A voxel-based lidar method for estimating crown base height for deciduous and pine trees

Sorin C. Popescu⁎, Kaiguang Zhao ¹
Spatial Sciences Laboratory, Department of Ecosystem Science and Management, Texas A&M University, 1500 Research Parkway, Suite B 223, College Station, TX 77845, United States
Received 27 July 2006; received in revised form 8 June 2007; accepted 8 June 2007

Abstract

The overall goal of this study was to develop methods for assessing crown base height for individual trees using airborne lidar data in forest settings typical for the southeastern United States. More specific objectives are to: (1) develop new lidar-derived features as multiband height bins and processing techniques for characterizing the vertical structure of individual tree crowns; (2) investigate several techniques for filtering and analyzing vertical profiles of individual trees to derive crown base height, such as Fourier and wavelet filtering, polynomial fit, and percentile analysis; (3) assess the accuracy of estimating crown base height for individual trees, and (4) investigate which type of lidar data, point frequency or intensity, provides the most accurate estimate of crown base height. A lidar software application, TreeVaW, was used to locate individual trees and to obtain per tree measurements of height and crown width. Tree locations were used with lidar height bins to derive the vertical structure of tree crowns and measurements of crown base height. Lidar-derived crown base heights of individual trees were compared to field observations for 117 trees, including 94 pines and 23 deciduous trees. Linear regression models were able to explain up to 80% of the variability associated with crown base height for individual trees. Fourier filtering used for smoothing the vertical crown profile consistently provided the best results when estimating crown base height.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Lidar; Tree height; Crown diameter; Crown base height; Height bins

1. Introduction

Reliable forest canopy structure metrics and individual tree crown characteristics, such as crown base height, tree height, crown dimensions, and crown bulk density, are required by forest resources and fire managers to support their management plans. Prediction of crown base height is important in several forestry problems. Crown ratio, which is the ratio of crown length to total tree height, is related to tree vigor and thus to the timing and potential response to thinning (Smith, 1986). Crown dimensions are strongly correlated with stem diameters and, therefore, to forest volume and biomass (Avery & Burkhart, 1994, p. 265). Crown metrics can be used to derive crown volume and crown surface area, which are measures of the photosynthetic potential (Sprinz & Burkhart, 1987). Crown dimensions are also useful for fire behavior analysis (Andersen et al., 2005), being used to calculate crown bulk density or total canopy fuel weight which are data layers for input into fire behavior models such as FARSITE (Finney, 1998). However, in the past, tree crowns have usually not been measured in the field due to difficulties in defining and measuring their changing dimensions and shapes (Sprinz & Burkhart, 1987). Tree crowns have also been measured on aerial photographs and aerial volume tables were derived to substitute the usual ground measurements of stem diameter (Avery & Burkhart, 1994, p. 267). Lately, advances in remote sensing have produced other tools that afford the estimation of crown dimensions, such high resolution aerial and satellite imagery and lidar.

The use of remote sensing for mapping the spatial distribution of canopy characteristics has the potential to allow an accurate and efficient estimation of tree dimensions and canopy properties from local to regional scales. In particular, lidar remote sensing has the capability to acquire direct three-dimensional measurements of the
forest canopy that are useful for estimating a variety of forest inventory parameters, including tree height, volume, and biomass, and also for deriving useful information for input into fire simulation models, such as FARSITE (Finney, 1998), e.g., (Mutlu et al., in press).

Previous lidar studies, whether using waveform or discrete return lidar data, attempted to derive measurements, such as tree height and crown dimensions, at stand level (Hall et al., 2005; Næsset & Bjerknes, 2001), plot level (Holmgren et al., 2003; Hyyppä et al., 2001; Lim & Treitz, 2004; Popescu et al., 2004), or individual tree level (Chen et al., 2006; Coops et al., 2004; Holmgren and Persson, 2004; Persson et al., 2002; Popescu, in press; Roberts et al., 2005; Yu et al., 2004), and then use allometric relationships or statistical analysis to estimate other characteristics, such as biomass, volume, crown bulk density, and canopy fuel parameters. Forest canopy structure was estimated using data from scanning lasers that provided lidar data with full waveform digitization (Harding et al., 1994, 2001; Lefsky et al., 1997; Means et al., 1999). Small-footprint, discrete-returns systems were used to estimate canopy characteristics, with many studies focusing on tree height (Magnussen & Boudewyn, 1998; Magnussen et al., 1999; Maltamo et al., 2004; McCombs et al., 2003; Næsset, 1997; Næsset & Økland, 2002; Popescu & Wynne, 2004; Popescu et al., 2002).

Few lidar studies focused on assessing canopy structure and characteristics, such as fuel weight, canopy and crown base height, and crown bulk density (Andersen et al., 2005; Holmgren & Persson, 2004; Pyysalo & Hyyppä, 2002; Riano et al., 2004). Among these studies, there seems to be a unanimous acceptance that lidar overestimates crown base height for individual trees or plot-level canopy base height, which is an intuitive finding given the fact that airborne lidar portrays crowns from above and lower branches have a reduced probability of being intercepted by laser pulses which might be blocked by higher branches. Holmgren and Persson (2004) calculated crown base height as the distance from ground to the lowest laser data height interval containing more than 1% of the total number of non-ground laser points within a crown area. Andersen et al. (2005) estimated plot-level canopy base height using linear regression with several percentile-based metrics of the lidar heights distribution and a canopy density metric calculated as the percentage of first returns within the canopy. Lovell et al., 2003, used both airborne and ground-based lidar to measure canopy structure in Australian forests.

