Improved Salient Feature-Based Approach for Automatically Separating Photosynthetic and Nonphotosynthetic Components Within Terrestrial Lidar Point Cloud Data of Forest Canopies

Lixia Ma, Guang Zheng, Jan U. H. Eitel, L. Monika Moskal, Wei He, and Huabing Huang

***Abstract*—Accurate separation of photosynthetic and nonphotosynthetic components in a forest canopy from 3-D terrestrial laser scanning (TLS) data is a challenging but of key importance to understand the spatial distribution of the radiation regime, photosynthetic processes, and carbon and water exchanges of the forest canopy. The objective of this paper was to improve current methods for separating photosynthetic and nonphotosynthetic components in TLS data of forest canopies by adding two additional filters only based on its geometric information. By comparing the proposed approach with the eigenvalues plus color information-based method, we found that the proposed approach could effectively improve the overall producer's accuracy from 62.12% to 95.45%, and the overall classification producer's accuracy would increase from 84.28% to 97.80% as the forest leaf area index (LAI) decreases from 4.15 to 3.13. In addition, variations in tree species had negligible effects on the final classification accuracy, as shown by the overall producer's accuracy for coniferous (93.09%) and broadleaf (94.96%) trees. To remove quantitatively the effects of the woody materials in a forest canopy for improving TLS-based LAI estimates, we also computed the "woody-to-total area ratio" based on the classified linear class points from an individual tree. Automatic classification of the forest point cloud data set will facilitate the application of TLS on retrieving 3-D forest canopy structural parameters, including LAI and leaf and woody area ratios.**

***Index Terms*—Light detection and ranging (lidar), pattern recognition, point classification, terrestrial laser scanning (TLS), woody-to-total area ratio.**

Manuscript received February 18, 2014; revised July 29, 2014, December 21, 2014, March 11, 2015, and May 14, 2015; accepted July 11, 2015. Date of publication September 14, 2015; date of current version January 19, 2016. This work was supported in part by the National Natural Science Foundation of China under Award 41201337, by the High Technology Research and Development Program of China (863 Program) under Grant 2012AA12A306, and by the Special Climate Change Fund under Grant CCSF201412.

L. Ma, G. Zheng, and W. He are with the International Institute for Earth System Science, Nanjing University, Nanjing 210023, China (e-mail: zhengguang@nju.edu.cn).

J. U. H. Eitel is with the Reveley Geospatial Laboratory for Environmental Dynamics, College of Natural Resources, University of Idaho, Moscow, ID 83844, USA.

L. M. Moskal is with the Remote Sensing and Geospatial Analysis Laboratory, Precision Forestry Cooperative, School of Environment and Forest Science, University of Washington Seattle, WA 98195 USA.

H. Huang is with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing 100101, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2015.2459716

I. Introduction

FOREST canopy structure, which is defined as the shape, size, orientation, and spatial distribution of foliage elements [1], controls the energy transfer between the atmosphere and terrestrial ecosystems by affecting photosynthesis and evapotranspiration processes [2]–[6]. Quantitative characterization of the forest 3-D structure, particularly differentiating the photosynthetic (i.e., leaf, flowers, etc.) and nonphotosynthetic components (i.e., stem, branches, etc.), is critical to retrieve biophysical parameters such as LAI defined as one half of the total green leaf area per unit ground surface area [7]. For mapping forest 3-D structure, light detection and ranging (lidar) is increasingly used. This relatively novel technology determines the x, y, and z location of a surveyed point $\boldsymbol{q}$ based on a laser-based distance measurement between the lidar and the surveyed object and two angle measurements (azimuth and zenith angles) [8]. Usually, the attributes of any point q from a given discrete point cloud data set generated from TLS could be summarized as

$$\boldsymbol{q}(x, y, z, i_1, i_2, \ldots, i_n, \text{ r, g, b, NIR, } l_1, l_2, \ldots, l_{n'}) \quad (1)$$

where x, y, z are the three Cartesian coordinates for each point, $i_1, i_2, \ldots, i_n$ denote the laser return intensity information associated with each point, and n depends on the number of laser beams with different wavelengths. Danson *et al.* [9] has recently developed a dual-wavelength lidar system to retrieve the forest canopy structure; r, g, b, NIR are red, green, blue, and near-infrared components for a color image captured with the built in digital camera of a lidar system, and $l_1, l_2, \ldots, l_{n'}$ are different labels for each point. For example, the points from leaf class could be labeled as l_{leaf}, and the curvature or normal vector of point q can be marked as l_{curv} and l_{normal}, respectively. Laser return intensity values are affected by the reflective properties of surveyed objects and hence might be useful for automatically mapping canopy components. For example, Yang *et al.* [10] reconstructed the different foliage element components of forest stands based on the different intensity returns collected by a dual-wavelength TLS. However, laser return intensity values are also affected by the distance between the lidar and the surveyed object (inverse distance square law) and the angle of incidence [11]. Hence, before using laser return information for classification purposes, laser

return intensity information has to be calibrated, which remains challenging [12]–[14]. In addition to the laser return intensity information, the digital number (DN) values associated with the r, g, b, and NIR information from the digital camera could be helpful to automatically classify forest structural parameters. However, the dimensionless and uncalibrated DN values are highly variable even within a single digital image resulted from a complex interaction between variable light environment to light exposure control. Due to the calibration issues associated with DN values and laser return intensity, many studies to date have been focusing on using geometric features (i.e., x, y, z information) contained in lidar point clouds for automatically mapping canopy components. Some studies [15]–[17] used ordered eigenvalues and eigenvectors for mapping canopy components based on the x, y, z information provided by lidar.

To classify photosynthetic and nonphotosynthetic components in a laser point cloud, many studies to date have used manual-based techniques where laser points associated with different canopy components are visually identified. For example, Watt *et al.* [18] manually separated the stem point cloud from the whole tree point cloud, which improved TLS-based estimates of diameter at breast height (DBH) and height of forest canopies. Similarly, in [19], the stem points were manually isolated from an individual tree point cloud to estimate the biomass and DBH. In [20], the leaf area density are obtained after manually separating the nonphotosynthetic components from the registered points cloud data generated using both ALS and TLS in the same plot; then, the results with field-based measurements were compared. In [21] and [22], it was found that isolation of the woody materials from the forest canopy could effectively improve the accuracy of tree biomass estimation compared with the traditional allometry-based method.

Isolating the photosynthetic and nonphotosynthetic canopy components is also critical in retrieving true LAI. LAI is usually estimated using the gap-fraction [23] and gap-size theories [24], [25]: the gap-fraction theory assumes that the foliage elements within the canopy are randomly distributed [1], [26], but this assumption could be removed based on the gap-size information in the canopy [25]. In recent years, many studies attempted to develop various algorithms for different lidar systems (i.e., discrete or waveform) to estimate LAI [27]–[29]. Zheng and Moskal [30], [31] developed two different approaches to retrieve effective LAI using discrete TLS through computational geometry and geometric projection-based methods, respectively, and they validated the results using the destructive sampling and digital hemispherical photograph (DHP) techniques. Hosoi and Omasa [32] found that nonphotosynthetic component could contribute ranging from 4.2% to 32.7% to LAI estimation errors for different tree canopies; thus, the accuracy of LAI estimation could be greatly improved if the photosynthetic and nonphotosynthetic components' points could be separated. The parameter of "woody-to-total area ratio" defined as the ratio of woody surface area and the sum of woody and total leaf surface area [33] is a necessary parameter for true LAI estimation by excluding the contribution of nonphotosynthetic components to the whole forest canopy. The leaf area can be obtained by multiplying the LAI by the projected area of a forest canopy on horizontal ground surface, and the woody surface area is the summation of the stem and branch surface areas.

TABLE I
CHARACTERISTICS OF TLS LEICA SCAN STATION 2

Single shot time-of-flight measurement	
Only first return recorded for each laser beam	
Scanning speed	50,000 pts/sec
Field of view	Horizontal: 0 ~ 360 degrees
	Vertical: -45 ~ 90 degrees
Laser beam spot size	4mm (FWHH-based)
	6mm(Gaussian-based)
Minimum angle resolution	0.3mm (3e-06 radians)
Wavelength	532nm (visible green light)
Detection range	300m @90% reflectivity

The labor-intensive and time-consuming manual-based methods for classifying canopy components within lidar point cloud data will limit the use of lidar data at broader spatial scales for estimating biophysical parameters such as LAI. Therefore, a robust and reliable classification method is needed for automatically classifying canopy components from lidar point cloud data. The objectives of this paper are as follows:

1) to improve the current method for separating the 3-D forest point cloud data into nonphotosynthetic canopy components (i.e., stems and branches), photosynthetic canopy components (i.e., leaves), and bare ground based on salient feature and evaluate its accuracy;
2) to develop a method to estimate the parameter of "woody-to-total area ratio" in order to retrieve the true LAI;
3) to explore the effects of forest density, tree species, and 3-D spatial searching radius on the final classification accuracy through sensitivity analysis.

II. MATERIALS

A TLS and Hemispherical Camera Data

Two study areas were chosen in this paper to test the proposed method and its applicability. The first study area was located in the Washington Park Arboretum (WPA) near Seattle ($47°38'08''$ N, $122°17'46''$ W), Washington, USA. In September 2008, three forest plots were established and scanned using the Leica Scan Station 2 (Leica Geosystem AG, St. Gallen, Switzerland) (see Table I). Each plot was scanned from one central scan location using the hemispherical scanning mode (horizontal: 0° to 360°; vertical: $-45°$ to 90°) with 0.1-m sampling spacing at 30 m. In addition to the center scan at plot-4, one additional scan of a Douglas fir (*Pseudotsuga menziesii*) tree and a New Mexican locust (*Robinia Neomexicana*) tree was collected for the purposes of testing the performance of the proposed method on different tree species and comparison with other methods. In addition, the hemispherical photos with an image resolution of 2048 by 1360 were captured with the digital camera (Nikon CoolPix 4500, Nikon, Inc., Melville, NY) mounted with a fish-eye lens (Nikon FC-E9, 8 mm, f/2. 8) for

each forest plot. The images were taken at the same height as the TLS with two stops lower than the automatic shutter speed in this paper [34].

In October 2009, a second experiment was conducted in a mature forest stand at the Wind River Experimental Forest (WREF)($45°13.76''$ N, $121°6.88''$ W), Washington, USA. At this site, a single plot was established and scanned from a central scan location using the full field-of-view range (horizontal: $0°-360°$, vertical: $-45°-90°$). The dominant tree species in the forest plot was mature Douglas fir (*Pseudotsuga menziesii*). The mean canopy height at this site is around 50 m with high canopy closure and only a few shrubs and small trees on the ground. The point cloud was sampled with a point spacing of 9 mm at 50 m.

B. Validation Data set

We clipped the point cloud data for four different forest plots from both WPA and WREF sites as circular plots with radius of 30 m, and we then visually identified points in the lidar point cloud associated to the following three classes: 1) photosynthetic canopy components (i.e., leaf, flowers, etc.); 2) nonphotosynthetic canopy components (i.e., stem and branches); and 3) ground surface. The visually classified data set was saved as reference data sets for validating the computer-based classification. We combined the visual inspection and hemispherical photo information to ensure the accuracy of manually selected validation data sets. For example, we ensure the accuracy of the bare ground and shrub point validation data by visually inspecting the TLS point cloud. To ensure the classification accuracy of the validation data set of the linear class, we used hemispherical photos to visually separate the branches from the leaves.

III. Methods: GAFPC Algorithm

The geometric-based automatic forest point classification (GAFPC) algorithm was developed by improving Lalonde's method [15] in this paper. Thirty training point data sets have to be manually selected for each of the three classes (photosynthetic canopy components, nonphotosynthetic canopy components, and ground) in the four forest plots. For each forest plot, five stem and five branch point training data sets were selected at the 30%, 60%, and 90% percentile heights of the bounding box of the forest plot. Further, ten bare ground point training data sets were selected at the annulus regions with the radii of 10, 20, and 30 m, respectively. For the random class, we selected ten leave training point data sets with $90°$ azimuthal angle interval in three different horizontal planes within the 30%, 60%, and 90% percentile height of the forest canopy, respectively. The proportions of the training data points to total forest points was about 15%. Based on the geometrical spatial distribution patterns (i.e., salient feature) of the manually selected training points for each class, we used the expectation–maximization (EM) algorithm to estimate the model parameters for Gaussian mixture models (GMMs) of each class, which served as a classifier for pointwise classification. In addition, six different filters were used to postprocess the preliminary classification. By doing this, the given forest point cloud data were separated into three different categories, including "photosynthetic canopy components" with random features (i.e., leaf and brush), "nonphotosynthetic canopy components" with linear features (i.e., stem and branch), and "ground surface" with surface features (i.e., bare ground). Finally, the contribution of woody canopy components to the overall forest canopy point cloud was quantitatively computed based on the classified linear class points.

