Estimating Plot-Level Forest Biophysical Parameters Using Small-Footprint Airborne Lidar Measurements

by

Sorin C. Popescu

Dissertation Submitted to Virginia Tech in Partial Fulfillment of the Requirements of The Degree of

Doctor of Philosophy
in Forestry

Dr. Randolph H. Wynne, Chairman
Dr. Richard G. Oderwald
Dr. James B. Campbell
Dr. Stephen P. Prisley
Dr. Ross F. Nelson
Dr. John A. Scrivani

April 12, 2002
Blacksburg, Virginia

Keywords: lidar, canopy height model, forest inventory, DEM, biomass, volume

Estimating Plot-Level Forest Biophysical Parameters Using Small-Footprint Airborne Lidar Measurements

By

Sorin C. Popescu

Abstract

The main study objective was to develop robust processing and analysis techniques to facilitate the use of small-footprint lidar data for estimating forest biophysical parameters measuring individual trees identifiable on the three-dimensional lidar surface. This study derived the digital terrain model from lidar data using an iterative slope-based algorithm and developed processing methods for directly measuring tree height, crown diameter, and stand density. The lidar system used for this study recorded up to four returns per pulse, with 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, coniferous, and mixed stands of varying age classes and settings typical of the southeastern United States (37° 25' N, 78° 41' W). Lidar processing techniques for identifying and measuring individual trees included data fusion with multispectral optical data and local filtering with both square and circular windows of variable size. The window size was based on canopy height and forest type. The crown diameter was calculated as the average of two values measured along two perpendicular directions from the location of each tree top, by fitting a four-degree polynomial on both profiles. The ground-truth plot design followed the U.S. National Forest Inventory and Analysis (FIA) field data layout. The lidar-derived tree measurements were used with regression models and cross-validation to estimate plot level field inventory data, including volume, basal area, and biomass. FIA subplots of 0.017 ha each were pooled together in two categories, deciduous trees and pines. For the pine plots, lidar measurements explained 97% of the variance associated with the mean height of dominant trees. For deciduous plots, regression models explained 79% of the mean height variance for dominant trees. Results for estimating crown diameter were similar for both pines and deciduous trees, with R^2 values of 0.62-0.63 for the dominant trees. R^2 values for estimating biomass were 0.82 for pines (RMSE 29

Mg/ha) and 0.32 for deciduous (RMSE 44 Mg/ha). Overall, plot level tree height and crown diameter calculated from individual tree lidar measurements were particularly important in contributing to model fit and prediction of forest volume and biomass.

Acknowledgments

I have been very privileged to find inspiration in the academic and mentorship culture that my major professor, Dr. Randolph H. Wynne promotes within the Department of Forestry at Virginia Tech. I would like, therefore, to express my warmest thanks and gratitude to him for his professional and personal support, for his energy and commitment to university values that enlightened my years as a graduate student.

My heartfelt appreciations are also extended to the other members of my advisory committee, Dr. Richard G. Oderwald, Dr. John A. Scrivani, Dr. James B. Campbell, Dr. Stephen P. Prisley, and Dr. Ross F. Nelson, for their guidance, support, and constructive challenges they raised. They all made my graduate study the most enriching professional experience I have had thus far.

My thanks also go to Dr. Harold E. Burkhart, with whom I started corresponding almost ten years ago regarding my admission as a graduate student at Virginia Tech. Without his trust and support, I would not be here today.

I appreciate the financial support I received from various sources, including the Department of Forestry, NCASI, and NASA Earth System Science Fellowship. I was privileged to have outstanding support for collecting my ground truth data, from friends, colleagues, and the Virginia Department of Forestry, through Dr. John Scrivani. I am thankful for the cooperation of the personnel at the Appomattox-Buckingham Forest Office.

I have been fortunate to have outstanding colleagues and friends upon which to rely for input, help with collecting field data or solving statistical mysteries, and companionship. I would like to thank Neil Clark, Olaf Kuegler, Jan van Aardt, Jared Wayman, Rebecca Musy, Christine Blinn, Dr. Mahadev Sharma, Dr. Jim Westfal, Dr. Bronson Bullock, Dr. Rien Visser, and all the graduate students within the department for their technical or moral support. Special thanks go to all my American fellow students, especially Jason Rodrigue, for sharing their traditions and hospitality that enriched my experience in the United States. A special thank-you goes to the “Flying Foresters,” the forestry soccer team, that kept me fit enough to go through the graduate student life.

I am also thankful to the professors at the “Transylvania” University of Brasov, which recommended me for the graduate study at Virginia Tech or contributed to my education in forestry. I would like to mention Dr. Stefan Tamas and Dr. Victor Stanescu.

Finally, my deepest appreciation goes to my family, especially to my wife, Oana, and my daughter, Alexandra, for their love, support, sacrifices, and understanding that allowed me to achieve my dreams. Most importantly, I would like to thank my parents, Elena and Ilie Popescu, for everything they meant for me throughout my life.

I thank you from all my heart!

Dedication

To my parents, Elena and Ilie Popescu, for whom education has a sacred value, and to my wife, Oana, and my daughter, Alexandra, for their love, inspiration, and support.

Table of Contents

List of Figures

List of Tables

1. Introduction

Airborne laser sensors allow scientists to analyze forests in a three-dimensional format over large areas. In contrast to traditional remote sensing methods, which yield information on horizontal forest pattern, modern lidar systems provide georeferenced information of the vertical structure of forest canopies. Laser altimetry or lidar (LIght Detection And Ranging) is an established technology for obtaining accurate, high-resolution measurements of surface elevations (Krabill *et al.*, 1984). Laser pulses from the sensor carried aboard an aircraft are directed toward the ground to collect ranging data to the top of the canopy, and in some instances, to subcanopy layers of vegetation and to the ground. The airborne lidars have been previously used to describe topographic relief (Krabill *et al.*, 1984; Schreier *et al.*, 1985; Bufton *et al.*, 1991; Ritchie, 1995), and forest vegetation characteristics, such as percent canopy cover, biomass (Nelson *et al.*, 1988b), and gross-merchantable timber volume (Maclean and Krabill, 1986).

In recent years, the use of airborne lidar technology to measure forest biophysical characteristics has been rapidly increasing. In addition to providing a characterization of ground topography, lidar data gives new knowledge about the canopy surface and vegetation parameters, such as height, tree density and crown dimensions, which are critical for environmental modeling activities. Airborne lidar data combines both surface elevations and accurate planimetric coordinates, and processing algorithms can identify single trees or groups of trees in order to extract various measurements on their three-dimensional representation.

Laser scanner systems currently available are in a fairly mature state of art, while the processing of airborne scanning lidar data still is in an early phase of development (Axelsson, 1999). Airborne laser scanning represents an emerging technology that is making the transition from the proof-of-concept to reliable uses (Flood and Gutelius 1997). It is a general feature of new technologies that technical potential opens the ground for new applications. Airborne laser scanning is presently in that process, spreading into other fields beyond the generation of terrain models (Ackermann, 1999).

