Geoinformatica (2013) 17:35–61
DOI 10.1007/s10707-011-0148-1

Computer-based synthetic data to assess the tree delineation algorithm from airborne LiDAR survey

Lei Wang · Andrew G. Birt · Charles W. Lafon · David M. Cairns · Robert N. Coulson · Maria D. Tchakerian · Weimin Xi · Sorin C. Popescu · James M. Guldin

Received: 7 December 2010 / Revised: 2 October 2011
Accepted: 2 November 2011 / Published online: 29 November 2011

Abstract Small Footprint LiDAR (Light Detection And Ranging) has been proposed as an effective tool for measuring detailed biophysical characteristics of forests over broad spatial scales. However, by itself LiDAR yields only a sample of the true 3D structure of a forest. In order to extract useful forestry relevant information, this data must be interpreted using mathematical models and computer algorithms that infer or estimate specific forest metrics. For these outputs to be useful, algorithms must be validated and/or calibrated using a sub-sample of 'known' metrics measured using more detailed, reliable methods such as field sampling. In this paper we describe a novel method for delineating and deriving metrics of individual trees from LiDAR data based on watershed segmentation. Because of the costs

L. Wang (✉)
Department of Geography & Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA
e-mail: leiwang@lsu.edu

A. G. Birt · R. N. Coulson · M. D. Tchakerian
Knowledge Engineering Laboratory, Department of Entomology, Texas A&M University, College Station, TX 77843, USA

A. G. Birt
e-mail: abirt@tamu.edu

R. N. Coulson
e-mail: r-coulson@tamu.edu

M. D. Tchakerian
e-mail: mtchakerian@tamu.edu

C. W. Lafon · D. M. Cairns
Department of Geography, Texas A&M University, College Station, TX 77843, USA

C. W. Lafon
e-mail: clafon@tamu.edu

D. M. Cairns
e-mail: cairns@tamu.edu

involved with collecting both LiDAR data and field samples for validation, we use synthetic LiDAR data to validate and assess the accuracy of our algorithm. This synthetic LiDAR data is generated using a simple geometric model of Loblolly pine (*Pinus taeda*) trees and a simulation of LiDAR sampling. Our results suggest that point densities greater than 2 and preferably greater than 4 points per m2 are necessary to obtain accurate forest inventory data from Loblolly pine stands. However the results also demonstrate that the detection errors (i.e. the accuracy and biases of the algorithm) are intrinsically related to the structural characteristics of the forest being measured. We argue that experiments with synthetic data are directly useful to forest managers to guide the design of operational forest inventory studies. In addition, we argue that the development of LiDAR simulation models and experiments with the data they generate represents a fundamental and useful approach to designing, improving and exploring the accuracy and efficiency of LiDAR algorithms.

Keywords Computers synthetic data · Tree delineation · Algorithm · LiDAR

1 Introduction

Measuring the biophysical properties of forest landscapes is an important challenge for modern forestry. The current condition or state of the forest is an essential input to decision making—for example through forest growth, harvesting, fire hazard, and pest and disease models. Recent trends towards sustainable and regional management of forest landscapes emphasize the need for accurate and extensive forest inventories [1]. Regional management necessitates data collection over broad spatial scales, even for areas that may not be subject to active management while sustainable forestry is driven by the efficient use of forest resources, which in turn requires detailed understanding of forest processes. Whereas in the past, measurements such as vegetation type, Basal Area (BA) and mean Diameter at Breast Height (DBH) were considered satisfactory descriptors of forest composition, today there is a drive towards finer scale measurements including the location, species, condition and size of individual trees. In short, modern forestry requires the measurement of increasingly detailed information over broader spatial scales.

LiDAR has been proposed as a tool for efficiently measuring forest landscapes to achieve both high levels of detail and broad spatial coverage. Discrete return LiDAR data acquired by small footprint airborne laser scanners (ALS) consist of a collection of spatially distributed points, each representing an intercepted target along the path of a laser pulse [2]. Each point represents an estimate of the location, and height of an object (for example a leaf or branch of a tree) intercepted by the laser while the collection of points approximates the surface of complex structures such as understory vegetation, canopy surface or whole trees. It follows

W. Xi
Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, Madison, WI 53706, USA
e-mail: wxi3@wisc.edu

S. C. Popescu
Spatial Sciences Laboratory, Department of Ecosystem Science and Management,
Texas A&M University, College Station, TX 77845, USA
e-mail: s-popescu@tamu.edu

J. M. Guldin
Southern Research Station, USDA Forest Service, Hot Springs, AR 71901, USA
e-mail: jguldin@fs.fed.us

that LiDAR measurements are affected by the fine scale structure of a sampled area. For example in trees the laser may penetrate a canopy such that points are recorded from within the tree crown rather than the surface of a canopy. In addition, for every laser pulse fired from the LiDAR equipment, only a fraction of the energy will return to the sensing equipment.

Because LiDAR produces only a 3D ‘sample’ of real world objects, algorithms are required to interpret the data and estimate or infer the properties of the underlying landscape. For forestry applications, algorithms have been developed to estimate stand level characteristics such as Mean Height [3], Basal Area [4, 5], Stand Volume [6], Canopy Fuel Load [7], Leaf Area Index [8], and canopy gaps caused by disturbances [9]. Many algorithms have been developed to detect and measure the characteristics of individual trees [3, 10–14]. In all cases, an overriding aim of LiDAR studies is to obtain information as efficiently and accurately as possible, and in common with all scientific measurement the goal should be to provide estimates plus the accuracy or confidence of these estimates. In most LiDAR studies, accuracy assessment usually involves comparing results from LiDAR algorithm to a subset of more detailed and intensively sampled ground measurements from the same area. However, given the cost of ground surveys, validation using high quality satellite or aerial images (and expert interpretation) has also been proposed [15].

Although numerous studies have focused on the use of small footprint LiDAR to derive detailed forest information by using a variety of data processing algorithms, there still lacks sufficient understanding of the uncertainty caused by many external and internal factors. These studies reported their results as individual cases, using LiDAR data with different sampling quality, and targeting to a variety of forest stands. It is very difficult to integrate these studies together to form a complete picture of the complex LiDAR data acquisition, processing, and interpretation system. In this paper, we make the very first attempt of the assessment procedure to investigate the major issues related to LiDAR forest studies, and initialize a research framework which could help answer some key research questions such as “Is the LiDAR sampling density a critical factor of individual tree delineation?”; “Is forest stand mean DBH a controlling factor of the accuracy of individual tree detection?”. Success of this research will add to our current knowledge of the use of LiDAR technology in forest studies, and set a basis for further exploration in similar directions.

In this research we use a computer program to mimic the LiDAR sampling characteristics and run controlled sampling tests over a synthetic forest stand simulated from geometric tree models. This provides data sets quantitatively similar to those created by real LiDAR measurements, but with the advantage that we have exact knowledge of the size, location and spacing of trees in the original stands. By comparing the output of LiDAR processing algorithms with the ‘known’ properties of the stand, the success of our algorithms can be measured directly and accurately. And by systematically varying the properties of synthetic stands and the simulated LiDAR sampling, it is possible to obtain a mechanistic understanding of the problems involved with collecting and processing LiDAR data that is a useful reference for all forest managers. In the following sections, we will present the procedure of creating the artificial forest stand model, and then describe the method of generating point clouds that resemble the major characteristics of airborne small-footprint LIDAR survey. In tree detection we use the watershed segmentation algorithm. In the following the delineated tree crowns and center locations are compared to the preset values so to evaluate the accuracy and uncertainties in the tree detection algorithm. Some discussions are then based on varying some of the controlling factors such as the point density, footprint size, and mean DBH. Finally, some conclusions will be drawn with previsions of further research works.

2 Methods

2.1 Synthetic forest stand

Real world forest landscapes are complex 3D aggregates of structures such as tree trunks, branches, needles and leaves, rocks, undulating ground surfaces, bushes, surface water etc. Each of these surfaces has unique characteristics such as texture and color. However, our goal is to generate ‘artificial’ forest stands made up of trees comprising aggregations of simple geometric shapes, including crowns, trunk, and ground level. These trees represent a level of detail appropriate for the purposes of deriving ‘synthetic’ LiDAR data (we deal with penetration caused by surface complexity in a following section). The tree species of interest is Loblolly Pine (and other Yellow Pines) grown in East Texas. Since our purpose is to test the LiDAR sampling scenarios, the forest stands are represented as a simplified model with following assumptions:

1) The shape of the stand is a 50×50 m square, and the ground surface is horizontal and flat. The ground level is used as the reference with zero height and all other vertical measures are made relative to this level.
2) A variable number of geometric trees comprising cylindrical trunk and ellipsoidal crowns in the stand. We assume that the crown and the trunk are made of the same material (with the same penetration properties—see next section). We define each tree by its x,y location within the stand and its DBH. The ratios of DBH to tree height, crown width and crown height are given by the following equations [16]:

$$Log(h) = 0.0785 + 0.869\ Log(DBH) \tag{1}$$

$$Dc = 0.904 + 0.168\ DBH \tag{2}$$

where h is tree height in meters, DBH is the diameter at breast height in centimeters. Dc is the diameter of tree crown in meters. The location, spacing, size and number of trees are defined by three stand variables—mean DBH, standard deviation of DBH and total Basal Area.