A. Salient Feature

If $\boldsymbol{q} = (x, y, z) \in \mathbb{R}^3$ is defined as a point in 3-D space, then $Q = \{\boldsymbol{q}_i \in \mathbb{R}^3 | i = 1, \ldots, N\}$ denotes the 3-D point cloud data of a forest stand. It implicitly contains the structural information of foliage elements based on the different spatial distribution patterns for local point sets. A local point set is a small subset of point cloud data within a certain distance around a given point and defined as

$$V_R(\boldsymbol{q}') = \left\{\boldsymbol{q} \in \mathbb{R}^3 | \text{dist}(\boldsymbol{q} - \boldsymbol{q}') \leq R\right\} \tag{2}$$

where $\boldsymbol{q}(x, y, z)$ is any point of n points for a local point set $V_R(\boldsymbol{q}')$, $\boldsymbol{q}'(x', y', z')$ is the center point, R is radius, and $\text{dist}(\boldsymbol{q} - \boldsymbol{q}')$ is the Euclidean distance between the point of $\boldsymbol{q}$ and $\boldsymbol{q}'$. The matrix format can be written as follows.

Let matrix

$$A = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix}_{n \times 3} \tag{3}$$

be a local point set $V_R(\boldsymbol{q}')$ and the center matrix of $V_R(\boldsymbol{q}')$ be

$$B = \begin{bmatrix} x' & y' & z' \\ x' & y' & z' \\ \vdots & \vdots & \vdots \\ x' & y' & z' \end{bmatrix}_{n \times 3}. \tag{4}$$

The covariance matrix C_{cov} of $V_R(\boldsymbol{q}')$ can be written as

$$C_{\text{cov}} = \frac{1}{n}(A - B)^T(A - B). \tag{5}$$

The ordered eigenvalues of this covariance matrix are $\lambda_0 \geq \lambda_1 \geq \lambda_2$, which can be used to quantitatively characterize the spatial distribution pattern of a given local point set $V_R(\boldsymbol{q}')$.

Usually, the point sets from different foliage elements of a forest stand has three different spatial distribution patterns (see Fig. 1), which has been summarized in [15] as follows: 1) The points from leaves, flowers, or brushes are randomly distributed with corresponding eigenvalues as $\lambda_0 \approx \lambda_1 \approx \lambda_2$ [see Fig. 1(a)]; 2) branches and stems tend to be the linearly distributed with the ordered eigenvalues as $\lambda_0 \gg \lambda_1 \approx \lambda_2$ [see Fig. 1(b)], and λ_0 is much larger than other two similar eigenvalues λ_1 and λ_2; and 3) as for ground points, the spatial distribution pattern could be characterized by the ordered eigenvalues as $\lambda_0 \approx \lambda_1 \gg \lambda_2$ [see Fig. 1(c)], and λ_2 is much smaller than the other two similar eigenvalues λ_0 and λ_1. Then, the salient

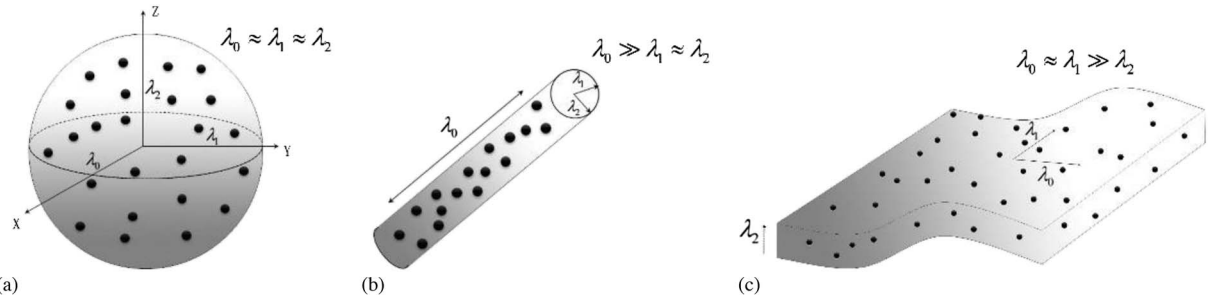


Fig. 1. Schematic to illustrate the salient features of three different spatial distribution patterns. $\lambda_0, \lambda_1, \lambda_2$ are the eigenvalues of the covariance matrix of a local point set $V_R(\boldsymbol{q}')$, respectively. (a) "Random" distribution with similar eigenvalues in three directions ($\lambda_0 \approx \lambda_1 \approx \lambda_2$). (b) "Linear" distribution with a bigger eigenvalue in one of three directions ($\lambda_0 \gg \lambda_1 \approx \lambda_2$). (c) "Surface" distribution with a smaller eigenvalue in one of the three directions ($\lambda_0 \approx \lambda_1 \gg \lambda_2$).

feature of a given forest local point set $V_R(\boldsymbol{q}')$ can be defined as the linear combination of the eigenvalues as

$$S\left(V_R(\boldsymbol{q}')\right) = (\lambda_2, \lambda_0 - \lambda_1, \lambda_1 - \lambda_2). \tag{6}$$

B. Feature Distribution Model

Statistically, any continuous probability density function $p(x)$ can be approximated by the linear combination of several probability density functions $p(x|k)$

$$p(x) = \sum_{k=1}^{m} w_k p(x|k) \tag{7}$$

where m is the number of probability density function, w_k is the weight for each probability density function, $\sum_{k=1}^{m} w_k = 1$, and $p(x|k)$ is the conditional probability of any point x belonging to the kth probability density function. The GMM, which is built by the linear combination of several Gaussian density functions (GDFs), is usually used to approximate the continuous probability density function [35]. For a given forest point cloud data set, the probability density functions of the salient features for each class can be approximated by the corresponding GMM. For any point $x_j (j = 1, 2, \ldots, N)$ of a given class point cloud data set, the probability density x_j belonging to the GMM can be defined as

$$\begin{aligned} p'(x_j) &= \sum_{k=1}^{m} w_k p(x_j|k) \\ &= \sum_{k=1}^{m} w_k \frac{1}{(2\pi)^{\frac{\bar{d}}{2}} \left|\sum_k\right|^{\frac{1}{2}}} \\ &\quad \times \exp\left(-\frac{1}{2}(x_j - \mu_k)^T \sum_k^{-1} (x_j - \mu_k)\right) \end{aligned} \tag{8}$$

where $p(x_j|k)$ is the probability x_j belonging to kth GDF, μ_k is the mean value of the kth GDF, $\sum_k$ is the covariance matrix of the kth GDF, and $\bar{d}$ is the number of dimension for the salient feature. We manually selected the training sample data set from a specific class $\{x_1, \ldots, x_m\}$ and computed the salient feature for each sample point set. Then, the EM algorithm was used to estimate the w_k, μ_k, and $\sum_k$ of the GMMs for linear, random, and surface classes, respectively. Usually, the GMM with three different GDFs can well approximate a continuous probability density function, which has been verified in this paper and suggested in [15].

C. Pointwise Classification

We used the pointwise classification method to classify the given forest point cloud data by only considering the geometric information. For each point from a local point set $V_R(\boldsymbol{q}') = \{\boldsymbol{q} \in \mathbb{R}^3 | \text{dist}(\boldsymbol{q} - \boldsymbol{q}') \leq R\}$, its salient feature $(\lambda_2, \lambda_0 - \lambda_1, \lambda_1 - \lambda_2)$ was input into three GMMs to compute the probability of the point belonging to one of the three classes: 1) photosynthetic canopy components (i.e., leaf, flowers, etc.); 2) nonphotosynthetic canopy components (i.e., stem and branches); and 3) ground surface. The point was then assigned to the class for which the GMMs determined the highest probability.

The determination of the search radius is a critical step in the pointwise classification since it may affect the classification accuracy. In this paper, the K-dimensional tree ($k - d$ tree) was used to spatially search the nearest neighbor points for a given point in 3-D space, and the radius of the searching ball was set to 0.45 m based on our sensitivity analysis to well balance the classification accuracy and computational efficiency.

The GMM for the u class (i.e., one of the three different classes including linear, surface, and random classes) can be written as

$$M^u = \left\{ \left(w_k^u, \mu_k^u, \sum_k^u \right) \right\}_{k=1,\ldots,m} \tag{9}$$

where m is the number of GDF for the u class, w_k^u is the weight of the kth GDF of u class, $\sum_k^u$ is the covariance matrix of the kth GDF of the u class, and μ_k^u is the mean value of kth GDF of the u class. Thus, the special condition for the conditional probability of the point belonging to one class can be computed by inputting the salient feature $S(V_R(\boldsymbol{q}')) = (\lambda_2, \lambda_0 - \lambda_1, \lambda_1 - \lambda_2)$ of each point as

$$\begin{aligned} P'\left(S\left(V_R(q')\right)|M^u\right) &= \sum_{k=1,\ldots,m} \frac{w_k^u}{(2\pi)^{\frac{\bar{d}}{2}} \left|\sum_k^u\right|^{\frac{1}{2}}} \\ &\quad \times e^{\frac{-1}{2}\left(S(V_R(\boldsymbol{q}')) - \mu_k^u\right)^T \sum^{-1}\left(S(V_R(\boldsymbol{q}')) - \mu_k^u\right)}. \end{aligned} \tag{10}$$

The point will be labeled as the class with the highest conditional probability $U_{\max}$ quantitatively expressed as follows:

$$U_{\max} = \arg\max_{u} \left\{ P' \left(S \left(V_R(\boldsymbol{q}') \right) | M^u \right) \right\}. \tag{11}$$

This process will be iteratively repeated for each point of a forest-stand-point cloud data set to obtain the preliminary classification.

D. Postprocessing Filtering

Following previous work [15], the preliminary classification results were postprocessed by using four filters: the edge filter, individual surface filter, isolated density filter, and ground filter. In addition to these filters, two new filters were introduced in this paper, i.e., the enhanced ground filter-1 and enhanced ground filter-2. The six different filters were orderly applied to the preliminary classification. The filters make use of search spheres with different search radii. The search radii were determined based on the sensitivity analysis of various radii on classification accuracy (more details in discussion section). Each of the filters applied is described in detail in the following.

Edge Filter (Filter-1): The purpose of this filter is to remove misclassified linear points. For this, a search sphere with a radius of 1 m was created around each point that was previously classified as a linear feature, and each point within the search sphere was assigned into the class with the highest counting number.

Isolated Surface Filter (Filter-2): The purpose of this filter is to remove misclassified ground points. For each of the points that were newly classified as bare ground points by filter-1, this filter creates a search sphere with a radius of 1.5 m. The point is then assigned into the class with the highest counting number within the search sphere.

Isolated Density Filter (Filter-3): The purpose of this filter is to remove isolated points (e.g., outliers), for example, caused by laser returns from flying insects or ghost points. It iteratively searches all points using a searching ball with the radius as 0.45 m and removes the point if the number of points within the searching ball is smaller than 5, as determined by sensitivity analysis.

Ground Filter (Filter-4): The purpose of filter-4 is to find ground points. For this, filter-4 iteratively scans the non-surface class points using a facing-down searching cone. Following [15], each point labeled as non-surface class after applying filter-3 is set as the apex point with top angle ranging from 10° to 20° [36], and the cone height was set as 8 m, which equals to the height difference between the crown base and ground surface in this paper. Theoretically, the apex point will be relabelled as surface class if there is more than one point within the cone; however, three were used due to the slope variation ranging from 0° to 10°.

Enhanced Ground Filter-1 (Filter-5): The purpose of this filter is to correct stem points misclassified into the surface class. For this, the filter algorithm creates a down-facing search cone, with an apex and search radius of 0.45 m for each ground point. The point is relabeled as a linear class if the number of points within the searching cone is greater than 15 and the height difference between the highest and lowest points within the searching cone was set as 0.1 m determined by the topographic variation (0°–10° in this paper). The threshold of 15 was determined by a histogram analysis of all points within the different search spheres.

Enhance Ground Filter-2 (Filter-6): The purpose of this filter is to remove canopy points previously misclassified as the surface class. An additional criterion was set as the height attribute of each point; the point with Z value (< 0) (i.e., higher than the scanner height) was labeled as random class. This criterion was determined based on our visual inspection of the topographic variation for this field plot.

E. Woody-to-Total Area Ratio

"Woody-to-total area ratio" is a parameter suggested in [37] to remove the contribution of woody biomass (i.e., nonphotosynthetic components) to the total forest surface area when estimating true LAI. Let N_p and N_{np} denote the number of points from the photosynthetic canopy components (i.e., leaves and shrubs) and nonphotosynthetic canopy components (e.g., stems and branches), respectively. s_i is the sampling spacing of a point at the distance d_i away from the TLS, the sampling spacing was predefined as 0.1 m with 4 mm spot size at 30 m in this paper. The quantitative relationship [see (12)] between the distances away from scanner and various sampling spacing can be expressed based on the similar triangle theorem as (units in meters) [see Fig. 2(a)]

$$\frac{s_2}{d_2} = \frac{s_1}{d_1} \tag{12}$$

where s_1 is the sampling spacing at the distance d_1 away from the scanner predefined by the scanner user. Accordingly, the sampling spacing s_2 at the point with d_2 distance away from scanner can be computed as

$$s_2 = (0.1 \cdot d_2/30) = d_2/300 \tag{13}$$

and the corresponding surface area represented by this point cloud, which is perpendicular to the laser beam direction, can be obtained by computing $s_2 \times s_2$.

