Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable Window Size for Estimating Tree Height

Article *in* Photogrammetric Engineering and Remote Sensing · May 2004

DOI: 10.14358/PERS.70.5.589

CITATIONS
353

READS
1,758

2 authors:

Sorin C. Popescu
Texas A&M University
92 PUBLICATIONS **4,396** CITATIONS

SEE PROFILE

Randolph Wynne
Virginia Polytechnic Institute and State University
171 PUBLICATIONS **8,061** CITATIONS

SEE PROFILE

Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable Window Size for Estimating Tree Height

Sorin C. Popescu and Randolph H. Wynne

Abstract

The main study objective was to develop robust processing and analysis techniques to facilitate the use of small-footprint lidar data for estimating plot-level tree height by measuring individual trees identifiable on the three-dimensional lidar surface. Lidar processing techniques included data fusion with multispectral optical data and local filtering with both square and circular windows of variable size. The lidar system used for this study produced an average footprint of 0.65 m and an average distance between laser shots of 0.7 m. The lidar data set was acquired over deciduous and coniferous stands with settings typical of the southeastern United States. The lidar-derived tree measurements were used with regression models and cross-validation to estimate tree height on 0.017-ha plots. For the pine plots, lidar measurements explained 97 percent of the variance associated with the mean height of dominant trees. For deciduous plots, regression models explained 79 percent of the mean height variance for dominant trees. Filtering for local maximum with circular windows gave better fitting models for pines, while for deciduous trees, filtering with square windows provided a slightly better model fit. Using lidar and optical data fusion to differentiate between forest types provided better results for estimating average plot height for pines. Estimating tree height for deciduous plots gave superior results without calibrating the search window size based on forest type.

Introduction

Laser scanner systems currently available have experienced a remarkable evolution, driven by advances in the remote sensing and surveying industry. Lidar sensors offer impressive performance that challange physical barriers in the optical and electronic domain by offering a high density of points at scanning frequencies of 50,000 pulses/second, multiple echoes per laser pulse, intensity measurements for the returning signal, and centimeter accuracy for horizontal and vertical positioning. Given a high density of points, processing algorithms can identify single trees or groups of trees in order to extract various measurements on their three-dimensional representation (e.g., Hyyppä and Inkinen, 2002).

The foundations of lidar forest measurements lie with the photogrammetric techniques developed to assess tree height, volume, and biomass. Lidar characteristics, such as high sampling intensity, extensive areal coverage, ability to penetrate beneath the top layer of the canopy, precise geolocation, and accurate ranging measurements, make airborne laser systems useful for directly assessing vegetation characteristics. Early lidar studies had been used to estimate forest vegetation characteristics, such as percent canopy cover, biomass (Nelson *et al.*, 1984; Nelson *et al.*, 1988a; Nelson *et al.*, 1988b; Nelson *et al.*, 1997), and gross-merchantable timber volume (Maclean and Krabill, 1986). Research efforts investigated the estimation of forest stand characteristics with scanning lasers that provided lidar data with either relatively large laser footprints, i.e., 5 to 25 m (Harding *et al.*, 1994; Lefsky *et al.*, 1997; Weishampel *et al.*, 1997; Blair *et al.*, 1999; Lefsky *et al.*, 1999; Means *et al.*, 1999) or small footprints, but with only one laser return (Næsset, 1997a; Næsset, 1997b; Magnussen and Boudewyn, 1998; Magnussen *et al.*, 1999; Hyyppä *et al.*, 2001). A small-footprint lidar with the potential to record the entire time-varying distribution of returned pulse energy or waveform was used by Nilsson (1996) for measuring tree heights and stand volume.

As more systems operate with high performance, research efforts for forestry applications of lidar have become very intense and resulted in a series of studies that proved that lidar technology is well suited for providing estimates of forest biophysical parameters. Needs for timely and accurate estimates of forest biophysical parameters have arisen in response to increased demands on forest inventory and analysis.

The height of a forest stand is a crucial forest inventory attribute for calculating timber volume, site potential, and silvicultural treatment scheduling. Measuring of stand height by current manual photogrammetric or field survey techniques is time consuming and rather expensive. Tree heights have been derived from scanning lidar data sets and have been compared with ground-based canopy height measurements (Næsset, 1997a; Næsset, 1997b; Magnussen and Boudewyn, 1998; Magnussen *et al.*, 1999; Næsset and Bjerknes, 2001; Næsset and Økland, 2002; Persson *et al.*, 2002; Popescu, 2002; Popescu *et al.*, 2002; Holmgren *et al.*, 2003; McCombs *et al.*, 2003). Despite the intense research efforts, practical applications of

Department of Forestry, Virginia Tech, 319 Cheatham Hall (0324), Blacksburg, VA 24061 (wynne@vt.edu).

S.C. Popescu is presently with the Spatial Sciences Laboratory, Department of Forest Science, Texas A&M University, 1500 Research Parkway, Suite B223, College Station, TX 77845-2120 (s-popescu@tamu.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 70, No. 5, May 2004, pp. 589–604.

0099-1112/04/7005–0589/$3.00/0

small-footprint lidar have not progressed as far, mainly because of the current cost of lidar data. However, with an anticipated decline of lidar data cost in the near future and promising current research efforts, lidar is expected to be used extensively in forest measurements.

A review of the rapidly growing literature on lidar applications emphasizes the need for optical data fusion in the processing phase of lidar data as a method to improve various feature extraction tasks. Previous studies using high-resolution digital images attempted to estimate tree heights, canopy density, and forest volume or biomass by individually mapping tree crowns (Gougeon, 1995; St-Onge and Cavayas, 1995; Brandtberg, 1997; Wulder *et al.*, 2000). A trial study by Næsset (2002) revealed that automated softcopy photogrammetric methods did not provide better results when compared to manual photogrammetric techniques for estimating tree heights. As opposed to such endeavors, lidar sensors allow analysts to directly portray forests in a three-dimensional format over large areas. Lidar sensors are clearly superior to photogrammetric instruments in their ability to see between the trees and through the canopy openings, but lidar sensors have their own shortcomings. Lidar data provide multiple return position and intensity measurements, but contain only limited information for deriving the correspondence to target objects. Optical imagery allows for feature identification; thus, the fusion of range and reflectance data provides additional support for the automatic feature measurement process. Optical data are particularly useful in forestry applications for differentiating between forest and non-forest areas and for discriminating between major tree species, such as coniferous and deciduous. Toth *et al.* (2001) examined the feasibility of combining lidar data with simultaneously captured digital images to improve the surface extraction process. Their investigation was limited to the conceptual level and was only intended to demonstrate the potential of lidar and optical data fusion. Lidar and multispectral data fusion were used by McCombs *et al.* (2003) to estimate stand density and mean tree height. They found that tree identification in regularly spaced loblolly pine plantations was more accurate when using spectral data rather than lidar data alone, and that a fused data set provided the best results for identifying peak values assumed to represent tree tops.

This study attempts to make a contribution to inventorying and measuring forest biophysical parameters using lidar data by developing and testing specific processing algorithms targeted towards forestry applications. The use of lidar remote sensing techniques for assessing forest biophysical parameters has been investigated by other researchers, but as of yet such approaches have met with little success for multiage, multispecies forests. Lidar studies published at this point have shown success in several forest types with large-footprint lidar, but applications of small-footprint lidar to forestry have not progressed as far (Means, 2000), being limited mainly to measuring even-aged coniferous stands. Thus, this study aims at presenting a new approach for using a fused lidar data set and multispectral optical data for assessing tree height in forest areas covered with both hardwoods and softwoods typical of the eastern United States, because there is an increasing need to improve the accuracy of forest estimates. The specific objectives provide a general outline of the study approach and are as follows: (1) to develop robust processing and analysis techniques to facilitate the use of small-footprint lidar data for predicting plot-level tree heights by directly measuring individual trees identifiable on the three-dimensional lidar surface, and (2) to compare the performance of different lidar processing techniques for estimating tree height, as follows: local filtering with variable window size and different window shapes, square and circular, each with and without data fusion with multispectral optical imagery.

Materials and Methods

Study Site

The study area is located in the southeastern United States, in the Piedmont physiographic province of Virginia. It includes a portion of the Appomattox-Buckingham State Forest that is characterized by deciduous, coniferous, and mixed stands of varying age classes (upper-left coordinates: 37°25′ N, 78°41′ W; lower-right coordinates: 37°24′ N, 78°39′ W). A mean elevation of 185 m, with a minimum of 159 m and a maximum of 238 m, and rather gentle slopes characterize the topography of the study area.

Ground Inventory Data

The ground-truth data were collected from November 1999 to April 2000. Six forest vegetation types were covered by the field sampling—pine-hardwoods; upland hardwoods; bottomland hardwoods; and stands of loblolly pine, Virginia pine, and shortleaf pine. Forest type is a plot-level classification defined by the relative stocking of tree species or species groups (Powell *et al.*, 1993). The stand age varied, being approximately 15 years for the majority of the pine stands, 35 to 55 years for the pine-hardwood mixed stands, 85 to 90 years for the bottomland hardwoods, and up to 100 to 140 years for the upland-hardwood stands. Three stands of loblolly pine were exceptionally old, with ages of 60 to 65 years, and with a multilayer vertical structure. The number of tree species identified per subplot ranged from one (pure stands) to ten, with an average of 4.6 species per subplot. A more detailed description of the tree species inventoried on the ground can be found in Popescu (2002) and Popescu *et al.* (2002).

The plot design followed the U.S. National Forest Inventory and Analysis (FIA) field data layout (Figure 1). An FIA plot consists of a cluster of four subplots approximately 0.017 ha (0.04 acres) each, with a radius of 7.32 m (24.0 ft) (USDA Forest Service, 2001). One plot is distributed over an area of approximately 0.4 ha (1 acre); thus, it represents a sample of the conditions within this area. The center plot is subplot 1. Subplots 2, 3, and 4 are located 36.58 m (120.0 ft) at azimuths 0, 120, and 240 degrees from the center of subplot 1. Subplots are used to collect data on trees with a diameter at breast height (dbh, diameter measured at 1.37 m (4.5 ft) above the ground) of 12.7 cm (5.0 in) or greater. For the purpose of this

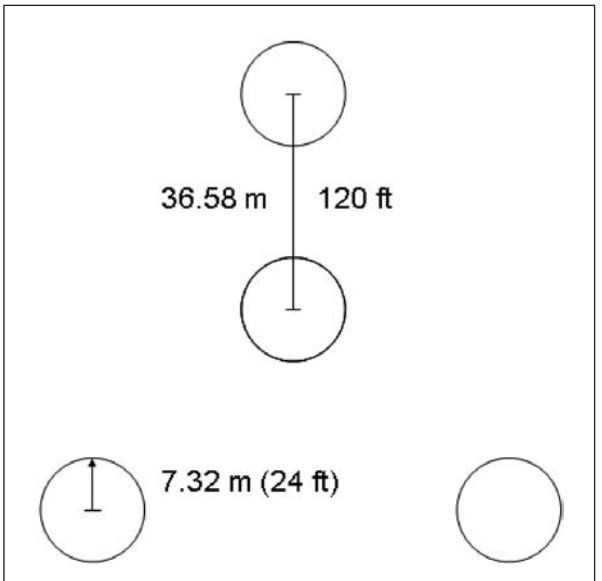


Figure 1. Layout of a single FIA plot with four subplots.