3) Regular distribution of the trees

Generation of a stand proceeds by drawing normally distributed random number that represents a diameter value for an individual tree. The use of the regular distribution of tree location can allow us to evaluate how these variables (DBH and BA) affect tree delineation. By setting the tree distance and crown sizes, the degree of crown overlapping is controlled so to examine the discrimination ability of the LiDAR algorithm. The trees are spaced on a grid, namely, the center of the tree trunk is equally spaced. Although this type of distribution is not realistic, it does not influence our analysis of the LiDAR sampling problem because the only factor that matters is the percent of crown overlapping. Another approach to tree positioning would be randomization. It might generate some realistic-looking stands. But it could also generate conditions that are unlikely present such as two very large trees located very close to each other. If there is too much overlapping, the tree delineation algorithms will stop working. Under the circumstances of regular distribution, the amount of overlapping highly depends on the stand variable parameter we select for generating the trees, but not controlled by the randomness of tree locations. New trees are created by the randomly drawn DBH and added to the stand until the target Basal Area of the stand is reached. Then the trees are put on a grid occupying the 50 m by 50 m stand. For example, if totally there are 15 tree generated, the grid will have 4 rows and 4 columns, leaving the last spot of the grid empty. Using the diameter of each tree the height, crown width and crown height are calculated using Eqs. 1 and 2. If there is crown overlap, the upper-most part is considered in the following sampling procedure.

2.2 LiDAR sampling

Data collected from LiDAR are the result of interactions between the technical specifications of the LiDAR equipment (including the beam divergence, sampling rate, scan geometry, laser power, pulse length, and wavelength), external factors (such as the flight speed, height of the aircraft, distance to the GPS base station, flight path overlap), and the complex structure of the forest and their surface properties. To efficiently acquire laser measures, a LiDAR system is usually equipped with a rotating/oscillating mirror which allows scanning ability.

In this study we are interested in the how the sampling quality is related to the stand parameters and tree characteristics. Such studies lead to a 3D cloud of points with each point (or sample) defined by x, y and z coordinates (where z is height). The sampling quality can be summarized into four types: sampling density, locational accuracy, vertical accuracy, and volumetric backscatter probability.

2.3 Sampling density

Sampling density can be measured by the count of points per unit sample area or by the average spacing between adjacent points. It is the major parameter of LiDAR surveys. The sampling density is controlled by many factors: sampling rate, flight speed, rotating/oscillation rate of the prism/mirror system. The higher the laser frequency, the higher density the sample points. As limited by the maximum laser pulse frequency and scanning patterns set by the manufacturer, obtaining higher sampling densities can be implemented by adjusting the fly speed and height.

2.4 Locational accuracy

The locational accuracy is related to two factors: the combined systematic errors (e.g. DGPS, IMU, timing errors, etc.) [17, 18] and the laser footprint size resulted from the laser beam divergence and flight height. The magnitude of the latter, nevertheless, is usually greater than the systematic errors. For example, the high-end differential GPS with custom installed base stations can achieve positioning accuracy of several centimeters, while the footprint sizes of the projected laser beam on the ground targets are about 30~60 cm [12]. The returned signal detected by the receiver will record the location at where the footprint is centered. The actual location where the major reflectance occurs can be anywhere inside the footprint. The location of the sample points is then determined in a stochastic way. Within the footprint of the laser illumination on the ground, the location is randomly selected following a Gaussian distribution, in which the center of the footprint has the highest probability of distribution.

2.5 Ranging uncertainty

The laser ranging is done by multiplying the speed of light and the time elapsed between the sending and receiving of the laser energy. It is related to the laser pulse length and the accuracy of the clock. Combined with the IMU and DGPS errors, the magnitude of the vertical measurement (Z) is about 10~30 cm [19]. The common standard of LiDAR vertical accuracy is 15 cm. We use this number to disturb and add a random component to the height values of the LiDAR sample points. For example, we assume the ground of the stand is flat with a zero height. The height of the ground samples follows a normal distribution with zero mean and standard deviation of 15 cm.

2.6 Backscatter probability

Not all laser pulses result in a LiDAR return. The surface property of materials and their orientation relative to the laser affect LiDAR backscatter (the proportion of the pulses returned to the sensor). Pulses that do not return to the sensor are termed clutter. Although a forest comprises a large number of basic, solid elements (trunks, branches, and needles or leaves), the arrangement of these objects in 3D space is complex. Although it is convenient to model trees with smooth well defined geometry, in reality they have complex, fractal like outlines and are permeable to the LiDAR pulse. The backscattering probability can be modeled as an exponential function of leaf density and the structure of the foliage area [20]:

$$p(Z) = e^{-\int_0^Z u(z)G(z)dz} \tag{3}$$

where p(Z) is the bidirectional gap probability at depth Z; u(z) is the density of leaf area at depth z; and G(z) is the Ross-Nilson G function [20]. We simplify this model by assuming a constant value for u(z) and G(z) which means the density and the geometric shape of the leaves does not change when laser beams go deeper into the tree canopy. Then the bidirectional probability function is an exponential function of the penetration depth:

$$p(Z) = ce^{-Z} \tag{4}$$

where c is a constant value. There are two effects of backscatter. First, since valid returns occur randomly in the x, y plane (ground surface), the distribution of points in this plane will also be random. Secondly, because of surface roughness and permeability, the z-measurement of a LiDAR return will not represent the local maximum height of the tree. Figure 1 illustrates the effects of backscatter on the sampling error of LiDAR returns.

Based on this mechanistic description, we created point clouds that mimic the sampling characteristics of small footprint LiDAR flown over our synthetic loblolly stands. The point clouds carry the characteristics of the LiDAR sampling errors and sampling density. The following steps are the procedure to create the point clouds:

1) Set a 'potential' coverage that determines the maximum number of LiDAR samples (footprints) per unit area (m2).
2) For each 'potential' LiDAR point, generate a random number from a uniform distribution that determines whether a valid return occurs. If this random deviate is less than a critical value, the point is omitted from the data set.
   Note that steps 1 and 2 determine the number of LiDAR points per unit area (m2).
3) Offset the x and y position of each valid return by a uniform random deviate between +/− half the footprint size.
4) Given the x, y location of each valid return, determine the height of the geometric object perpendicular to this point.
5) To simulate penetration and/or roughness of the simple 3D stands, the actual height of the scene is offset by a random variable. If the x, y location coincides with a tree crown, offset the height of the point by subtracting a random variable drawn from an exponential distribution. If the x, y location corresponds to a point on the ground, a random variable drawn from a normal distribution is used.

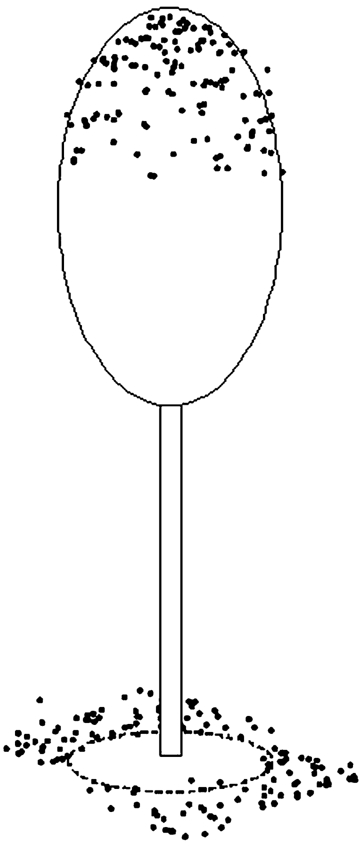


Fig. 1 Diagram showing simple geometric trees and the distribution of synthetic LiDAR data sampled from the trees. The tree on the left shows LiDAR samples with x,y error and backscatter error only. The tree on the right shows LiDAR samples with additional penetration error

The above steps produce data representing synthetic point clouds (see Fig. 1 for an example of a point cloud surrounding a single tree). Steps 1–5 are repeated for each synthetic stand, leading to sets of paired of data—the known location and size of each tree in a stand and the LiDAR sample of the stand. This data is analogous to a LiDAR study plus a detailed field sample of an area. The forest stand can be re-created with other sets of parameters so to study the influence of the stand characteristics.

In the next section, we describe a method for processing this synthetic point cloud to estimate (back calculate) locations and size of trees within the stand. For most forest managers, this is the raison d'etre of a LiDAR study—to extract valuable ecological or forestry information from the data. However, this paper addresses more fundamental issues —i.e. the accuracy associated with LiDAR processing algorithms, the pitfalls and design of LiDAR studies and therefore the amount of information that can be reliably extracted from LiDAR. By using synthetic data, the location and sizes of trees in each stand are known exactly making it easy to assess the accuracy and performance of a LiDAR processing algorithm. In addition, it is possible to vary both the physical properties of the stand (the sizes, density and juxtaposition of trees) and details of the LiDAR sampling process (footprint size, backscatter, and penetration) and explore their effects on the performance of the algorithm. This information provides valuable insight into the utility and management of LiDAR studies and the development of effective LiDAR processing algorithms.

2.7 Tree detection algorithm

In this section we describe a series of algorithms to obtain the location and size of individual trees from LiDAR data. This is a common requirement of LiDAR processing algorithms. Many different paradigms, algorithms and software exist for processing forestry LiDAR data. The method described here is an enhanced inverse watershed segmentation method, which has been used by many researchers [e.g. 13, 21–23]. In outline, the method proceeds as follows:

1) A local maxima filter is used to remove non-canopy or under-canopy returns.
2) An interpolation algorithm rasterizes point data model into a canopy height model (CHM)
3) A watershed segmentation algorithm delineates the location and size of the tree from the CHM.
4) Post-processing algorithms are used to further delineate trees.

We discuss these steps in detail in the next sections. All of the methods described in this paper are coded in C++ programming language and compiled using Microsoft Visual Studio and tested under a Microsoft Windows environment.

2.8 Local maxima filter

Local maxima filters (also sometimes known as box-maxima operators) have been widely used in forestry based LiDAR algorithms, e.g. [12]. The local maxima filter is used to separate points representing the canopy surface from those that have penetrated the canopy. The local maxima filter proceeds by dividing the x, y plane of the study area (stand) into a grid of 0.2×0.2 m cells. The value of each cell is set to the highest (maximum z-value) LiDAR point whose x, y location falls within the cell boundary. As such, points representing LiDAR reflections from (or very near to) the canopy surface are preserved and used for further analysis. In local maxima filters, the size and shape of the cells determine the results of the analysis and should be selected according to LiDAR point density and size of the features (trees) to be derived from the data. In this case, based on our experiments, the cell size was set to approximately 1/20th the size of a tree crown.