However, the apparent leaf area varies with changes in leaf orientation and direction of the incident laser beam. We assume that the computed surface area represented by a laser point is the projected area of the real foliage element's surface on the plane perpendicular to the incident laser beam direction. The orientation of each TLS laser point could be obtained through computing its normal vector within a neighbor region around this point [38]. As shown in Fig. 2(b), p_i is one sample point from a leaf, AP_i is the surface area represented by the point p_i with sampling spacing as s_2 on the projection plane perpendicular to the direction of incident laser beam ($\boldsymbol{OP}_i$), AR_i is the real leaf surface area represented by the point p_i in the real 3-D space, θ_1 is the intersection angle of the projected surface area and the real leaf surface, and θ_2 is the intersection angle of the lines in which normal vectors ($\boldsymbol{r}_i$ and $\boldsymbol{OP}_i$) of the real leaf surfaces and projected surfaces stay $(0-90^\circ)$; it is easily known that $\theta_1 = \theta_2$. Since we assumes that the surface

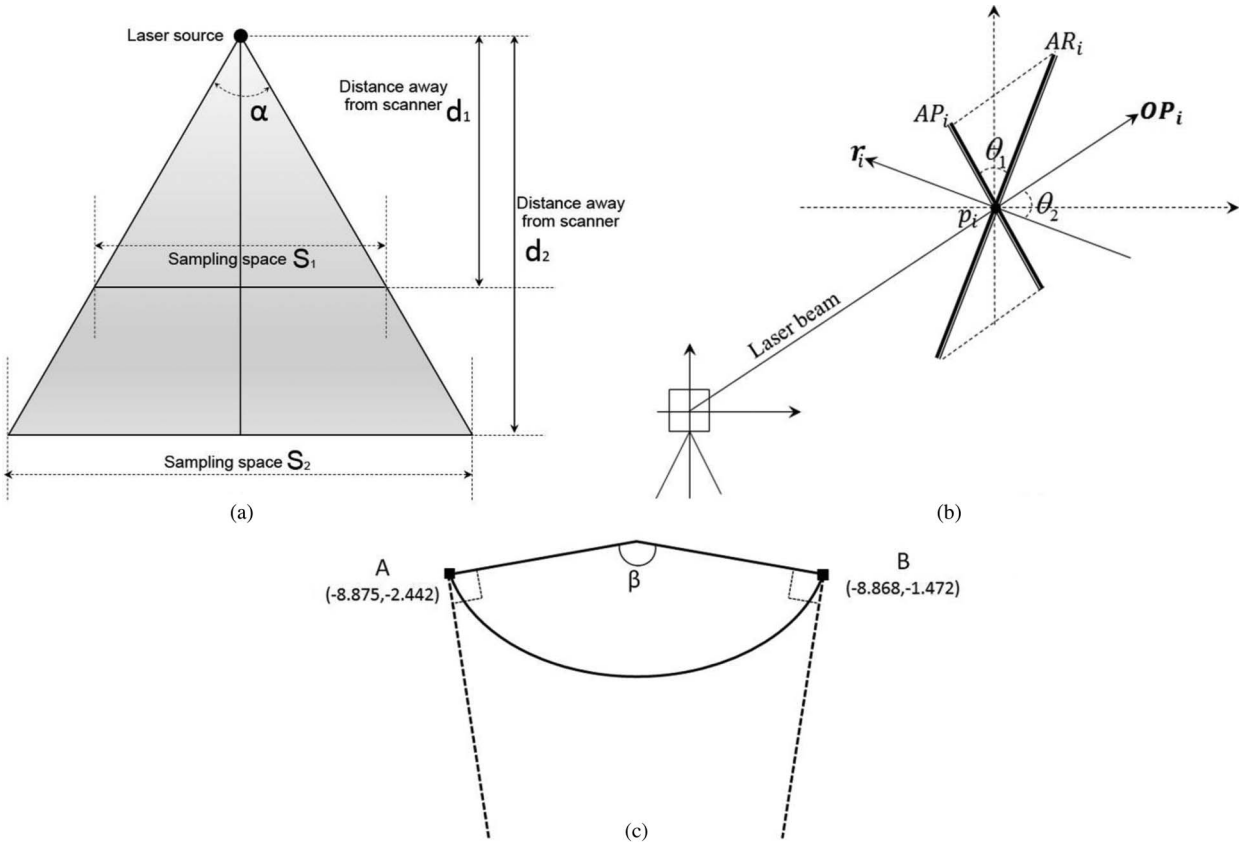


Fig. 2. Schematic to illustrate the (a) relationship between the distances away from scanner and corresponding sampling spaces, (b) geometric relationship between the surface area represented by a discrete point in 3-D space, and (c) proportion of the partially sampled stem or branch surface. In (a), the laser beams are emitted from a point light source with an interval angle of α. S_1 and S_2 are the corresponding sampling spaces at d_1 and d_2 distances away from scanner, respectively. In (b), p_i is one sample point from a leaf, AP_i is the surface area represented by the point p_i with sampling space s_2 as on the projection plane perpendicular to the direction of incident laser beam ($\boldsymbol{OP}_i$), AR_i is the real leaf surface area represented by the point p_i in the real 3-D space, θ_1 is the intersection angle of the projected surface area and the real leaf surface, and θ_2 is the intersection angle of the normal vectors ($\boldsymbol{r}_i$ and $\boldsymbol{OP}_i$) of the real leaf surfaces and projected surfaces. (c) Angle of chord consisted from the center point of the imaged cross-sectional circle of the stem and the end edge points A and B from the partially sampled stem surface; the dashed lines denote the directions of incoming laser beams.

AP_i is the projected area of the surface AR_i in the projected plane perpendicular to the incident laser beam direction, we computed the real leaf surface area represented by the point p_i with the sampling spacing as s_2 using the following: equations:

$$AR_i = \frac{AP_i}{\cos(\theta_1)} \tag{14}$$

$$\theta_2 = \arccos\left|\frac{\boldsymbol{r}_i \cdot \boldsymbol{OP}_i}{|\boldsymbol{r}_i| \times |\boldsymbol{OP}_i|}\right|. \tag{15}$$

Mathematically, a boundary condition $(\cos(\theta_1) = 0.1$, if $\cos(\theta_1) < 0.1)$ was set when computing AR_i to avoid the infinitely large values as the angle θ_1 can be beyond 84.26° in this paper.

By doing this, the number of points for different tree components was converted into surface area, and the "woody-to-total area ratio (ψ)" was calculated as follows:

$$\psi = \frac{A_{\mathrm{np}}}{A_{\mathrm{p}} + A_{\mathrm{np}}} \tag{16}$$

where A_{p} is the total leaf surface area (i.e., photosynthetic canopy components), and A_{np} is the total surface area of stems and branches (i.e., nonphotosynthetic canopy components). The total area of the leaves (e.g., random class points) was calculated as

$$A_{\mathrm{p}} = 2 \times \sum_{i=1}^{N_{\mathrm{p}}} A_{\mathrm{random}}(\mathrm{i}) \tag{17}$$

where A_{random} is the surface area from the points labeled as random class (i.e., photosynthetic canopy component points); coefficient 2 was used here to convert the one-sided broadleaf surface area to a total surface area. The total surface area of nonphotosynthetic canopy components A_{np} was calculated as follows:

$$A_{\mathrm{np}} = \tau \times \sum_{i=1}^{N_{\mathrm{np}}} A_{\mathrm{linear}}(\mathrm{i}) \tag{18}$$

where A_{linear} are the surface area from points classified as branch and stem classes. The coefficient τ converts the partial points into the comprehensive points for area estimation purpose based on the assumption that the stem and branches are cylindrically shaped. The conversion coefficient might be different depending on the extent of comprehensive for the sampled points of forest canopies. In order to determine the

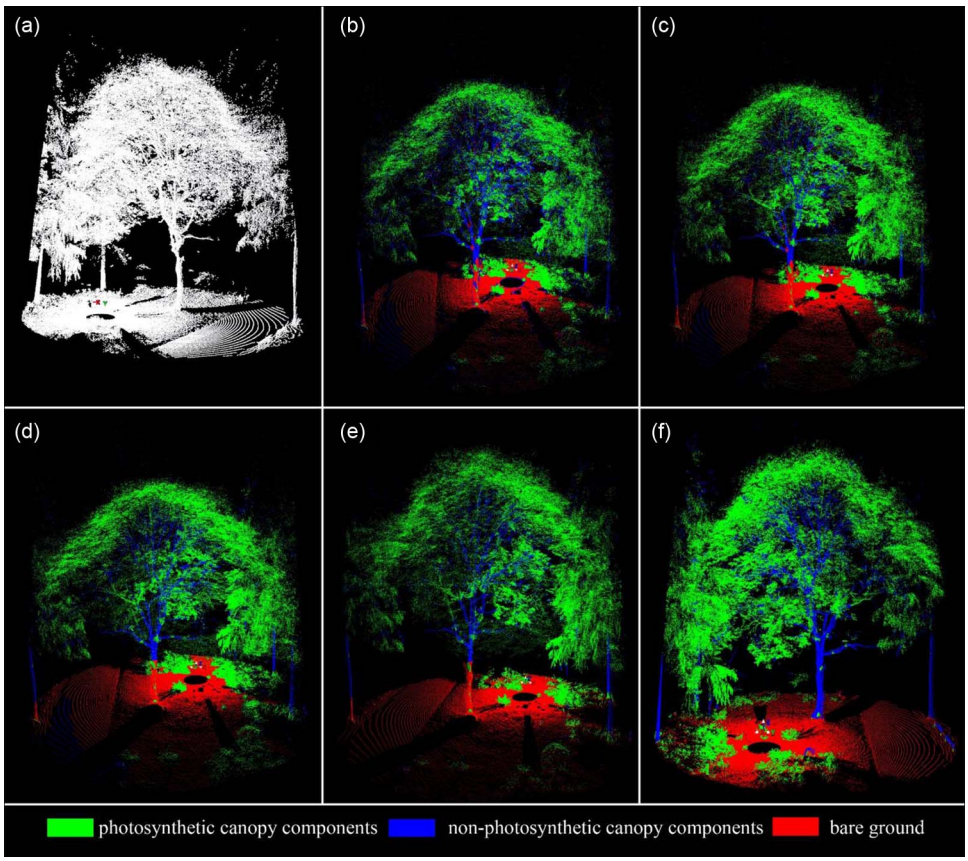


Fig. 3. Different classification stages of an individual tree point cloud data set in WPA forest plot-4. (a) Original point cloud data. (b) Preliminary classification with random class points in green color (i.e., leaves and brushes), linear class points in blue color (i.e., branches and stems), and surface class points in red color (i.e., ground surface). (c) Adjusted classification after removing some misclassified linear points by filter-1. (d) Adjusted classification after removing some misclassified surface points by filter-2. (e) Improved classification after removing the isolated points and misclassified ground points by filter-3 and filter-4. (f) Final classification after correcting misclassified stem surface points by filter-5 and filter-6.

coefficient τ, we extracted a horizontal sliced layer points around the height 1.3 ± 0.5m from one tree of the four WPA forest plots, and computed the angle points A and B [see Fig. 2(c)]. The averaged distance (9.054 m) of distances between scanner to points A and B (i.e., OA $= 9.118$ m and OB $= 8.989$ m) was used for estimating the proportion of the sampled stem cross-sectional arc for the whole stem surface. The computed angle of β around 173.863° allows us to determine the coefficient τ as 2 for converting the discrete point to the surface area; other appropriate parameters should be used according to the comprehensiveness of the collected point cloud data using TLS. We also computed the proportion of the sampled branch using the same method, showing that it is good enough to use the conversion coefficient of 2.

F. Stem Surface Meshing

We reconstructed the mesh surface of major classified stem component points using the Delaunay refinement paradigm [39]. The computed mesh surface area was used to validate the developed method of converting discrete points to surface area by incorporating the point orientation and sampling spacing information, as described in Section III-E.

IV. Results

A Individual Tree Level

1) Visual Accuracy Assessment: After applying the GAFPC algorithm to an individual tree TLS point cloud data in the WPA forest plot-4 [see Fig. 3(a)], we obtained the classification for different stages. Fig. 3(b) shows the preliminary classification with various misclassified points as follows: 1) Some points from the stem have been labeled as surface class; 2) few edge points of surface or random class were misclassified into linear class; 3) some points belonging to random class were labeled as surface class; and 4) some points from ground surface were misclassified as random class. It was shown in Fig. 3(c)–(f) that the classification accuracy could be improved by using different filters. Some points belonging to the surface or random class that were misclassified as linear were correctly relabeled by using filter-1 [see Fig. 3(c)]. Points misclassified as surface were correctly relabeled as random or linear points by using filter-2 [see Fig. 3(d)]. Filter-3 removed outliers, and ground points misclassified as linear or random class were correctly relabeled as surface class by using filter-4[see Fig. 3(e)]. Two additional filters (filter-5 and filter-6) proposed in this paper were applied to obtain the final improved classification [see Fig. 3(f)].