Most of the studies dealing with crown base height estimation analyze the vertical profile of laser hits within a single crown or at plot level. Holmgren and Persson (2004) use a binary series of 0.5 m height layers to estimate crown base height. Each layer with less than 1% of the total number of vegetation hits was set to zero and the others to one. Reutebuch et al. (2005) proposed various types of lidar-derived products useful for multiple resource inventories. One of the products, namely canopy cover maps, consists of images that provide a direct measurement of cover by height aboveground. Mutlu et al. (in press) and Griffin et al. (in review) used this concept by creating height bins of laser hits frequency at various height intervals above ground for mapping surface fuels, leaf area index (LAI), and percent canopy cover. The number of laser hits above an area, in various height bins, indicates the canopy cover in each of the height bins.

With discrete lidar points collected over the forest canopy, laser hits at various height intervals or within height bins above ground elevation contain information about the reflective surfaces that exist in the vertical canopy space, i.e., in a “slice” of canopy space above ground. The lidar height bins are essentially conventional, multiband, two-dimensional representations of lidar-derived voxels, or volumetric pixels, as they contain the frequency of laser returns within a three-dimensional space. Other studies have used a binary approach to characterizing the vertical space of forest canopies with voxels, by considering occupied or unoccupied voxels, i.e., voxels with or without laser hits within their volumetric space (e.g., Chasmer et al., 2004; Parker, 1995; Weishampel et al., 1997). A concept similar to the height bins was used by Næsset (2004) to derive independent variables for regression models of total tree height. In his study of 2004, Naesset divided the range between the lowest laser canopy height and the maximum canopy height into 10 fractions of equal length and computed canopy densities for each fraction.

Although morphological computer vision algorithms have been used to automatically identify tree crown structures visible on lidar-derived three-dimensional canopy height models (CHM) and to measure tree height and crown diameter, results are usually reported at plot level, as in Popescu and Wynne, 2004, or stand level, e.g., Hall et al., 2005. The main reason is the difficulty of validating results for individual trees, when an objective correspondence needs to be established between field- and lidar-measured individual trees, e.g., Yu et al., 2004. This difficulty arises due to uncertainties with individual tree mapping on the ground, e.g., GPS locations, closed canopy conditions, vertical tree position in the canopy, etc. The current study attempts to estimate individual tree crown parameters and reports results at individual tree levels.

1.1. Objective

The overall goal of this study was to develop new representations of lidar data for assessing the crown base height for individual trees using airborne lidar data in forest settings typical for the southeastern United States. More specific objectives were to:

1. (1) develop new lidar-derived features such as multiband height bins and processing techniques for characterizing the vertical structure of individual tree crowns;
2. (2) investigate several techniques for filtering and analyzing vertical profiles of individual trees for deriving crown base height, such as Fourier filtering and wavelet shrinkage, polynomial fit, and percentile analysis;
3. (3) assess the accuracy of estimating crown base height for individual trees; and
4. (4) investigate which type of lidar data, point frequency or intensity, provides the most accurate estimate of crown base height.
2. Materials and methods

2.1. Study site

The study area is located in the southern United States (30° 42’ N, 95° 23’ W), in the eastern half of Texas (Fig. 1), and has approximately 4800 ha. The study area covered with scanning lidar includes pine plantations in various developmental stages, old growth pine stands in the Sam Houston National Forest, many of them with a natural pine stand structure, and upland and bottomland hardwoods. Much of the southern U.S. is covered by forest types similar to the ones included in our intensive study areas, with similar forest types, productivity, and patterns of land use change. A mean elevation of 85 m, with a minimum of 62 m and a maximum of 105 m, and gentle slopes characterize the topography of the study area.

2.2. Lidar data set and multispectral imagery

The lidar data were acquired from an average of 1000 m above ground level (AGL) during the leaf-off season in March 2004 by M7 Visual Intelligence, of Houston, Texas. The lidar system (Leica-Geosystems ALS40) utilizes advanced technology in airborne positioning and orientation, enabling the collection of high-accuracy digital surface data, while recording two returns per laser pulse, first and last. The reported horizontal and vertical accuracies with the Leica-Geosystems ALS40 system for the mission specifications of this project are 20–30 cm and 15 cm, respectively.

For our data set, the lidar system provided a 10° swath from nadir, for a total scan angle of 20°. With a cross-hatch grid of flight lines, the average laser point density was 2.6 laser points per m². The point density translates into an average distance between laser points of about 0.62 m, for the entire point cloud. The average swath width was 350 m, with 19 and 28 flight lines in a north–south direction and east–west direction, respectively. A digital surface model (DSM) was created by interpolating lidar point elevations to a regular grid with a spatial resolution of 0.5 m using the triangulated irregular network (TIN) method. Lidar points used for creating the DSM included only the highest points in 0.5 by 0.5 m cells, to allow for an accurate characterization of the top canopy surface, as explained in Popescu et al., 2004. The CHM, with a subset shown in Fig. 2, has a spatial resolution of 0.5 m and it represents a three-dimensional surface that characterizes vegetation height across the landscape. The CHM was created by subtracting the ground elevation from the DSM. A multispectral Quickbird image (DigitalGlobe) of the study area is shown in Fig. 1.