2.9 Inverse Distance Weighted interpolation

Once the LiDAR points have been filtered, the points are interpolated to a 0.2 m resolution grid to produce a canopy height model (CHM). Various interpolation methods have been used to create a CHM from LiDAR samples. These include Nearest Neighbor [24], Inverse Distance Weighted Interpolation [25], Minimum Curvature Splining [26, 27], and Kriging [28, 29]. The general problem of interpolation is to infer the value of a line or surface between a finite set of measured points. As such, interpolation methods have no unique solution and each has specific limitations and advantages. Kriging is widely used [e.g. 25, 30–32]. However, none of the existing interpolation method is capable of dealing with the discontinuity between canopy points and ground points in LiDAR data. In this research, we developed a modified IDW interpolation method by adding a localized median filter to remove the effects of discontinued data. We call it the Edge-Preserving IDW method. In canopy-on LiDAR data (as comparing to bare-earth data), the discontinuity is very common and therefore it is more important to deal with the edge problem, if any, than other features of the interpolation algorithm. With its great simplicity and fast computation speed, the Inverse Distance Weighted (IDW) can be preferred at the circumstances when the data sample density is high and the distance decaying property is obvious, such as the airborne LiDAR data. The Edge-Preserving IDW algorithm can preserve the edge of canopy well by adding a robust median filtering function to its neighborhood function. The filter takes a robust statistic offered by the calculation of Median About Deviation (MAD) [33]. The filter removes the values enclosed by the IDW neighborhood, which is considered as outliers by the MAD method. As such, when the edge of tree crown is encountered, the MAD filter can remove the edge-blurring effects of the outliers (e.g. those ground points near the crown edge). These boundaries are important for the next stage of the tree delineation algorithm. Our observation also revealed that this robust IDW interpolation algorithm can reduce the well-known “bull’s eye effect” [34] of the IDW interpolation which occurs when discrepancy in the data occurs in very short distance. Figure 2 shows CHMs derived from synthetic data sets (described above) at point densities of 8, 4, 2, and 1 point(s) per m2.

2.10 Constrained watershed segmentation algorithm

Watershed segmentation algorithms are widely used in image processing, pattern recognition and topology segmentation. They are based on the idea of watersheds in natural terrain models. Imagine that a drop of water falls to the ground and follows gravity downhill. If there is a boundary where water drops flow to different sides of it, this boundary is called a “divide” or “ridge” and the areas delineated by divides or ridges are called “watersheds”. Applied to forest LiDAR and the delineation of trees, the crowns are flipped upside down so that crown apexes become watershed basins. The watershed algorithm proceeds by identifying minima that represent outlets of watersheds. Then, a connected component searching algorithm is used to locate a neighborhood of image pixels that ‘contribute’ to the minima. We modify the method of Bieniek and Moga [35] which is a more efficient version of an original method described by Soille [36]. In our algorithm we assume tree crowns are symmetric, compact shapes. By incorporating this information into the data processing, we add a useful constraint to the algorithm that prevents delineated areas becoming unrealistically stretched, asymmetric, and oversized. Another constraint for the algorithm is a parameter (slope threshold) that determines when a local change of height in the CHM is considered indicative of the edge of a crown. This is based on the assumption that in real canopies, heights drop quickly (large slopes) at the edge of crowns. In our algorithm, we set the slope constraint to 100%, which means if the height change is larger than the length of the data pixel size, it is considered as the crown edge. As a given watershed is expanded from a local minima, the boundary of the watershed is marked wherever a slope (ratio of the heights in adjacent pixels) greater than 100% is encountered. One problem associated with watershed algorithms is over-segmentation where small watersheds are enclosed by larger ones and may be misclassified as trees. Over-segmentation also leads to under-estimation of crown area. To overcome this problem, we use a strategy similar to the marker-controlled procedure [37]. First, local maxima are found and marked as candidates to grow into watersheds.

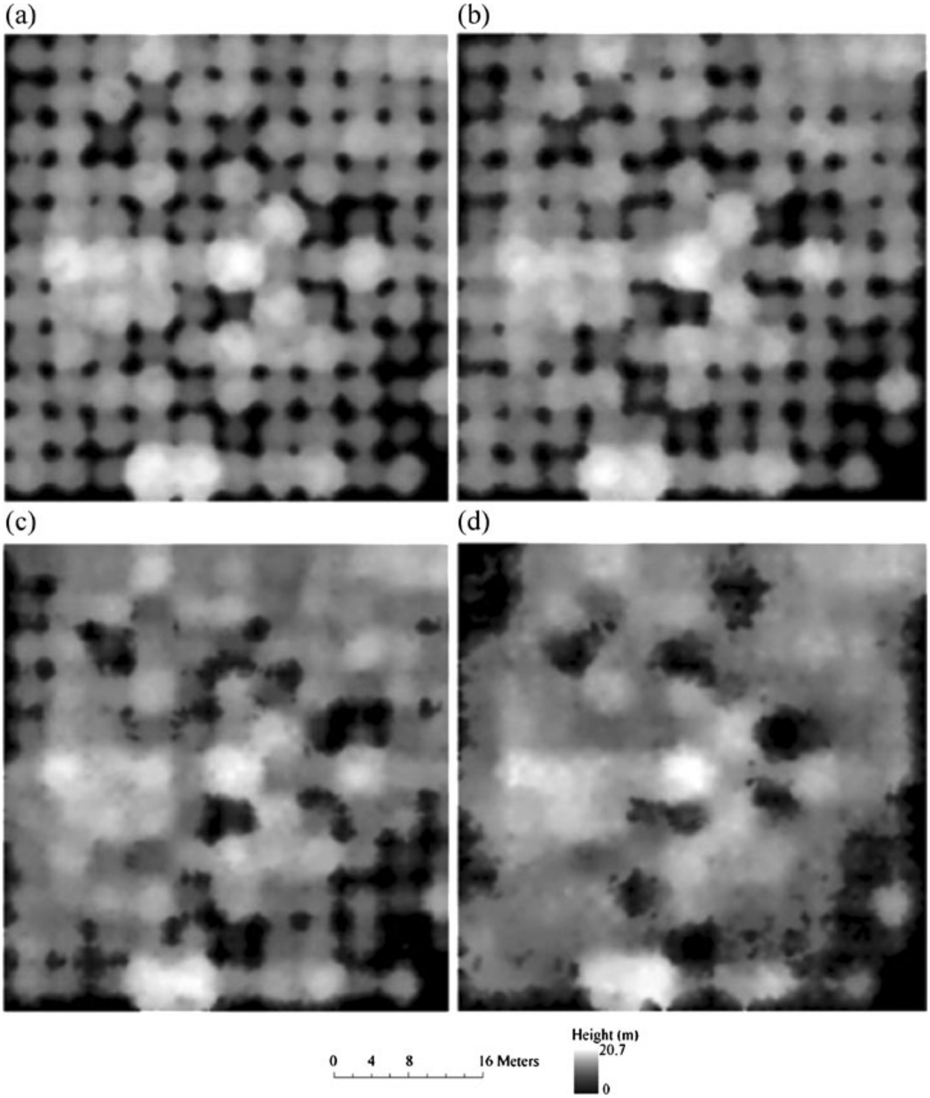


Fig. 2 Canopy height models derived from synthetic LiDAR data simulated to yield densities of 8 (**a**), 4 (**b**), 2 (**c**) and 1 (**d**) points per m^2

From each of these local maxima the nearest 4 neighbors will be checked to see if any are likely to be a crown covering this point and if so, the current point is marked as a sub-watershed and is merged into the larger one.

2.11 Post-processing

The watershed segmentation algorithm effectively delineates crown boundaries but may lead to fragmented shapes and underestimation of crown sizes. To solve this problem we implemented shape morphology operations. These morphology operations include two basic actions: dilation and erosion. The dilation operation expands the area of the feature in the image while an erosion operation shrinks the area of the feature in the image. The structural element used in the operations is a 3 by 3 square shape. The erosion is performed after the dilation (also called closing operation). The sequence of operation is performed twice to ensure fully removal of boundary irregularities.

Figure 3 shows the results of the steps described above using synthetic stand data and LiDAR points. Centroids of delineated regions are used to estimate tree location, the maximum height of the region is used to estimate tree height and the mean diameter of the region is used as an estimate of the crown diameter. The next section discusses how these values are compared with known locations and crown diameters in the synthetic stand in order to assess the success of the algorithms.