Previous studies that focused on estimating forest stand characteristics with scanning lasers used lidar data with either relatively large laser footprints, 5-25 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, 1997b, Magnussen and Boudewyn, 1998, Magnussen *et al.* 1999). 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.

Forestry applications of small-footprint lidar have a bright future that is currently being explored. Existing processing algorithms for lidar data are implemented through proprietary software and generally aim at filtering vegetation to obtain the terrain elevation model. Potential future uses that have been foreseen in the literature (e.g., Means, 2000) include the assessment of forest biomass, measurement of forest structural attributes critical to understanding forest ecosystem condition, automated processing, and integration with co-registered multi- and hyperspectral digital imagery. Small-footprint lidars are available commercially and more research on their potential for forestry applications is needed. Applications of small-footprint lidar have not progressed too 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. This study explores the feasibility of using multiple return, small-footprint lidar data for estimating forest biophysical parameters. The primary purpose of this research is to develop a lidar processing procedure that takes advantage of the ability of small footprint scanning lasers to portray the canopy structure down to the individual tree level and to capture land surface topographic elevations with high accuracy. Basic tasks in processing the lidar data include the separation of the bare ground surface from forest vegetation, i.e., defining a digital elevation model (DEM) as a subset of the measured digital surface model (DSM), and estimating forest biophysical parameters of interest for biomass assessment.

1.1. Importance of forest biomass in carbon cycles in temperate zones

The above ground biomass of a forest is closely related to crown metrics that can be accurately sensed by airborne lidar. The stand biomass is defined as the sum of the biomass of the individual trees that comprises the stand (Parresol, 1999). Therefore, most of the models for estimating stand biomass incorporate tree dimension variables. The most common procedure for estimating tree biomass consists in the use of regression.

Some commonly used tree dimension variables with linear and nonlinear regression are diameter at breast height, total height, age, and live crown length. Modeling dbh as a function of tree height can mitigate the lack of tree diameter information in the lidar data.

Estimation of forest biophysical parameters, such as tree height, crown diameter, and number of trees per unit area, from lidar observations can make a unique contribution to pressing environmental issues. Quantitative descriptions of landcover and global productivity could benefit from the characterization of forest canopy with high-resolution airborne lidar measurements. Biomass in forests contains a major reservoir of carbon in terrestrial ecosystems that can be easily affected by natural disturbances, climate change, and land use change. Forest biomass is not only important for commercial purposes and for national development planning, but also for scientific studies of ecosystem productivity, energy and nutrient flows, and for assessing the contribution of changes to the global carbon cycle (Parresol, 1999).

One of the primary ecosystem services provided by the temperate forests in the United States, both natural and managed, may be carbon sequestration, providing a negative feedback to the accumulation of greenhouse gases and, thus, global warming (Schlesinger, 1995). Early attempts to quantify carbon sources and sinks revealed a "missing" carbon sink (Detwiler and Hall, 1988, Tans *et al*., 1990) that was likely located in the temperate areas of the northern hemisphere (Ciais *et al*., 1995). Birdsey *et al*. (1993), using United States Department of Agriculture (USDA) Forest Service Forest Inventory and Analysis (FIA) data, found that carbon stored on US timberland has increased by 38% since 1952. This carbon sink, primarily in the northeastern and southern U.S., may account for over 20% of the missing carbon globally. The bulk of this sequestration has arisen from forest growth and the conversion of approximately 10 million hectares of agricultural lands to forests. Hardwood forests with the highest biomass densities are mostly located in the Appalachian Mountains, including the Commonwealth of Virginia. Forests in the southeastern U.S. have a wide range of biomass densities reflecting in part the influence of intensive management of pine plantations and natural forests. While the total biomass of eastern hardwood forests spans a wide range, their average biomass density is less than half of what it could be, because they lack the typical structure of old-growth forests with many large diameter trees

(Brown *et al.*, 1999). The biological possibility of storing additional carbon may not be fully satisfied because of many competing uses and management objectives for forest land. Thus, there is a strong impetus to use remote sensing techniques, such as lidar, to improve the accuracy of the forest inventory estimates at suitable scales.

1.2. Lidar versus photogrammetry

The principal overlap between lidar and photogrammetry lies in the 3-D measurement of surfaces. Baltsavias (1999a) presents a comparison between lidar and photogrammetry with a short overview of the major differences, advantages and disadvantages of each, and major applications.

Lidar affords the ability to "see" the ground in three dimensions. Even with a very dense canopy cover, there are often small openings in the canopy where, because of the high sampling intensity, the laser beam will manage to reach the ground and produce a return. In contrast, photogrammetric methods, particularly automated cross-correlation techniques using infrared photographs, are often unable to accurately compute parallax in these small gaps due to substantially reduced illumination. As a result, the photogrammetrist can often obtain elevations only at the top of dense canopies. Though lidar can have low penetration rates through a dense forest canopy, it offers a direct measurement of the ground elevation beneath the tree crowns.

Photogrammetry is based on processing of images, analog or digital, with the main products being topographic maps, thematic maps, DEMs, DSMs, orthoimages, and visualizations. Processing of films is made by analytical plotters, while digital data are processed by specific digital photogrammetric methods and image analysis techniques that have been thoroughly researched and developed over decades. Thus, users can produce custom-products themselves by making use of image processing packages and rather affordable data. On the opposite side, lidar data and its processing remain mostly a provided service. There are no commercial packages available for processing of lidar data and thorough research is still needed to prove the advantages of scanning lasers. In short, lidar has some strengths over photogrammetry, such as an increased density of points with known ground elevation beneath forest vegetation under certain conditions, independence of illumination conditions (as it is an active system), mapping of surfaces

with poor texture and definition (ice, sand in coastal areas), fast response applications (e.g., natural disasters), direct acquisition of 3-D coordinates, and, as some studies suggest (Gomes Pereira and Wicherson, 1999, Petzold *et al.*, 1999), lower costs in comparison to photogrammetry, mainly for large-area projects. There appears to be an emerging consensus in the mapping community that lidar is a cost-effective alternative to conventional technologies for the creation of DEMs at vertical accuracies of 15 to 100 centimeters. Lidar data is also used to orthorectify digital camera imagery, aerial photography, and high-resolution satellite imagery, such as that from Space Imaging Inc.'s IKONOS satellite (Hill *et al.*, 2000).

While the two technologies are competing for many applications, they are not mutually exclusive. In fact, optical sensors are able to overcome the difficulty of classifying and identifying objects with laser scanning. The integration of the two technologies can lead to more accurate applications.

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. In a broader context, forests play an important role in regional and global issues that range from economics to carbon cycles. The use of remote sensing techniques for assessing forest biomass has been investigated by other researchers, but as of yet such approaches have met with little success for multi-age, multi-species forests, and only with limited success in forests with few species and age classes (Wu and Strahler, 1994). 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). Thus, the study aims at presenting a new approach for assessing forest volume and biomass for both hardwoods and softwoods typical for eastern United States, as there is an increasing need to improve accuracy of forest estimates.