2.3. Ground inventory data

The in situ data were collected during May–July 2004 by measuring trees and canopy characteristics on circular plots, including 36 large plots of 404.7 m² or 0.1 acre) and 26 small...
plots of 40.5 m² (0.01 acre) plots. The smaller plots were used with young unthinned pine plantations, with little variability of tree size. Individual tree measurements were taken for tree height, crown width, height to crown base, and diameter at breast height (dbh). A total of 1004 trees were measured on these plots to investigate multiple research objectives, including the objectives presented in this study.

On each plot, the heights of all trees were measured using a Vertex Forestor hypsometer. The same instrument was used to measure the distance from plot center to every tree. The height measurement recorded the total length of the tree, to the nearest 0.30 m from ground level to the top of the tree. Azimuth was recorded by using a Suunto compass (KB-14) with sightings from plot center to every tree. Diameter at breast height (dbh) was measured on all trees using a diameter tape. The actual diameter was recorded for each tallied tree to the nearest 0.25 cm (0.1 in.).

Crown width was determined as the average of two perpendicular crown diameters measured with a tape along cardinal directions, north–south and east–west. Field measurements considered crowns to their full extent, therefore measured overlapping crown diameters.

In the field, crown base height was considered the distance between the ground and the lowest live branches in the crown of a tree. Small and isolated branches with leaves, separated from the main crown, were not considered as indicating crown base height. The Vertex Forestor hypsometer was used to record crown base height.

Tree locations were mapped by recording GPS coordinates for the plot center, with azimuth and distance readings for every tree. In addition to the measurements mentioned above, the crown class (Kraft) was recorded for each tree, such as dominant, co-dominant, intermediate, and overtopped (USDA Forest Service FIA National Core Field Guide, 2005, p. 78). We excluded overtopped trees from our analysis, as they have very little chances, if at all, of being measured with lidar from above.

Out of 743 measured trees in the dominant, co-dominant, and intermediate crown classes, 504 were Loblolly pines (Pinus taeda L.) and 239 were deciduous trees, such as water oak (Quercus nigra L.), red oak (Quercus falcata Michx.), sweetgum (Liquidambar styraciflua L.), and post oak (Quercus stellata Wangenh.). Errors with GPS locations of plot centers, tree mapping, and location differences between tree tops as observed with lidar and stem bases mapped on the ground make lidar-field tree matching difficult. Out of all trees measured on the ground, only 117 field trees were paired with lidar-observed trees, or approximately 12% of the measured trees. Trees were paired by visual analysis of TreeVaW- and field-located trees as shown in Fig. 3b. Matched trees had to be located within the same crown as seen on the canopy height model. When multiple field trees were present, we matched the closest location to the TreeVaW-identified tree. Table 1 shows measurement statistics of the lidar-field matched trees. Out of these matched trees, 94 were pines and 23 were deciduous. Lidar, however, is able to detect a higher percentage of trees, but out of the detected trees, only 12% were paired with high confidence by visually analyzing the location of the lidar-mapped trees, the CHM, and the location of the ground-mapped trees, as seen in Fig. 3.

The crown class (Kraft) distribution for the 117 matched trees was as follows: 18% dominant, 72% co-dominant, and 11% intermediate.

Taking into account the positional accuracy of the differential GPS unit for determining the location of the subplot centers, the error of a tree’s position is expected to be approximately up to 1.5–2.0 m. This error only refers to the position of the base of the tree, without considering the deviation of the tree top relative to the base. The CHM is expected to have a horizontal accuracy of about 0.5 m, given the horizontal accuracy of the laser system of about 20–30 cm. Therefore, for trees where lidar hits intercept the tree top and not the crown shoulders, the location of the tree tops on the CHM is expected to be identified with higher accuracy than the location of tree tops derived from ground measurements and GPS.

2.4. Lidar-derived tree dimensions: tree height and crown diameter

Individual trees were located and measured on the lidar-derived CHM by automated processing using the TreeVaW software, with algorithms described in Popescu and Wynne (2004) and Popescu et al. (2004). To summarize the approach implemented in TreeVaW, tree height estimates were based on single tree identification using an adaptive technique for local maximum focal filtering.

The CHM is a lidar-derived three-dimensional surface that contains information on vegetation height above the ground surface. The lidar point cloud in the lidar data exchange binary format (LAS) was processed to derive the digital surface model (DSM) of the canopy. For this purpose, only the highest laser hits within 0.5 × 0.5 m grid cells were used with the TIN interpolation method to derive the DSM. The CHM was obtained by subtracting the terrain elevation from the digital surface model of the canopy, with a spatial resolution of 0.5 m.
The algorithm for measuring crown diameter uses the tree locations identified with local maximum filtering and analyses two perpendicular profiles of the CHM centered on the tree top to find out the crown extent. Total tree height and crown width are important parameters for extracting individual crown vertical profiles, as explained in the next section.