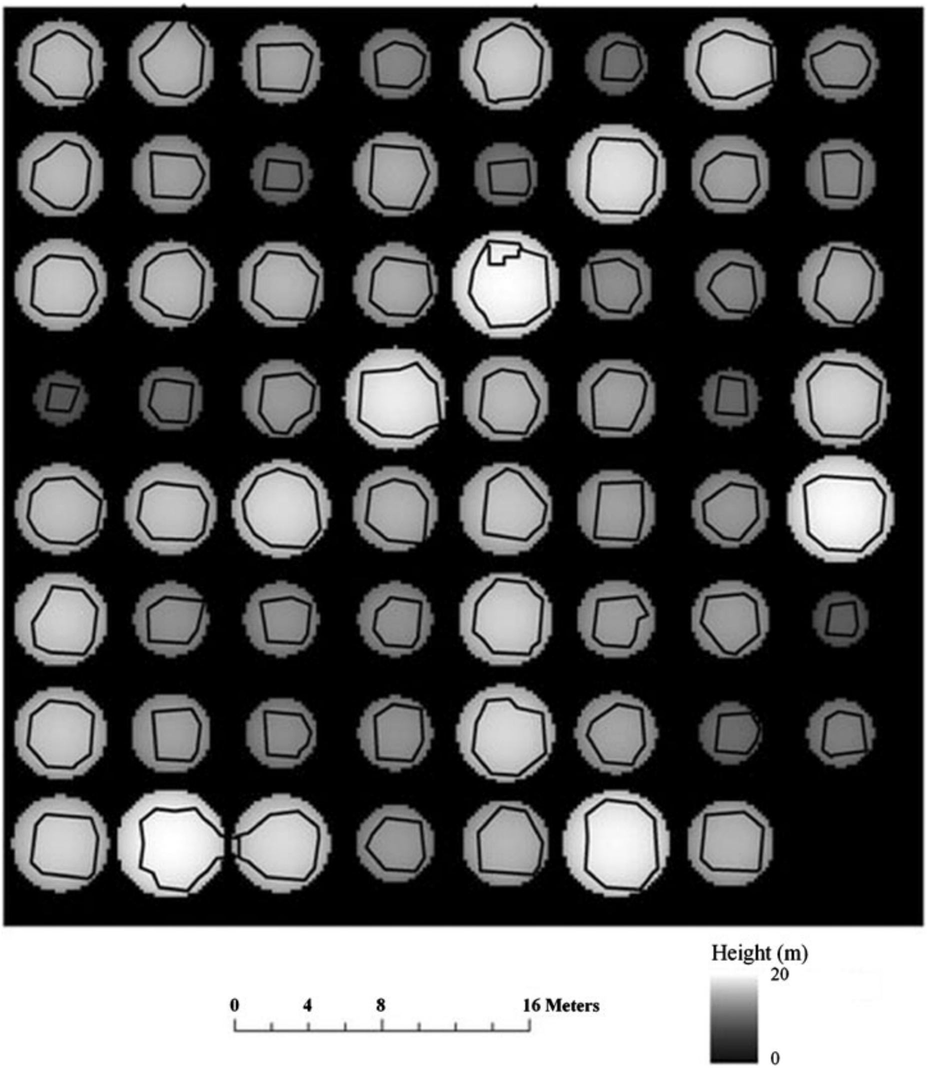


Fig. 3 Delineated crowns compared to actual tree crowns using synthetic LiDAR data at 4 points per m^2. Note the consistent under-estimation of crown diameter and a single case of over-segmentation

2.12 Validation

Our primary goal is to determine the effects of forest structure and LiDAR sampling quality on the accuracy of the tree delineation algorithms described above. Although there are a number of interrelated factors that we could analyze using this system (including the effects of tree geometry, tree size, stand management or LiDAR penetration) the two factors that we concentrate on are LiDAR point density and the Basal Area of the stand. To investigate these factors we repeat the following steps, controlling for Basal Area and LiDAR point density:

1) Create a stand with a given Basal Area.
2) Simulate a set of sample points from the stand with a specified density
3) Use our LiDAR processing algorithm to calculate the location and the crown diameter of each tree in the stand.
4) Determine the success of the algorithm by compare the results of Step 3 with the 'known' location and crown diameters of trees in the stand.
5) Repeat the process by altering the Basal Area in 1) or the sampling density in 2)

We use two methods for assessing the success of the interpretation algorithm—omission and commission error. Omission error is the inability of the algorithm to detect a real tree at a defined level of accuracy. In our case, we consider a tree to be successfully detected if the algorithm estimates a tree location less than 4 m away from the tree's true location (i.e. known exactly from the synthetic stand) and if the estimated height and crown diameter is within 20% of the true height and crown diameter. We use 4 m as the threshold because it is the about crown diameter of a loblolly pine with 10 cm DBH according to Eq. 2. 10 cm DBH is about the size of smallest tree in our simulation. Since a stand consists of many trees, the omission error for the stand is as follows:

$$E_o = \frac{N_m}{N_c + N_m} \tag{5}$$

where E_o is the omissions error, N_m is the number of real trees that are missed, and N_c is the number of trees that are correctly identified. Commission error occurs if the algorithm estimates trees estimated by the algorithm when no actual tree exists (using the same criteria as for omission):

$$E_c = \frac{N_m}{N_c + N_e} \tag{6}$$

where E_c is the commission error, N_e is the number of derived trees that are not correct, and N_c is the number of trees that are correct.

3 Results

We used these procedures to investigate algorithm given stands with BA ranging from 10 to 30 m2/ha and for LiDAR sampling parameters giving 0.5–20 points/m2. For each stand the LiDAR penetration rate (10%) and the DBH of trees (normally distributed with mean= 21.4 cm, std. dev.=4 cm) were fixed.

Figure 3 shows delineated watershed regions compared to the original canopy model for the same stand. In this case, the watershed algorithm successfully derived most of the actual tree locations, but consistently underestimated the diameter of tree crowns. Figure 4 shows

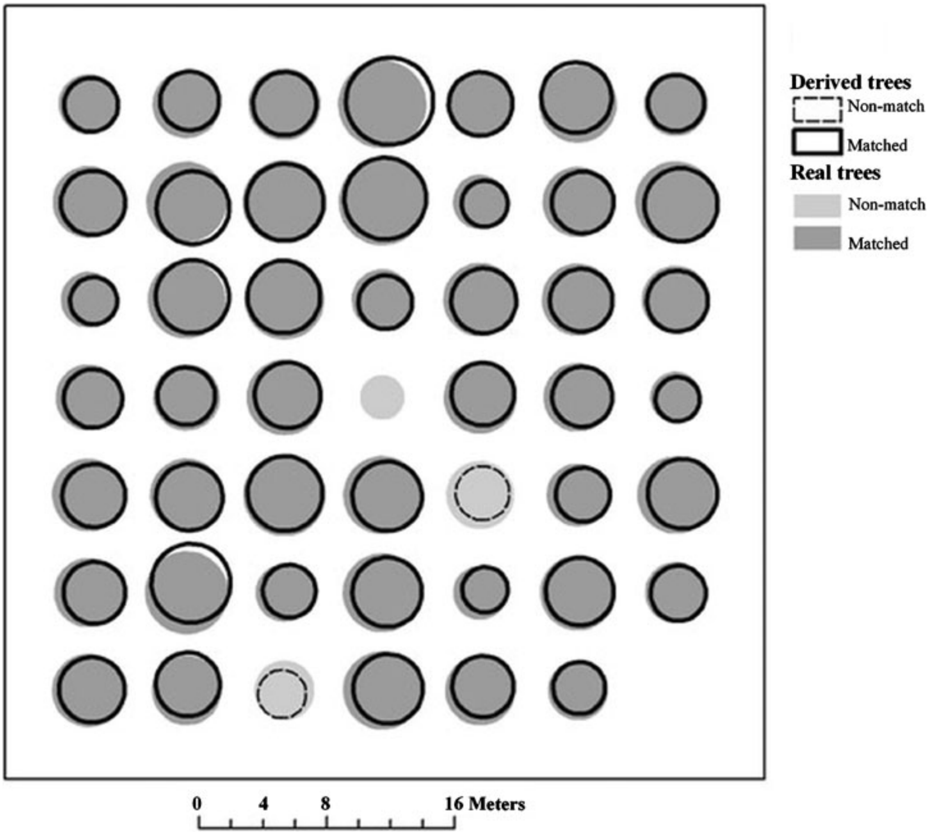


Fig. 4 Mapped output showing cases of omission error. Note that 2 trees are fairly successfully derived but do not meet the threshold set for successful detection (i.e. a<20% difference in crown or location error). Note also that this 20% threshold, when viewed visually is actually fairly stringent

a comparison of estimated tree location and area versus the actual stand model for the same LiDAR data/synthetic stand pair. In this case, 3 trees were dropped from the watershed regions (and therefore contribute to omission error) because they did not meet the accuracy criteria based on crown diameter.

The interaction between footprint size and estimation accuracy is shown in Fig. 5. Typical footprint size for discrete LiDAR systems are between 0.3 and 2 m. The graph shows that over this range, footprint size has little effect on the success of the algorithm. Based on these results, for the rest of the analysis we fixed the footprint size to 0.5 m.

Figure 6 shows the interaction between Basal Area and estimation accuracy. For a fixed LiDAR point density, Basal Area has a large effect on the estimation accuracy. Similarly, Fig. 7 shows the effect of LiDAR point density on estimation accuracy for low (10 m2/ha) and high (20 m2/ha) Basal Area stands. Both graphs show a negative relationship between LiDAR point density and omission error, but that the nature of this relationship is different for the high and low BA stands. In particular, the relationship between point density and accuracy is linear for high density stands, but is non-linear and asymptotic for low density stands. In both graphs, the scatter of points around the best-fit line indicates that omission

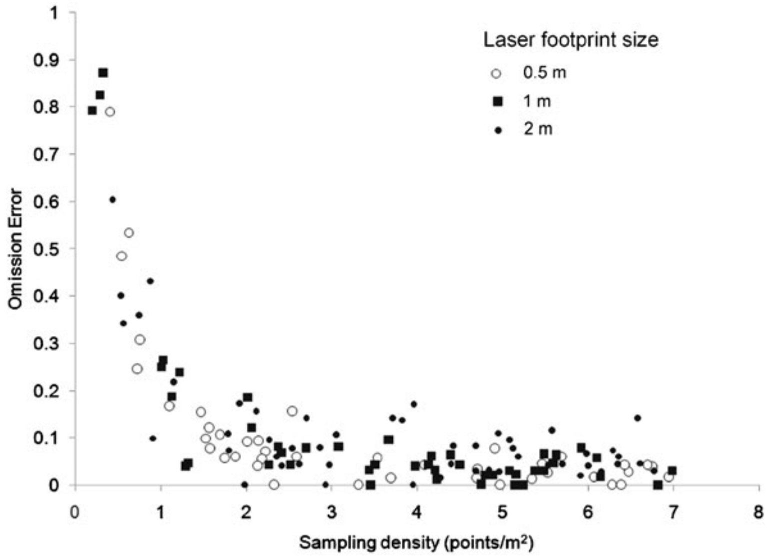


Fig. 5 Effect of LiDAR footprint size on tree identification. The graph shows the success of the tree identification algorithm over a range of sampling densities and assuming footprint sizes of 0.5 m (*open circles*), 1 m (*closed squares*) and 2 m (*closed circles*). The graph suggests that footprint size has a minimal effect on the success of tree detection

error is related to the specific juxtaposition of trees within a stand in addition to a simple measurement of Basal Area.