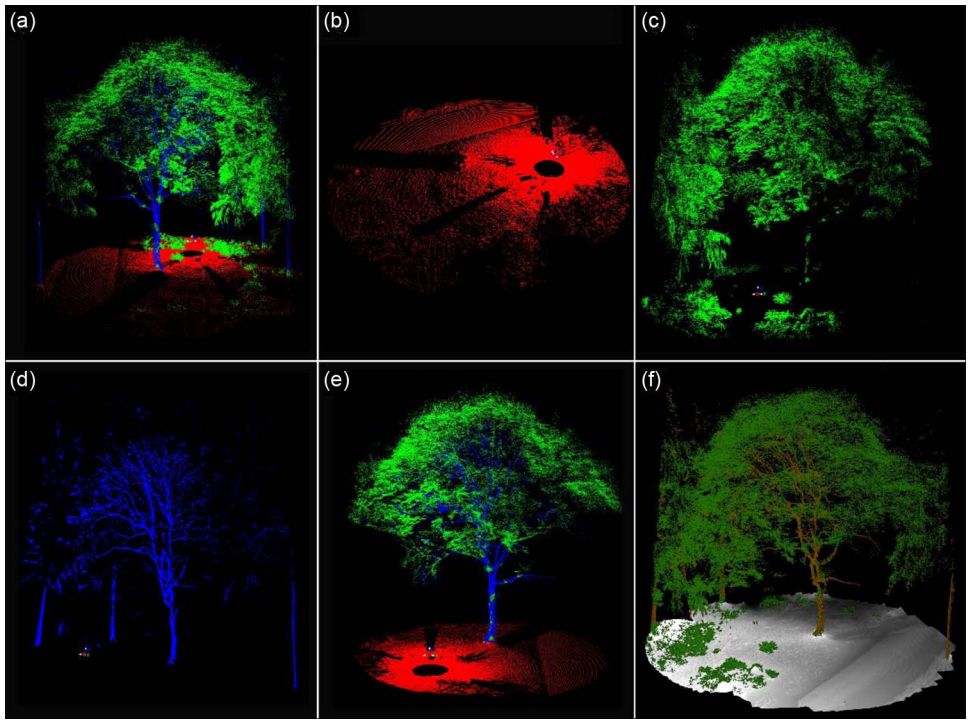


Fig. 4. Visualization of final classification for different components of an individual tree in WPA forest plot-4. (a) Final classification with surface class in red, linear class in blue, and random class in green, respectively. (b), (c), and (d) Classified point cloud for ground surface, green leaves, and woody materials, respectively, (e) Individual tree point cloud data after excluding the points from other trees used for computing the parameter of "woody-to-total area ratio." (f) Visualization of final classification with the DEM generated with the classified ground surface points.

After applying filters 1–6 to the pointwise classification (see Section III-C), we obtained the final classification with three different classes [see Fig. 4(a)], including bare ground points [see Fig. 4(b)], photosynthetic canopy components (i.e., leaves and shrubs) [see Fig. 4(c)], and nonphotosynthetic canopy components (i.e., stem and branches) [see Fig. 4(d)]. An individual broadleaf tree point cloud data from the forest plot-4 was used for accuracy assessment and computation of the parameter of "woody-to-total area ratio" [see Fig. 4(e)]. In addition, based on the final classification, we produced the digital elevation model (DEM) and colored the photosynthetic and nonphotosynthetic components as green and brown, respectively [see Fig. 4(f)].

The classification accuracy was assessed by visually inspecting the projections of classified points in the XOZ and YOZ planes for each class (see Fig. 5). The initial classification for the linear class denoted by red dots showed that many points were misclassified [see Fig. 5(a) and (b)]. By applying filters-1–6 (see Section III-D), the classification accuracy could be considerably improved. Some points belonging to the surface class had been misclassified into the random class, and there were some "ghost points" over the tree canopy [see Fig. 5(c) and (d)]. The postprocessing filters newly introduced in this paper show how to effectively improve the classification of the bare ground class [see Fig. 5(e) and (f)].

2) Quantitative Accuracy Assessment: The forest point cloud from forest plot-4 (794,320 points) was manually classified into the following classes: random (47.33%), surface (38.15%), and linear (10.4%). A small fraction (4.13%) of points were difficult to classify and thus were not used as reference data for validating the computer-based classification. We computed both the user's and producer's accuracy to quantitatively assess the classification accuracy of each class at each filtering stage (see Tables II–IV). User's accuracy is defined as the ratio of the number of correctly classified points to the total number of this class after each classification stage, whereas the producer's accuracy is defined as the ratio of the number of correctly classified points to the number of manually selected sample points for each class. For the surface class, the user's accuracy and producer's accuracy of preliminary classification were 92.92% and 82.56%, respectively. The user's and producer's accuracy could be significantly improved by applying filters 1–6, particularly filter-4, which improved the producer's accuracy from 86.78% to 99.6% (see Table II). For the random class, the preliminary user's and producer's accuracy (82.89% and 82.66%, respectively) could be considerably improved (88.77% and 97.56%) by applying filter-1 (81.96% and 96.01%, respectively). (Table III). Similarly, filter-1 greatly improved the user's accuracy of the linear class from 36.65% to 71.97%. However, the producer's accuracy decreased from 66.66% to 56.69% after using filter-1, and filter-2, filter-3, and filter-4 did not improve the producer's accuracy until using the newly proposed ground filters (filter-5 and filter-6). The producer's accuracy increased from 60.28% to 66.11% with the user's accuracy being 76.89% (see Table IV).

The overall classification accuracy is defined as the ratio of the number of correctly classified points to the number of total points. Table V shows the overall classification accuracy for different classification stages. The continuous improvement of the overall classification accuracy suggests that the postprocessing filters significantly improve the classification accuracy.

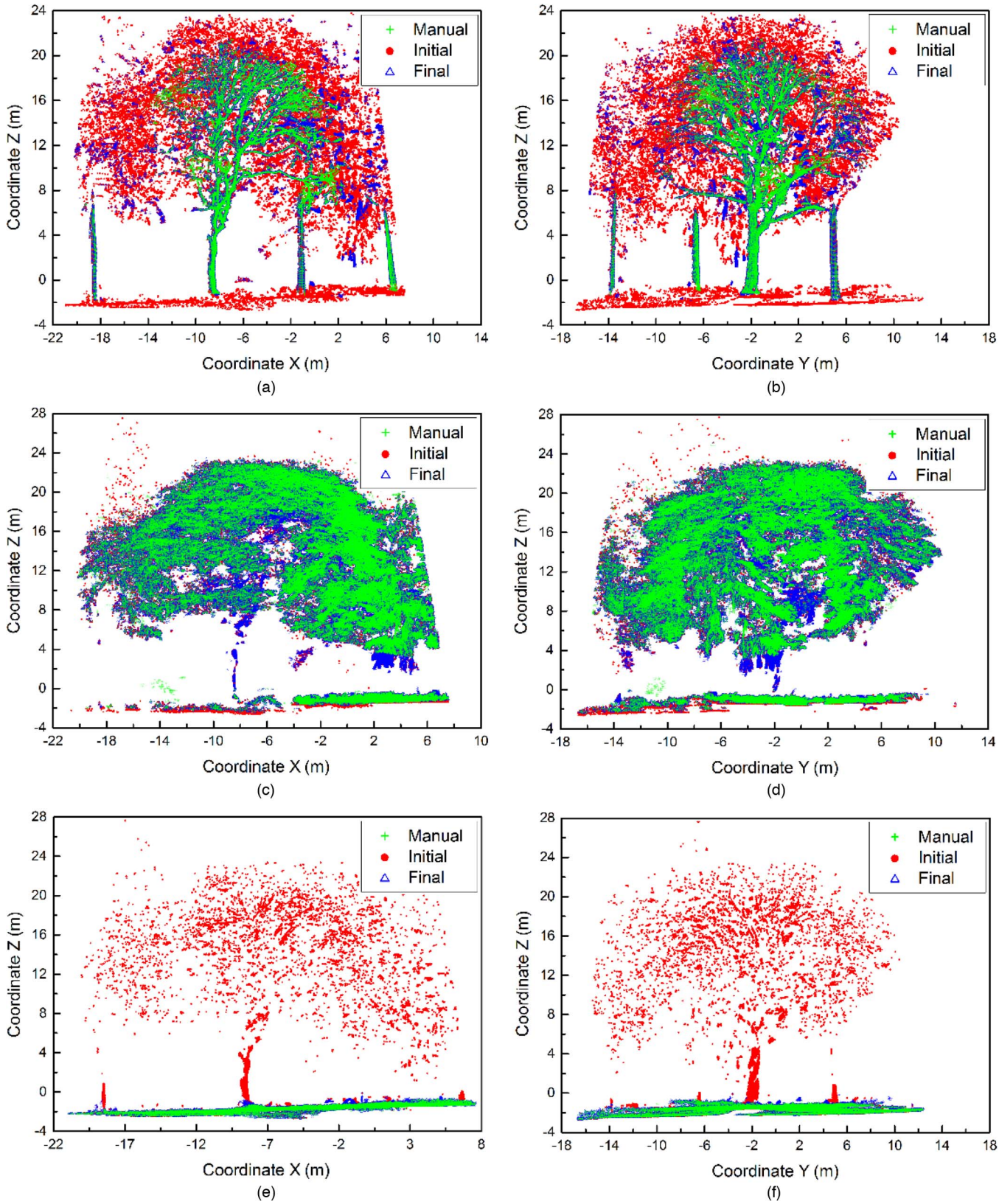


Fig. 5. Scatter plots of the classified point cloud in the XOZ and YOZ projection planes at three different stages of each class for an individual tree in WPA forest plot-4 (green crosses are the manually selected validation points, red dots are the preliminary classified points, and blue triangles are the final classified points). (a) and (b) Scatter plots of classified linear class points (i.e., stems and branches) on XOZ and YOZ projection planes, respectively. (c) and (d) Scatter plots of classified random class points (i.e., leaves, brushes, and grass) on XOZ and YOZ projection planes, respectively. (e) and (f) Scatter plots of classified surface class points (i.e., ground) on XOZ and YOZ projection planes, respectively.

TABLE II
ACCURACY ASSESSMENT OF CLASSIFICATION AT DIFFERENT STAGES FOR SURFACE CLASS FROM AN INDIVIDUAL TREE IN FOREST PLOT-4

	Initial classification	Filter-1	Filter-2	Filter-3	Filter-4	Filter-5	Filter-6
Total number of points	269,251	288,865	272,691	272,627	311,907	306,601	306,559
Correctly classified number of points	250,199	266,449	262,987	262,969	301,836	301,838	301,838
User's accuracy	92.92%	92.24%	96.44%	96.46%	96.77%	98.44%	98.46%
Producer's accuracy	82.56%	87.93%	86.78%	86.78%	99.6%	99.6%	99.6%

TABLE III
ACCURACY ASSESSMENT OF CLASSIFICATION AT DIFFERENT STAGES FOR RANDOM CLASS FROM AN INDIVIDUAL TREE IN FOREST PLOT-4

	Initial classification	Filter-1	Filter-2	Filter-3	Filter-4	Filter-5	Filter-6
Total number of points	374,877	440,397	453,180	450,705	412,542	413,101	413,143
Correctly classified number of points	310,723	360,953	366,703	366,703	366,703	366,730	366,743
User's accuracy	82.89%	81.96%	80.92%	81.36%	88.89%	88.77%	88.77%
Producer's accuracy	82.66%	96.02%	97.55%	97.55%	97.55%	97.55%	97.56%

TABLE IV
ACCURACY ASSESSMENT OF CLASSIFICATION AT DIFFERENT STAGES FOR LINEAR CLASS FROM AN INDIVIDUAL TREE IN FOREST PLOT-4

	Initial classification	Filter-1	Filter-2	Filter-3	Filter-4	Filter-5	Filter-6
Total number of points	150,192	650,58	684,49	673,81	662,64	710,11	710,11
Correctly classified number of points	550,52	468,19	497,85	497,85	497,85	545,99	545,99
User's accuracy	36.65%	71.97%	72.73%	73.89%	75.13%	76.89%	76.89%
Producer's accuracy	66.66%	56.69%	60.28%	60.28%	60.28%	66.11%	66.11%
Misclassified surface	19,002	1,119	1,149	1,124	14	7	0
Misclassified scatter	58,782	8,504	8,639	8,639	8,639	8,497	3,621

TABLE V
OVERALL ACCURACY ASSESSMENT FOR CLASSIFICATION AT DIFFERENT STAGES FROM AN INDIVIDUAL TREE IN FOREST PLOT-4

	Initial classification	Filter-1	Filter-2	Filter-3	Filter-4	Filter-5	Filter-6
Total number of points	794,320	794,320	794,320	790,713	790,713	790,713	790,713
Correctly classified number of points	615,974	674,221	679,475	679,457	718,324	723,167	723,180
Overall accuracy	77.55%	84.88%	85.54%	85.93%	90.85%	91.46%	91.46%

3) Circular Point Density Distribution: The variation of point density along the azimuthal angles reflects the spatial distribution pattern of foliage elements in four different quadrants. We decomposed the curve of total point circular point density into three different ones for each class [see Fig. 6(a)], and plotted the circular point density distributions for the surface class [see Fig. 6(b)], and random class [see Fig. 6(c)] and linear class [see Fig. 6(d)], respectively.