2. Objectives

The overall objective of this research was to develop robust processing and analysis techniques to facilitate the use of airborne laser data for forest biomass and volume assessment. The specific objectives provide a general outline of the study approach and are as follows:

1. To develop an algorithm to characterize ground elevation with multiple-return lidar data.

2. To develop robust processing and analysis techniques to facilitate the use of small-footprint lidar data for predicting plot level tree heights, stem density, and average crown diameter by directly measuring individual trees identifiable on the three-dimensional lidar surface.

3. To relate lidar-derived forest biophysical parameters to diameter at breast height, forest biomass, basal area, and stand volume.

4. To compare the performance of different lidar processing techniques for estimating forest biophysical parameters. Investigate the effect of local filtering with variable window size based on tree height and forest type. Examine lidar data fusion with multispectral optical imagery.

3. Literature review

Airborne lidar is a remote sensing technique that can accurately depict the earth surface in a three-dimensional format by measuring the distance from the sensor to the ground. Lidar is the newest method for DEM development that provides high-accuracy, high-resolution direct surface elevation measurements (Renslow, 1999). Different configurations of airborne lidar systems are currently being used in surveying, geosciences, and vegetation assessment. Airborne lidar data have also been used to study sea ice, ocean surface characteristics, and for sounding atmospheric volumes to measure aerosol characteristics (Yegorov and Potapova, 2000).

This literature review analyzes publications on the lidar research achievements related to vegetation assessment and characterization of ground elevation. After introducing basic lidar principles, the review first examines the body of knowledge for assessing forest vegetation characteristics and distinguishes between profiling and scanning lidar systems, using either small- or large-footprint lasers. Second, the review considers published algorithms for filtering vegetation laser points and interpolating the ground elevation as a prerequisite for estimating vegetation height. The purpose of the review is not only to present the state-of-the-art of lidar research and applications in characterizing forest vegetation, but also to offer a justifying background for the proposed study of using data from a small-footprint, multiple return lidar system to characterize the terrain elevation and to derive canopy metrics of interest for volume and biomass estimation.

3.1. Lidar background

Lidar is the optical equivalent of radar and uses laser energy to measure the distance to a target. An airborne lidar sensor sends laser pulses to the earth's surface and measures the distance associated with the time difference between pulse generation and pulse return. Laser ranging in a repetitively pulsed mode in a near-nadir direction is also called laser altimetry. When the sensor is flown over the forest canopy, the laser energy interacts with leaves and branches and reflects back to the instrument. A portion of the initial pulse may continue through the canopy to lower canopy layers, and possibly to the ground. Some of the lidar sensors are capable of monitoring not only the first or the last

of the laser returns from a single pulse, or both, but also the secondary returns from within the multi-layered canopy (Figure 1). Modern lidar systems are composed of a laser sensor, a GPS (Global Positioning System) receiver, and an INS (Inertial Navigation System) or an IMU (Inertial Measurement Unit). By accurately recording the roll, pitch and heading of aircraft with a time stamp coincident with the laser measurements and the GPS readings, the motion of the aircraft can be corrected and precise positions of the laser hits on the ground surface can be calculated. This assemblage is also coupled with a data acquisition system and sometimes with a video or mapping camera. In lidar range measurements, two major ranging principles are applied, namely the pulse ranging and the phase difference. The latter is applied with lasers that continuously emit light. These lasers are called continuous wave (CW) lasers. In current ranging laser systems, mostly pulsed lasers are used.

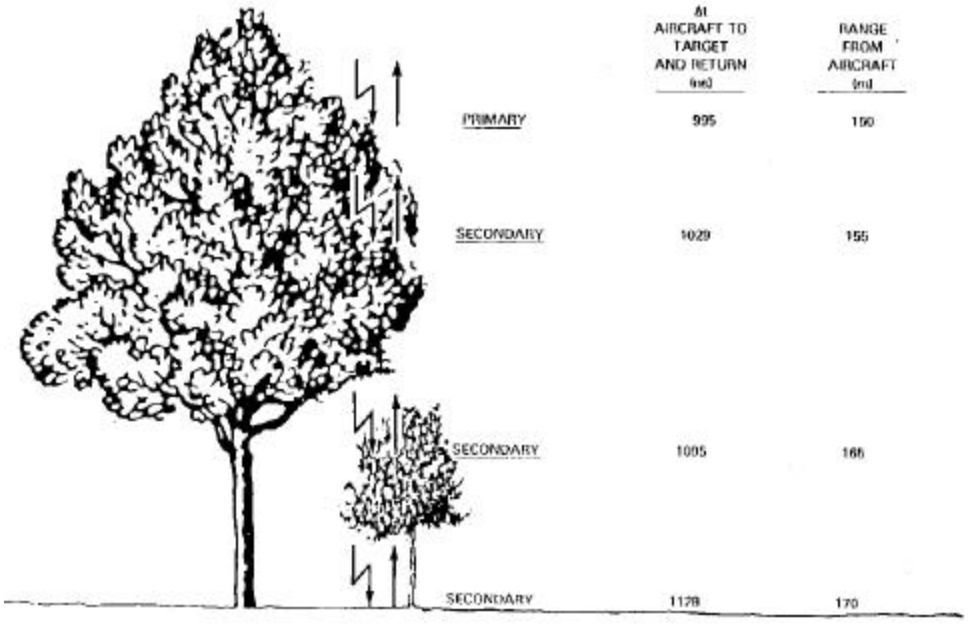


Figure 1: Secondary returns from a multi-layered forest canopy (Nelson, 1988a)

3.2. Airborne lasers for vegetation assessment

The foundations of lidar forest measurements lie with the photogrammetric techniques developed to assess tree height, volume, and biomass. Aerial stand volume tables are based on estimates of two or three photographic characteristics of the dominant-codominant crown canopy: average stand height, average crown diameter, and percent of crown closure (Avery and Burkhart, 1994). Such tables are derived by multiple regression analysis with independent variables measured on photographs by skilled interpreters. Forest measurements on photographs covering large areas can become a tedious endeavor and rely to some degree on the interpreter's ability. Since it is generally not feasible to measure and count every tree in the area of interest, a sampling process analogous to field procedures is often used. The height and density of forest stands can also be estimated on large-scale digital airborne imagery because there exists a close link between the three dimensional organization of the canopy and image texture (St-Onge and Cavayas, 1995). The image spatial structure is only a two-dimensional representation of forest structure. In contrast, lidar pushes traditional remote sensing image processing for forest applications in the three-dimensional domain by being able to provide a unique metric, the vertical dimension of the canopy. 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.

The first generation of lidar sensors used for remote sensing of vegetation was designed to measure the range to the first surface intercepted by the laser, typically along singular transects defined by the flight line (Nelson *et al.*, 1988a, 1997, Ritchie *et al.* 1993, Weltz *et al.*, 1994). More advanced laser altimeters, imaging or scanning lidars, are capable of scanning the ground surface beneath the airborne platform, resulting in a true three-dimensional data set. Commonly, for such lidar sensors the laser beam sampling area, or footprint, is relatively small, usually less than 1 m in diameter. An alternate type of laser altimeter, also known as surface lidar, utilizes the complete time-varying distribution of returned pulse energy, or waveform, that results from the reflection of a single pulse with a large footprint (up to 25 meters).