Figure 8 shows the biases associated with the LiDAR processing algorithm. Crown diameter is consistently underestimated even at high point densities. Although the mean

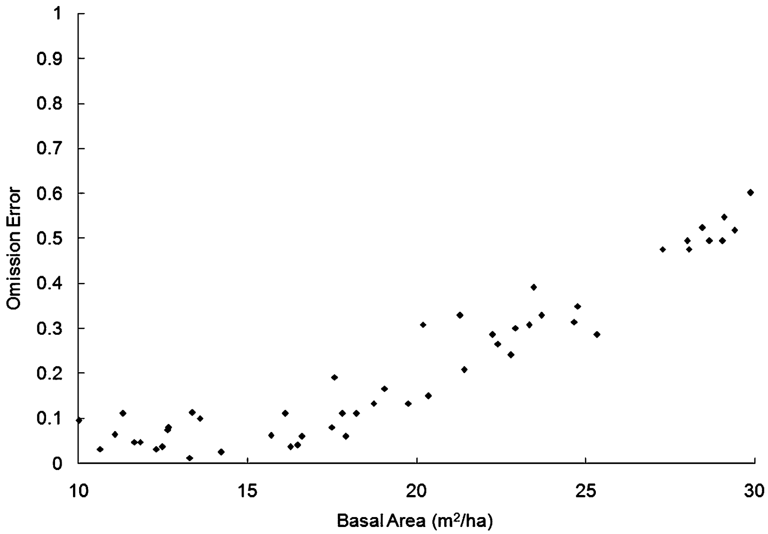


Fig. 6 Graph showing the effect of stand Basal Area on algorithm performance. Each point represents the omission error for one synthetic stand. Omission error increases with stand Basal Area. Note that for the species we are simulating (Loblolly pine), canopy closure occurs at Basal Areas greater than approximately 20 m^2/ha. All results were obtained at a point density of 4 per m^2

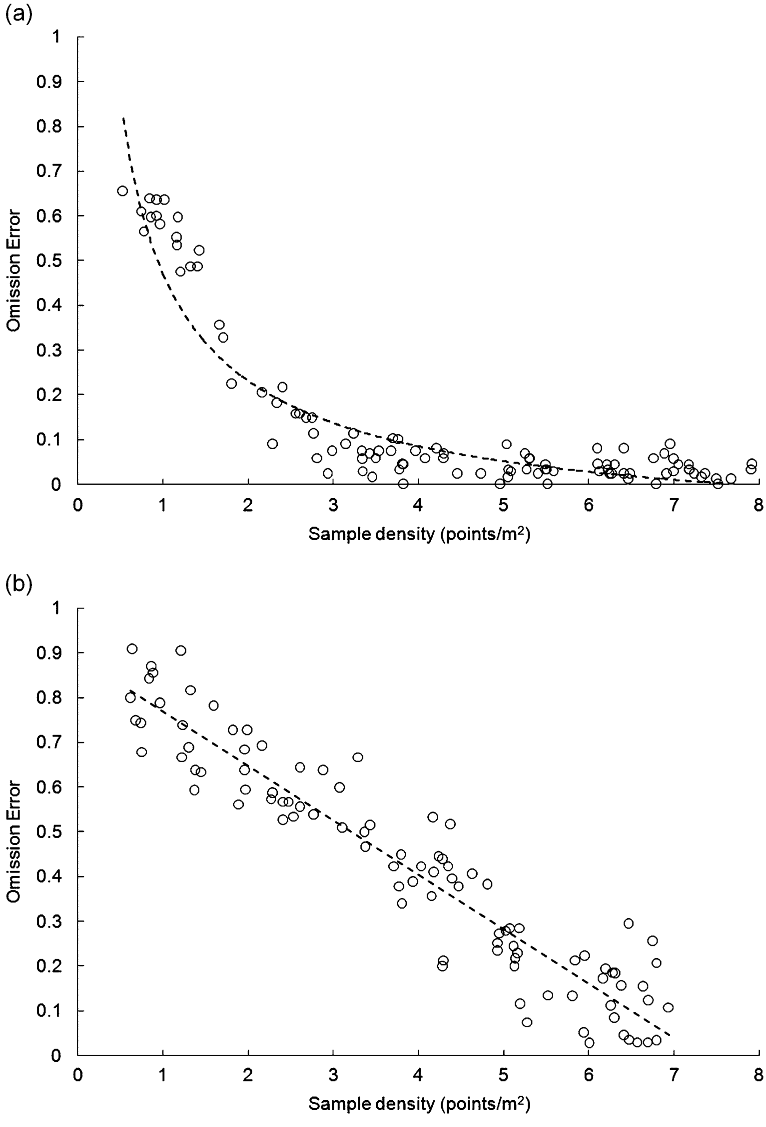


Fig. 7 Graphs showing the effect of LiDAR sampling density (points per m^2) for **a** low (10 m^2/ha) and **b** high (20 m^2/ha) stands. Each point shows omission error for a single stand. In both cases, omission error decreases with point density, but the shape of the response is different for low versus high BA stands. A practical interpretation suggests that for low density stands there is likely to be a threshold point density above which extra sampling does not contribute to increased accuracy. However, for high BA stands, extra points will have a direct effect on the accuracy of results

error (i.e. the average over all trees in the stand) is largely unaffected by the Basal Area of the stand, the standard deviation of error (the variation in error for each tree in the stand) is larger for high Basal Area stands. In both cases, the error associated with crown estimation (both the mean and standard deviation) decreases with increased point density, but the graphs suggest that the algorithm will consistently underestimate crown diameter irrespective of point density.

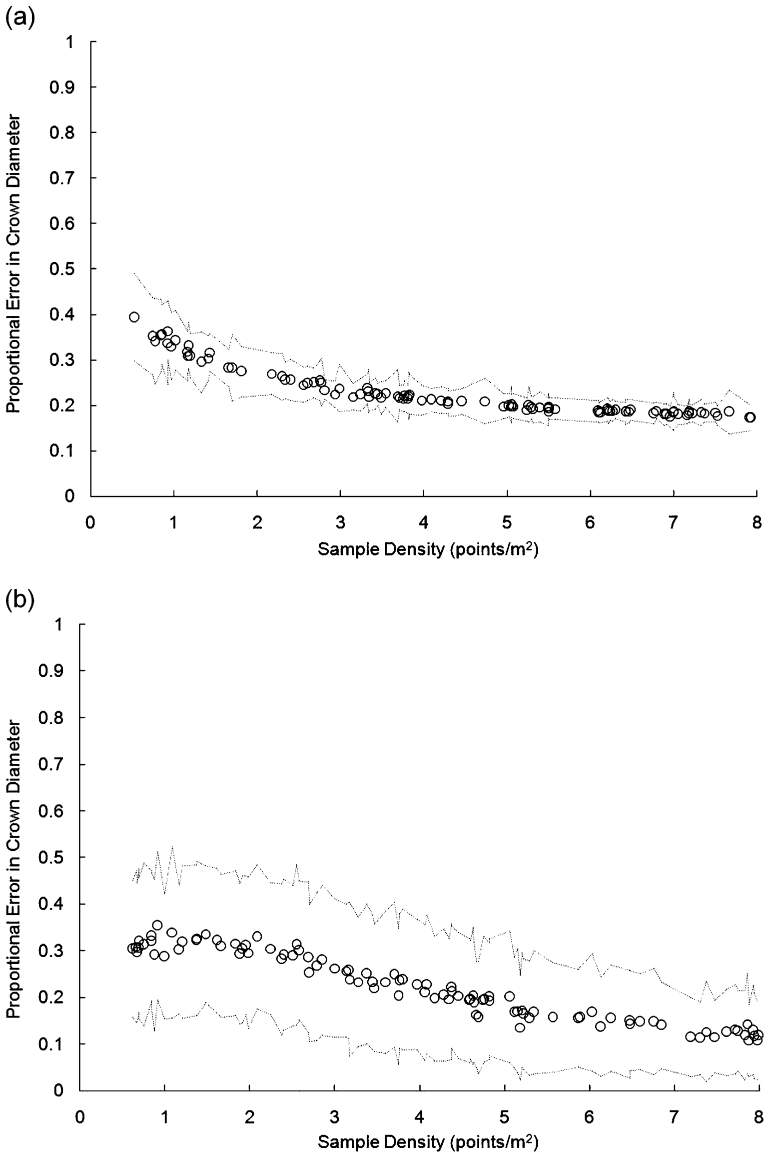


Fig. 8 Graphs showing the bias in estimating crown diameters. Each point shows the mean error (per stand) in crown estimation for trees that have been correctly classified based on their x, y location. The outer lines show the standard deviation of error for estimated crown diameter for each stand. The graphs illustrate the relationship between crown error and point density for Basal Areas of **a** 10 and **b** 20 m^2/ha

Finally, an assessment of algorithm 'accuracy' depends on a subjective definition of successful detection. In the examples above, we define a successful detection as a location accuracy of 4 m and a crown diameter and height accuracy of less than 20%. Figure 9 shows the effects of changing this 'success' criteria. As might be expected, stricter definitions of accuracy lead to increased omission error. The shape of the relationship between omission error and point density also differ according to the