The surface class points fell in all four quadrants with two deep valley values in quadrant-1 and quadrant-4 and two shallow valley values in quadrant-3. These "valley values" were attributed to the "occlusion effect" since the laser beam cannot penetrate the tree trunk surface. The random class points distributed in four quadrants with higher density in quadrant-3 and quadrant-4 than quadrant-1 and quadrant-2. In Fig. 6(d), two peak values were found in quadrant-3 and quadrant-4, suggesting that two trunks were located in these directions.

4) Angular Point Density Distribution: We divided the angular point density versus inclination angle plot into three different regions with inclination angle intervals of 45° to examine spatial distribution of the points, and labeled the regions above and below the scanner height with "$< 0°$" and "$> 0°$," respectively (see Fig. 7). The angular density distribution of the raw point cloud data set of WPA forest plot-4 was decomposed into three different curves of each classes to examine the angular distribution of foliage elements for linear, surface and random classes, respectively [see Fig. 7(a)]. Few points located in the region with inclination angle $< 0°$ suggested that some canopy points were misclassified as surface class [see Fig. 7(b)]. The valley area of the angular point density curve from the random class [see Fig. 7(c)] provided information about crown base height. The angular point density curve of the linear class distributed in both negative $(-45°-0°)$ and positive $(45-90°)$ angular regions [see Fig. 7(d)], and the peak value around the region with an inclination angle of $-45°$ indicates that many points of the linear class points were classified as ground points in the preliminary classification. The continuous increase in the angular point density curve suggests that the tree had a relative high crown base height and little understory grew on the ground surface [see Fig. 7(c)]. This was confirmed by a visual evaluation of the point cloud.

5) Woody Materials Percentage: We separated the photosynthetic (249,022 points) and nonphotosynthetic points

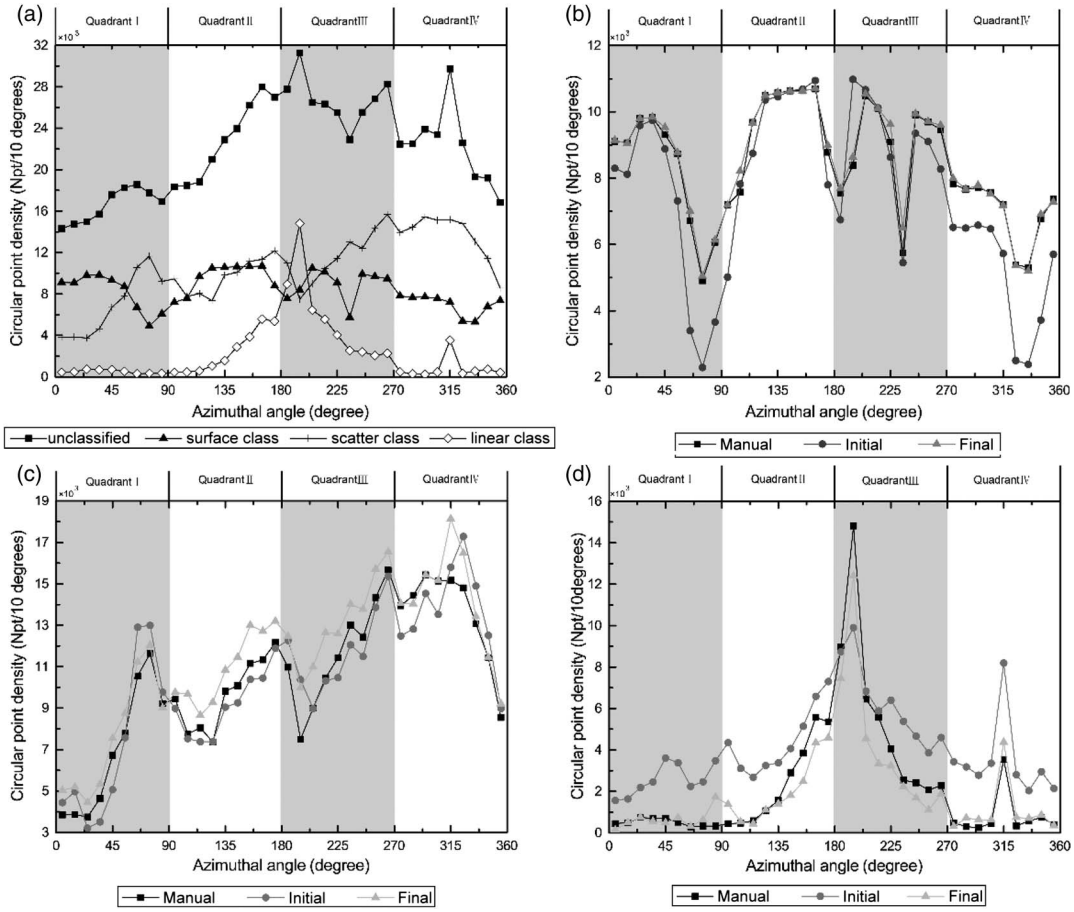


Fig. 6. Circular point density variations with changed azimuthal angles for different point cloud data sets of an individual tree from WPA forest plot-4. (a) Circular point density distribution along the azimuthal angles for the manually selected validation points and the classified points from linear, surface and random classes, respectively. (b), (c), and (d) Circular point density distributions along the azimuthal angles at different classification stages (manually selected validation points, preliminary classification points, and final classification points) for surface, random, and linear classes, respectively.

(59,235 points) from the whole forest stand points using the pointwise classification method. The contribution of nonphotosynthetic components to the forest canopy characterized by the parameter of "woody-to-total area ratio" was computed as around 12% [see (16)].

6) Stem Surface Area Validation: For validation purposes, we reconstructed the stem surface mesh based on the classified stem surface points [see Fig. 8(a)–(d)]. The surface area of the reconstructed stem surface mesh was 6.755 m^2, which was close to the surface area (i.e., 5.902 m^2) calculated based on the method that converts the discrete point into the surface area described in Section III-E.

B. Forest Stand Level

1) Forest Stand Classification: We applied the proposed method to three forest plots at WPA site with different canopy densities, including forest plot-1 with LAI as 4.15, forest plot-2 with LAI as 3.65, and forest plot-3 with LAI as 3.13, respectively. Fig. 9(a), (c), and (e) were the preliminary classification for three different plots, whereas the Fig. 9(b), (d), and (f) were the final classification with good accuracy. The overall producer's accuracy resulting from the preliminary classification was 65.82%, 70.58%, and 75.19%, respectively. Some edge points of leaves and ground surface had been misclassified as the linear class. Some stem surface points had been mislabeled as surface class due to the partial sampling using one location TLS scanning setup. In addition, the ground points further away from the scanner were prone to being mislabeled as linear points due to increasing spacing between samples with increasing distance. After applying the six different filters, the overall producer's accuracy for final classification of the three forest plots were improved to 84.28%, 94.40%, and 97.80%, respectively.

V. Discussions

A. Forest Stand Density Effects

Both the user's and producer's accuracy of the final classification increased as the forest densities decreased due to the decreasing likelihood of ghost returns and the occlusion of canopy components (see Table VI). The user's accuracy and producer's accuracy were as low as 85.50% and 84.28%, respectively, for WPA forest plot-1 characterized by a high density $(\mathrm{LAI} = 4.15)$ [see Fig. 9(a)]. In the case of WPA forest plot-2 characterized by medium density $(\mathrm{LAI} = 3.65)$, the user's and producer's accuracy increased to 93.62% and 94.40%, respectively [see Fig. 9(b)]. Finally, for the WPA forest

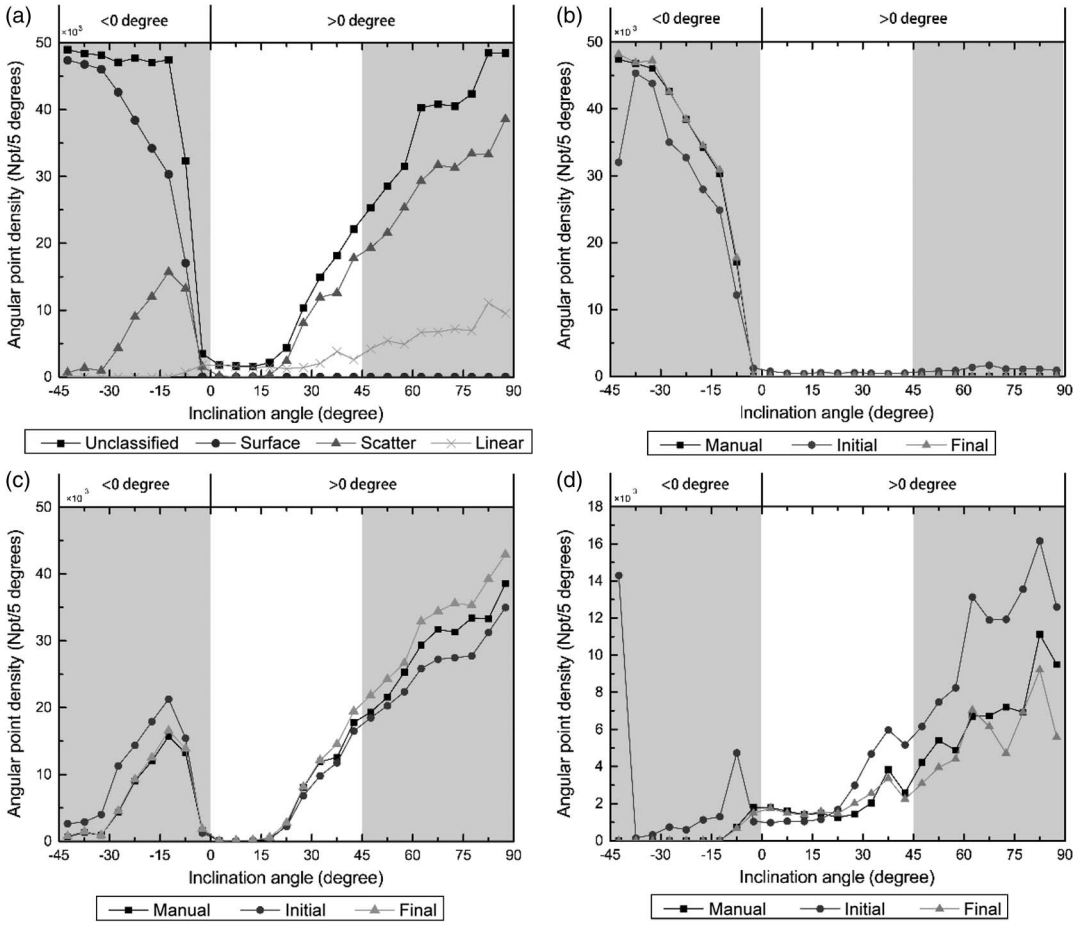


Fig. 7. Angular point density variations with inclination angles for different point cloud data sets for an individual tree from WPA forest plot-4. (a) Angular point density distribution along the inclination angles for the manually selected validation points. Three different classified classes points: linear, surface, and random. (b), (c), and (d) Angular point density distributions along the inclination angles of three different stages (i.e., manually selected validation points, preliminary classification points, and the final classifications after using the filters) for surface, random, and linear classes, respectively.

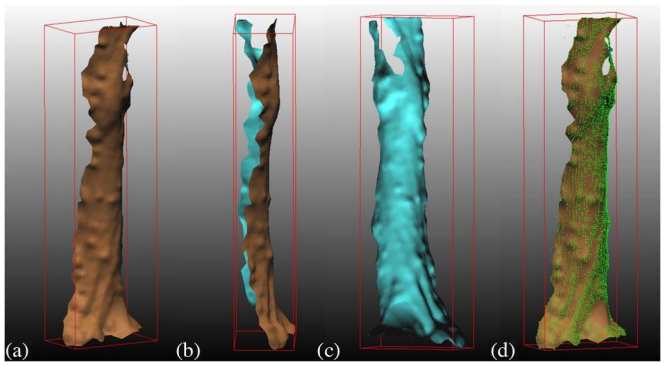


Fig. 8. Reconstructed stem surface mesh from its discrete points. (a) 3-D perspective view of the reconstructed mesh and its (b) side-lateral view and (c) backside views. (d) Overlaid view of the reconstructed mesh and discrete points for a stem surface.

plot-3 characterized by low density (LAI = 3.13), the user's and producer's accuracy reached up to 97.68% and 97.80%, respectively [see Fig. 9(c)]. Among the three classes, the linear class showed the lowest final producer's accuracy with 69.43%, 77.10%, and 80.48% for forest plot-1, forest plot-2, and forest plot-3, respectively. The low accuracy of the linear final classification was likely due to the incomplete sampling of the stem surface and branches with many small leaves.