3.2.1. Profiling lidar

Lidar systems that sample along a single track defined by the flightline are known as single-transect or profiler lidars. The laser beam is pointed from the aircraft in a near-nadir direction and normally operated in a repetitively pulsed mode. The resulting series of pulses can be used to derive the surface elevation profile. Bufton *et al.* (1991) provide a complete description of an airborne lidar system for profiling of surface topography that is able to measure laser pulse time-of-flight and the distortion of the pulse waveform for reflection from Earth surface terrain features.

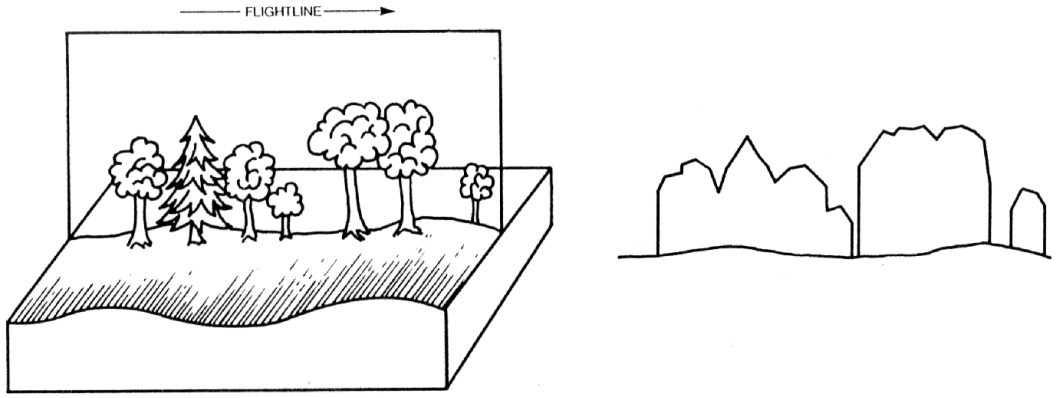


Figure 2. Canopy profile area (CPA) observed with a profiling lidar. (Nelson *et al.* 1984).

Profiling lidar has been used for quantifying the vertical properties of the forest vegetation. Single-track lidars were among the first systems to demonstrate the potential of airborne laser data to measure canopy structure and properties over large forested areas quickly and quantitatively. The canopy profile area, conceptualized in Figure 2 and defined as the area between a trace of the top of the forest canopy and the ground over a given distance (Nelson, 1994), was initially estimated with photogrammetric methods (Maclean and Martin, 1984) and was correlated with timber volume. Nelson *et al.* (1984) analyzed airborne laser data acquired over an oak-hickory forest in south-central Pennsylvania and found out that lidar estimates of canopy heights underestimated photogrammetric measurements by 60 cm, on average, but the laser estimates were more precise. Their results indicated that that canopy closure was most strongly related to the penetration capability of the laser pulse. Schreier *et al.* (1985) used laser height, laser reflection and reflection variability parameters to develop a semiautomated classification

technique which allowed the distinction between conifer and broadleaf forests in northeastern Ontario, Canada. Their laser system was operating in a profiling mode and major modifications were made to measure both the time and amplitude of the laser beam that had a footprint size of 50 cm and 6.7 laser points per square meter. With this recording rate, it was possible to obtain a trace of individual trees. They concluded that lasers produce adequate terrain height profiles and allow for an accurate measurement of tree height. In addition, mean laser reflection and reflection variability measurements may be used to differentiate between coniferous and broadleaf trees.

Maclean and Krabill (1986) investigated the utility of an airborne lidar system for the assessment of gross-merchantable timber volume using the NASA Airborne Oceanographic Lidar (AOL) system. Their research builds upon the previous photogrammetric work on using the canopy profile area as a variable to estimate timber volume (Maclean and Martin, 1984). They found that the laser estimated canopy profile area is a very good indicator of the total gross-merchantable timber and that a stratification by species, effectively a conifer-hardwood distinction, improved the strength of the mathematical relationship. In their 1986 study, Maclean and Krabill employed multiple regression to conclude that the logarithmic transformation of timber volume was more appropriate than the untransformed timber volume to improve the overall correlation with profile area. Canopy profiles defined by heights above ground level from 10 m to 20 m, at 2.5 increments and the entire profile area were analyzed for the best relationship with volume. Partial profile variables were introduced to overcome possible drawbacks from using the entire profile cross-section alone. The mensurational significance of the canopy profile area is twofold. First, it incorporates three factors that significantly affect merchantable timber volume, total tree height, crown diameter, and crown closure into a single variable. Second, the variable measures directly the major contributors to merchantable timber, the dominant and co-dominant trees in a stand. Tree height is a major indicator of merchantable timber volume and the canopy profile area alone cannot distinguish between stands with the same profile area, but with a very different vertical distribution of it. They also expected, without researching the hypothesis, that the canopy profile area would be a very good indicator of total biomass or total woody biomass.

Potential uses of laser profiling data for estimating plant biomass and the repeatability of the laser observations were explored by Nelson *et al.* (1988a). A key component of this study was to accurately locate the laser transects on the ground. They found 3 to 6 percent variation between comparable sections of different overpasses along the same flight line. Differences between the laser and ground estimates of biomass and volume proved to be approximately 7 to 8 percent. Differences between the two sets of laser flight lines were only 0.9 and 1.2 percent for biomass and volume. They concluded that the advantages of using an airborne profiling laser to collect forest mensuration data are twofold: first, canopy height data can be collected quickly along transects hundreds of miles long and, second, laser data can be used to extend a limited ground sampling effort over areas that may not be easily accessible by ground inventory crews. In their study of 1988b, Nelson *et al*. predicted ground heights, biomass and volume using six laser height variables, four of them reflecting canopy profile area (mean canopy height, and three modified canopy profiles with exclusion limits of 2, 5, and 10 meters, i.e., the area between the top of the canopy and a line drawn 2, 5, and 10 m respectively, above the ground trace) and two more directly reflecting actual tree heights (average of the largest three canopy heights and the mean plot height). In addition, one laser canopy density measure was selected for subsequent study. However, the four variables reflecting canopy profile area have coefficients of correlation exceeding 0.95 and the same high correlation was found between the two variables reflecting actual tree height. They tested several regression models with two logarithmic equations in an attempt to determine which model best described variation in the ground measurements. The best model explained between 53% and 65% of the variability noted in the ground measurements of forest biomass and volume. The results of this study showed that species stratification did not consistently improve regression relationships for four southern pine species.

