.

González-Aguilera, D., Rodríguez-Gonzálvez, P., Gómez-Lahoz, J., 2011. An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. ISPRS J. Photogramm. Remote Sens. 64, 308–316.

Han, A.R., Lee, S.K., Suh, G.U., Park, Y., Park, P.S., 2012. Wind and topography influence the crown growth of *Picea jezoensis* in a subalpine forest on Mt. Deogyu, Korea. Agric. For. Meteorol. 166–167, 207–214.

Hauglin, M., Gobakken, T., Astrup, R., Ene, L., Nesset, E., 2014. Estimating single-tree crown biomass of Norway spruce by airborne laser scanning: a comparison of methods with and without the use of terrestrial laser scanning to obtain the ground reference data. Forests 5, 384–403.

Heikkinen, V., Tokola, T., Parkkinen, J., Korpela, I., Jääskeläinen, T., 2010. Simulated multispectral imagery for tree species classification using support vector machines. IEEE Trans. Geosci. Remote Sens. 48, 1355–1364.

Heinzel, J., Koch, B., 2011. Exploring full-waveform LiDAR parameters for tree species classification. Int. J. Appl. Earth Obs. Geoinf. 13, 152–160.

Heinzel, J., Koch, B., 2012. Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. Int. J. Appl. Earth Obs. Geoinf. 18, 101–110.

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

Holopainen, M., Kankare, V., Vastaranta, M., Liang, X., Lin, Y., Vaaja, M., Yu, X., Hyyppä, J., Hyyppä, H., Kaartinen, H., Kukko, A., Tanhuanpää, T., Alho, P., 2013. Tree mapping using airborne, terrestrial and mobile laser scanning – a case study in a heterogeneous urban forest. Urban For. Urban Green. 12, 546–553.

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.

Hosoi, F., Omasa, K., 2006. Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. IEEE Trans. Geosci. Remote Sens. 44, 3610–3618.

Kantola, T., Vastaranta, M., Yu, X., et al., 2010. Classification of defoliated trees using tree-level airborne laser scanning data combined with aerial images. Remote Sens. 2, 2665–2679.

Korpela, I., Orka, H.O., Maltamo, M., Tokola, T., Hyyppa, J., 2010. Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fennica 44, 319–339.

Krahwinkler, P., Rossmann, J., 2013. Tree species classification and input data evaluation. Eur. J. Remote Sens. 46, 535–549.

Kunert, N., Schwendenmann, L., Potvin, C., Hölscher, D., 2012. Tree diversity enhances tree transpiration in a Panamanian forest plantation. J. Appl. Ecol. 49, 135–144.

Li, J., Hu, B., Noland, T.L., 2013. Classification of tree species based on structural features derived from high density LiDAR data. Agric. For. Meteorol. 171–172, 104–114.

Liang, X., Hyyppä, J., 2013. Automatic stem mapping by merging several terrestrial laser scans at the feature and decision levels. Sensors 13, 1614–1634.

Lin, Y., Hyyppä, J., Kukko, A., Jaakkola, A., Kaartinen, H., 2012. Tree height growth measurement with single-scan airborne, static terrestrial and mobile laser scanning. Sensors 12, 12798–12813.

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.

Mascaro, J., Hughes, R.F., Schnitzer, S.A., 2012. Novel forests maintain ecosystem processes after the decline of native tree species. Ecol. Monogr. 82, 221–228.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1778–1790.

Monnier, F., Vallet, B., Soheilian, B., 2012. Trees detection from laser point clouds acquired in dense urban areas by a mobile mapping system. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. I-3, 245–250.

Nagy, M.T., Janssens, I.A., Yuste, J.C., Carrara, A., Ceulemans, R., 2006. Footprint-adjusted net ecosystem CO_2 exchange and carbon balance components of a temperate forest. Agric. For. Meteorol. 139, 344–360.

Naidoo, L., Cho, M.A., Mathieu, R., Asner, G., 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. ISPRS J. Photogramm. Remote Sens. 69, 167–179.

Othmani, A., Voon, L.L.Y., Stolz, C., Piboule, A., 2013. Single tree species classification from terrestrial laser scanning data for forest inventory. Pattern Recognit. Lett. 34, 2144–2150.

Pfeifer, N., Briese, C., 2007. Geometrical aspects of airborne laser scanning and terrestrial laser scanning. Int. Arch. Photogramm. Remote Sens. 36 (3/W52), 311–319.

Pinard, M.A., Putz, F.E., Rumiz, D., Guzman, R., Jardim, A., 1999. Ecological characterization of tree species for guiding forest management decisions in seasonally dry forests in Lomerio, Bolivia. For. Ecol. Manage. 113, 201–213.

Pirotti, F., Guarnieri, A., Vettore, A., 2013. Vegetation filtering of waveform terrestrial laser scanner data for DTM production. Appl. Geomat. 5, 311–322.

Puttonen, E., Suomalainen, J., Hakala, T., Räikkönen, E., Kaartinen, H., Kaasalainen, S., Litkey, P., 2010. Tree species classification from fused active hyperspectral reflectance and LIDAR measurements. For. Ecol. Manage. 260, 1843–1852.

Roxby, G., Howard, T., Lee, T., 2015. Effects of whole-tree harvesting on species composition of tree and understory communities in a northern hardwood forest. Open J. For. 5, 235–253.

Seidel, D., Hoffmann, N., Ehbrecht, M., Juchheim, J., Ammer, C., 2015. How neighborhood affects tree diameter increment – new insights from terrestrial laser scanning and some methodical considerations. For. Ecol. Manage. 336, 119–128.

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., Ni-Miester, W., Boykin-Morris, W., 2008. Retrieval of forest structural parameters using a ground-based lidar instrument (echidna). Can. J. Remote Sens. 34, 426–440.

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.

Vastaranta, M., Melkas, T., Holopainen, M., Kaartinen, H., Hyyppä, J., Hyyppä, H., 2009. Laser-based field measurements in tree-level forest data acquisition. Photogramm. J. Finl. 21, 51–61.

Yao, W., Krzystek, P., Heurich, M., 2012. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. Remote Sens. Environ. 123, 368–380.

Zhang, K., Hu, B., 2012. Individual urban tree species classification using very high spatial resolution airborne multi-spectral imagery using longitudinal profiles. Remote Sens. 4, 1741–1757.

Zhou, H., Lin, Y., 2013. Estimation of the height of the first live branch (FLBH) from stop-and-go mobile laser scanning data. In: SilviLaser2013, Beijing, China, October, pp. 108–115.