2.5. Height bins and crown vertical profiles

Two sets of lidar height bins were generated as multiband images of 1 m height intervals and 0.5 × 0.5 m pixel dimensions, i.e., 0.5 × 0.5 × 1 m voxels. We generated 31-band images, with the first 30 bands corresponding to height bins from 0 to 30 m.
The last height bin includes all laser pulses above 30 m. In the first image every band represents a height bin layer, and a pixel value represents the number of laser points as a percentage of the total number of points summed up for all pixels in the stack at this position — concept illustrated in Fig. 4. The second image is generated by using the sum of intensity values of all laser points in each voxel. Canopy lidar height bins are generated automatically using IDL (Interactive Data Language,ITT Visual Systems) programming developed by the authors.

Height bins are an innovative lidar-derived product that has the potential to become a standardized imagery product for lidar applications in ecosystem studies, for generating images of the vertical structure of forest vegetation. Fig. 5 shows the lidar height bins generated by using authors’ software developments. The height bins in Fig. 5 were generated for variable height intervals, as indicated under each image.

When combining individual trees located on the CHM with the height bins, it is possible to construct the vertical profile of laser hits for each tree and derive a pseudo-waveform for each tree identifiable on the CHM, similarly to the modeling of large-footprint lidar waveforms with discrete-return lidar done by Blair et al. (1999). The modeled pseudo-waveform in this case is available for each tree, rather than for each large footprint laser beam, and is derived by knowing the tree top location and crown diameter. A cylinder with the diameter equal to the crown width and centered on the tree top is used like a cookie-cutter to isolate vertical profiles of individual trees from the height bin images, with the concept illustrated in Fig. 6 c). This method works well for trees with symmetrical crowns, but may be less accurate for crown vertical profiles of trees with asymmetric crowns.

The height-bin image derived from intensity values more closely resembles the waveform of large footprint lasers. The return waveform of large footprint lasers provides a record of the height distribution of the returned intensity from the surfaces illuminated by the laser pulse (Harding et al., 1994, 2001; Lefsky et al., 2002). As such, the waveform can be conceptualized as the height- or time-varying intensity which is the sum of intensity reflections from individual surfaces within laser footprints, distributed over the total height of the intercepted surfaces. With a high density of small-footprint laser hits, in our case with an average point density between 3 and 4 points per m², the vertical distribution of the sum of intensity values from individual laser hits is closely related to the waveform recorded by large footprint lasers. However, the pseudo-waveforms generated in our study for each tree do not attempt to completely simulate a true lidar waveform with a diameter equal to that of tree crowns. As Blair et al. (1999) mentioned in their study, the true waveform represents the spatial distribution of the laser beam intensity across and along the beam path, given the usual definition of the laser beam width as “full width at half maximum”.

2.6. Estimating crown base height

As opposed to other studies attempting to characterize the canopy vertical structure (e.g., Parker, 1995; Weishampel et al., 1997), our approach uses both frequency and intensity values of laser hits to create, respectively, frequency and intensity height bins and to derive the height to crown base. The basic concept behind measuring height to crown base is to detect the height where the intensity or the frequency of laser hits drops abruptly on the vertical profile. The shape of the vertical profile depends on the vertical position of laser hits on the exposed crown length. According to Magnussen and Boudewyn (1998), laser crown hits mimic sampling with probability proportional to the projected tree crown sizes. In their study, they concluded that about 2% of the laser return signals can be expected to have been reflected off a tree top, while a majority of laser hits would be returned from the side of the crowns of dominant trees. Half the laser pulses hitting the canopy in a plot are presumably returned at or above a height that coincides with the height of mean leaf area index (LAI) (Magnussen & Boudewyn, 1998). In their study of 1999, Lefsky et al. proposed a volumetric method of waveform analysis by using a matrix of 10-m adjacent waveforms with 1 m vertical bins, which are essentially cylindrical voxels. They identified an “euphotic” zone of each waveform, corresponding to the uppermost canopy profile returning 65% of the total laser energy reflected back from the canopy. This zone intercepts the bulk of available light and laser hits and intuitively corresponds to the peak of the crown vertical profile (CVP) generated from either laser point frequency or intensity values, while the drop following this maximum should indicate the base of the crown.

For this study, the height where a drop in the vertical profile occurs can be automatically detected by fitting a fourth-degree polynomial curve to the vertical profile, in a first step, and by analyzing the shape of the fitted function to estimate where the drop occurs (Fig. 7), in a second step. With height values along the x axis, the drop is considered the first inflection point after the curve reaches its maximum. A fourth-degree polynomial allows the corresponding function to have a concave shape along the crown profile of a single tree, with three extreme values.

An extreme value corresponds in most of the cases to either a local maximum or minimum of the fitted function (Gillett, 1984, p. 188). Local minima occur at the top and bottom of the crown, where few laser hits intercept the crown, while the local maximum indicates the crown height with the highest frequency of laser hits. The first derivative of the fitted polynomial is equal to zero at extreme values, while the sign of the second derivative, negative or positive, indicates respectively whether the graph of the fitted function is concave or convex and whether a critical
An inflection point occurs when the first derivative of a curve, i.e., the slope, takes a local extreme, while the second derivative is zero. A similar method of fitting a fourth degree polynomial was used by Popescu et al., 2003, to derive crown width on perpendicular height profiles extracted on canopy height models.