TABLE 1. NUMBER OF SUBPLOTS DIFFERENTIATED BY FOREST COVER TYPE

Forest Cover Type	Number of Subplots
Pine-hardwood	14
Upland hardwood	17
Bottomland hardwood	7
Total hardwood and mixed pine-hardwood	38
Loblolly pine	16
Virginia pine	6
Shortleaf pine	4
Total pines	30
Total number of subplots	64

study, the FIA standard protocol was modified and data were collected on trees with a dbh of at least 6.35 cm (2.5 in). A total of 16 plots were measured in the study area, each with four subplots. FIA plot centers (subplot 1 centers) were located systematically on a 200- by 200-m (656- by 656-ft) grid, with rows oriented east-west and columns oriented north-south (Plate 1). The origin of the grid relative to the map was randomly selected. Plots were selected to ensure representation of the forest-cover types in the study area while maintaining approximate equivalence between the number of coniferous and deciduous plots (Table 1).

To simplify the analysis relative to tree species, subplots were categorized as either "hardwoods" or "pines." For the pine-hardwoods mixed stands, the species group of the subplot was named after predominant tree species. Predominance was established by basal area (Eyre, 1980), and the subplot category was assigned to the species comprising more than half of the stocking. The ground-truth data set contained 33 hardwood subplots and 31 pine subplots.

The centers of subplots 1, for most of the plots, were laid out in the field using a navigational GPS unit, a Rockwell Collins PLGR. Centers of subplots 2, 3, and 4 of the same plot were located by bearing and distance from subplot 1. Four out of the 16 plots were set by bearing and distance from previously located plots. In addition, all FIA subplot locations were determined using 60-second static measurements with a 12-channel GPS receiver, an HP-GPS-L4 with a PC5-L data collector (Corvallis Microtechnology, Inc., URL: http://www.cmtinc.com/nav/frprod.html, last accessed 26 January 2004). For a better GPS satellite signal reception, the antenna was raised on a height pole fixed to a staff equipped with a level for maintaining the antenna vertically over the plot location. The GPS antenna height varied between 1.8 m and 7.6 m, with an average height of 3.8 m. All measurements were collected during the leaf-off season for the hardwood stands. The lack of canopy foliage and the raised antenna in the denser pine stands reduced the error effects of forest canopies on GPS measurements. The reported mapping accuracy for the HP-GPS-L4 unit, obtained under open sky for 60 seconds of static measurements is 30 cm (Corvallis Microtechnology, Inc.). Under forest canopy, GPS systems tend to yield from 1.5 to 3 times less accurate solutions (Craig Greenwald, Corvallis Microtechnology, Inc., Technical Support, personal communication, 2001). Therefore, we estimate submeter accuracy for locating the plot centers. Depending on the data availability, the following National Geodetic Survey continuously operating reference stations in Virginia were used for the differential correction: Blacksburg, Driver, Charlottesville, and Richmond, all within the baseline distance of 300 km (187.5 miles) for this type of GPS receiver from the location of the study area.

On each subplot, the heights of all trees were measured using a Vertex Forestor hypsometer. Some of the pine tree heights that were less than 7.62 m (25 ft) were measured using a height pole. Several heights less than 7.62 m were measured with both methods and the height difference never exceeded 15 cm (0.5 ft). Tree heights on three plots were measured using a Suunto clinometer (PM-5) and a distance tape. The height measurement recorded the total length of the tree, to the nearest 0.30 m (1.0 ft) from ground level to the top of the tree. Diameter at breast height (dbh) was measured on all trees within the subplots using a diameter tape. The actual diameter was recorded for each tallied tree to the last whole 0.25 cm (0.1 in). Crown width was measured on all trees with a dbh larger than 12.7 cm (5.0 in) and was used in a separate study to investigate the performance of lidar data for estimating crown diameter (Popescu *et al.*, 2003). Crown width was determined as the average of four perpendicular crown radii measured with a tape from the tree bole towards the subplot center, away from it, to the right and to the left. The location (X, Y) of each tree relative to the subplot center was determined by bearing and distance using a distance tape and a Suunto compass (KB-14), with an expected standard error of up to 30 cm (1 ft), depending on the distance to the subplot center. Bearing was measured from the subplot center sighting to the center of the base of each tree. The horizontal distance was recorded to the nearest 0.03 m (0.1 ft) from the subplot center to the pith at the base of the tree. 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 of 1.5 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.

Subplot averages were calculated from individual tree measurements and were used to assess the performance of the lidar processing algorithms. Descriptive statistics of subplot values for the pines and deciduous plots are given in Table 2.

The standards for FIA data collection (i.e., acceptable errors in quality checks, though check crews were not used in this study) are as follows: tree height: ±10 percent of the total

TABLE 2. DESCRIPTIVE STATISTICS OF THE FIELD INVENTORY DATA FOR PINES AND DECIDUOUS SUBPLOTS

Statistic	Dbh (cm)	Height (m)	Crown Width (m)	Number of Trees/Plot
		Pines Subplots		
Mean	13.22	10.56	4.04	21.90
Minimum	7.57	5.03	1.97	3
Maximum	26.67	17.37	10.12	68
Standard Deviation	4.21	2.98	1.79	13
		Deciduous Subplots		
Mean	17.18	12.99	5.98	11.3
Minimum	8.42	8.58	3.79	3
Maximum	28.88	18.64	8.85	22
Standard Deviation	4.30	2.18	1.27	5

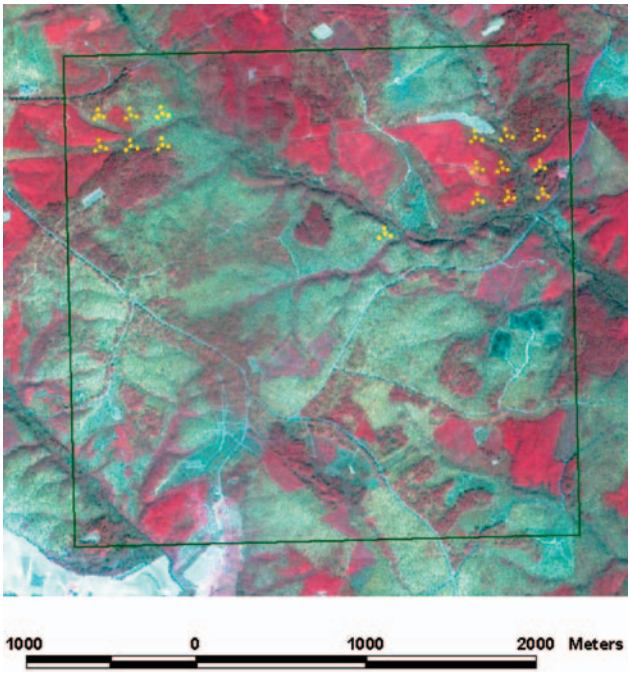


Plate 1. Location of study plots (yellow dots) on a leaf-off color infrared ATLAS image (NASA's Airborne Terrestrial Land Applications Scanner, 4-m resolution, 1998). The green square shows the lidar data coverage. (Copyright 2002, American Society for Photogrammetry and Remote Sensing, 2002 Annual Conference Proceedings.)

height; tree mapping: ±3 degrees for azimuth and 0.3 m (1.0 ft) for distance; and dbh: 0.25 cm (0.10 inch) for trees 50.8 cm (20 inches) or less and 0.51 cm (0.2 inches) for trees larger than 50.8 cm. For crown width there is no FIA standard but the error is estimated to be about 0.6 to 0.9 m (2 to 3 ft).

Lidar Data Set

The lidar data were acquired on 02 September 1999 over an area of 1012 ha (2500 acres) located in the Appomattox-Buckingham (AB) State Forest in Virginia. The lidar system (AeroScan, EarthData, Inc.) utilizes advanced technology in airborne positioning and orientation, enabling the collection of high-accuracy digital surface data. The scanning system uses an oscillating mirror with a scanning rate of 10 Hz and a scanning angle that can be adjusted from 1° to 75°. For the Appomattox-Buckingham data set the scanning angle was 10°, giving a total field of view of 20°. The average ground swath width was 699 m and the entire research area was covered by 21 parallel flight lines. The mission was designed with up to 70 percent side overlap to increase the point density on the ground and to correct for the typical zig-zag lidar scanning pattern. A more detailed description of the lidar data set and the sensor characteristics is given in Popescu *et al.* (2002a). The laser beam divergence was 0.33 mrad, yielding a footprint of 0.65 m from the flying height of 1980 m.

The AeroScan system used for this study was capable of recording up to four returns for each laser pulse, depending on the ground cover. For this study, only the first and the last returns were used. The last return could coincide with the first, if there is only one return per pulse, or could be any other return from the second to the fourth, depending on the number of returns for a particular pulse.

TABLE 3. BASIC STATISTICAL MEASURES FOR THE NUMBER OF LIDAR POINTS PER 1 m^2

Statistic	Value
Number of 1-m^2 cells analyzed	435,600
Mean	1.35
Mode	1
Median	1
Standard Deviation	1.89
Range	54
Interquartile Range	2

To investigate the laser point density, a regular grid of 660 by 660 meters was overlaid with the lidar points located in the upper right corner of the study area. This portion of the study area is covered by nine FIA-type plots with a mixture of pine and hardwood stands and is representative for the range of scanning patterns. The grid cell size was 1 by 1 m; therefore, the statistical measures were reported directly per 1 m^2. The area included laser points from seven adjacent flight lines, though some of the flight lines only partially covered the area. The distribution of the number of points in each 1-m^2 cell was analyzed for the entire grid, and the results are summarized in Table 3. Figure 2 shows the frequency distribution of the number of lidar points per 1 m^2. By pooling all the laser points from adjacent swaths into the same point file, the average interpoint distance decreased to 0.7 m.

The provider performed an evaluation of the lidar data, including a comparison of the data from flight line to flight line. This comparison showed high relative accuracy and no anomalies in the data. All ranges were postprocessed by EarthData, Inc., and corrected for atmospheric refraction and transmission delays. The reported accuracies for the AeroScan lidar system flying at less than 2400 m above ground, over open homogeneous flat terrain, are as follows: an elevation or vertical accuracy of ±25 cm and a horizontal accuracy of ±50 cm (EarthData, Inc., URL: http://www.earthdata.com/lidar.htm, last accessed 26 January 2004).

Optical Data

In addition to the lidar data, spatially coincident optical data used for this study include a leaf-off ATLAS image (NASA's Airborne Terrestrial Land Applications Scanner; 4-m spatial resolution; flown 17 March 1998 at 2100 m above ground level), shown in Plate 1. The multispectral ATLAS image was acquired in the leaf-off season of 1998. Only the first eight bands, covering the visible, near-, and mid-infrared portion of the spectrum, were used for this study.