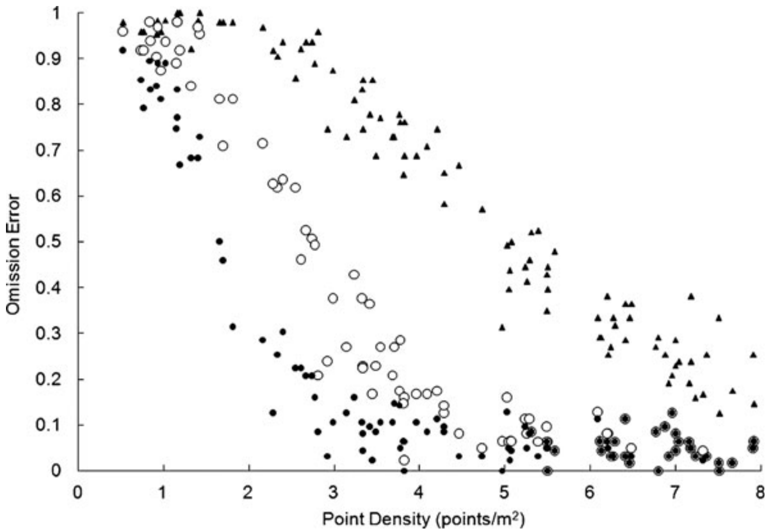


Fig. 9 Graph showing how changes to the definition of success criteria influence omission errors. Omission error (whether a tree is successfully identified) is defined by an acceptable level of accuracy between estimated vs actual tree characteristics. The graph shows acceptance criteria of 20 (*triangles*), 25 (*open circles*) and 30 (*closed circles*) percent

definitions used for successfully detected trees. Practically, this suggests that the performance (hence the comparability) of different LiDAR algorithms is sensitive to predefined success criteria.

4 Discussions

The results from this study show that our LiDAR processing algorithm derives useful, individual tree level information from synthetic data sets. They also provide quantitative and qualitative information about the factors that influence the utility of LiDAR for forest managers and that may help the development of better algorithms and the most appropriate use of LiDAR technology. In particular, omission error is strongly related to the LiDAR sampling density and to obtain operationally useful information about the location and sizes of trees in typical loblolly pine stands our results suggest that point densities greater than 2 and preferably greater than 4 per m^2 are required. Using real life data, Popescu [38] was able to successfully detect approximately 50% of trees in Loblolly pine study using a point density of 2.6 per m2. However, our results stress that the density of points needed to extract useful information also depends on the structure of the stand including tree species, tree and crown sizes and the juxtaposition of trees within a stand. These conclusions are consistent with findings from other studies. Using real data, Maltamo et al. [39] and Persson et al. [40] suggest that stand structure affects the success of tree finding algorithms. For example Morsdorf et al. [41] achieved point densities of approximately 30 per m2 but were only able to detect 64% of a conifer species with a much smaller crown diameter (1–3 m). In our study, canopy closure occurs at Basal Areas greater than 20 m2/ha. This has the effect of reducing height differences between local maxima (tree apexes) and local minima (gaps between trees) of the canopy surface, which in turn affects the success of the watershed algorithm, and the density of points required to detect trees.

Our results also illustrate consistent biases in the detection of individual tree locations and sizes. Small trees are less likely to be detected than large trees (although the location errors for detected trees are unbiased), while commission error (detection of trees that do not actually occur) is almost non-existent (except for over-segmentation). In addition, there is a consistent error towards underestimation of crown diameters. This can be explained by the inherently 'biased' nature of LiDAR methods. Backscattering (a valid return) is possible only if an object lies within the path of the laser, but is unlikely to occur if an object is not present (i.e. returns from mid-air are unlikely except for location error associated with footprint size). Consequently, returns are much more likely to register within a tree crown than at its surface or outside of the crown. Our algorithm is driven by 'tell-tale' changes in height (for example valleys between tree crowns) between adjacent data points. Consequently, the "lost" points due to the penetration error (Eq. 4) lead to an underestimation of crown sizes by the tree delineation algorithm. Low point densities exacerbate this problem because the true boundaries of crowns are less likely to be represented by the LiDAR sample.

A paradox of using LiDAR to measure forest properties is that results have little scientific or operational value unless an independent assessment of accuracy has been made. In practice, this is done by comparing LiDAR estimates with measurements using more reliable (but also more resource intensive) techniques such as field inventories and expert interpretation of aerial and satellite imagery. Therefore it could be stated that LiDAR is used to extrapolate results from a detailed and reliable (but resource intensive) sampling methodology in order to efficiently sample a broader spatial area. The errors and biases that we present here are essential to understanding this extrapolation process. We have shown that estimation errors are likely to vary according to stand composition and local point density. Practically, these findings emphasize that the interaction between LiDAR sampling and estimation bias dictates that a broad subset (i.e. encompassing the full range of forest composition) of validation data is collected in order to independently assess the accuracy of estimates. Although synthetic data may help determine the nature of these biases, we argue that the collection of appropriate validation/calibration data during an operational study is essential. However, by rigorously assessing the algorithm using synthetic data, its output and limitations are easier to understand when applied to a real world study.

The cost of intensive field inventories, diversity of forest types (e.g. species composition, ages, density) and the variety of potentially relevant forest measurements (e.g. Basal Area, individual tree locations, DBH, canopy diameter, canopy density) suggest that synthetic data is a valuable tool for investigating the success of LiDAR interpretation algorithms and consequently LiDAR studies. For example, in a real life study into individual tree detection, Maltamo et al. [39] were able to measure approximately 700 trees (in 10 different plots where a plot represents a different stand structure) in order to assess the accuracy of their LiDAR study; Brandtberg [42] used 200 individual trees and in one study [43] ground truth data could only be obtained 2 years after the LiDAR measurements. In contrast, the synthetic canopy model as the reference data is inexpensive, can be simulated for any forest type, is 100% accurate and can be used to investigate estimation errors for any type of forest measurement. We argue that a synergistic approach to LiDAR algorithm design and testing, using both real and synthetic data, will greatly benefit the advancement of operational LiDAR methods.

Although the focus of this paper is to use synthetic data to assess the accuracy of a tree finding algorithm, the concept could easily be extended to assess optimal methodology and the allocation of resources for a complete operational study. A complete study will be defined by the type and accuracy requirements of forest, the spatial extent and heterogeneity of forest types in a study area and the resources available to conduct both LiDAR and complimentary ground survey measurements. The appeal of detecting individual trees is that they are the fundamental units of forest management. In theory individual tree metrics can be used to infer other stand level forest metrics such as average stand height, DBH distributions, Basal Area, species composition (given known allometric relationships of different species) or biomass. Here, errors in individual tree detection algorithms will clearly affect estimates of these aggregate metrics. For example, many tree finding studies estimate individual tree metrics such as height, crown diameter or DBH but only assess the accuracy of estimates for detected trees only [e.g. 38, 44]. In such cases, aggregate measures will be biased because of missing information surrounding non-detected trees. In turn, our results show that non-detection is likely to be biased towards smaller trees. It is suggested that the mechanisms by which individual tree errors translate to aggregate measurements are complex but could be understood using a synthetic data and a theoretical approach.

Other algorithms and methods that have been used to detect individual trees from LiDAR data include 2-Dimensional CHM segmentation methods including edge detection, region growing, Variable Window Filters [e.g. 45] and variations of the watershed algorithm used here [e.g. 22]. A number of other methodologies have been described that use the 3D point cloud directly including: contour analysis of CHM's; k-means clustering of 3D points [41]; the detection of locally high point densities corresponding to crown centers [14] and Spatial Wavelet Analysis [24]; and some of these algorithms have been developed into practically useful, procedural software that can be used with little expert knowledge. However, the task of efficiently utilizing raw LiDAR data and assessing the accuracy of algorithms remains an outstanding research issue. In particular we argue that LiDAR methods are most useful to foresters if they yield both estimates of particular forest properties and the error associated with these estimates. In this study we have used synthetic data to understand these errors, and a number of other authors have taken similar approaches. For example, Lovell et al. [46] used simple geometric stands and simulated small footprint LiDAR data to determine the optimal acquisition parameters for measuring forest height and Goodwin et al. [47] describe a model that can be used to generate synthetic forest LiDAR data. Ni-Meister et al. [48], Sun and Ranson [20], Koetz et al. [49] used simulated 3D forests to investigate large footprint LiDAR analysis. Garcia et al. [50] used real life LiDAR data to compare the efficiency of different algorithms. In all cases, an analysis of the estimated forest properties versus known properties (obtained either using synthetic data or through intensive field sampling) reveals essential information about the performance of the methods and the errors associated with estimation.

One noticeable characteristic of many current tree detection algorithms (including the one presented here) is that the results usually comprise the metrics of successfully estimated trees. In some cases, these results may be presented along with an analysis (a comparison of LiDAR estimated vs. field sampled measurements) that can be used to assess the accuracy of these estimates. However, output does not usually include standard statistical metrics most familiar to scientists such as confidence or prediction intervals or standard errors of estimates. And as demonstrated by Fig. 9, successful detection requires a subjective definition that makes comparisons between studies difficult. We argue that for a tree detection algorithm, the most useful output would include best estimate of location, height or crown diameter with a meaningful confidence or prediction interval. In this way, aggregate measurements (derived from all the detected trees) such as stem density and DBH distributions could also be calculated with appropriate levels of confidence. Crucially, confidence or prediction intervals have a universally accepted (and practically useful) definition based on the reliability of an estimate. For example, an appropriate confidence interval for a tree location would be defined as a range that, in a given percentage of cases, would contain the true location if the entire analysis (collection and interpretation of LiDAR data) were repeated a large number of times. Although Hirata [51] provides an example of repeated measurements to investigate the effects of scanning height on LiDAR utility, conducting a large number of LiDAR replicates is likely to be prohibitively expensive for operational studies. Another option for exploring interpretation accuracy would be to resample LiDAR real-life data with good coverage and validation data using sub-sample [52], bootstrap or jackknife methods. However, we argue that synthetic data has a potentially important role for developing algorithms with more readily interpretable and useful assessments of accuracy.