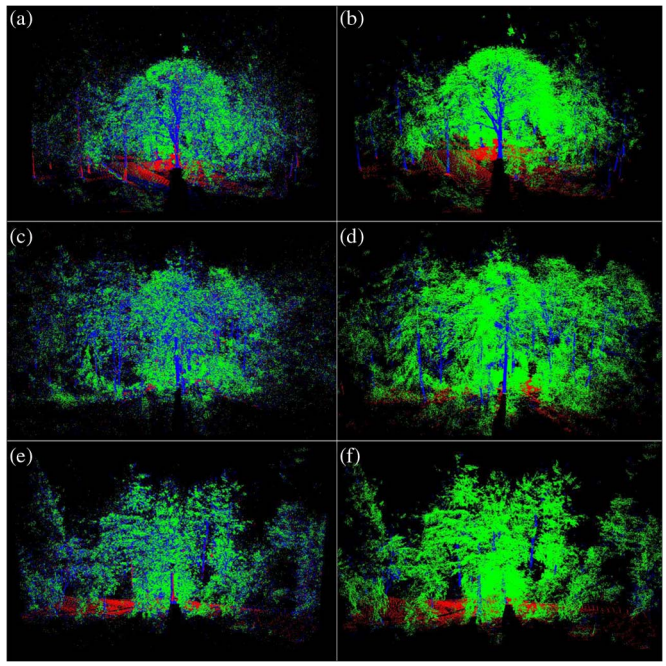


Fig. 9. Classifications of forest plots at WPA site with high density (forest plot-1 with LAI = 4.15), medium density (forest plot-2 with LAI = 3.65), and low density (forest plot-3 with LAI = 3.13) at WPA site, respectively. (a), (c), and (e) Preliminary classifications. (b), (d), and (f) Final classifications.

TABLE VI
COMPARISONS BETWEEN THE PRELIMINARY AND FINAL CLASSIFICATIONS FOR THREE FOREST PLOTS WITH HIGH DENSITY (FOREST PLOT-1, LAI = 4.15), MEDIUM DENSITY (FOREST PLOT-2 WITH LAI = 3.65) AND LOW DENSITY (FOREST PLOT-3 WITH LAI = 3.13), RESPECTIVELY

		Initial classification				Final classification			
		Total number of points	Correctly classified number of points	User's accuracy	Producer's accuracy	Total number of points	Correctly classified number of points	User's accuracy	Producer's accuracy
Plot1	Surface	268,770	236,372	87.95%	68.80%	315,664	310,078	98.23%	90.25%
	Random	607,992	494,096	81.27%	63.83%	784,656	650,701	82.93%	84.07%
	Linear	383,217	88,621	23.13%	69.86%	126,366	88,078	69.70%	69.43%
	All	1,259,979	819,089	65.01%	65.82%	1,226,686	1,048,857	85.50%	84.28%
Plot2	Surface	120,603	91,490	75.86%	50.60%	170,481	162,834	95.51%	90.06%
	Random	672,455	618,770	92.02%	72.57%	863,945	844,308	97.73%	99.02%
	Linear	464,967	147,869	31.80%	81.08%	191,503	140,609	73.42%	77.10%
	All	1,258,025	858,129	68.21%	70.58%	1,225,929	1,147,751	93.62%	94.40%
Plot3	Surface	231,541	211,054	91.15%	74.70%	272,157	270,043	99.22%	95.58%
	Random	562,720	530,973	94.36%	75.46%	714,341	702,990	98.41%	99.90%
	Linear	266,766	36,644	13.74%	74.10%	50,406	39,799	78.96%	80.48%
	All	1,061,027	778,671	73.39%	75.19%	1,036,904	1,012,832	97.68%	97.80%

B. Forest Species Effects

Tree species showed to have a negligible effect on the final classification accuracy using the proposed method. For WPA plot-4, the final producer's accuracy for only broadleaf tree species was 94.96% [see Fig. 10(c)] and for only conifer tree species 93.09% [see Fig. 10(d)]. The linear class showed the lowest user's and producer's accuracy for the final classification for both broadleaf and coniferous tree species with 66.11% and 51.28%, respectively, (see Table VII). In particular, the user's accuracy of the linear class was as low as 9.24% due to many of points from other classes misclassified as linear class. Moreover, some ghost points floating above the canopy were mislabeled as linear class. For some old growth coniferous trees, shoots associated with long and narrow branches tended to be misclassified as linear class (see a failure example presented in Section V-F).

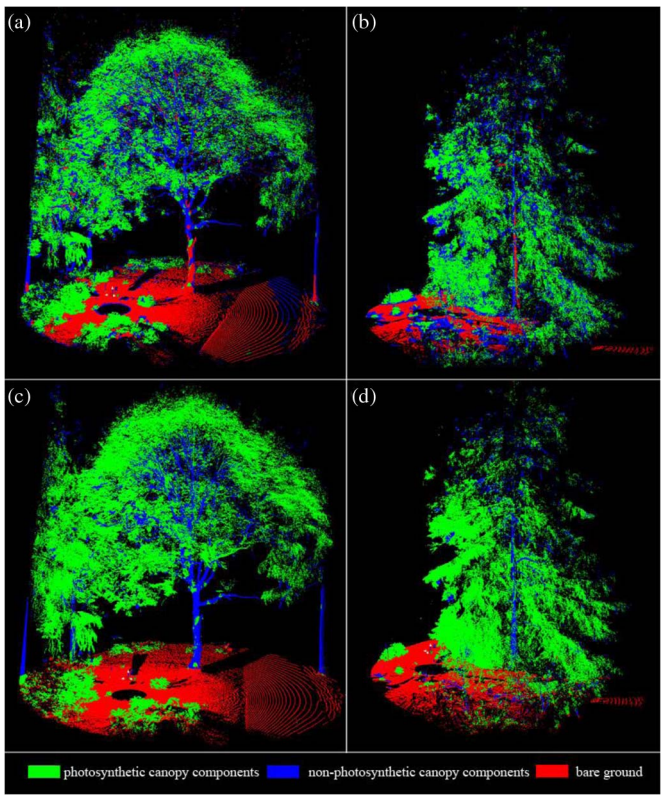


Fig. 10. Comparisons between the preliminary and final classifications for different forest types from forest plot-4 at WPA site, respectively. (a) and (c) Preliminary and final classifications for the broadleaf tree. (b) and (d) Preliminary and final classifications for the coniferous tree.

C. Comparison With Existing Classification Methods

3-D point clouds have been traditionally classified based on their color information provided by the digital camera [see Fig. 11(a)], local geometric features, and laser return intensity [see (1)] [40]–[43]. To test the performance of the proposed method, we compared the proposed method against the methods using following features: 1) eigenvalues plus color-based information; and 2) salient feature only. For method-1, color information showed to lower the classification accuracy likely due to incorrect exposure values [see Fig. 11(b)]. For method-2, our results showed many misclassified stem points [see Fig. 11(c)]. When compared with method-1 and method-2, the proposed method showed the highest classification accuracy [see Fig. 11(d)].

The overall producer's accuracy of method-1 was 62.12%, and the producer's accuracy of the surface, random, and linear classes were 49.63%, 73.15%, and 49.65%, respectively.

TABLE VII
COMPARISONS BETWEEN THE PRELIMINARY AND FINAL CLASSIFICATIONS FOR BROADLEAF AND CONIFEROUS TREE SPECIES, RESPECTIVELY

		Initial classification				Final classification			
		Total number of points	Correctly classified number of points	User's accuracy	Producer's accuracy	Total number of points	Correctly classified number of points	User's accuracy	Producer's accuracy
Broad leaf	Surface	269,251	250,199	92.92%	82.56%	306,559	301,838	98.46%	99.60%
	Random	374,877	310,723	82.89%	82.66%	413,143	366,743	88.77%	97.56%
	Linear	150,192	55,052	36.65%	66.66%	71,011	54,599	76.89%	66.11%
	All	794,320	615,974	77.55%	80.88%	790,713	723,180	91.46%	94.96%
Conifer	Surface	180,586	174,414	96.58%	68.93%	246,700	236,762	95.97%	93.57%
	Random	232,955	212,547	91.24%	74.16%	285,332	276,010	96.73%	96.31%
	Linear	153,633	14,203	9.24%	56.72%	35,226	12,840	36.45%	51.28%
	All	567,174	401,164	70.73%	71.05%	567,258	525,612	92.66%	93.09%

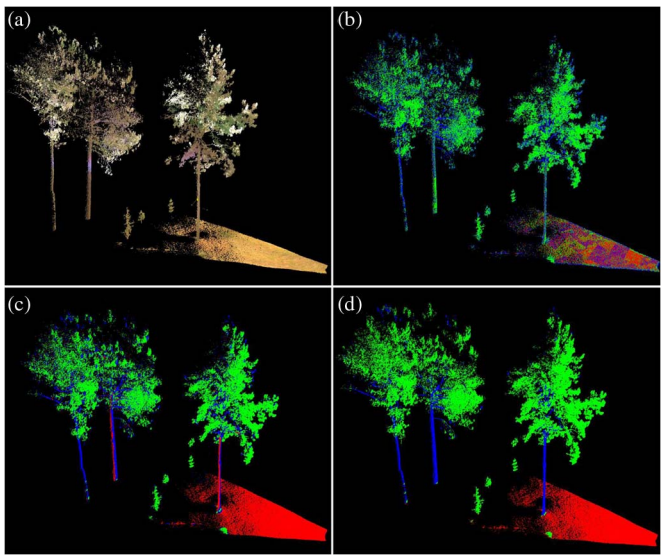


Fig. 11. Comparisons between the classifications using different methods for forest plot-3. (a) Raw point cloud data colored by R(ed), G(reen), B(lue) information. (b) Classification combining the eigenvalue and RGB information. (c) Classification using salient-feature-based method. (d) Classification using the proposed method by combining the salient feature and postprocessing filters.

Method-2 showed overall classification accuracy of 88.95% with the producer's accuracy of 99.31%, 85.52%, and 77.66% for the surface, random, and linear classes, respectively. For the proposed method, the overall producer's accuracy was 95.45% with producer's accuracy of 99.69%, 96.98%, and 79.70% for the surface, random, and linear classes, respectively (see Table VIII).

D. Sensitivity Analysis of Search Radius

The search radius showed to affect both the preliminary and final classification. To better understand the effect of search radius on the classification accuracy, a sensitivity analysis was conducted using the data from three WPA plots. The sensitivity analysis showed that the correctly classified percentage (CCP) curves for all canopy features, the random class, the linear class, and the surface class changed asynchronically as the search radius increased. As shown in Fig. 12, the CCP increased as the searching radius increased from 0.1 to 0.3 m for all the canopy features, the random class, and the linear class. However, the CCP for the surface class showed low classification accuracy for the search radius of 0.25 m (the lowest CCP was observed when using a search radius of 0.8 m). The CCP for the linear class showed to increase when increasing the search radius from 0.1 to 0.25 m. Between 0.25 and 0.35 m, the CCP for the linear class showed to decrease probably due to misclassified points from leaves. Overall, the 0.45-m search radius showed to result in the highest classification accuracy (see Fig. 12). This finding is in agreement with results shown in [15].

For the individual tree in WPA forest plot-4, the CCP of the linear class first decreased as the search radius increased from 0.3 to 0.6 m and then stayed relatively stable as the search radius increased to 1.2 m [see Fig. 13(a)]. The CCP of other classes continuously increased up to a search radius of 1.1 m. However, the changes of CCPs were very small ($<= 0.2\ \%$) throughout the entire range of radius variation [see Fig. 13(b)]. Our results suggest that 1.0 and 1.5 m were the best radii for filter-1 and filter-2, respectively.

E. Terrestrial Laser Scanner Setup

The TLS setup likely affects the classification accuracy. For example, in this paper, plots were only scanned from a single scan position. As a result, some stem surface points were misclassified as surface class. Further, increasing distance between the scanner and the surveyed tree showed to result in the misclassification of edge ground points as linear class. Branches with only a few leaves or shoots showed to lower the final classification accuracy of the photosynthetic canopy components, whereas the leaves or shoots were often misclassified as linear class. Further, tree stems or branches surrounded by shrubs or grasses showed to be often misclassified as random class points. To increase the classification accuracy, we recommend using multiple scan locations to better capture the 3-D structural characteristics of a forest plot.