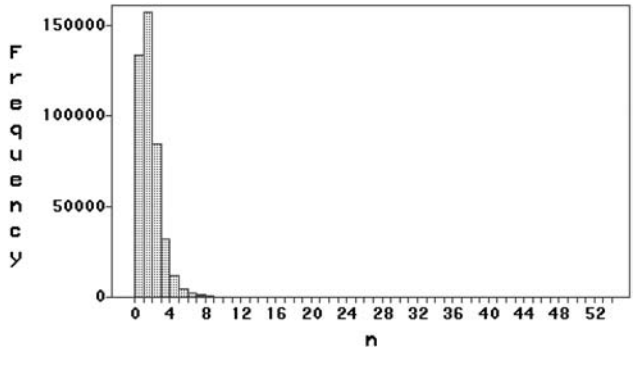


Figure 2. Frequency distribution of the number of lidar points (n) per 1 m^2.

Canopy Height Model (CHM)

The tree canopy height model was computed as the difference between tree canopy hits and the corresponding lidar-derived terrain elevation values. An iterative approach was used to construct the terrain model from the raw lidar data points using a slope-based algorithm. The ground-points classification algorithm was implemented in IDL (Interactive Data Language), with details provided in Popescu (2002).

Tree canopy hits or first-return lidar points are usually interpolated to a regular grid that corresponds to the digital surface model (DSM). To take advantage of the lidar point density that allows a 3D surface representation of individual trees, the grid size of the DSM of first-return lidar points was 0.5 meters. The lidar point density per 0.25 m^2 was investigated by overlaying a grid of 0.5- by 0.5-meter cells over the first-return lidar points. The number of lidar points per 0.25 m^2 ranged from 0 to 32. The average elevation difference of lidar points in the same cell was 0.44 m, with a range between 0 and 29.73 m and a standard deviation of 1.8 m. This large elevation difference for a small area is most likely due to overlaying lidar points with different incidence angles from adjacent flight lines or to different penetration heights for laser pulses with similar incidence angles (Figure 3).

When situations like the one depicted in Figure 3 occur, it is difficult to anticipate what elevation values are used to interpolate lidar heights to a regular grid. To measure tree height, processing techniques must accurately derive the top vegetation surface. Therefore, to have a better control over the interpolation results, only the highest lidar elevations in each of the 0.25 m^2 cells were used with kriging to derive the top DSM. A comparison with the interpolated surface obtained from all first-return lidar heights shows that the top DSM is, on average, higher by 0.17 m. The largest height difference between the top DSM and the first-return surface was 25.19 m. Therefore, the use of all first-return lidar points to interpolate the DSM creates a slightly lower surface that can contribute to underestimating tree heights with lidar. The underestimation of tree height by lidar has been attributed in recent studies (Nilsson, 1996; Næsset, 1997a; McCombs *et al.*, 2003) to the more frequent laser sampling of the crown shoulders than the tree apex. As such, canopy heights are biased toward low values, but the method of deriving the lidar surface can further contribute to this effect.

To obtain the tree canopy height model (CHM), the terrain elevation was subtracted from the top DSM. Therefore, the accuracy of deriving the ground elevation directly affects the accuracy of measuring tree heights.

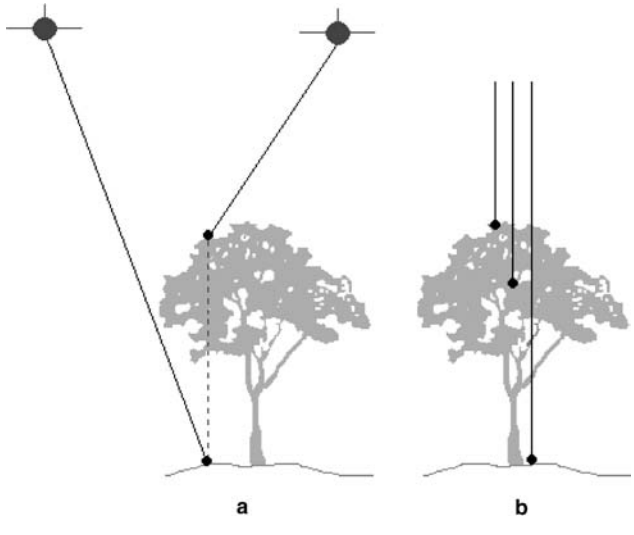


Figure 3. Difference in elevation over the same horizontal area due to combining data from adjacent flight lines (a) or to different laser penetration heights for the same incidence angle (b).

Tree Dimensions

Differentiation between Conifers and Hardwoods

The forest biometrics relationship between tree height and crown width was used in the processing of lidar data to locate individual trees and to measure their crown diameter. Because such a relationship is highly dependent on the tree species, it is of interest in the processing phase to differentiate between coniferous and deciduous species. Lidar data with only height measurements do not offer adequate information to distinguish between tree species. Therefore, data fusion with the leaf-off ATLAS image in Plate 1 was used to differentiate between the two categories of species, deciduous and coniferous.

The image was classified into three classes—open ground, deciduous, and coniferous—using a supervised approach and the maximum-likelihood classification decision rule (Plate 2). Only the first eight bands, covering the visible, near-, and mid-infrared portion of the spectrum (0.45 to 2.35 μm), were used in the classification process. The regions that represent each class are distinct and clearly identifiable on the ATLAS image; therefore, training samples were selected on this image for each of the three classes using the digitized polygon method. Several polygons digitized for each class were merged to create each spectral signature. Once the image was classified, the dominant and codominant trees from all the ground plots, which are the trees most likely seen from above, were used to assess the accuracy of the classified image. Tree locations and their species are known from the ground inventory. Dominant trees are considered the trees that have their dbh larger than the quadratic mean diameter for each plot. The quadratic mean diameter is also known as the diameter of the tree of mean basal area and is slightly larger than the arithmetic mean dbh (Avery and Burkhart, 1994). From the 64 FIA-type subplots, 425 dominant and codominant trees (133 deciduous and 292 pines) were selected for the accuracy assessment. In addition, the seven GPS points collected in

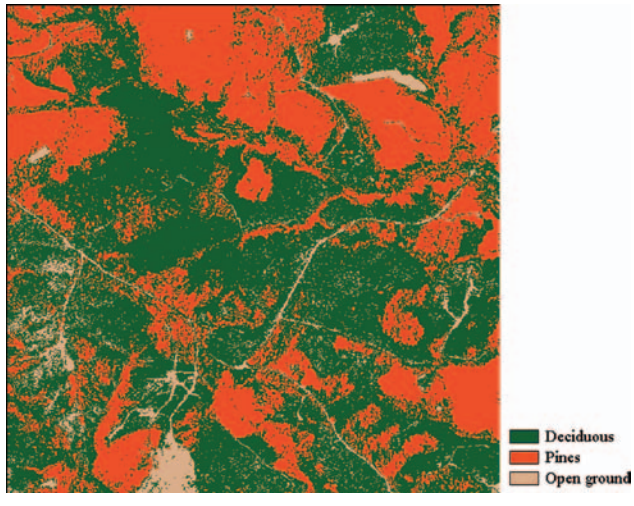


Plate 2. Classified ATLAS image.

TABLE 4. ERROR MATRIX FOR THE MAXIMUM-LIKELIHOOD CLASSIFICATION OF THE ATLAS IMAGE

	Reference Data			
	Open Ground	Deciduous	Pines	Row Total
Classified Data				
Open ground	6	4	1	11
Deciduous	1	97	36	134
Pines	0	32	255	287
Column Total	7	133	292	432

open ground were added to the reference data. Table 4 shows the error matrix associated with the classification process. The accuracy report is given in Table 5.

The users accuracy gives an indication on the reliability of the classified image as a predictive device relative to what species are on the ground (Campbell, 1996). Of the pine-forested area on the image, 88.85 percent is actually covered by pine species, while of the deciduous area, 72.39 percent. The Kappa statistic expresses the difference relating the observed agreement between the classified image and the ground reference, compared to the agreement of a completely random classification (Congalton and Green, 1999, p. 49). The 0.66 value for pines and 0.60 for deciduous species would imply that the classification process achieved an accuracy that is better than would be expected from chance assignment of pixels to the three categories (Congalton and Green, 1999, pp. 51–53).

The ATLAS image was georeferenced to the color-infrared ortho-image and, therefore, spatially coregistered to the lidar data. The initial 4-meter spatial resolution of the classified ATLAS image did not match the 0.5-m pixel size of the lidar CHM, but the image was resampled to the smaller grid size of the lidar surface using the nearest-neighbor method. The resampled image with three classes allows a pixel-level data fusion with the lidar surface. For a visual analysis of the coregistration between lidar data and the ATLAS image, the multiband image was draped over the three-dimensional CHM, as shown in Plate 3. Large individual deciduous crowns are visible in the foreground in Plate 3, while pine plantations show a smooth texture and are covered with shades of red, due to their strong reflection in the infrared portion of the spectrum on the leaf-off ATLAS image. In the absence of multispectral imagery acquired simultaneously with the lidar data, the ATLAS image offers an adequate source of information to differentiate between the two major tree species, deciduous and coniferous, which can be subsequently used for processing lidar data.

Tree Heights

Popescu *et al.* (2002) used two approaches to estimate the tree height on the same circular areas covered by the FIA-type subplots. The first approach was based on the height of all laser pulses within the area covered by the ground-truth data. The second method to estimate tree heights was based on single-tree identification using a variable window technique with local maximum (LM) focal filtering. Their results showed that the technique of estimating mean tree height by identifying the location of individual trees performed better than the first technique, which makes use of all laser height values within the subplots. For their study, Popescu *et al.* (2002) used a variable square shaped window. A similar technique with variable window size and texture analysis was used by Daley *et al.* (1998) with high-resolution optical images (MEIS-II) to estimate crown position in stands of Douglas fir (*Pseudotsuga menziesii* (Mirb.) Franco). Variable window sizes were also used by Wulder *et al.* (2000) for the extraction of tree locations and estimation of basal area from high spatial resolution imagery for stands of Douglas fir and western red cedar (*Thuja plicata*). With an image spatial resolution of 1 m, they used window sizes of 3 by 3, 5 by 5, and 7 by 7 m. The variable window sizes assigned to each pixel were based on the semivariance range or local breaks in slope.

The LM technique used for this study operates with two shapes of the search window, such as a square *n* by *n* window and a circular window that is more appropriate for identifying tree crowns. The algorithm was implemented in IDL Version 5.5 (Research Systems, Inc.). The LM technique operates on the assumption that high laser values in a spatial neighborhood represent the tip of a tree crown. Successful identification of the tree location using the LM technique depends on the careful selection of the filter window size. If the filter size is too small or too large, errors of commission or omission respectively, occur. Thus, the variable window LM technique functions under the supposition that there are multiple tree crown sizes and that the moving LM filter should be adjusted to an appropriate size that corresponds to the spatial structure found on the lidar image and on the ground. The LM filter works best for trees with a single well-defined apex, such as coniferous species.

Tree crown form has been associated with different geometric shapes, such as conical, parabolic, ellipsoidal, or combinations of such geometric shapes. Although the form of a tree crown does not follow exactly a Euclidean geometric shape, but is stochastic in nature (Biging and Gill, 1997) and, when seen from above, the tree crown most closely can be projected within a circle. Doruska and Burkhart (1994) investigated the circular distribution of branches for crowns of loblolly pine trees and found that in most cases a circular uniform distribution was appropriate. Therefore, it is evident that searching for the LM to identify individual crowns with a circular window of variable diameter is more appropriate than filtering with a square window.