For synthetic data to be useful it must bear some resemblance to real-life LiDAR data. However, we suggest that it does not have to be perfect. In this study, we have used a simple, intuitive mechanistic model of LiDAR sampling and forest properties. We also qualitatively compared the pattern of simulated points used in a real life LiDAR data set of Loblolly pines. Although we also acknowledge that many details of LiDAR sampling and of simulated forest structure have not been represented by our model, the methods could easily be modified to include them. Most notably, we have omitted mechanisms involving scan angle, discrete versus continuous data, ground terrain (important for height estimation) and any number of details concerning the complexity of stand structure (for example species mixes, juxtaposition of trees, presence of understory or mid-storey). In some cases, details of a LiDAR simulation considered unimportant for a particular system could be omitted from the methods. The decisions concerning the amount of detail necessary to adequately represent a system are an intrinsic and valuable part of all modeling efforts. Along these lines, we argue that although synthetic LiDAR data are directly useful for exploring the biases inherent in existing tree detection algorithms, the insights gained from model building also provide far reaching benefits for the improvement of existing algorithms and the design of new ones.

Given our experience of using synthetic data, there are two notable features of current algorithms that might be addressed. The first is that most current tree detection algorithms rarely use all the information provided by the LiDAR samples to estimate metrics—for example segmentation algorithms (including ours) often filter points representing the upper portion of the forest canopy before interpolating them (in 2-Dimensions) to a convenient gridded surface. In doing so, many points that might be used to estimate other useful forest metrics such as understory or ground cover vegetation are discarded [12]. Secondly, algorithms (including ours) rarely attempt to directly estimate the intuitive, mechanistic properties of LiDAR sampling (such as backscatter and penetration). This observation is intrinsically linked to conceptual models of LiDAR sampling and the process of generating synthetic data. For example, despite using a simple conceptual model to generate synthetic LiDAR data, our algorithm does not attempt to 'back estimate' all of the parameters that generated the data. Instead, it estimates only the tree locations and crown diameters. In contrast, most statistical estimation involves the development of a simple conceptual model of the underlying system (including probability functions and known parameters) followed by methods to estimate 'unknown' parameters using available data. In other words, estimation is a reverse process to generating synthetic data and vice versa. We suggest that one valuable use of synthetic data would be to 'generalize' the LiDAR problem into one of estimating the location and arrangement of geometric shapes given known 'sampling' errors. The applicability of this idea to LiDAR will likely depend on whether the conceptual models from which synthetic LiDAR data are generated can be developed to be both simple enough that relevant parameters can be back estimated directly; and realistic enough that this process yields useful interpretable practical estimates of forest structure. For example, in our LiDAR simulation, the components of the system can be divided into known and unknown parameters. The former category might include footprint size, scan angle (fixed to perpendicular in this case), flight speed, the pulse rate of the equipment and prior assumptions concerning the shape and size of trees. Using this known information and the data, the task would be to estimate all unknown parameters including the location and size of each tree but also backscatter and penetration rates. Although the location and size of trees is perhaps the most relevant output to forest managers, backscatter and penetration (as an integral part of the model that generated the data) may be crucial for increasing the accuracy of these estimates. For example, edge detection algorithms often underestimate crown diameter in closed canopies because, although they might be able to distinguish 'valleys' between crowns, they cannot determine the extent to which crowns overlap [e.g. 21, 45]. In such cases, 'local' changes in penetration (the height distribution of returns in an x,y plane) and of backscatter might improve accuracy by signifying different kinds of surface property; or denser, overlapping vegetation; or by correcting otherwise biased height measurements that can then be used in (a priori specified) models of tree allometry. Although the methods and results we present here do not address such hypotheses, it is suggested that there is much progress to be made in the development of paired algorithms that create realistic 3D, LiDAR like data sets and then attempt to back estimate the parameters of interest. Such experiments and the relevance of LiDAR simulation to real world LiDAR data and forestry applications will be the focus of our future work. In addition, some "area-based" methods other than the individual-tree based algorithms have been proposed to use LiDAR to estimate forest stand information [53–56]. We could also extend our framework towards evaluating these methods.

5 Summary and conclusion

This article presented a procedure and framework to evaluate several key parameters in LiDAR forest surveys, including forest stand characteristics (total basal area and mean and standard deviation of DBH), and the LiDAR sample point density, by using a synthetic forest stand and computer generated sample points. The results were consistent with previous research which used real world LiDAR data. The advantage of using computer-aided analysis is that it can generate a large amount of replicas of randomized sample datasets under certain controlled parameter sets. The results from the consequential data processing can be directly compared to the preset ground truth. As such, the accuracy of the forest stand metrics along with their uncertainty and confidence levels can be obtained, which is an unprecedented feature of our approach to understanding the usability of the LiDAR survey in forest study.

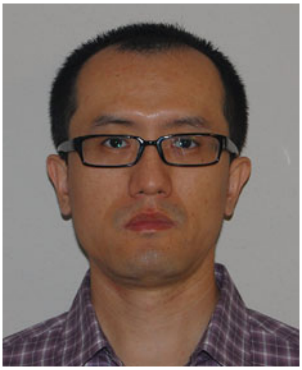
We also acknowledge there are still some limitations in our method. Because this is a proof-of-concept study, some simplifications have to be assumed so that the feasibility of the computer models can be evaluated. The assumptions include the regular distribution of trees in the stand, uniform foliage density, flat ground terrain, constant laser pulse permeability in the canopy, symmetric shape of crown, and perpendicular looking angle of the sensor. Upon success of this research, we shall put more insights into variability of these factors by relaxing some of the assumptions at a time so that more useful guidelines for airborne LiDAR survey for foresters are generalized from the synthetic scenarios.

References

1. Coulson RN, Stephen FM (2006) Impacts of insects in forest landscapes: Implications for forest health management. In: Paine TD (ed) Invasive forest insects, introduced forest trees, and altered ecosystems. Ecological pest management in global forest of a changing world. Springer, New York, pp 101–125
2. Zhao K, Popescu SC, Nelson RF (2009) Lidar remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers. Remote Sens Environ 113(1):182–196
3. Magnussen S, Boudewyn P (1998) Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. Can J For Res 28(7):1016–1031
4. Lefsky MA, Harding D, Cohen WB, Parker G, Shugart HH (1999) Surface lidar remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA. Remote Sens Environ 67(1):83–98
5. Means J, Acker S, Fitt B, Renslow M, Emerson L, Hendrix C (2000) Predicting forest stand characteristics with airborne scanning lidar. Photogramm Eng Remote Sens 66(11):1367–1371
6. Naesset E (1997) Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens Environ 61(2):246–253
7. Anderson MC, Neale CMU, Li F, Norman JM, Kustas WP, Jayanthi H, Chavez J (2004) Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery. Remote Sens Environ 92(4):447–464
8. Jensen JLR, Humes KS, Vierling LA, Hudak AT (2008) Discrete return lidar-based prediction of leaf area index in two conifer forests. Remote Sens Environ 112(10):3947–3957
9. Zhang K (2008) Identification of gaps in mangrove forests with airborne LIDAR. Remote Sens Environ 112(5):2309–2325
10. Hyyppä J, Hyyppä H, Litkey P, Yu X, Haggrén H, Rönnholm P, Pyysalo U, Pitkanen J, Maltamo M (2004) Algorithms and methods of airborne laser-scanning for forest measurements. In: Thies M, Koch B, Spiecker H, Weinacker H (eds), Laser-Scanners for Forest and Landscape Assessment: Proceedings of the ISPRS Working Group VIII/2, Freiburg, Germany, 3–6 October
11. Lim K, Treitz P, Wulder M, St-Onge B, Flood M (2003) LiDAR remote sensing of forest structure. Prog Phys Geogr 27(1):88–106
12. Popescu SC, Wynne RH, Nelson RF (2002) Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. Comput Electron Agric 37(1):71–95
13. Anderson HE (2009) Using airborne Light Detection and Ranging (LIDAR) to characterize forest stand condition on the Kenai Penisula of Alaska. West J Appl For 24:95–102
14. Rahman MZA, Gorte B Individual Tree Detection Based on Densities of High Points of High Resolution Airborne LiDAR. In: Hay GJ, Blaschke T, Marceau D (eds) GEOBIA 2008 – Pixels, Objects, Intelligence. GEOgraphic Object Based Image Analysis for the 21st Century, Calgary Alberta, Canada, 2008. International Society for Photogrammetry and Remote Sensing (ISPRS), pp 350–355
15. Yang L, Wu X, Ma X (2009) Tree detection from aerial imagery. In: ACM GIS '09, Seattle, WA, USA, November 4–6
16. Kasischke ES, Christensen NL Jr, Haney E (1994) Modelling of geometric properties of loblolly pine tree and stand characteristics for use in radar backscatter models. IEEE Trans Geosci Remote Sens 32:800–822
17. Wehr A, Lohr U (1999) Airborne laser scanning—an introduction and overview. J Photogramm Remote Sens 54:68–82
18. Habib AF, Bang KI, Kersting A, Lee DC (2009) Error budget of LiDAR systems and quality control of the derived data. Photogramm Eng Remote Sens 75(9):1093–1108
19. Baltsavias EP (1999) Airborne laser scanning: basic relations and formulas. ISPRS J Photogramm Remote Sens 54(3):199–214
20. Sun G, Ranson KJ (2000) Modeling lidar returns from forest canopies. IEEE Trans Geosci Remote Sens 38(6):2617–2626
21. Chen DBQ, Gong P, Kelly M (2006) Isolating individual trees in a savanna woodland using small footprint lidar data. Photogramm Eng Remote Sens 72:923–932
22. Kwak DA, Lee WK, Lee JH, Biging GS, Gong P (2007) Detection of individual trees and estimation of tree height using LiDAR data. J For Res 12(6):425–434
23. Schardt M, Ziegler M, Wimmer A, Wack R, Hyyppa J (2002) Assessment of forest parameters by means of Laser Scanning. Int Arch Photogramm Remote Sens Environ XXXIV(part 3A):302–309