Measurements of canopy heights, percent canopy cover, and spatial patterns were estimated using profiling airborne lidar by Ritchie *et al.* (1993). Lidar derived canopy heights of forested and rangeland areas were not significantly different from measurements made on the ground, as indicated by a paired *t*-test. Both the field and laser measurements were considered site attributes, since no attempts were made to locate and measure the same tree. Interpretation of the lidar measurements allowed inferences about

spacing, type, and maturity of trees in the forest. Indications of forest canopy closure, spacing, and gaps in the canopy were inferred by considering means, maximums, medians, and coefficients of variation. In order to detect the ground elevation, the authors assumed that the minimum elevations along a laser flight path correspond to ground hits. In areas of vegetation, the minimum values assumed to reach the ground surface were determined by calculating a moving minimum elevation for 50 laser measurements. Vegetation heights were subsequently computed as the difference between the calculated ground surface and the actual laser measurement. Rangeland vegetation properties were also measured by Weltz *et al.* (1994) using the same lidar sensor. Plant height and canopy cover for vegetation higher than 0.3 m were not significantly different than field measurements made using the line-intercept transect method at seven of the eight sites evaluated. The major limitation of using lidar for estimating canopy cover on semiarid rangelands proved to be laser penetration of open canopy structure.

Most of the studies mentioned above require ground observations coincident with the lidar observations in order to develop regression equations needed for prediction. High-resolution lidar profiling data was used to estimate tree heights, canopy density, basal area, and biomass of tropical forests (Nelson *et al.*, 1997) from laser measurements that include average canopy height and coefficients of variation, without relying on lidar-ground transect colocation. This study used fixed area ground plots to simulate the height characteristics of the tropical forests and to simulate canopy laser measurements. Multiple linear regression was used to establish ground-laser relationships for basal area, volume, and biomass as a function of simulated, laser-measured variables. Their results indicated that the untransformed multiple regression models forced through the origin were most apt. The simulated laser measurements that proved to be significant were: average canopy height for all pulses (including ground hits), average canopy height for canopy hits only, and coefficients of variation of average canopy height, for all pulses and, respectively, canopy hits only.

Landscape topography for estimating water balance components and for distinguishing between landscapes has been investigated with high-resolution profiling laser altimetry (Ritchie, 1995). Lidar measurements of the microroughness of the landscape surface can be used to predict soil moisture, runoff, and soil erosion at the landscape scale. Pachepsky *et al.*

(1997) used fractal dimensions of laser altimetry data to distinguish between grass and shrub landscapes.

3.2.2. Scanning lidar

In the early 1990s, profiling lidar sensors were gradually replaced by scanners, while GPS was combined with Inertial Navigation Systems (INS) or Inertial Measurement Unit (IMU). Airborne laser scanning represents a new and independent technology for the highly automated generation of digital terrain models (DTM) and digital surface models (DSM). Laser scanning systems provide overall vertical accuracy in the order of tenths of a meter, by operating usually at flying heights of up to about 1000 m above ground. The scan angle is generally less than ± 30°, in most cases ± 20°. Present measuring rates, i.e. pulse repetition frequencies, situate between 2 kHz and 25 kHz. The actual sampling density is a function of flying speed, pulse rate, scan angle, and flying height. For a given flying height above ground *h*, the laser footprint (LF) mainly depends on the divergence of the laser beam γ (rad) and the instantaneous scan angle θ_{inst} (deg) and is given by (Wehr and Lohr, 1999, Baltsavias, 1999b):

$$LF = h\gamma / \cos^2(\theta_{inst})$$

The swath width (SW) depends on the scan angle θ (deg), which also defines the field of view (FOV):

$$SW = 2h \tan(\theta/2)$$

The high measuring rate of laser scanning is an important characteristic in airborne laser scanning. The ground point density is strongly dependent on the type of scanning system and the speed of the aircraft. The area sampled depends not only on the laser ground footprint, but also on the point- or post-spacing across and along flight direction. A complete survey of existing commercial laser scanning systems and firms, including detailed systems parameters, is provided by Baltsavias (1999c).

Table **1** synthesizes different technical parameters of small- and large-footprint lidars that can be found in more detail in Dubayah *et al.*, 1997, Baltasavias (1999c), Blair *et al.* (1999), and Means (2000).

Table 1. Comparative technical specifications of small- and large-footprint lidar systems

Technical parameter	Small-footprint (discrete returns / pulse)		Large-footprint (waveform return / pulse)
	Minimum-Maximum	Typical values	
Point spacing (m)	0.1 - 10	0.3 - 2	10-25
Flying height - h (m)	20 - 6100	100 - 2000	4000-400000
Footprint diameter (m)	0.1 - 2	0.3 - 0.6	10-25
Returns recorded / pulse	1 - 5	1	Full waveform digitization
Swath width, m	0 - 1.5 h*	0.25 - 0.75 h	800-8000 (discontinuous)
Pulse rate (kHz)	0.1 - 83	5 - 15	0.1-0.5
Platform	Airborne (helicopter, airplane)		Airborne (airplane); Spaceborne (satellite)

*flying height

3.2.3. Scanning lidar applications in forest inventory

Previous investigations of using various configurations of lidar for forest assessment have returned positive conclusions regarding the estimation of tree heights, stand volume and biomass, and canopy cover. Most of the commercial laser scanners use a small-footprint laser beam. Non-commercial and research systems mainly use large-footprint scanning laser systems, such as SLICER (Scanning Lidar Imager of Canopies by Echo Recovery) and LVIS (Laser Vegetation Imaging Sensor) developed by NASA and used for validation of future spaceborne lidar missions. Since the lidar data set used in this study was acquired with a small-footprint laser sensor, this review will concentrate more on this type of system and their applications in estimating forest parameters.

3.2.3.1. Mapping terrain topography with scanning lidar

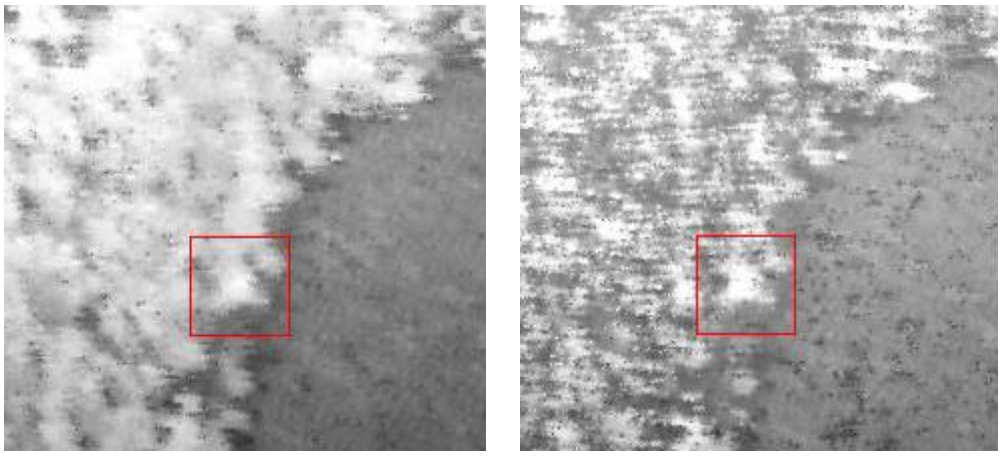
The basic processing task that needs to be accomplished before attempting to estimate forest parameters is the characterization of the terrain elevation and creation of a DEM, as a subset of the digital surface model obtained from raw laser points. Small-footprint lidars are readily used to create high resolution DEMs, 1x1 m to 3x3 m (Means, 2000), and characterize vegetation characteristics of relatively small areas. One of the major advantages that lidar offers over traditional photogrammetry is the ability to directly measure ground elevation in forested areas. Some of the laser pulses find openings in the canopy and penetrate to the ground or to lower layers of vegetation. These irregularly dispersed laser points assumed to correspond to the ground are used with appropriate interpolation methods to derive the high-accuracy DEM. Lam (1983) offers a comprehensive review of spatial interpolation methods classified as point and

areal interpolation techniques. For point interpolation, the numerous methods can be further classified into exact and approximate, depending on whether they preserve the original sample point values. Exact methods include distance-weighting, Kriging, spline interpolation, interpolating polynomials, and finite-difference methods. Approximate methods include power-series trend models, Fourier models, distance-weighted least squares, and least-squares fitting with splines. Previous attempts to characterize the terrain elevation with lidar preferred to use exact interpolation methods in order to preserve the raw lidar data values (e.g., Young *et al.* 2000).