The derivation of the appropriate window size to search for tree tops is based on the assumption that there is a relationship between the height of the trees and their crown size. The higher the trees, the larger the crown size. Thus, tree height and crown size data from the field inventory were used to derive a relationship between tree height and crown size. Crown size was considered the dependent variable, and linear

TABLE 5. ACCURACY ASSESSMENT REPORT FOR THE MAXIMUM-LIKELIHOOD CLASSIFICATION OF THE ATLAS IMAGE

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy %	Users Accuracy %	Kappa Statistic
Open ground	7	11	6	85.71	54.55	0.5380
Deciduous	133	134	97	72.93	72.39	0.6011
Pines	292	287	255	87.33	88.85	0.6559
Total	432	432	358			

Overall classification accuracy = 82.87%
Overall Kappa statistic = 0.6236

and nonlinear regression models were tested separately for deciduous trees, pines, and the combined data. Data on height and crown diameter from 235 pines and 189 deciduous trees were used in this analysis. The best prediction for crown width (m) using tree height (H m) was obtained when using linear regression with a quadratic model as shown below:

$$\textit{Deciduous:}\quad \textit{Crown width} = 3.09632 + 0.00895\,H^2 \quad (R^2 = 0.54,\ S_{y\cdot x} = 1.49) \qquad (1)$$

$$\textit{Pines:}\quad \textit{Crown width} = 3.75105 - 0.17919\,H + 0.01241\,H^2 \quad (R^2 = 0.58,\ S_{y\cdot x} = 1.20) \qquad (2)$$

$$\textit{Combined:}\quad \textit{Crown width} = 2.51503 + 0.00901\,H^2 \quad (R^2 = 0.59,\ S_{y\cdot x} = 1.45) \qquad (3)$$

As expected, using only height as the predictor variable, the relationship is not as strong as between dbh and height, but it offers a base to continuously vary the LM filter size when moved across the grid of laser height values. The regression models are different for pines and deciduous trees, because height proved to be non-significant at the 0.05 level in the regression model for deciduous trees. The regression model for pines had a higher R^2 value and a reduced standard error of the estimate when compared with the deciduous model. Consequently, it is advantageous to differentiate between deciduous trees and pines when relating lidar heights with window size for the LM filter.

Based on the CHM heights and Equations 1, 2, and 3 above, the window size varied between 3 by 3 and 31 by 31 pixels, which corresponds to crown sizes between 1.5 m and 15.5 m. The maximum crown diameter measured on the ground, 13.8 m, belonged to a white oak tree. In the case of the circular window for the LM filter (Figure 4), the window diameter varied between the same limits mentioned above for the size of the regular square windows.

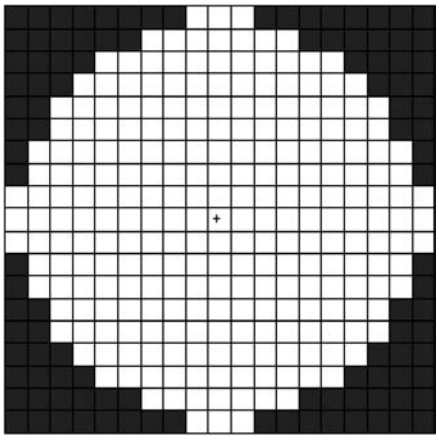


Figure 4. Circular window (white background) compared to a square window (19 by 19 pixels—9.5 by 9.5 m).

The algorithm (Figure 5) reads the height value at each pixel and calculates the window size to search for the local maximum. The concept of variable windows is illustrated in Figure 6, which shows a portion of the CHM with the filtering windows that identified tree tops, with either square or circular shape. When filtering the image with variable windows, if the current pixel corresponds to the local maximum, it is flagged as a tree top (Plate 4). Once the location of each identified tree crown has been established, the canopy 3D surface of laser heights (CHM) is sampled only at the positions of the tree

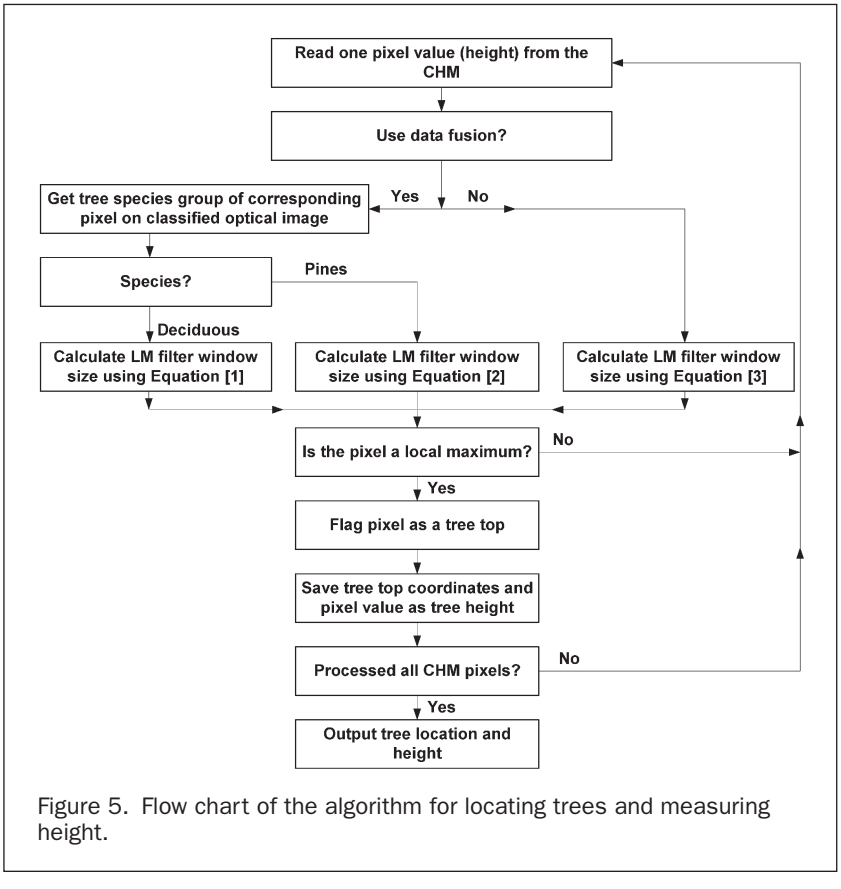
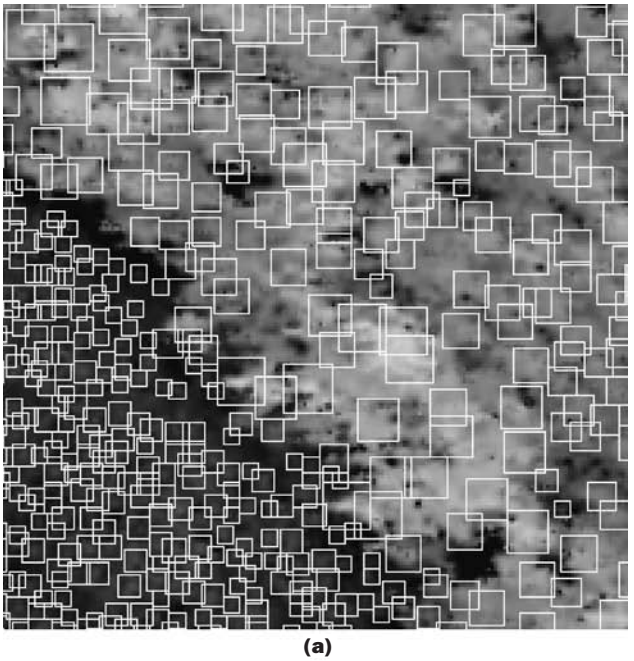
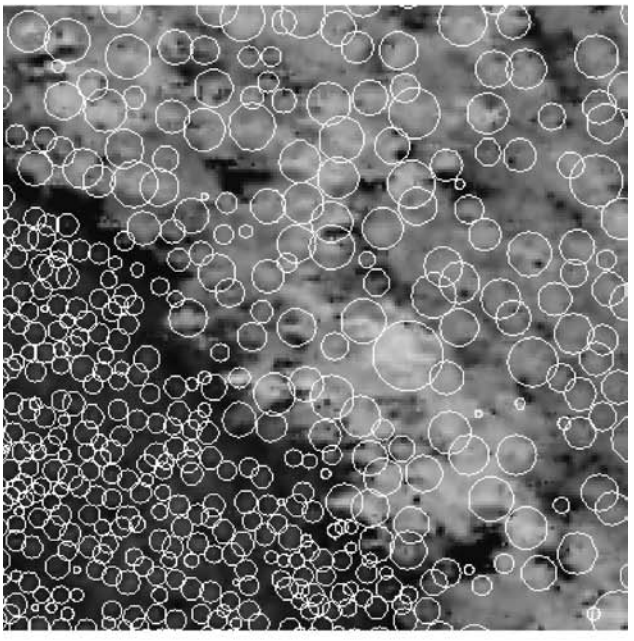


Figure 5. Flow chart of the algorithm for locating trees and measuring height.



(a)



(b)

Figure 6. Portion of the CHM with variable windows of square (a) and circular (b) shape. Only the windows that identified tree tops are shown in the figure.

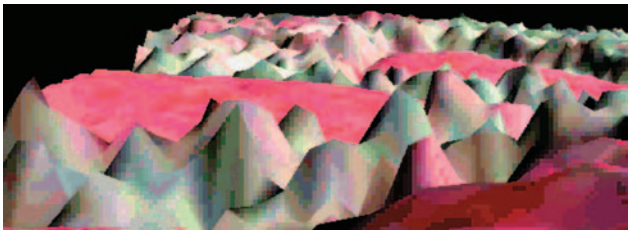
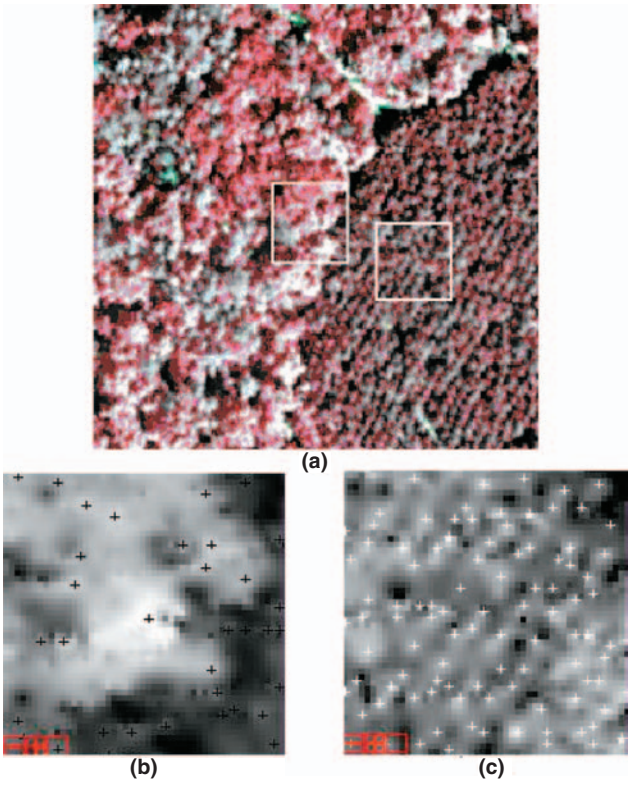


Plate 3. Multispectral ATLAS image draped over the CHM.