24. Falkowski MJ, Smith AMS, Hudak AT, Gessler PE, Vierling LA, Crookston NL (2006) Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. Can J Remote Sens 32(2):153–161
25. Clark ML, Clark DB, Roberts DA (2004) Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sens Environ 91(1):68–89
26. Eckstein BA (1989) Evaluation of spline and weighted average interpolation algorithms. Comput Geosci 15(1):245–257
27. Hutchinson MF, Gessler PE (1994) Splines - more than just a smooth interpolator. Geoderma 62:45–67
28. Kitanidis PK (1997) Introduction to geostatistics. Cambrsidge University Press
29. Oliver MA (1990) Kriging: a method for interpolation for Geographical Information Systems. Int J Geogr Inf Syst 4:313–332
30. Phillips DL, Dolph J, Marks D (1992) A comparison of geostatistical procedures for spatial analysis of precipitations in mountainous terrain. Agr Forest Meterol 58:119–141
31. Voltz M, Goulard M (1994) Spatial interpolation of soil moisture retention curves. Geoderma 62:109–123
32. Tabios GQ, Salas JD (1985) A comparative analysis of techniques for spatial interpolation of precipitation. Wat Resour Bull 21(3):365–380
33. Liu H, Wang L, S-j T, Jezek K (2011) Robust multi-scale image matching for deriving ice surface velocity field from sequential satellite images. Int J Remote Sens 24(7):1015–1041
34. Yang C-S, Kao S-P, Lee F-B, Hung P-S (2004) Twelve different interpolation methods: A case study of Surfer 8.0. In: Proceedings of the XXth ISPRS Congress. pp 778–785
35. Bieniek A, Moga A (2000) An efficient watershed algorithm based on connected components. Pattern Recogn 33:907–916
36. Soille P (1999) Morphology image analysis. Springer, Berlin
37. Meyer F, Beucher S (1990) Morphological segmentation. J Vis Commun Image Represent 1:21–46
38. Popescu SC (2007) Estimating biomass of individual pine trees using airborne lidar. Biomass Bioenergy 31(9):646–655
39. Maltamo M, Eerikäinen K, Pitkänen J, Hyyppä J, Vehmas M (2004) Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sens Environ 90(3):319–330
40. Persson A, Holmgren J, Soderman U (2002) Detecting and measuring individual trees using an airborne laser scanner. Photogramm Eng Remote Sens 68(9):925–932
41. Morsdorf F, Meier E, Kotz B, Itten KI, Dobbertin M, Allgwaer B (2004) LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. Remote Sens Environ 92(3):353–362
42. Brandtberg T (2007) Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. ISPRS J Photogramm Remote Sens 61(5):325–340
43. Takahashi T, Yamamoto K, Miyachi Y, Senda Y, Tsuzuku M (2006) The penetration rate of laser pulses transmitted from a small-footprint airborne lidar: a case study in closed-canopy, middle-aged pure sugi (Cryptomeria japonica D. Don) and hinoki cypress (Chamaecyparis obtusa Sieb. et Zucc.) stands in Japan. J For Res 11:117–123
44. Suarez JC, Ontiveros C, Smith S, Snape S (2005) Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. Comput Geosci 31(2):253–262
45. Popescu SC, Wynne RH, Nelson RF (2003) Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can J Remote Sens 29:564–577
46. Lovell JL, Jupp DLB, Newnham GJ, Coops NC, Culvenor DS (2005) Simulation study for finding optimal LiDAR acquisition parameters for forest height retrieval. For Ecol Manag 214(1–3):398–412
47. Goodwin NR, Coops NC, Culvenor DS (2007) Development of a simulation model to predict LiDAR interception in forested environments. Remote Sens Environ 111(4):481–492
48. Ni-Meister W, Jupp DLB, Dubayah R (2001) Modeling LiDAR waveforms in heterogeneous and discrete canopies. IEEE Trans Geosci Remote Sens 39(9):1943–1958
49. Koetz B, Morsdorf F, Sun G, Ranson KJ, Itten K, Allgower B (2006) Inversion of a LiDAR waveform model for forest biophysical parameter estimation. Geosci Remote Sens Lett IEEE 3(1):49–53
50. García R, Suárez JC, Patenaude G (2007) Delineation of individual tree crowns for LiDAR tree and stand parameter estimation in Scottish woodlands. In: Fabrikant SI, Wachowicz M (eds) The European information society: Leading the way with geo-information. Lecture notes in geoinformation and cartography. Springer, Berlin, pp 55–85
51. Hirata Y (2004) The effects of footprint size and sampling density in airborne laser scanning to extract individual trees in mountainous terrain. In: SPRS working group 8/2 Laser-Scanners for Forest and Landscape Assessment, Freiburg, Germany. pp 102–107

52. Goodwin NR, Coops NC, Culvenor DS (2006) Assessment of forest structure with airborne LiDAR and the effects of platform altitude. Remote Sens Environ 103:140–152
53. Lefsky MA, Harding DJ, Keller M, Cohen WB, Carabajal CC, Espirito-Santo FDB, Hunter MO, de Oliveira R Jr (2005) Estimates of forest canopy height and aboveground biomass using ICESat. Geophys Res Lett 32:L22S02
54. Ioki K, Imanishi J, Sasaki T, Morimoto Y, Kitada K (2010) Estimating stand volume in broad-leaved forest using discrete-return LiDAR: plot-based approach. Landsc Ecol Eng 6(1):29–36. doi:10.1007/s11355-009-0077-4
55. Beets PN, Reutebuch S, Kimberley MO, Oliver GR, Pearce SH, McGaughey RJ (2011) Leaf area index, biomass carbon and growth rate of radiata pine genetic types and relationships with LiDAR. Forests 2:637–659
56. Ni-Meister W, Lee S, Strahler AH, Woodcock CE, Schaaf C, Yao T, Ranson KJ, Sun G, Blair JB (2010) Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. J Geophys Res 115:G00E11

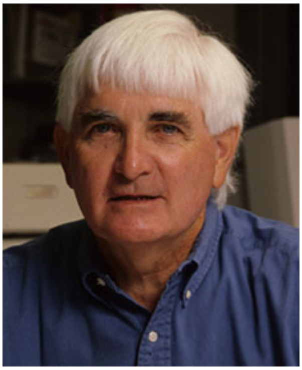


Lei Wang Received his Ph.D. in Geography from the Texas A&M University in 2006. He worked as a postdoc research associate at the Knowledge Engineering Lab of Texas A&M University in 2007. He is currently an assistant professor at the Department of Geography & Anthropology, Louisiana State University. His research focuses on Lidar remote sensing, climate change, and flood hazard.

Andrew G. Birt is a Postdoctoral Research Associate within the Department of Entomology at Texas A&M University. His research interests center around population modeling and its practical application to landscape ecology, life history theory and the interface between humans and ecological systems. He is currently researching the population dynamics of SPB and its impacts on the forested landscape of the USA. He holds a PhD from the University of Birmingham, England, and has experience as a risk assessor and researcher in the pesticide industry.

Charles W. Lafon is an Associate Professor in the Department of Geography at Texas A&M University. He is interested in multiple the influence of multiple disturbances on forest landscapes. He is an expert on the southern Appalachian Mountains and has recently been involved in fire history reconstructions for those environments. His Ph.D. is from the University of Tennessee.

David M. Cairns is a Professor of Geography at Texas A&M University. His research interests are in vegetation response to climate change, ecological modeling and the influence of herbivores on forest landscapes. He has a Ph.D. from the University of Iowa.



Robert N. Coulson is a professor in the Department of Entomology at Texas A&M University. He directs the Knowledge Engineering Laboratory (KEL) http://kelab.tamu.edu. The focus of KEL research and development centers on integration, interpretation, and use of both qualitative and quantitative information for the purpose of landscape planning, problem-solving and decision making.

Maria D. Tchakerian is an associate research scientist in the Knowledge Engineering Laboratory, Entomology Department, Texas A&M University. She earned her B.S. in biochemistry, M.S. in conservation and ecology (Monterrey Institute of Technology and Higher Education), and Ph.D. in Geography (Texas A&M University). Her research interests are landscape ecology, spatial analysis, and geographic information system technology for research and problem solving.

Weimin Xi Received his PhD in Biology from the University of North Carolina at Chapel Hill in 2005. He is currently a senior research associate at the Department of Forest and Wildlife Ecology, University of Wisconsin-Madison. Dr. Xi's research focus on the interface of forest ecology, disturbance ecology, landscape ecology, restoration ecology, ecological modeling, eco-informatics, and climate change biology.

Sorin C. Popescu received his Ph.D. in 2002 from Virginia Tech University in Blacksburg, Virginia, USA. He worked at Virginia Tech for a year as a Postdoctoral Research Associate and in 2003 joined the Department of

Ecosystem Science and Management at Texas A&M University as an Assistant Professor. He was promoted to Associate Professor with tenure in 2009. His research interests include remote sensing applications in earth sciences, with focus on lidar sensors and algorithm developments for assessing forest ecosystems structure, species habitat, fire risk, biomass, and carbon.

James M. Guldin is Project Leader for Ecology and Management of Southern Pines, in the Southern Research Station of the USDA Forest Service. His current research focuses on ecology and silviculture of naturally regenerated forest stands, with an emphasis on southern pines. His other research interests are theory and practice of uneven-aged silviculture, ecological restoration of forested ecosystems, and application of silviculture under changing ecological, economic, and social forcing factors.