TABLE VIII
COMPARISONS BETWEEN CLASSIFICATION ACCURACY USING DIFFERENT POINTWISE CLASSIFICATION-BASED METHODS

Methods	Category	Total number of points	Correctly classified number of points	User's accuracy	Producer's accuracy
Eigenvalue+ RGB +	Surface	117,343	113,667	96.87%	49.63%
	Random	378,542	270,791	71.54%	73.15%
	Linear	200,972	48,462	24.11%	49.65%
	All	696,857	432,920	62.12%	62.12%
Saliency Feature	Surface	237,606	227,466	95.73%	99.31%
	Random	333,222	316,604	95.01%	85.52%
	Linear	126,029	75,811	60.15%	77.66%
	All	696,857	619,881	88.95%	88.95%
Saliency Feature+ Post-processing filters	Surface	228,463	228,324	99.94%	99.69%
	Random	379,350	359,012	94.64%	96.98%
	Linear	87,100	77,797	89.32%	79.70%
	All	694,913	665,133	95.71%	95.45%

F. Limitation of Proposed Method

The proposed method might provide poor classification results if the spatial distribution patterns of foliage elements violates salient feature definitions described in Section III-A. For example, for the WREF site, our results showed that random points (i.e., photosynthetic active tissue) were frequently mislabeled as linear points and that the final classification could not be improved even after using the different filters [see Fig. 14(a)]. In this case, the spatial distribution pattern of the leaves usually shows the linear characteristic [see Fig. 14(b)]; we recommend to incorporate additional information such as calibrated intensity information or multispectral and hyperspectral information to differentiate the green leaves from the woody materials. For example, the multispectral or hyperspectral information collected by spectral imaging instruments such as Headwall (Headwall Photonics Inc. Fitchburg, MA) could be used to improve the final classification. Yet, the classification improvements achievable by fusing TLS data with spectral information from optical imaging systems (e.g., multispectral and hyperspectral imaging systems) might also be limited since the measured reflectance signal is often confounded by: 1) the variation of leaf orientation; 2) variations in leaf reflectance properties; 3) diffuse reflection between the foliage elements; and 4) shadowing between different foliage elements.

G. Training Data

We carefully selected 30 training data sets for each of the three classes. When selecting training data for the random class, leaves at different heights of a forest canopy with different densities and width should be included. Similarly, for the linear class, training data should be selected from branches and stems at different canopy heights. The training data sets for the ground points should be selected to obtain a representative sample of different elevations and slopes within the surveyed area. The computed salient feature of the selected sample data was compared with the definition described in Section III-A to determine class to which the selected sample data are qualified. The use of training data with 'pure salient features' for each specific class would make the classification more effective, but incorporating them would degrade the quality of the preliminary classification results and the effectiveness of the postprocessing filters. By doing this, training sample data sets were manually selected for each of the three different classes. We recommend sampling the training data at 30%, 60%, and 90% percentile heights of the forest canopy and the bounding box of a forest canopy point cloud data for random and linear class. In addition, a 90° azimuthal angle interval is preferred to collect the typical training data set.

In this paper, training sample data sets were created for each of the three classes by manually selecting points that clearly belonged to a given class. However, boundary points covering the area between different classes might contain valuable information for other classification approaches not used in this paper. For example, support-vector-machine approaches could make use of the boundary points as support vectors to better differentiate classes. Although beyond the scope of this paper, incorporating such an approach into the GMM could potentially improve the classification accuracy and possibly make the filters used in this paper unnecessary.

H. Conversion From Point to Surface Area

The leaf orientation and direction of the incident laser beam has to be considered when converting laser points to a surface area for calculating the "woody-to-total area ratio." In this paper, we only calculated the surface area of leaves from broadleaf trees from TLS data but not for coniferous trees. The calculation of the surface area of coniferous leaves from TLS data is problematic given the laser point size is often not small enough to resolve individual needles. For coniferous trees, we recommend to treat each shoot as a leaf with an additional parameter named "needle-to-shoot area ratio" [33] to characterize the gap and clumping effects within a single shoot structure. This is a direction deserved for future research.

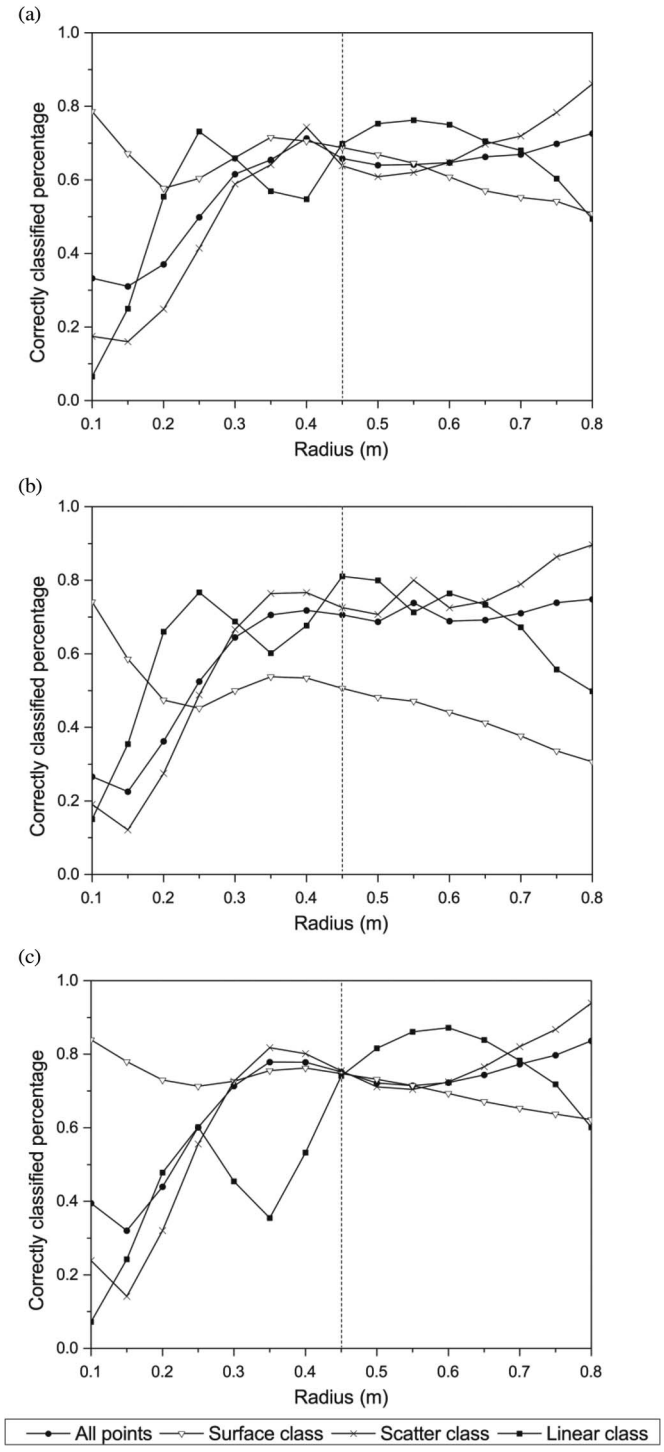


Fig. 12. Sensitivity analysis of the searching radius on the classification accuracy of both the all points and each individual class for three forest plots: (a) high (forest plot-1 with LAI $= 4.15$); (b) medium (forest plot-2 with LAI $= 3.65$); and (c) low (forest plot-3 with LAI $= 3.13$).

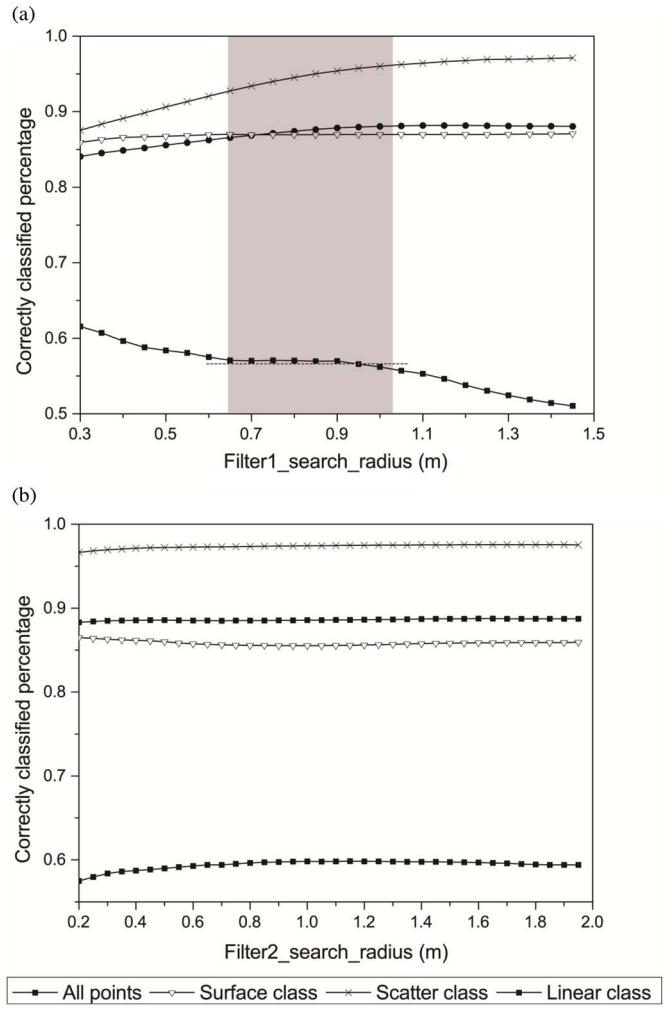


Fig. 13. Sensitivity analysis of the searching radius of (a) filter-1 and (b) filter-2 on the CCP for all points, linear, random, and surface classes in the forest plot-4, respectively.

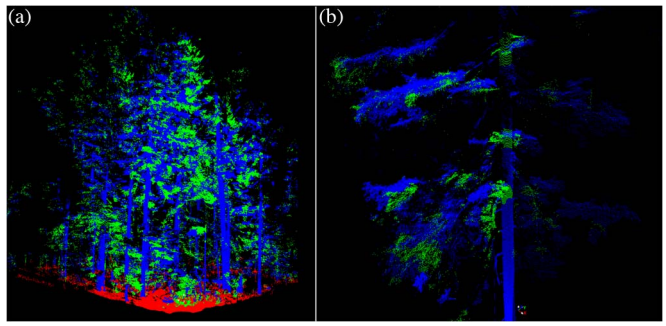


Fig. 14. Failure example of the proposed method for a mature heterogeneous forest stand in WREF site (surface class points in red color, linear class points in blue color, random class point in green color). (a) Forest stand with Douglas fir (*Pseudotsuga. menziesii*) as the dominant tree species. (b) Zoom-in view of an individual Douglas fir tree.

VI. Conclusion

This paper has described an improved GAFPC method to classify the 3-D forest lidar points into three different classes: photosynthetic canopy components (i.e., leaves, brushes or grasses), nonphotosynthetic canopy components (i.e., branches or stems), and bare ground. Based on the classification from the improved GAFPC, we calculated the parameter of "woody-to-total area ratio." The main conclusions from this study are as follows. First, the salient feature expressed by linear combination of eigenvalues allows capturing the different spatial distribution patterns associated with different canopy

components, including photosynthetic (i.e., leaves, brushes or grasses), nonphotosynthetic (i.e., branches or stems), and bare ground. Further, the GMM with three GDFs allows approximating the probability density distribution functions of a given forest-point cloud data set for each of the three classes. Second, it is feasible to classify the 3-D forest point cloud data only using the geometric (i.e., x, y, z) information. For example, the GAFPC approach obtained the producer's accuracy for preliminary classification as 77.55% and achieved the overall producer's accuracy for final classification as 91.46% after using the postprocessing filters for an individual tree point cloud data in forest plot-4. The producer's accuracy of final classification will decrease with increasing forest stand density because the dense forest stands usually have strong and increasing likelihood of occlusion and ghost points. Finally, the tree species has negligible effects on the proposed improved GAFPC method in classifying a forest point cloud data. Separating the photosynthetic and nonphotosynthetic components of a forest canopy will provide better estimates of "woody-to-total area ratio," which will ultimately help to improve our ability to derive more accurate estimates of LAI from TLS data and make use of TLS data for modeling radiative transfer through tree canopies.

ACKNOWLEDGMENT

The authors would like to thank the International Institute for Earth System Science, Nanjing University, X. Zhou for the valuable discussion and assistance, Dr. K. Bible for supporting our fieldwork at the Wind River Canopy Crane Research Facility located within the Wind River Experiment Forests, Dr. A. Kato of Chiba University, Japan for the assistance in designing and conducting our experiments, and the two anonymous reviewers for their great help in this paper.

REFERENCES

[1] J. M. Norman and G. S. Campbell, "Canopy structure," *Plant Physiol. Ecol.*, pp. 301–325, 1989.

[2] J. S. Amthor, D. S. Gill, and F. H. Bormann, "Autumnal leaf conductance and apparent photosynthesis by saplings and sprouts in a recently disturbed northern hardwood forest," *Oecologia*, vol. 84, no. 1, pp. 93–98, Aug. 1990.