(a)

(b)

(c)

Plate 4. Ortho-image (a) and tree tops identified in the hardwood stand (b) and the pine plantation (c). Rectangles on the ortho-image shows approximate location of zoom windows (b) to the left, and (c) to the right. Plantation row pattern oriented SW-NE is visible in (a) and (c). The ortho-image was provided by EarthData, Inc., derived from 1:13,800-scale color-infrared photography acquired by NASA in the fall of 1999 (0.5-m spatial resolution).

apex to find out the height of each tree. To avoid identifying local maxima, i.e., trees, in areas with low vegetation heights, a minimum threshold was used to flag a location as a tree top. The threshold value was set to the minimum tree height inventoried on the ground (3.96 m).

The algorithm was run with both circular and square window LM filters, with and without data fusion with optical data. When no optical data were used to differentiate between deciduous and pines when calculating the width of the search window, the filter size was calculated based on the relationship between height and crown size derived from all inventoried trees (Equation 3).

Regression Analysis

The algorithm for identifying the location of individual trees was used in conjuction with an algorithm for measuring the crown diameter of each identified tree. The crown diameter is the average of two values measured along two perpendicular directions from the location of the tree top. Details about the steps of this algorithm and its performance are given in Popescu *et al.* (2003). The variable window size LM technique that identifies tree tops was also used to estimate the number of trees per plot. The total number of local maxima within one plot is an indicator of the number of stems per plot. Linear regression models were used to develop equations relating lidar-derived parameters, such as tree height, stand density,

TABLE 6. REGRESSION VARIABLES

Independent Variables (lidar measured)	Predicted Variables (field measured)
Tree height • Average height/subplot • Minimum height • Maximum height • Standard deviation of individual tree heights	Tree height • Average tree height/subplot • Maximum height
Crown diameter • Average crown diameter/subplot • Minimum crown diameter • Maximum crown diameter • Standard deviation of individual tree crown diameters	
Number of trees	

and crown width, with field inventory data of tree height for each of the FIA subplots. Subplots were pooled together in two categories, deciduous trees and pines. Stepwise multiple regression models with 0.15 significance level were developed separately for each of the two forest type categories. The independent variables (Table 6) were the lidar measurements for each subplot, including the number of trees, average height, minimum and maximum height, average crown diameter, minimum and maximum crown diameter, and the standard deviation of height and crown diameter. Lidar measurements were obtained for each of the four methods of filtering the CHM—square and circular variable windows, each with and without data fusion. Each set of lidar estimates was compared to the same set of field measurements for each FIA subplot, which includes mean and maximum height (Table 2).

The study of Popescu *et al.* (2002) confirmed that lidar is better suited to measure trees in the upper layer of the canopy, mainly the dominant and codominant trees. Therefore, the field-measured dependent variables for height, crown diameter, dbh, and number of trees were separated into three categories, based on the dbh: (1) all trees inventoried on the ground (includes trees with a dbh larger than 6.35 cm or 2.5 inch), (2) all trees traditionally measured using FIA standards (trees with dbh larger than 12.7 cm or 5.0 inch), and (3) dominant and codominant trees (trees with dbh larger than the quadratic mean diameter). Intermediate and overtopped trees, with small values for dbh and height, have a small contribution to the total subplot volume and biomass, and thus, ground measured volume, basal area, and biomass were not separated into the three categories above. Instead, these values were calculated using all ground-inventoried trees.

The presence of multicollinearity effects was investigated using eigenvalues and eigenvectors of the correlation matrices. Multicollinearity can be measured in terms of the ratio of the largest to the smallest eigenvalue, which is called the condition number of the correlation matrix (Myers, 1990, p. 370). A condition number that exceeds 1,000 raises concerns for multicollinearity effects. The highest condition number of all the regression models for each of the biophysical parameters, for both pines and deciduous data, was equal to 305.99, considerably lower than the value that raises concerns, i.e., 1000 (Myers, 1990, p. 370).

Because the ground-truth data were split into pine and deciduous plots, it was not practical to split it again for validation purposes. Therefore, the PRESS statistic (Prediction Sum of Squares) was used as a form of cross-validation, very much in the spirit of data splitting (Myers, 1990, p. 171–178). To calculate the PRESS statistic, one observation, in this case one subplot ground value, is set aside from the sample, and the remaining observations are used to estimate the coefficients for a particular candidate model. Each observation is therefore removed one at a time and the model is fit n times, n being the number of observations in the data set. The observation set aside is predicted each time, resulting in n prediction errors or PRESS residuals. These residuals are true prediction errors, because one observation is not simultaneously used for fit and model assessment. The PRESS statistic was calculated for the models obtained for each of the four filtering methods. In addition, the range of PRESS residuals, their mean, and their standard deviation are reported for each model. For the choice of the best model, one might favor the model with the smallest PRESS.

Identification of Outlying Observations

Maximum tree heights measured with lidar and on the ground for the same subplots are not affected by the number of trees "seen" on the lidar CHM, unlike, for example, the average height. Therefore, the maximum height per plot is a good indicator of the correspondence between the two sets of measurements, lidar and ground. Maximum height is only affected by the inclusion of the highest tree on the subplot within the subplot boundaries that tie the ground and lidar measurements. Therefore, it mitigates the positional errors of both the field and lidar data.

Linear regression with the ground maximum height as the dependent variable was used to identify outliers or observations that are well separated from the remainder of the data. Such observations involve large residuals and have a dramatic effect on the fitted least-squares regression model, not only for regressing maximum height, but also for the rest of the estimated parameters. Externally studentized residuals, also called R-Student, (Neter *et al.*, 1983, pp. 406–407; Montgomery and Peck, 1992, pp. 174–177) were used for the outliers diagnostic. The R-student residuals follow the t distribution with $n - p - 1$ degrees of freedom, where n is the number of observations and p is the number of regression parameters in the model, including the intercept term. Tail areas of 0.05 on each side of the t distribution were considered extreme; therefore, absolute values of the R-Student residuals were compared with $t\,(.95, n - p - 1)$.

Large differences between maximum tree heights can occur due to misregistration between the lidar CHM and the FIA-type plots located with GPS. Also, very large trees located in the plot neighborhood may overtop inventoried trees. Besides, their top could be identified on the lidar CHM as being inside the plot. Errors in the derivation of the CHM and the terrain DEM can also lead to large differences between the lidar and ground measurements. Due to the size of the FIA-type subplots (0.017 ha or 0.04 acres), large differences between lidar and ground measured tree heights are more likely to occur in stands with a complex vertical and horizontal canopy structure, such as the deciduous stands. Outliers were also investigated by analyzing the CHM and the ground data to gather nonstatistical evidence for discarding extreme values.

Results and Discussion

Outlying Observations

All four filtering methods—circular window with and without data fusion and square windows with and without data fusion—gave similar results with respect to the residual analysis. Residuals and R-Student residuals were investigated when regressing maximum height for both pines and deciduous plots against all lidar-derived variables. For one of the pine plots, the lidar processing was not able to identify any trees; therefore, the results for the pine plots are reported for only 30 plots. The pine plot with no lidar trees has five dominant trees ranging in height from 21.95 to 34.75 m, with an

average of 29.3 m. The corresponding filtering window sizes for these tree heights (Equation 1) range between 5.55 and 12.27 m, with an average of 9.12 m. Such window sizes for LM filtering are large when compared to the plot radius (7.32 m) and could lead to errors of tree top identification over the plot area.

As explained in the previous section, the R-Student residuals follow a *t* distribution; therefore, the *t* statistic can be used to ascertain outlying values. Tail areas of 5 percent on were considered extreme, and absolute values of the R-Student residuals were compared with $t_{95\%,27} = 1.703$ for pine plots and $t_{95\%,30} = 1.697$ for deciduous plots. By using this formal procedure for outlier detection employing hypothesis testing based on the *t* distribution, four of the pine plots and three of the deciduous plots were identified as outliers. The magnitude of the residuals for pine and deciduous outliers are very different. While pine outliers have absolute values for residuals of about 2 m, deciduous outliers have much larger absolute values, between 7.5 and 11.8 m. To investigate the influence of outliers on the regression models for both pines and deciduous plots, new regression equations for the maximum height were fitted with outliers deleted from the data set. For the method of filtering with circular windows and data fusion, a comparison of the summary statistics from the two models, with and without outliers, for the pines and deciduous plots are given in Table 7. The other three methods of processing the CHM gave similar results when comparing regression models with and without outliers.

Deleting the outliers from the pines data set had almost no effect on explaining the variance associated with the maximum height. The increase in R^2 is almost negligible, with only a slight reduction in the standard error of the estimate. There was, however, a significant increase in the R^2 value for the deciduous data set along with a substantial reduction in the RMSE. Figures 7a and 7b show the plots of residuals versus the predicted maximum height for pines and the normal probability plot of the residuals. These plots do not indicate any serious departures from the normality assumption. The points on the normal probability plot lie approximately on a straight line, while for the deciduous data set with outliers (Figure 7d), they indicate a skewed distribution. For the deciduous data set with outliers out, the range of residuals decreased considerably and the normal probability plot (Figure 7f) indicates a closer approximation of normality.

The two deciduous plots that had large negative residuals, i.e., fitted height much larger than the actual maximum height measured on the ground, had very large trees right next to the plot. The radius for the FIA-type subplots is 7.32 m and one of the plots (ground maximum height 10.97 m) had a large mockernut hickory tree (*Carya tomentosa* (Poir.) Nutt) with a height of 18.90 m at 7.56 m from the plot center and a southern red oak (*Quercus falcata* Michx.) with a height of 27.74 m at 9.20 m from the plot center. A similar situation was found on the second plot (ground maximum height of 17.07 m) that had a white oak (*Quercus alba* L.) with a height of 25.30 m at 7.60 m from the plot center and chestnut oak (*Quercus prinus* L.) with a height of 23.16 m at 8.35 m, respectively. Such tall trees located next to the plot have large crowns extending over the plot and can have their top vertically located inside the plot boundary. The third outlier plot had a large positive residual with the lidar-measured maximum height being lower than the ground observed height. The reason why lidar failed to measure the tallest tree on this plot is not apparent. However, the average height of the dominant-codominant trees on this plot is 27.3 m, while the lidar-measured maximum height is 27.87 m. The situation might be explained by inaccuracies in the lidar DEM. The vegetation profile of the plot estimated on the ground reveals a cover height of 3.65 m for 35 percent of the subplot area that, along with the dense overstory canopy, might prevent laser pulses to reach the ground.

TABLE 7. SUMMARY STATISTICS FOR REGRESSING MAXIMUM HEIGHT WITH AND WITHOUT OUTLIERS

	With Outliers In		With Outliers Out	
Statistics	Pines	Deciduous	Pines	Deciduous
R^2	0.98	0.36	0.99	0.61
RMSE (m)	0.99	3.71	0.68	2.33

Despite the conclusion of the statistical testing for outliers of the pines data set, the examination of residuals ranges and normality plots fails to reveal strong reasons for discarding the four subplots from further analysis. Therefore, the subsequent results presented for the pines data set were obtained by using all 30 subplots. Linear regression with stepwise elimination and 0.15 significance level was used to predict subplot-level average tree tree height measured on the ground.