When fitting a polynomial to a tree’s vertical lidar profile, the polynomial was forced to go through the tree top along the x-axis. This constraint leaves only four unknown coefficients to be determined for fitting a fourth-degree polynomial but still allows enough flexibility to capture the shape of lidar-derived crown vertical profiles (CVPs). A gradient-based non-linear...
least square method was used to iteratively estimate the four unknowns (Press et al., 2002, p. 662).

Despite a relatively high average laser point density of 3 to 4 points per m², the vertical profiles of individual trees derived from either the frequency or intensity height bins may not approximate a smooth curve due to several reasons. One group of factors affecting the vertical profile smoothness is vegetation-dependent — individual trees have a very diverse crown architecture that depends on species, age, position within the canopy, leaf-on/leaf-off conditions, position of small gaps within tree crowns, etc. The other group of factors depends on laser scanning characteristics, scanning patterns, point density, footprint size, number of returns per pulse, and minimum separable distance between returns of the same pulse. Also, voxel dimensions, mainly the vertical resolution of height bins, affect the smoothness of the vertical profile. In order to have more suitable vertical distributions for the fourth order polynomial fitting, we used two methods for smoothing the profiles prior to the fitting. We investigated two techniques for smoothing the vertical profile: Fourier filtering and wavelet shrinkage. In addition to fitting a polynomial function to the smoothed profile, we fitted the polynomial function to the original, unfiltered, vertical profile, to compare the effects of filtering.

Fourier and wavelet analysis are powerful tools commonly used to analyze digital 1-D signals. The Fourier and wavelet transforms can decompose a signal into a series of coefficients in the domains of frequency and scale-translation, respectively. By appropriately enhancing or reducing certain coefficients in the new representations, the signal reconstructed through inversion transforms can reveal certain desirable features, or suppress undesirable features, e.g., noises. For the Fourier transform, signals are expanded in terms of cosine or sine basis functions while the wavelet transform is based on scaled and shifted versions of mother wavelet. Their digital versions are typically implemented by Fast Fourier Transform (FFT) and

---

**Fig. 6.** a) Laser hits within over a tree crown, with frequency and intensity crown vertical profiles; b) CHM and the location of the tree crown shown in a); c) automatically identifying laser hits over the tree crown.
filter banks respectively. The roughness or the local abrupt variations present in CVPs may mislead the polynomial fitting that is originally intended to capture the smooth trends of profiles, therefore, FFT filters and the wavelet shrinkage can be used to de-noise vertical profiles. When smoothing vertical crown profiles by Fourier filtering, the upper half high frequency components are directly removed. The choice of the mid-frequency threshold was determined by observing that the magnitude of the upper half FFT spectra becomes insignificant and almost levels off for most of the crown vertical profiles derived from our data.

In the wavelet shrinkage, one-level wavelet decomposition was performed on raw vertical profiles using the 2nd order Daubechies wavelet function (Walker, 1999, p. 29). Back-transformation of this wavelet gives a smoothed profile, as seen in Fig. 7 b). For a reduced signal length of the canopy vertical profile, i.e., 31 height bins, a 2nd order Daubechies wavelet function was considered appropriate. FFT and Daubechies-based wavelet transformations and filtering were implemented in IDL (ITT Visual Systems), with examples of filtering algorithms and IDL implementations given in Canty, 2006, p. 52 and 65.

On some tree vertical profiles it is impossible to detect where the drop occurs, as there is no acceptable mathematical solution. In such situations, the crown base height was estimated as the height corresponding to the 25th percentile of the vertical distribution of laser heights within the tree crown. The study of Kimes et al. (2006) discusses the typical forest waveform based on NASA’s Laser Vegetation Imaging Sensor (LVIS) and focused on four quartile heights: 25, 50, 75, and 100%. They found that about 25% of the returned energy occurs from the ground elevation, with the base of the crown corresponding approximately to the 50% quartile. As such, we eliminate all ground laser points (25% quartile) and when curve fitting cannot provide a crown base height estimate, we consider the 25% quartile of the remaining laser points as an indicator of

Fig. 7. A case of lidar frequency crown vertical profile and the fitted polynomial (a) the original profile (b) the profile smoothed by FFT filtering (c) the profile smoothed by wavelet-based filtering.
crown base height. Out of the 117 matched trees, fitting a polynomial on the original frequency-derived vertical profile could not provide a crown base height estimate for eight trees (6.8%), while fitting the polynomial on the Fourier and wavelet-filtered profile could not estimate crown base height for two (1.7%) and three trees (2.6%), respectively. The polynomial fitting method could not provide crown base measurements for only one tree (0.85%). In such situations, the crown base height is estimated from the percentile method.