Small-footprint lidars need to have a high intensity of sampling, i.e., laser hits per unit area, in order to sample the tops of relatively broad trees and to reach the ground in areas of dense canopy closure. A high laser point density allows an adequate filtering of vegetation hits for the derivation of terrain elevation, since only a relatively low number of laser pulses are able to entirely penetrate the canopy. As expected, Krabill *et al.* (1984) found differences between penetration rates obtained under summer conditions, i.e., the leaf-on season for deciduous vegetation, and under winter conditions for deciduous stands located in Tennessee, U.S.A. Penetration rates for the NASA Airborne Oceanographic Lidar (AOL) system operated in a profiling mode with waveform digitization capabilities over various terrain conditions were in the range of 10 to 63% under summer conditions, and 39 to 68% under winter conditions, for both deciduous and coniferous stands. Evidently, under summer conditions fewer laser pulses were able to directly penetrate the canopy and in the case of the AOL system, the ground elevation had to be extracted almost entirely from the temporarily recorded laser return waveforms. Their observations also showed much lower energy in returns from the ground due to the energy losses through canopy penetration. In addition, return signals from low-lying vegetation occurred rather close in time and space to ground returns so that the receiver could not separate them, thus masking the ground returns. In other words, if the lidar detects a return at 1.5 – 2.0 m above the ground and the minimum separable distance between returns is larger than that, then the sensor will not detect the ground. This means that low vegetation layers with foliage can produce returns that are virtually impossible to distinguish from the true ground level. However, for computing a high accuracy DEM, it is necessary to distinguish the laser points into ground points, the points lying on the

terrain with the accuracy of the lidar measurement, and vegetation points. In areas with low penetration rates, filtering vegetation points proved to be a difficult task. Recent surveys in the U.S. Pacific Northwest carried out using the Optech ALTM 1020 scanning system indicated a minimum 20-30 % penetration of coniferous canopies (Flood and Gutelius, 1997). In the same region, with conifer-dominated stands and dense overstory, Means (2000) experienced a very low penetration to the ground, only 1-5%, for a small-footprint lidar. Kraus and Pfeifer (1998) estimated a penetration rate of less than 25% for their lidar study in the Vienna Woods (Wienerwald) in Austria, though they do not offer any description of the forest stand characteristics.

A low penetration rate for single-return lidar systems can be partially compensated by the ability of the system to record the last return or multiple returns. But the separation of laser points into terrain and vegetation hits still remains a difficult task even for the last laser return, as seen in Figure 3. A two-dimensional black-and-white representation of lidar data depicts higher points with lighter pixels and lower points with darker pixels. The images in Figure 3 were created by interpolating the cloud of first-return (a) and last-return (b) lidar points to a regular grid. Both images show the same area with a deciduous stand in the upper-left part of the scene and a young coniferous stand (pine stand) in the lower right corner and were obtained from lidar data collected with the AeroScan system over the Appomattox-Buckingham state forest in Virginia, USA. Though it is evident that Figure 3 (b) shows points of lower heights, with darker pixels, parts of the deciduous tree crowns are still apparent. Assumed ground returns are also noticeable in the young coniferous stand, as they appear as small areas of darker pixels. It is thus clear that the last return does not necessarily penetrate dense canopy layers to record the ground elevation and most of the returns are still vegetation points.



a) b)

Figure 3. First- (a) and last-return (b) lidar image of a deciduous-coniferous wooded area (lidar data set acquired over the Appomattox-Buckingham state forest in Virginia, USA, with the AeroScan system)

Few papers published on the subject of sorting out lidar data into vegetation and terrain points provide all the algorithmic details. The filtering of lidar points in ground points and non-ground points, such as the laser hits intercepted by vegetation, is performed using either raw lidar points or an interpolated surface as a regular grid. Using raw data has the advantage of preserving original laser elevation values, but most of the filtering algorithms are complex and implemented through proprietary software developments. Filtering ground points on an interpolated surface is influenced by interpolation errors, but a wide array of filtering methods is already available with popular image processing software.

A robust iterative algorithm for filtering vegetation laser hits is presented by Kraus and Pfeifer (1998). At the first run, the algorithm computes an average surface of all points. The true terrain points are expected to have negative residuals, while the vegetation points are more likely to have small negative or positive residuals. These residuals are used to compute weights ranging from 0 to 1 for each laser point. Details about the weighting function are given in Kraus and Pfeifer (1998). To identify points that receive a maximum weight of one, they used three methods to identify the threshold of the residuals that classifies a laser return as a terrain point. All three methods are based on analyzing the histogram of the residuals. The first method uses the expected accuracy (σ_T) of the terrain points and shifts the terrain/vegetation threshold on the histogram of

the residuals until the standard deviation of the negative branch of the residuals reaches σ_T. The second method is more robust in the presence of blunders and it calculates the standard deviation of the negative branch of residuals for all possible shifts of the terrain/vegetation threshold. The minimum value indicates the appropriate threshold. The third method makes use of a rough estimation of the penetration rate. For 40% terrain points, the threshold is at the position where the first 20% of the residuals are found. The result of their algorithm is a high quality DEM for large-scale applications, though it lacks a quantitative estimation of its accuracy. However, Kraus and Pfeifer (1998) foresaw modeling opportunities for forest applications of the volume between the top of the canopy surface and the terrain surface.

Magnussen and Boudewyn (1998) and Magnussen *et al.* (1999) derived stand heights from airborne laser data of douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) stands on Vancouver Island in Canada using the Optech ALTM 1020 laser scanning system. They first needed to derive a DEM from the ground hits. Two flight lines were flown and, for the second one, the sensor was configured to register the last return from each pulse, which is assumed to have hit the ground. The post-spacing of the ground points averaged 2.3 m. The model was obtained from locally weighted thin-plate splines using a tri-cube weight function.

Axelsson (1999) gives a brief description of his approach to finding the ground surface using a small footprint, multiple return, scanning lidar (Saab TopEye). His implementation analyzes each scan line at a time and not the whole surface and interpolates the terrain elevation using the assumed low ground surface points.