For the deciduous data set, the residuals analysis and ground data investigations offer a robust motivation for discarding the three subplots from subsequent analysis. The results that follow were obtained after removing the three deciduous subplots.

Tree Height

The current research results show that lidar can accurately estimate tree height, which is one of the key parameters in forest inventories. These findings are particularly important, taking into account that height measurements on the ground are considered more difficult and costly to collect than dbh, especially in tall dense stands. As a result, some forest inventories measure all trees on the plot for dbh and subsample for heights. For this study, all four processing methods, such as square and circular filtering windows, each with and without data fusion, were able to explain a high percentage of the variance associated with the average tree height. Results show a rather intuitive behavior for both pines and deciduous data sets (Table 8) by obtaining better R^2 values for estimating the height of dominant and codominant trees and for trees measured by FIA standards, i.e., trees with a dbh larger than 12.7 cm (5 in). The upper layer of dominant trees intercepts most of the laser shots, and thus, estimates better correlate with their mean height. Part of the unexplained variance could be attributed to the terrain DEM, to the collocation of the lidar CHM and field subplots, and to the lidar limitations for constructing an accurate CHM.

For pines, the LM technique with a variable window size of circular shape gave the best results with data fusion. This method explained 97 percent of the variance associated with the mean height of dominant trees with a small standard error for the estimate (1.14 m). The PRESS statistic (Table 9) for this method is almost three times smaller than the one obtained when performing the LM filtering with squared windows. The PRESS residuals have a standard deviation of 1.33 m. Næsset and Økland (2002) report a standard deviation of PRESS residuals for individual tree heights of 3.15 m. They conducted their study in an uneven-aged spruce forest.

Results for estimating mean height for pines show that the LM window shape plays the most important role in the accuracy of measuring height. The use of circular windows of variable radius for identifying tree tops brings an 11 percent improvement in R^2 values for the height of all trees measured on the plots (from 0.75 to 0.85) and 7 percent for the height of dominant trees (from 0.90 to 0.97). The cross-validation revealed that filtering with circular windows, without data fusion, gave the best prediction for mean height of all trees. The PRESS statistic in this case is less than half of the value

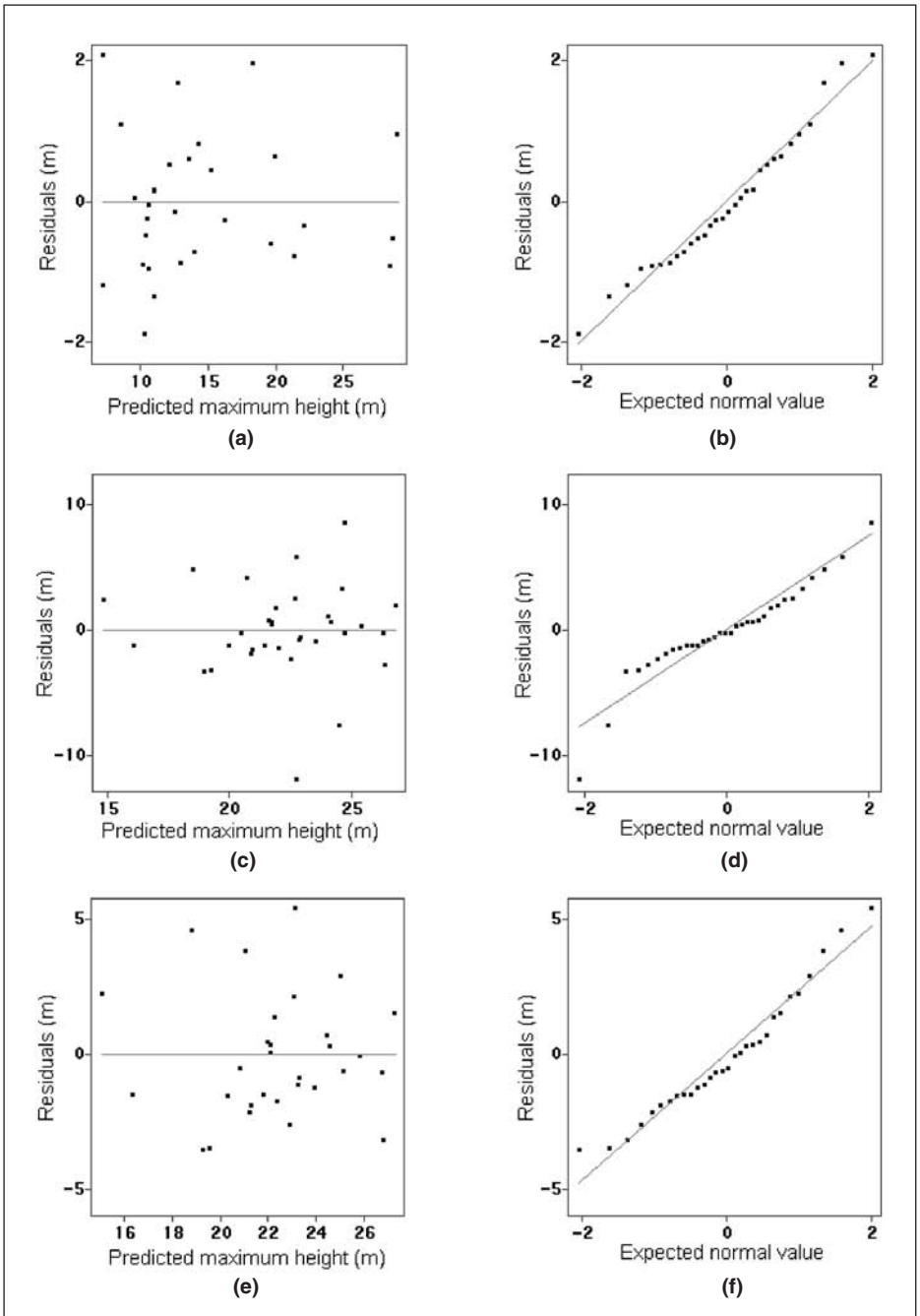


Figure 7. Plot of residuals versus fitted values (a) and normal probability plot (b) for pines, with outliers in. (c) Plot of residuals versus fitted values and (d) normal probability plot for deciduous plots, with outliers in. (e) Plot of residuals versus fitted values and (f) normal probability plot for deciduous plots, with outliers out.

obtained by the model for filtering with squared windows. The gain in explaining the mean height variance for trees measured by FIA standards (dbh greater than 12.7 cm or 5 in) with circular LM filters is not that substantial, but all methods provide R^2 values above 0.94. Data fusion and circular LM filters brought the standard error for estimating pines mean height with the FIA threshold down to 1.07 m, for an R^2 value of 0.95. The independent variables included in the regression model in this case were the number of trees, maximum height, and maximum crown diameter. Parameter estimates are positive for maximum height and crown diameter, while for the number of trees the parameter estimate is negative—the larger the trees, the fewer there are.

Results were different for deciduous plots that have a very complex horizontal and vertical structure. The square LM filter performed better overall, though for the dominant trees the difference between the two window shapes was small. Regression models explained 79 percent of the mean height variance for dominant trees, with a 1.91-m root-mean-squared error (RMSE). Data fusion only proved to be useful for assessing the height of dominant trees. The cross-validation indicated better model prediction for filtering with square windows,

TABLE 8. REGRESSION RESULTS—DEPENDENT VARIABLE: AVERAGE HEIGHT (M)/SUBPLOT

Trees	Method*	Significant Independent Variables**	$S_{y.x}$	R^2	Model***
		Pines			
Dominants	SQ	H_{max}	1.87	0.90	$0.82647 + 0.80818H_{max}$
	SQF	H_{max}	1.87	0.90	$0.82647 + 0.80818H_{max}$
	CW	H_{max}, CD_{ave}, CD_{std}	1.26	0.96	$-0.77630 + 0.57257H_{max} + 0.91903CD_{ave} + 2.38845CD_{std}$
	CWF	H_{ave}, H_{min}, H_{std}, CD_{ave}	1.14	0.97	$-0.21885 + 0.78538H_{ave} - 0.37937H_{min} + 0.50150H_{std} + 1.78613CD_{ave}$
All	SQ	H_{ave}, CD_{std}	1.50	0.74	$4.52287 + 0.37786H_{ave} + 1.14151CD_{std}$
	SQF	H_{ave}, CD_{std}	1.48	0.75	$4.31588 + 0.39436H_{ave} + 1.16062CD_{std}$
	CW	H_{ave}, CD_{std}	1.14	0.85	$3.93120 + 0.39505H_{ave} + 1.64873CD_{std}$
	CWF	H_{ave}, CD_{std}	1.18	0.84	$4.20656 + 0.37934H_{ave} + 1.48877CD_{std}$
FIA standard	SQ	H_{ave}, H_{max}, CD_{max}	1.18	0.94	$3.31532 + 0.23535H_{ave} + 0.57874H_{max} - 0.54824CD_{max}$
	SQF	H_{ave}, H_{max}, CD_{max}	1.11	0.95	$3.79115 + 0.22589H_{ave} + 0.60009H_{max} - 0.70397CD_{max}$
	CW	H_{max}, CD_{max}	1.09	0.95	$1.08697 + 0.55224H_{max} + 0.69952CD_{max}$
	CWF	N, H_{max}, CD_{max}	1.07	0.95	$3.52918 - 0.08886N + 0.48535H_{max} + 0.62258CD_{max}$
		Deciduous			
Dominants	SQ	H_{max}, CD_{ave}	2.06	0.76	$4.20097 + 0.83107H_{max} - 0.58420CD_{ave}$
	SQF	H_{max}, CD_{ave}	1.91	0.79	$3.67791 + 0.82446H_{max} - 0.41608CD_{ave}$
	CW	H_{max}	2.08	0.73	$4.22851 + 0.68477H_{max}$
	CWF	H_{max}	2.07	0.74	$4.39941 + 0.67972H_{max}$
All	SQ	CD_{max}	1.22	0.50	$8.20149 + 0.66389CD_{max}$
	SQF	H_{ave}	1.58	0.40	$8.36490 + 0.22249H_{ave}$
	CW	H_{max}, CD_{ave}	1.60	0.37	$8.24679 + 0.28781H_{max} - 0.33682CD_{ave}$
	CWF	CD_{max}	1.58	0.36	$8.26377 + 0.66760CD_{max}$
FIA standard	SQ	H_{max}, CD_{max}	1.48	0.75	$7.31106 + 0.63421H_{max} - 0.62351CD_{max}$
	SQF	H_{max}, CD_{max}	1.57	0.74	$6.68390 + 0.64977H_{max} - 0.55472CD_{max}$
	CW	H_{max}, CD_{max}	1.57	0.71	$7.41716 + 0.53921H_{max} - 0.33535CD_{max}$
	CWF	H_{max}	1.66	0.66	$7.01565 + 0.45870H_{max}$

*Method refers to LM filtering technique: SQ (square window), SQF (square window with data fusion), CW (circular window), and CWF (circular window with data fusion).
**H_{ave}, average height of all lidar identified trees per plot; H_{min}, minimum height; H_{max}, maximum height; H_{std}, height standard deviation; CD_{ave}, average crown diameter; CD_{min}, minimum crown diameter; CD_{max}, maximum crown diameter; CD_{std}, crown diameter standard deviation; and N, number of trees.
***All units, except for the number of trees, are meters (m).

with standard deviations of PRESS residuals between 1.30 and 2.20.