[3] P. Blanken *et al.*, "Energy balance and canopy conductance of a boreal aspen forest: Partitioning overstory and understory components," *J. Geophys. Res., Atmos. (1984–2012)*, vol. 102, no. D24, pp. 28 915–28 927, Dec. 1997.

[4] C. Brümmer *et al.*, "How climate and vegetation type influence evapotranspiration and water use efficiency in Canadian forest, peatland and grassland ecosystems," *Agricul. Forest Meteorol.*, vol. 153, pp. 14–30, Feb. 2012.

[5] S. J. DeWalt, S. K. Maliakal, and J. S. Denslow, "Changes in vegetation structure and composition along a tropical forest chronosequence: Implications for wildlife," *Forest Ecology Manage.*, vol. 182, no. 1, pp. 139–151, Sep. 2003.

[6] L. Chasmer *et al.*, "Influences of vegetation structure and elevation on CO2 uptake in a mature jack pine forest in Saskatchewan, Canada," *Can. J. Forest Res.*, vol. 38, no. 11, pp. 2746–2761, Aug. 2008.

[7] J. M. Chen and T. A. Black, "Defining leaf area index for non-flat leaves," *Plant Cell Environ.*, vol. 15, no. 4, pp. 421–429, May 1992.

[8] J. U. H. Eitel, L. A. Vierling, and T. S. Magney, "A lightweight, low cost autonomously operating terrestrial laser scanner for quantifying and monitoring ecosystem structural dynamics," *Agricultural Forest Meteorol.*, vol. 180, pp. 86–96, Oct. 2013.

[9] F. M. Danson *et al.*, "Developing a dual-wavelength full-waveform terrestrial laser scanner to characterize forest canopy structure," *Agricultural Forest Meteorol.*, vol. 198/199, pp. 7–14, Nov./Dec. 2014.

[10] X. Yang *et al.*, "Three-dimensional forest reconstruction and structural parameter retrievals using a terrestrial full-waveform lidar instrument (Echidna®)," *Remote Sens. Environ.*, vol. 135, pp. 36–51, Aug. 2013.

[11] J. U. H. Eitel, T. S. Magney, L. A. Vierling, and G. Dittmar, "Assessment of crop foliar nitrogen using a novel dual-wavelength laser system and implications for conducting laser-based plant physiology," *ISPRS J. Photogramm. Remote Sens.*, vol. 97, pp. 229–240, Nov. 2014.

[12] J. U. H. Eitel, L. A. Vierling, D. S. Long, and E. R. Hunt, "Early season remote sensing of wheat nitrogen status using a green scanning laser," *Agricultural Forest Meteorol.*, vol. 151, no. 10, pp. 1338–1345, Oct. 2011.

[13] D. N. Donoghue, P. J. Watt, N. J. Cox, and J. Wilson, "Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data," *Remote Sens. Environ.*, vol. 110, no. 4, pp. 509–522, Oct. 2007.

[14] S. Kaasalainen, E. Ahokas, J. Hyyppa, and J. Suomalainen, "Study of surface brightness from backscattered laser intensity: calibration of laser data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 255–259, Jul. 2005.

[15] J. F. Lalonde, N. Vandapel, D. F. Huber, and M. Hebert, "Natural terrain classification using three-dimensional ladar data for ground robot mobility," *J. Field Robot.*, vol. 23, no. 10, pp. 839–861, Oct. 2006.

[16] N. Vandapel, D. F. Huber, A. Kapuria, and M. Hebert, "Natural terrain classification using 3-D ladar data," in *Proc. IEEE ICRA*, New York, NY, USA, 2004, vol. 1–5, pp. 5117–5122.

[17] M. Hebert and N. Vandapel, "Terrain classification techniques from ladar data for autonomous navigation," in *Proc. Collaborative Technology Alliances Conf.*, 2003. [Online]. Available: http://repository.cmu.edu/cgi/viewcontent.cgi?article=1410&context=robotics

[18] P. J. Watt and D. N. M. Donoghue, "Measuring forest structure with terrestrial laser scanning," *Int. J. Remote Sens.*, vol. 26, no. 7, pp. 1437–1446, Apr. 2005.

[19] M. Hauglin *et al.*, "Estimating single-tree branch biomass of Norway spruce with terrestrial laser scanning using voxel-based and crown dimension features," *Scandinavian J. Forest Res.*, vol. 28, no. 5, pp. 1–14, Apr. 2013.

[20] F. Hosoi, Y. Nakai, and K. Omasa, "Estimation and error analysis of woody canopy leaf area density profiles using 3-D airborne and ground-based scanning lidar remote-sensing techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2215–2223, May 2010.

[21] V. Kankare *et al.*, "Individual tree biomass estimation using terrestrial laser scanning," *ISPRS J. Photogramm. Remote Sens.*, vol. 75, pp. 64–75, Nov. 2013.

[22] V. Kankare *et al.*, "Single tree biomass modelling using airborne laser scanning," *ISPRS J. Photogramm. Remote Sens.*, vol. 85, pp. 66–73, Sep. 2013.

[23] T. Hirose, "Development of the Monsi-Saeki theory on canopy structure and function," *Ann. Botany*, vol. 95, no. 3, pp. 483–494, 2005.

[24] J. M. Chen and T. A. Black, "Foliage area and architecture of plant canopies from sunfleck size distributions," *Agricultural Forest Meteorol.*, vol. 60, no. 3/4, pp. 249–266, Aug. 1992.

[25] J. M. Chen and J. Cihlar, "Quantifying the effect of canopy architecture on optical measurement of leaf area index using 2 gap size analysis methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 777–787, May 1995.

[26] J. Welles and J. Norman, "Instrument for indirect measurement of canopy architecture," *Agronomy J.*, vol. 83, no. 5, pp. 818–825, Sep. 1991.

[27] J. J. Richardson, L. M. Moskal, and S. H. Kim, "Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR," *Agricultural Forest Meteorol.*, vol. 149, no. 6/7, pp. 1152–1160, Jun. 2009.

[28] A. S. Antonarakis, K. S. Richards, J. Brasington, and E. Muller, "Determining leaf area index and leafy tree roughness using terrestrial laser scanning," *Water Res. Res.*, vol. 46, Jun. 2010, Art. ID. W06510.

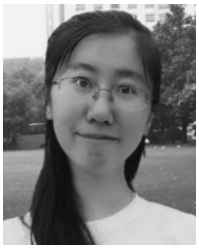
[29] K. G. Zhao and S. Popescu, "Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1628–1645, Aug. 2009.

[30] G. Zheng, L. M. Moskal, and S. H. Kim, "Retrieval of effective leaf area index in heterogeneous forests with terrestrial laser scanning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 777–786, Feb. 2013.

[31] G. Zheng and L. M. Moskal, "Computational-geometry-based retrieval of effective leaf area index using terrestrial laser scanning," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3958–3969, Oct. 2012.

[32] F. Hosoi and K. Omasa, "Factors contributing to accuracy in the estimation of the woody canopy leaf area density profile using 3D portable

lidar imaging," *J. Experiment. Botany*, vol. 58, no. 12, pp. 3463–3473, Oct. 2007.
[33] J. M. Chen, "Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands," *Agricultural Forest Meteorol.*, vol. 80, no. 2–4, pp. 135–163, Jul. 1996.
[34] Y. Zhang, J. M. Chen, and J. R. Miller, "Determining digital hemispherical photograph exposure for leaf area index estimation," *Agricultural Forest Meteorol.*, vol. 133, no. 1, pp. 166–181, Nov. 2005.
[35] J. F. Lalonde, N. Vandapel, and M. Hebert, "Automatic three-dimensional point cloud processing for forest inventory," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-06-21, p. 334, Jul. 2006.
[36] N. Vandapel, R. Donamukkala, and M. Hebert, "Experimental results in using aerial LADAR data for mobile robot navigation," *Field Service Robotics: Recent Advances in Research and Applications*, ser. Springer Tracts in Advanced Robotics, S. Yuta, H. Asama, E. Prassler, T. Tsubouchi, and S. Thrun, Eds. New York, NY, USA: Springer-Verlag, Jul. 2006, pp. 103–112.
[37] J. M. Chen and J. Cihlar, "Retrieving leaf area index of boreal conifer forests using landsat TM images," *Remote Sens. Environ.*, vol. 55, no. 2, pp. 153–162, Feb. 1996.
[38] G. Zheng and L. M. Moskal, "Leaf orientation retrieval from Terrestrial Laser Scanning (TLS) data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3970–3979, Oct. 2012.
[39] L. Rineau and M. Yvinec, "A generic software design for Delaunay refinement meshing," *Comput. Geometry*, vol. 38, no. 1/2, pp. 100–110, Sep. 2007.
[40] H. Gross, B. Jutzi, and U. Thoennessen, "Segmentation of tree regions using data of a full-waveform laser," *Int. Archives Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 36, p. W49A, 2007.
[41] J. Secord, and A. Zakhor, "Tree detection in urban regions using aerial lidar and image data," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 196–200, Apr. 2007.
[42] C. Mallet, F. Bretar, M. Roux, U. Soergel, and C. Heipke, "Relevance assessment of full-waveform lidar data for urban area classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. S71–S84, Dec. 2011.
[43] C. H. Lin, J. Y. Chen, P. L. Su, and C. H. Chen, "Eigen-feature analysis of weighted covariance matrices for LiDAR point cloud classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 94, pp. 70–79, Aug. 2014.



Lixia Ma received the B.Eng. degree in geographic information science from Northwest University, Xi'an, China, in 2012 and the M.Sc. degree in cartography and geographic information science from Nanjing University, Nanjing, China, in 2015. She is currently working toward the Ph.D. degree at the International Institute for Earth System Science, Nanjing University, Nanjing, China.

Her research interests include the application of light detection and ranging in retrieving forest canopy structural parameters and in the field of forest ecosystems.

Guang Zheng received the Ph.D. degree in forest resources management from the University of Washington, Seattle, WA, USA in 2011.

He is currently an Associate Professor of remote sensing with the International Institute for Earth System Science, Nanjing University, Nanjing, China. His research interests include the application of multisource remotely sensed data, particularly light detection and ranging technology in retrieving forest structural parameters at multiple spatial scales and forest aboveground biomass mapping and update using a process-based ecological model.

Jan U. H. Eitel received the Pregraduate degree in forest ecology and management from the University of Freiburg, Freiburg im Breisgau, Germany, in 2002 and the M.Sc. degree in forest resources and the Ph.D. degree in natural resources from the University of Idaho, Moscow, ID, USA, in 2008.

Currently, he is an Assistant Research Professor with the Reveley Geospatial Laboratory for Environmental Dynamics, College of Natural Resources, University of Idaho. His primary research interests include developing and using novel remote sensing tools and techniques to advance our understanding of processes that control the environment we live in.

L. Monika Moskal received the Ph.D. degree from the University of Kansas, Lawrence, KS, USA, in 2005.

She is currently an Associate Professor of remote sensing with the School of Environmental and Forest Sciences, College of the Environment, University of Washington (UW), Seattle, WA, USA, where she is also the Director of the Precision Forestry Cooperative and the Remote Sensing and Geospatial Analysis Laboratory. She is also affiliated with the UW BioEnergy Integrative Graduate Education and Research Traineeship and the UW Interdisciplinary Ph.D. Program. Her research focuses on the multiscale and multidimensional dynamics of landscape change through the application of hyperresolution remote sensing. Her research approach exploits the spatial, temporal, and spectral capabilities of remotely sensed data and focuses on the application of hyperspatial resolution remote sensing (LiDAR and imagery) to investigate vegetation structures, specifically the utilization of the leaf area index in heterogeneous canopies. Her other research themes involve multiresolution and multisensor data fusion, spatiotemporal object-based image analysis, and geovisualization techniques to communicate research results. Her research has been applied to ecosystem services and functions, bioenergy/biomass, forest health and inventories, spatiotemporal assessments and habitat mapping, spatiotemporal wetland assessment, the geostatistical analysis of prairie vegetation communities, urban growth, and forest fragmentation.

Wei He received the B.Sc. degree in geographic information science from Hubei University, Wuhan, China, in 2009 and the M.Sc. degree in cartography and geographic information science from Beijing Normal University, Beijing, China, in 2013. He is currently working toward the Ph.D. degree at Nanjing University, Nanjing, China, and at the University of Groningen, Groningen, Netherlands.

His research interests include quantitative remote sensing of land surfaces, application of remote sensing, and atmospheric measurements in global carbon cycle and climate change.

Huabing Huang received the B.S. degree in soil and water conservation and desertification control from Northwest A&F University, Yangling, China, in 2002, and the M.S. and Ph.D. degrees in cartography and geographic information systems from the Graduate School of the Chinese Academy of Sciences (CAS), Beijing, China, in 2005 and 2008, respectively.

He is currently an Associate Research Scientist with the State Key Laboratory of Remote Sensing, Institute of Remote Sensing and Digital Earth, CAS. His research interests focus on light detection and ranging applications in forest.