For the laser hit frequency-profile of crowns, linear regression for predicting field-measured crown base height by using valid lidar estimates of crown base height as the independent variable, i.e., when the polynomial fit provided an estimate, ranked the methods as follows: (1) Fourier filtering $R^2 = 0.79$; (2) wavelet-filtered profile $R^2 = 0.78$; (3) original unsmoothed profile $R^2 = 0.76$; and (4) 25th percentile $R^2 = 0.11$. Therefore, Fourier filtering provided the best estimates and was the first step in our algorithm for estimating crown base height. Whenever the polynomial fit could not provide an estimate when used with the Fourier-filtered vertical profile, crown base height would be determined from the wavelet-filtered profile. If polynomial fit with the wavelet-filtered profile could not provide an estimate for crown base height, fitting the curve to raw profiles would be preferred (1.7% of the trees). The worst
crown base height derived on the intensity CVP. FCBH = crown base height derived on the frequency CVP; and ICBH = crown base height derived on the percentile method (0.85% of the analyzed trees). The algorithm is shown in Fig. 8. A similar trend in the strength of the correlation between lidar and field measured crown base height, would lead to estimating crown base height most of the variance associated with crown base height.

### 2.7. Statistical analysis

We used ANOVA (ANalysis Of Variance) with repeated measures (SAS Help and Documentation, 2007; Repeated Measures ANOVA Using SAS PROC GLM, University of Texas Statistical Services, 2007) to investigate differences in the performance of the four methods for analyzing the vertical profiles — Fourier filtering, wavelet filtering, original unsmoothed profile, and the 25th percentile, and for two different species — pines and deciduous, and interactions between these factors. Repeated measures ANOVA is used when all members of a sample, in our case the trees, are measured repeatedly under a number of different conditions. As the tree is exposed to each condition in turn, the measurement of the dependent variable (CBH) is repeated for each processing method. Using a standard ANOVA in our case is not appropriate because it fails to model the correlation between the repeated measures — the same crown vertical profile of a tree is exposed to different processing methods. Also, in situations where there is a large variation between sample members (trees), error variance estimates from standard ANOVA are large. Using ANOVA with repeated measures provides a way of accounting for this variance, thus reducing error variance. In repeated measures ANOVA, a sample member, a tree in our analysis, is called a subject. The dependent variable CBH is measured repeatedly for all sample members (trees) with a set of conditions. Conditions known as trials in repeated measures ANOVA constitute the within-subject factors and in our case are the four types of processing methods — Fourier filtering, wavelet filtering, original unsmoothed profile, and the 25th percentile method. When a dependent variable (CBH) is measured on independent groups of sample members, i.e., species, conditions are known as between-subjects factors.

The following hypotheses were tested with repeated measures ANOVA: (1) does the processing method influence CBH measurements? This is the test for within-subjects, i.e., trees, main effect of processing method with repeated measures ANOVA. (2) Does species type influence CBH measurements? This is the test for between-subjects, i.e., trees, main effect of species type. (3) Does the influence of species on CBH measurements depend upon processing method? This is the test for between-subjects by within-subjects interaction of species type by processing method. In other words, should we use different processing methods depending on species type? We tested these hypotheses on CBH measurements obtained using frequency and intensity lidar data, before using regression to predict ground measurements.

We also investigated whether there are significant differences between crown base heights measured by the methods that ranked highest based on their polynomial fit using deviations from field-measured values. Therefore, we tested three paired combinations of the three methods: Fourier filtering, wavelet filtering, and unsmoothed original profile.

Linear regression models with a significance level of 0.05 were used to develop equations relating lidar- and field-measured crown base height in two instances: (1) using the lidar-derived crown base height as the only independent variable and (2) using all lidar-derived measurements as independent variables, including total tree height, crown width, and crown base height. Linear regression with the two approaches mentioned above was used with both types of crown vertical profiles, laser point frequency- and intensity-derived profiles, for all trees, and separately for pines and deciduous trees.

### Table 2

Results for estimating individual tree heights and crown width

<table>
<thead>
<tr>
<th>Species</th>
<th>$R^2$</th>
<th>RMSE (m)</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (H, m), lidar-measured height (LH) is the independent variable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.95</td>
<td>1.55</td>
<td>$H=1.43+0.99LH$</td>
</tr>
<tr>
<td>Pines</td>
<td>0.96</td>
<td>1.38</td>
<td>$H=1.22+1.00LH$</td>
</tr>
<tr>
<td>Deciduous</td>
<td>0.90</td>
<td>2.15</td>
<td>$H=2.04+0.96LH$</td>
</tr>
</tbody>
</table>

Crown width (CW, m), lidar-measured crown width (LCW) is the independent variable

| All     | 0.53  | 1.84     | CW=1.96–0.71 LCW |
| Pines   | 0.57  | 1.68     | CW=1.77–0.70 LCW |
| Deciduous | 0.59 | 2.08     | CW=1.92+0.95LCW |

$^a$ Where lidar independent variables are: LH = tree height; LCW = crown width; FCBH = crown base height derived on the frequency CVP; and ICBH = crown base height derived on the intensity CVP.