As explained when documenting outliers in the previous section, maximum height gives an indication of how well the CHM portrays vegetation height over one plot. The circular LM filter gave very accurate results for pines (Table 10). This method explained 98 percent of the variance with a sub-meter standard error of the estimate using only lidar-measured maximum height as the independent variable. For deciduous plots, the best R^2 value was 0.69 (RMSE 2.07 m).

Figures 8a and 8b show scatterplots of lidar-measured versus field-measured height and observed versus predicted values for height, for the pine and deciduous models that gave the best results for estimating ground-measured height.

Comparison between Processing Techniques

Processing techniques were compared based on two statistics, the variance of field-based estimates that each regression model was able to explain and the PRESS statistic. The conclusions are not surprising. Filtering for local maximum with circular windows gives better fitting models for pines, because one would expect that a circular window shape is more appropriate for identifying individual tree crowns. For deciduous trees, filtering with square windows provided a slightly better model fit.

All pine regression models, with only one exception when estimating average height for all trees, proved to explain a higher percentage of the variance associated with field-measured average height when the size of the filtering

TABLE 9. PRESS STATISTICS FOR PREDICTING AVERAGE HEIGHT (M)/SUBPLOT

Trees	Method*	PRESS	Range of PRESS Residuals		Mean of PRESS Residuals	Standard Deviation of PRESS Residuals
			Min	Max		
			Pines			
Dominants	SQ	141.66	−8.22	5.50	0.00	2.21
	SQF	141.66	−8.22	5.50	0.00	2.21
	CW	90.53	−5.23	3.33	−0.05	1.77
	CWF	51.62	−2.53	2.13	−0.01	1.33
All	SQ	119.35	−4.87	7.14	0.13	2.02
	SQF	108.41	−5.16	6.35	0.10	1.93
	CW	52.94	−2.30	3.24	0.04	1.35
	CWF	58.22	−2.41	3.77	0.04	1.42
FIA standard	SQ	63.85	−4.29	3.48	0.07	1.54
	SQF	55.54	−4.10	3.42	0.04	1.43
	CW	48.39	−3.48	2.24	−0.08	1.34
	CWF	49.40	−3.04	3.21	−0.03	1.35
			Deciduous			
Dominants	SQ	120.26	−3.73	4.52	0.01	2.19
	SQF	109.68	−3.87	4.19	0.01	2.05
	CW	139.99	−4.08	4.69	0.01	2.20
	CWF	138.70	−4.10	4.64	0.01	2.19
All	SQ	42.31	−2.35	1.79	−0.03	1.30
	SQF	75.15	−4.70	4.71	−0.02	1.70
	CW	120.71	−4.80	6.53	0.10	2.04
	CWF	79.77	−3.06	5.25	−0.03	1.66
FIA standard	SQ	63.97	−2.19	3.20	0.00	1.60
	SQF	73.15	−2.38	3.36	−0.02	1.68
	CW	78.82	−2.29	3.37	−0.00	1.65
	CWF	88.58	−2.60	3.54	0.01	1.75

*Method refers to LM filtering technique: SQ (square window), SQF (square window with data fusion), CW (circular window), and CWF (circular window with data fusion).

TABLE 10. REGRESSION RESULTS—DEPENDENT VARIABLE: MAXIMUM HEIGHT (M)/SUBPLOT*

Trees	Method	Significant Independent Variables	$S_{y \cdot x}$	R^2	Model
			Pines		
All	SQ	H_{max}, CD_{max}	0.91	0.98	$1.42214 + 1.02681H_{max} - 0.54600CD_{max}$
	SQF	H_{max}, CD_{max}	0.91	0.98	$1.58459 + 1.02089H_{max} - 0.56401CD_{max}$
	CW	H_{max}, CD_{std}	0.96	0.98	$0.41051 + 0.89765H_{max} + 0.60232CD_{std}$
	CWF	H_{max}	0.99	0.98	$0.54189 + 0.92320H_{max}$
			Deciduous		
All	SQ	H_{max}	2.24	0.64	$9.35914 + 0.56096H_{max}$
	SQF	H_{max}	2.07	0.69	$8.75699 + 0.59639H_{max}$
	CW	H_{max}	2.37	0.60	$9.19615 + 0.57294H_{max}$
	CWF	H_{max}	2.33	0.61	$9.21868 + 0.57395H_{max}$

*Method and variable abbreviations are the same as in Table 8.

windows was calibrated for the tree species groups, i.e., when using data fusion in conjunction with the lidar processing techniques. For deciduous plots, all regression models for estimating field-based average height, with one exception when estimating the average height of the dominants, had a better fit without using optical data. Still, the optical data (ATLAS multispectral imagery) has a spatial resolution of 4 m, while the lidar CHM has a grid size of 0.5 m. Previous lidar studies (e.g., Maclean and Krabill, 1986; Nelson *et al.*, 1988b; Næsset 1997b) reached the conclusion that, prior to fitting regression models for estimating forest parameters, it is necessary to differentiate between forest types. Therefore, it is expected to obtain a better fit for regressing field estimates when high spatial and spectral resolution optical data are used to differentiate between forest types in the processing phase of the lidar data. For practical forestry application of lidar, existing maps of forest types can be used to distinguish between forest types. However, coregistered optical data with a spatial resolution comparable to the lidar sampling density can be used not only for calibrating the lidar filtering window size, but also in the process of deriving the ground DEM and the lidar CHM.

The cross-validation showed the same situation as the R^2 values with respect to the shape of the LM filtering windows for pines. All pine models provided smaller PRESS residuals when the lidar estimates were obtained by identifying individual trees with circular search windows. While pine models

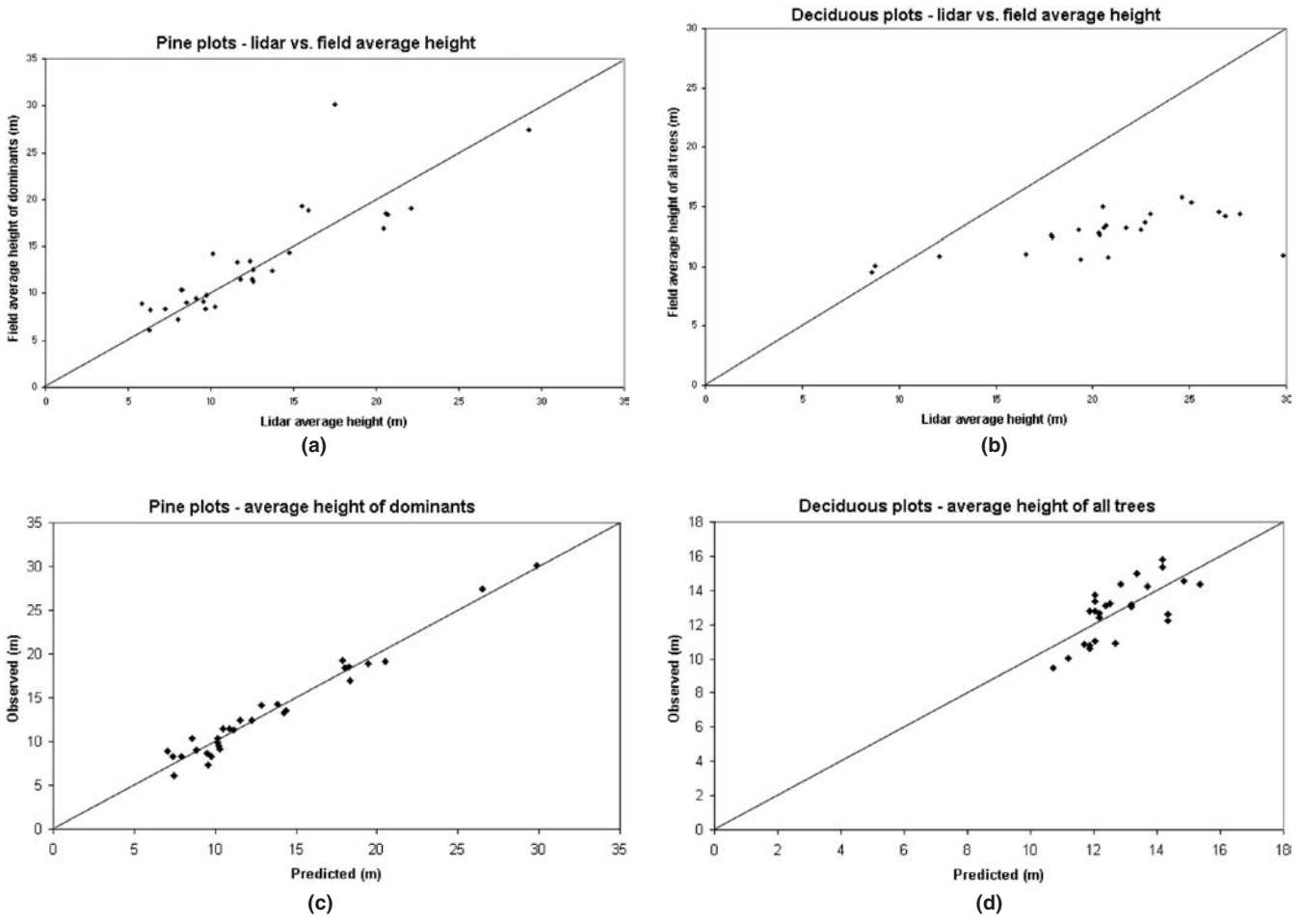


Figure 8. Scatterplots of predicted versus observed and lidar versus field height values.

work better when using data fusion, all deciduous models, with the exception of one model for dominant trees, better predicted field estimates without using optical data.

Conclusions

To conclude, using circular filtering windows to locate individual trees and optical data to differentiate between forest types provides better results for estimating biophysical parameters for pines. Given the spatial resolution of the optical data used for this study, estimating forest parameters for deciduous plots seems to give superior results without calibrating the search window size based on forest type.

The results of the current study show that lidar data could be used to accurately estimate plot-level tree height by focusing on the individual tree level. The generation of individual tree crown forest inventories from high spectral and spatial resolution imagery, although still a research subject, is coming of age (Gougeon *et al.*, 2001). In this context, lidar proves to be the best suited technology to derive accurate models of the terrain elevation and measure the height of the dominant and codominant trees in the forest canopy.

Overall, this research proved that small footprint airborne lidar data in conjunction with spatially coincident optical data are able to accurately predict tree heights of interest for forest inventory and assessment. The main objective of this research was to develop robust processing and analysis techniques to facilitate the use of lidar data for predicting tree height by focusing on the individual tree level. The algorithm used for measuring forest height provides individual tree heights for the entire forested area covered by lidar. These results have profound implications in forest management, because tree height in relation to tree age has been found the most practical, consistent, and useful indicator of site quality. In forestry, site index is estimated by determining the average total height and age of dominant and codominant trees in even-aged stands. For pine plantations and even-aged stands, stand age is commonly well documented. Much of the forest inventory data, including stand age, is available through GIS-stored maps and, by combining lidar-derived tree height and stand boundaries, site index can be mapped within stands. Therefore, seeing the trees in the forest and, more importantly, measuring them brings an important contribution to concepts such as precision forest inventory and automated data processing for forestry applications.