Jaafar *et al.* (1999) describe three methods for the construction of a DEM from scanning lidar data that include morphological filtering, conventional statistical mean filtering and the application of artificial neural networks to identify and remove above-surface features that consist of trees and man-made objects. All methods were applied to lidar data interpolated to a regular grid with a spacing of 2 meters. The morphological filtering approach comprises two passes that identify, first, the local minimum and, second, the local maximum, in order to eliminate negative outliers. Critical for this method is using the appropriate window size to search neighboring pixels for either local minima or maxima. Therefore, experimenting with the window size is a key factor for an

adequate filtering. For their study, Jaafar *et al.* (1999) used a window size of 18 meters (9 x 9 pixels). The second method, filtering and statistical approach, aims at isolating surface features by subtracting the original lidar digital surface model with any reference surface established from a 9 x 9 smoothing filter image onwards. It was found that the standard deviation for the above-surface features tend to be distinct from the terrain surface at smoothing filter size 9 x 9 onwards, for a spatial resolution of 2 meters. In this study, the 25 x 25 filter size (50 x 50 m) gave the best result. The artificial neural networks method was used to classify the lidar digital surface model into three main classes that include land, buildings, and trees. The input parameters used to train the artificial neural networks with a back-propagation algorithm were: area, standard deviation, mean slope, and maximum height. Both the filtering and statistical and the artificial neural networks approaches mask the above-ground features onto the lidar digital surface model and construct the DEM by filling the empty areas using interpolation with the surrounding pixel values. However, the morphological filtering and the filtering and statistical methods gave better results than the artificial neural networks approach. Crucial factors were identifying the correct window size in the morphological filtering method, choosing the appropriate filter size for the filtering and statistical method, and ascertaining optimum input descriptors for the artificial neural networks.

An algorithm that combines some of the filtering and thresholding methods presented above for removing vegetation points is given in Petzold *et al.* (1999) (used by TopScan). It is also an iterative procedure that first computes a rough terrain model from the lowest points found in a moving window of a rather large size (no indications are given regarding the exact window size). All points with residuals exceeding a given threshold are filtered out, and a new DEM is calculated from the remaining points. This step is repeated several times, reducing the window size with each iteration. Critical parameters influencing the accuracy of the final DEM are window size and threshold values for each iteration. A small window size and a large threshold lead to many vegetation points classified as ground points, while a large window and a small threshold smooth the terrain and remove small discontinuities. Evidently, these parameters need to be adapted for various terrain configurations, such as flat, hilly, and mountainous. Their final DEM was generated with a grid spacing of 10 meters for an average ground point

distance of 5 m. The mean ground point distance increased to 20 meters in areas with dense coniferous forests. The DEM quality check, using contours, shaded reliefs, orthoimages, and field checks was surprisingly good, especially for wooded areas. The accuracy proved to be better than the one achieved by photogrammetric stereo compilation, though the authors do not offer a direct comparison. Even in coniferous forest, the number of ground points retained after the filtering process was considered sufficient for the derivation of a DEM. The cost-benefit analysis for the derivation of a new DEM based on calculations made by the Surveying and Mapping Agencies of the Federal States of Germany showed that laser scanning requires only 25% to 33% of the budget that was needed for photogrammetric compilation (Petzold *et al.*, 1999).

A slope based filtering of lidar points is described by Vosselman (2000). The basic idea behind his algorithm is that a large height difference between two nearby points is unlikely to be caused by a steep slope in the terrain, therefore the higher point has a high probability of being a non-ground point, such as a vegetation hit. According to the filter definition, the height of a point needs to be compared with the heights of other points in its neighborhood. To determine neighboring points, the initial lidar points are organized in a Delauney triangulation.

Despite intense efforts in the creation of high-resolution DEMs from lidar data, driven by either commercial or scientific purposes, the characterization of terrain topography under forest conditions is still a challenge.

3.2.3.2. Deriving forest biophysical parameters with scanning lidar

The literature on airborne scanning lidar is mainly divided into describing two types of systems, small- and large- footprint lidars. As mentioned before, this literature review will focus more on small-footprint lidars and their processing algorithms and applications for forest vegetation assessment and will consider in less detail large-footprint systems. Means (2000) gives a good comparison of the two systems by examining them with respect to their design, capabilities and uses. The two biggest differences of the rationale for using small- and large-footprint lidars are in scale and resolution of terrain and vegetation characterization.

Small-footprint scanning lasers

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 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, 1997b; Magnussen and Boudewyn, 1998; Magnussen *et al.*, 1999; Young *et al.* 2000).

Recent studies show that, in moderate to dense forest, small-footprint lidars tend to underestimate stand height (Nilsson, 1996; Næsset, 1997a). It is rather intuitive to expect a more frequent laser sampling of the crown shoulders than the tree apex, thus canopy heights are biased toward low values. Finally, as mentioned before, the frequency of ground returns can be low and the characterization of terrain elevation might degrade the accuracy of the derived canopy heights.

Two approaches to the assessment of mean height, which aim to reduce the underestimation bias, are presented by Næsset (1997a). In his study, laser hits were classified as vegetation and ground hits and a DEM was generated, though details on the algorithms are not provided. The tree canopy height was computed as the difference between tree canopy hits and the corresponding DEM values. Observations with height values less than 2 m were excluded from the dataset in order to eliminate the effect of understory vegetation. The first approach applied three techniques to compute the mean stand height, as follows: (1) the average value for each stand was calculated simply as the arithmetic mean of the individual laser observations assumed to have hit the canopy; (2) the mean stand height was computed using the original height values weighted by themselves or (3) by the square of the original values. These mean heights were denoted h_{h1}, h_{h2}, and h_{h3}, respectively. The second approach to computing the mean stand height consisted in laying a grid over the laser data and selecting only the largest laser height value within each grid cell. Finally, the average stand height was computed as the mean value of the selected observations within each cell. To increase the significance of each of the laser heights, the highest laser values were weighted with the number of laser

measurements within the cell they represented. Three grid cell sizes were used: 15x15, 20x20 and 30x30 m, and the corresponding estimated mean stand heights were denoted h_{15x15}, h_{20x20}, and h_{30x30}, respectively. All the lidar estimated stand heights were compared to the mean stand heights determined from field measurements. Ground truth heights were expressed by Lorey's height, which is the mean height weighted by basal area. It was found that h_{h1} underestimated the ground truth by 4.1-5.5 m. The weighted approach (h_{h2}, and h_{h3}) reduced the underestimation of the ground truth height to 2.1-3.6 m. The mean difference between laser heights estimated using the grid approach (h_{15x15}, h_{20x20}, and h_{30x30}) and ground truth were between 0.4-1.9 m. Næsset (1997a) concluded that small-footprint lasers might produce unbiased estimates of stand heights provided that only the largest laser height values are selected. In the same study, he also investigated the effect of the laser scan angle on the stand height estimation by using regression. He found that the coefficients for the off-nadir scan angle were not significant, but they were still indicative of increased underestimation of the true stand height when measuring laser heights at an angle. However, the problems of off-nadir scanning can be reduced by imposing an acceptable scanning angle and properly calibrating the laser system for forestry applications.