### Table 3

Results for estimating crown base height

<table>
<thead>
<tr>
<th>Species</th>
<th>$R^2$</th>
<th>RMSE (m)</th>
<th>Model $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency CVP, all lidar independent variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.80</td>
<td>2.03</td>
<td>CBH=1.01+0.31 LH+0.41 FCBH</td>
</tr>
<tr>
<td>Pines</td>
<td>0.79</td>
<td>2.03</td>
<td>CBH=1.12+0.58 LH</td>
</tr>
<tr>
<td>Deciduous</td>
<td>0.74</td>
<td>1.88</td>
<td>CBH=1.53+0.47 LH</td>
</tr>
</tbody>
</table>

| Frequency CVP, crown base height as the only independent variable |       |          |            |
| All     | 0.78  | 2.08     | CBH=1.42+0.85 FCBH |
| Pines   | 0.78  | 2.08     | CBH=1.75+0.84 FCBH |
| Deciduous | 0.73 | 1.90     | CBH=1.51+0.76 FCBH |

| Intensity CVP, all lidar independent variables |       |          |            |
| All     | 0.78  | 2.07     | CBH=0.91+0.57 LH |
| Pines   | 0.79  | 2.03     | CBH=1.12+0.58 LH |
| Deciduous | 0.74 | 1.88     | CBH=1.53+0.47 LH |

| Intensity CVP, crown base height as the only independent variable |       |          |            |
| All     | 0.74  | 2.27     | CBH=2.68+0.76 ICBH |
| Pines   | 0.77  | 2.12     | CBH=2.62+0.78 ICBH |
| Deciduous | 0.49 | 2.63     | CBH=3.54+0.55 ICBH |

$^a$ Where lidar independent variables are: LH = tree height; LCW = crown width; FCBH = crown base height derived on the frequency CVP; and ICBH = crown base height derived on the intensity CVP.
Linear regression was also used to investigate the strength of the relationship between lidar- and field-measured total tree height and crown width for individual trees. We used a paired $t$-test to investigate whether there is any difference between predicted CBH measurements when using frequency and intensity lidar data.

Fig. 9. The lidar-derived CBH from frequency vertical profiles compared to reference field measurements, for the four different methods. Note: the regression line for estimating field-measured CBH is shown as the dashed line.

Fig. 10. The lidar-derived CBH from intensity vertical profiles compared to the reference field measurements, for the four different methods. Note: the regression line for estimating field-measured CBH is shown as the dashed line.
3. Results

3.1. Repeated measures ANOVA

For our within-subject main effect test, the null hypothesis is that the mean CBH does not change across different processing methods. The $p$-value for this test obtained using Wilke’s test (Repeated Measures ANOVA Using SAS PROC GLM, University of Texas Statistical Services, 2007) was 0.0507, thus we do not reject the null hypothesis and conclude that CBH does not change significantly with the processing method. However, the $p$-value is only marginally greater than the critical value for the 5% significance level.

Does species type influence CBH measurements? The between-subjects test indicates that species type influences CBH measurements significantly, as the null hypothesis of no species influence on CBH was rejected. In this instance, the $p$-value associated with the interaction was low (0.0003). Next, we were interested in finding out whether any of the processing methods are better suited for a species type; hence, the next hypothesis was tested.

Does the influence of species on CBH measurements depend upon processing method? We marginally do not reject the null hypothesis of no interaction ($p$-value=0.0694). A similar situation for all three hypotheses was obtained when using intensity-derived lidar height bins.

Paired $t$-tests for deviations of results from the three ranked methods and field-measured crown base heights indicated significant differences only between the Fourier method and the other two, i.e., wavelet filtering and original profile ($p$-values of 0.008 and 0.014, respectively).

3.2. Individual tree measurements

Linear regression results for estimating field-measured individual tree height and crown diameter for all trees and separately for pines and deciduous trees are shown in Table 2. Results for estimating crown base height are shown in Table 3. Figs. 9 and 10 show plots of lidar vs. field estimates of crown base height when using frequency and intensity CVP, respectively.

4. Discussions

Our results for measuring individual tree height and crown width are comparable to other findings in the lidar literature. Using similar algorithms, but with results reported at plot level for a study in the Piedmont region of Virginia, USA, Popescu and Wynne (2004) found that for the pine plots, lidar measurements explained 97% of the variance associated with the mean height of dominant trees (RMSE 1.14 m). For deciduous plots, regression models explained 79% of the mean height variance for dominant trees, with a 1.91 m RMSE.

Results for estimating crown diameter for individual trees are comparable to results reported at plot level in the study of Popescu et al. (2003), that reported $R^2$ values of 0.62–0.63 for the dominant trees (RMSE 1.36 to 1.41 m). Part of the unexplained variance associated with crown diameter can be attributed to the fact that the algorithm for calculating crown diameter on the lidar canopy height model aimed at measuring the non-overlapping crown diameter, while the field measurements considered crowns to their full extent, therefore measured overlapping crown diameters.

The repeated measures ANOVA indicated that, marginally, there were no differences between results obtained with different processing techniques used before polynomial fit to assess crown base height. However, as explained in the Materials and methods section, a limited number of tree vertical profiles are noisy and they do not provide an acceptable mathematical solution when fitting the polynomial function to derive crown base height. This situation occurred for 6.8% of the trees when fitting the polynomial on the raw, unfiltered, crown vertical profile, compared to 1.7% of the vertical profiles smoothed with Fourier filtering. Also, regression fit was slightly better for crown base height estimates obtained on the filtered vertical profiles. As such, although the repeated measures ANOVA indicated no differences between processing methods, processing the crown vertical profiles using filtering techniques proved to be beneficial for improving crown base height estimates.