The integration with co-registered multi- and hyperspectral digital imagery makes lidar a realistic precision forestry alternative to traditional measurements in forest inventory. Even without the same high spatial resolution as the lidar data, optical data used for this study demonstrated the ability of data fusion to improve the estimates of tree height, especially for the pine plots. Lidar and image data fusion can bring dramatic gains in characterizing the three-dimensional structure of the forest canopy, and it would accelerate the transition of lidar applications from scientific interests to reliable

commercial implementations. An ideal system would incorporate lidar and optical data for species recognition and tree measurements. Future investigations could consider using high spatial resolution multi- or hyperspectral data not only for species group identification, but also for processing lidar data for vegetation removal, individual tree location, and crown measurements. The focus of this research on the individual tree level and the innovative processing techniques, mainly the variable-radius circular window used for tree top filtering with optical data fusion, demonstrates that airborne laser scanners can reliably measure tree height.

Acknowledgments

We gratefully acknowledge the help provided with the field data collection by Dr. John Scrivani at the Virginia Department of Forestry; Jan van Aardt and Rebecca Musy at Virginia Tech; Neil Clark at the USDA Forest Service; Jared Wayman at Questerra, Inc.; Karsten Nitsch at the University of Berlin; FIA crew members at the Virginia Department of Forestry; and Wayne Bowman, David Houttekier, Donald Jamerson, and Ralph Totty at the Appomattox-Buckingham Forest Office. This research has been supported by the NASA Earth System Science Fellowship Program (NGT5-30198), NASA Laboratory for Terrestrial Physics, NCASI, McIntire-Stennis research program (VA-136589), Virginia Tech Department of Forestry, and USDA Fund for a Rural America (97-36200-5231).

References

Avery, T.E., and H.E. Burkhart, 1994. *Forest Measurements, Fourth Edition*, McGraw-Hill, Inc., New York, N.Y., 408 p.

Biging, G.S., and S.J. Gill, 1997. Stochastic models for conifer tree crown profiles, *Forest Science*, 43(1):25–34.

Blair, B.J., D.L. Rabine, and M.A. Hofton, 1999. The Laser Vegetation Imaging Sensor: A medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography, *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2–3):115–122.

Brandtberg, T., 1997. Towards structure-based classification of tree crowns in high spatial resolution aerial images, *Scandinavian Journal of Forest Research*, 12:89–96.

Campbell, J.B., 1996. *Introduction to Remote Sensing, Second Edition*, The Guilford Press, New York, N.Y., 622 p.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, Inc., New York, N.Y., 137 p.

Daley, N.M.A, C.N. Burnett, M. Wulder, L.O. Niemann, and D.G. Goodenough, 1998. Comparison of fixed-size and variable-sized windows for the estimation of tree crown position, *Proceedings IGARSS 1998*, 06–10 July, Seattle, Washington (IEEE), pp. 1323–1325.

Doruska, P.F., and H.E. Burkhart, 1994. Modeling the diameter and locational distribution of branches within the crowns of loblolly pine trees in unthinned plantations, *Canadian Journal of Forest Research*, 24:2362–2376.

Eyre, F.H., 1980. *Forest Cover Types of the United States and Canada*, Society of American Foresters, Bethesda, Maryland, 148 p.

Gougeon, F., 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images, *Canadian Journal of Remote Sensing*, 21(3):274–284.

Gougeon, F.A., B.A. St-Onge, M. Wulder, and D.G. Leckie, 2001. Synergy of airborne laser altimetry and digital videography for individual tree crown delineation, *Proceedings of the 23rd Canadian Symposium on Remote Sensing*, 21–24 August, Sainte-Foy, Québec, Canada (l'Association Québécoise de Télédétection), unpaginated CD ROM.

Harding, D.J., J.B. Blair, J.B. Garvin, and W.T. Laurence, 1994. Laser altimetry waveform measurement of vegetation canopy structure, *Proceedings of the International Geoscience and Remote Sensing Symposium—IGARSS '94*, 08–12 August, Pasadena, California (ESA Scientific & Technical Pub., Noordwijk, The Netherlands), 2:1251–1253.

Holmgren, J., M. Nilsson, and H. Olsson, 2003. Estimation of tree height and stem volume on plots using airborne laser scanning, *Forest Science*, 49(3):419–428.

Hyyppä, J., O. Kelle, M. Lehikoinen, and M. Inkinen, 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners, *IEEE Transactions on Geoscience and Remote Sensing*, 39(5): 969–975.

Hyyppä, J., and M. Inkinen, 2002. Detecting and estimating attributes for single trees using laser scanner, *The Photogrammetric Journal of Finland*, 18:43–53.

Lefsky, M.A., W.B. Cohen, S.A. Acker, T.A. Spies, G.G. Parker, and D. Harding, 1997. Lidar remote sensing of forest canopy structure and related biophysical parameters at the H.J. Andrews experimental forest, Oregon, USA, *Natural Resources Management Using Remote Sensing and GIS* (J.D. Greer, editor), American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, pp. 79–91.

Lefsky, M.A., D. Harding, W.B. Cohen, G. Parker, and H.H. Shugart, 1999. Surface lidar remote sensing of basal area biomass in deciduous forests of eastern Maryland, USA, *Remote Sensing of Environment*, 67:83–98.

Maclean, G.A., and W.B. Krabill, 1986. Gross-merchantable timber volume estimation using an airborne LIDAR system, *Canadian Journal of Remote Sensing*, 12(1):7–18.

Magnussen, S., and P. Boudewyn, 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators, *Canadian Journal of Forest Research*, 28:1016–1031.

Magnussen, S., P. Eggermont, and V.N. LaRiccia, 1999. Recovering tree heights from airborne laser scanner data, *Forest Science*, 45(3):407–422.

McCombs, J.W., S.D. Roberts, and D.L. Evans, 2003. Influence of fusing lidar and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation, *Forest Science*, 49(3):457–466.

Means, J.E., 2000. Comparison of large-footprint and small-footprint lidar systems: design, capabilities, and uses, *Proceedings: Second International Conference on Geospatial Information in Agriculture and Foretry*, 10–12 January, Lake Buena Vista, Florida (ERIM International), 1:85–192.

Means, J.E., S.A. Acker, D.J. Harding, J.B. Blair, M.A. Lefsky, W.B. Cohen, M.E. Harmon, and W.A. McKee, 1999. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon, *Remote Sensing of Environment*, 67:298–308.

Montgomery, D.C., and E.A. Peck, 1992. *Introduction to Linear Regression Analysis, Second Edition*, John Wiley & Sons, Inc., Hoboken, New Jersey, 527 p.

Myers, R.H., 1990. *Classical and Modern Regression with Applications, Second Edition*, The Duxbury Advanced Series in Statistics and Decision Sciences, PWS-Kent, Boston, Massachusetts, 488 p.

Næsset, E., 1997a. Determination of mean tree height of forest stands using airborne laser scanner data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 52:49–56.

_____, 1997b. Estimating timber volume of forest stands using airborne laser scanner data, *Remote Sensing of Environment*, 61(2):246–253.

_____, 2002. Determination of mean tree height of forest stands by digital photogrammetry, *Scandinavian Journal of Forest Research*, 17:446–459.

Næsset, E., and K.-O. Bjerknes, 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data, *Remote Sensing of Environment*, 78:328–340.

Næsset, E., and T. Økland, 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve, *Remote Sensing of Environment*, 79:105–115.

Nelson, R.F., W.B. Krabill, and G.A. Maclean, 1984. Determining forest canopy characteristics using airborne laser data, *Remote Sensing of Environment*, 15:201–212.

Nelson, R.F., R. Swift, and W. Krabill, 1988a. Using airborne lasers to estimate forest canopy and stand characteristics, *Journal of Forestry*, 86:31–38.

Nelson, R.F., W. Krabill, and J. Tonelli, 1988b. Estimating forest biomass and volume using airborne laser data, *Remote Sensing of Environment*, 24:247–267.

Nelson, R.F., R.G. Oderwald, and T.G. Gregoire, 1997. Separating the ground and airborne sampling phases to estimate tropical forest basal area, volume, and biomass, *Remote Sensing of Environment*, 60:311–326.

Neter, J., W. Waserman, and M.H. Kutner, 1983. *Applied Linear Regression Models*, Richard D. Irwin, Inc., Homewood, Illinois, 547 p.

Nilsson, M., 1996. Estimation of tree heights and stand volume using an airborne lidar system, *Remote Sensing of Environment*, 56:1–7.

Persson, A., J. Holmgren, and U. Soderman, 2002. Detecting and measuring individual trees using an airborne laser scanner, *Photogrammetric Engineering & Remote Sensing*, 68(9):925–932.

Popescu, S.C., 2002. *Estimating Plot-Level Forest Biophysical Parameters Using Small-Footprint Airborne Lidar Measurements*, Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 155 p.

Popescu, S.C., R.H. Wynne, and R.H. Nelson, 2002. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size, *Computers and Electronics in Agriculture*, 37(1–3):71–95.

———, 2003. Using lidar for measuring individual trees in the forest: an algorithm for estimating the crown diameter, *Canadian Journal of Remote Sensing*, 29(5):564–577.

Powell, D.S., J.L. Faulkner, D.R. Darr, Z. Zhou, and D.W. MacCleerry, 1993. *The Forest Resources in the United States, 1992*, General Technical Report RM-234, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, 132 p.

St-Onge, B.A., and F. Cavayas, 1995. Estimating forest stand structure from high resolution imagery using the directional variogram, *International Journal of Remote Sensing*, 16(11):1999–2021.

Toth, C.K., S. Berning, J. Leonard, and D.A. Grejner-Brzezinska, 2001. Integration of lidar data with simultaneously acquired digital imagery, *Proceedings: ASPRS 2001: Gateway to the New Millennium*, 23–27 April, St. Louis, Missouri (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), unpaginated CD-ROM.

USDA Forest Service, 2001. *Forest Inventory and Analysis National Core Field Guide, Volume 1: Field Data Collection Procedures for Phase 2 Plots, Version 1.5*, U.S. Department of Agriculture, Forest Service, Washington Office. Internal report, on file with: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, 201 14th St., Washington, DC 20250.

Weishampel, J.F., D.J. Harding, J.C. Boutet, Jr., and J.B. Drake, 1997. Analysis of laser altimeter waveforms for forested ecosystems of central Florida, *Proceedings, Advances in Laser Remote Sensing for Terrestrial and Oceanographic Applications*, 21–25 April, Orlando, Florida (SPIE, Bellingham, Washington), 3059:184–189.

Wulder, M., K.O. Niemann, and D.G. Goodenough, 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery, *Remote Sensing of Environment*, 73:103–114.

(Received 23 August 2002; accepted 26 November 2002; revised 03 June 2003)