These results compare favorably to other studies. As with most lidar measurements, the correlation between field and lidar-derived estimates is lower for deciduous trees than pines, but all $R^2$ values, with one exception, are in the range from 0.73 to 0.80. Riano et al. (2004) obtained $R^2$ values for estimating crown base height of 0.65 to 0.68 for individual trees and found that lidar produced much higher crown base height values than the field data. Holmgren and Persson (2004) found that crown base height for individual trees was on average overestimated by 0.75 m, with an $R^2$ of 0.71 (correlation coefficient $r=0.84$). Andersen et al. (2005) estimated plot-level canopy base height with a coefficient of determination of 0.77 using linear regression with several percentile-based metrics.

Our results indicate that lidar-derived crown base height overestimates field-measured crown base height, which is consistent with previous studies, e.g., the studies of Holmgren and Persson (2004), Riano et al. (2004), and Andersen et al. (2005). On average, with either CVPs, frequency- or intensity-derived, our method overestimated crown base height by 0.36 and 0.12 m, respectively. The difference can be attributed to both the penetrating characteristics of laser pulses and the polynomial fitting effects as seen in Fig. 7 a). When there is an abrupt change on the CVP at the base of the crown, the fitted curve will register a short lag and go through an inflection point at a slightly upper height. However, this bias can be removed through the prediction equations in Table 3. Errors for estimating crown base height are also expected to be related to errors for estimating crown diameter, since the CVP is generated from pixel values of lidar height bins within the vertical space defined by a cylinder with radius equal to the lidar-estimated crown width. The methodology for assessing crown width aims at measuring non-overlapping crown diameters.

Table 3 shows little difference between results obtained with the frequency- and intensity-based CVP. The frequency CVP
produces consistent $R^2$ values for each tree group, between 0.73 and 0.78, while the intensity CVP has a low $R^2$ values for deciduous trees, 0.49. In addition to the similar $R^2$ values for frequency and intensity CVP results, a paired t-test implemented in SAS (SAS Institute, Inc.) indicated that the difference between the two sets of crown base height measurements is only marginally non-significant at 0.05 level ($p$-value 0.06).

5. Conclusions

Results of this study show that lidar data can be used to accurately estimate biophysical parameters of individual trees, such as total tree height, crown width, and height to crown base. The lidar height-bins approach that we propose has high potential for becoming a standardized method for processing and exchanging forestry lidar data. As opposed to processing individual lidar points with in-house developed algorithms of limited availability to researchers and operational users, a multiband-type of lidar data is much easier to process and analyze using remote sensing image processing software. More so, the lidar height bins can be used for a variety of applications, such as habitat assessment, modeling photosynthetically-active radiation (PAR) through the canopy, and mapping surface and crown fuels, etc. As opposed to previous attempts to use lidar height bins as binary products, i.e., containing laser hits or not (Holmgren & Persson, 2004; Lefsky et al., 1999), the lidar height bins used in this study contain more detailed information that can be used in forest canopy studies. Lidar voxels and height bins of varying dimensions can be adapted to particular needs of a study, while processing methods are readily available from multispectral or hyperspectral image processing software packages, such as ENVI (ITT Visual Solutions) or ERDAS Imagine (Leica Geosystems).

Our approach is unique for being able to automatically identify individual trees on lidar canopy height models at local scales and for characterizing the vertical structure of each tree crown. The Forward Fourier Transform filtering proved to be an effective technique for smoothing the vertical profile of crowns before fitting a polynomial curve to mathematically analyze the vertical profile. These methods allowed us to predict field measured crown base height with high accuracy, given the limitations of lidar sensing of the base of the crown from above the forest canopy.

With respect to using either the frequency- or intensity-derived CVP, despite the fact that results showed few differences, we would consider the lidar frequency information more appropriate for future studies. With current lidar sensors reaching pulse frequencies of more than 100 kHz, future forestry lidar data sets are likely to have a high point density that better portray the canopy architecture and, therefore, frequency voxels or height bins may be appropriate for standardization of forestry lidar data products.

Overall, deriving information on the height of the crown base proved to improve results for assessing other biophysical forest parameters, such as height and crown width. Such individual tree measurements derived by automated processing of lidar-derived canopy height models have the potential to improve the accuracy of other forestry estimates, such as volume and biomass. The results of this study and the potential for integrating new lidar products with co-registered multi- and hyperspectral digital imagery makes lidar a realistic alternative to deriving traditional forest measurements with high accuracy. Moreover, lidar can bring dramatic gains in characterizing the three-dimensional structure of the forest canopy for a variety of ecological studies that could improve our understanding of the forest environment and management practices.

Acknowledgements

We gratefully acknowledge the support provided by the Texas Forest Service (award #: 02-DG-11083148-050) and the graduate student support provided by NASA (award #: NNG04GM34G). We are very thankful for the help provided with the field data collection by Curt Stripling, all forestry personnel with the Texas Forest Service, and by graduate students, Muge Mutlu and Alicia Griffin.

We would also like to thank the three anonymous reviewers who helped us improve our manuscript.

References