In another study completing the previous one, Næsset (1997b) attempted to estimate timber volume using the lidar mean stand height, the mean stand height of all laser pulses within a stand, and the canopy cover density as determined from the laser data. The mean stand height was calculated using the 15x15 grid approach explained above. The mean height of all laser pulses within a stand (h_a) was computed as the sum of the height values of the laser pulses assumed to be canopy hits divided by the total number of transmitted pulses. The reason for dividing by the total number of transmitted pulses and not the number of canopy hits is not apparent. The mean canopy cover density was computed using the grid approach. For each grid cell, the crown cover was computed as the number of canopy hits divided by the number of transmitted pulses. The average crown cover of a stand was computed as the mean value for individual cells. The coefficients of determination (R^2) were in the range of 0.456 to 0.887. Low R^2 values were obtained for a test site dominated by Scots pine (*Pinus sylvestris* L.), whereas high values were for found for Norway spruce stands (*Picea abies* Karst.). These results may indicate that a

stratification of the observations should be carried out prior to fitting of volume equations. Stratification criteria could include species composition, age, and stand density.

The expected difference between mean tree height and the laser-based mean canopy height is the central point in the study of Magnussen and Boudewyn (1998) on measuring heights of Douglas fir trees. The idea behind estimating this height difference is the size of the exposed treetops that receive most of the laser hits. The exposed crown size was computed through a model of crown shape. For a sampling intensity of one point per 5 m^2 (post-spacing of 2.3 m) and a footprint of 0.35 m, the authors expected only 2% of the return signals to have been reflected off a treetop. This observation is based on the assumption that the treetop has the same size as the footprint. The majority of returns would be reflected from the side of the crowns of dominant trees as laser canopy hits mimic sampling with probability proportional to the projected crown sizes. Geometrical probabilities were used to estimate the average vertical positions of laser hits on the exposed crown length in order to finally derive the estimated difference between the true tree height and laser-based height. The mean tree height is then computed by adding the calculated difference to the laser estimated tree height. To predict the latter, the authors made use of a grid approach by dividing the area into hexagons of a given size and then retaining the maximum canopy height value in each hexagon. This procedure in essence selects a quantile that can be controlled by changing the grid size. Modifying the cell area affects the number of canopy heights in each cell, thus the base of selection for the highest laser hit. Adding the estimated difference to the laser-based height improved the correlation between field and laser estimates from 0.61 to 0.83. An interesting finding was that half of the laser pulses hitting the canopy were returned at or above a height that coincided with the height of mean leaf area index (LAI). Paired *t*-test on a per plot basis rejected the null hypothesis of no difference between laser-based and field-based estimates of plot heights. However, testing the same hypothesis, but spatially nonexplicitly, using bootstrap samples led to the acceptance of the null hypothesis. A reduced influence of outlier plots (the chance that all outliers are in the bootstrap sample is low) and an increase in variability (due to bootstrapping) were the main factors

explaining this result. However, for operational testing procedures, the stand level spatially nonexplicit attributes are usually desired.

Two models for recovering tree heights from airborne laser scanner data are presented by Magnussen *et al.* (1999). The first height recovery model is based on the assumption that the canopy height is sampled with a probability proportional to the crown area. In sum, the PDF (probability distribution function) of laser canopy heights would be the result of a PPS (probability proportional to size) sampling process. The authors attempted to remove the effect of PPS on the empirical PDF of canopy heights to recover an unbiased approximation of the PDF of tree heights. The second recovery model, as in Magnussen and Boudewyn (1998), attempted to recover the distribution of canopy heights through estimates of the vertical distance between the point of a laser canopy hit and the top of the tree hit. The difference between a tree height and the height registered by the laser hit of the same tree is controlled by numerous factors, such as the vertical and horizontal distribution of branches and foliage, local stand conditions, height and shape of the tree, and view angle (Nelson, 1997). From the recovered distributions of canopy heights, using both models, the arithmetic mean height, the variance and the upper quantiles (75, 85, and 95) were obtained and compared to the corresponding ground-based estimates. The quantiles are considered substitutes for dominant and codominant tree heights. Laser canopy heights averaged over the sample plots were 3 m below the mean tree height measured on the ground. Application of either of the two models brought the laser estimated mean height to within 0.5 m of the ground-based average. Correlation to ground-based plot values was only 0.6. Elimination of suppressed trees from field data increased the correlation between ground- and laser-estimates to 0.7, since trees growing in the shadow of dominant trees are not "seen" by the lidar. Also, the lack of any estimate of the lower limit of the live canopy was clearly impeding the recovery of below average tree heights, as witnessed by poor predictions of plot minimums and variances.

Young *et al.* (2000) estimated tree heights of mid-rotation loblolly pine (*Pinus taeda* L.) stands in Mississippi, USA, by attempting to first find the location of dominant trees. Focal filtering and clump analysis were used to find the geometric center of each clump in the lidar data that was assumed to locate the terminal of undamaged trees. Once the

treetops were located, image subtraction using first and last return interpolated data was employed to extract tree heights.

A similar approach to estimating tree heights was used by Popescu *et al.* (2000), in a broader effort to assess forest biomass. In addition to estimating tree heights, this study also predicted crown closure, crown diameter and tree density using a simulated set of lidar data. The simulator that generates the 3-D top-of-canopy model constructs individual tree crowns based on ground-measured parameters, such as total tree height, height to first branch, crown diameter, and crown shape, and places these crowns into a fixed-area plot using mapped stand coordinates (Nelson, 1997, Nelson *et al.*, 1998). The simulation process can be adjusted to generate single return lidar data sets with different post-spacings and laser footprints. Results indicated that laser data with small to medium footprints might produce good estimates of stand biophysical parameters that can be further used to derive biomass. For coniferous stands, results showed that an appropriate laser footprint size could be between 0.75 and 1.0 m. The study concluded that the choice of appropriate airborne lidar sensor characteristics and processing algorithms depends on forest type and structure.

Large-footprint scanning lasers

An alternate type of laser altimeters, also known as surface lidar, utilizes the complete time-varying distribution of return pulse energy, or waveform, that results from the reflection of a single pulse with a large footprint (Figure 4). Multiple targets with different reflective properties and varying heights may occur within the area covered by the laser footprint, usually with a diameter on the order of one to two times the typical crown width. Each laser pulse has a high probability to intercept the crown apex and to penetrate to the ground through intra- and inter-crown openings. Vegetation height can be extracted from the waveform data based on the time difference between first and last returns. Recent NASA instruments (SLICER, LVIS) provide means for measuring topography and canopy vertical structure by analyzing the waveform return of a medium-large footprint laser (5-25 m diameter) (Harding *et al.*, 1994, Blair *et al.*, 1999). Recent studies using the SLICER instrument have demonstrated that large-footprint lasers can make accurate measurements of stand height, above ground biomass, basal area and LAI

in forest of varied cover types in eastern United States (Lefsky *et al*., 1997, 1999, Weishampel *et al*., 1997) and in the Pacific Northwest (Means *et al.*, 1999).

Data from large-footprint lidar may become publicly available with the launch of the Vegetation Canopy Lidar (VCL) in the year 2004, as a result of the collaboration between NASA and the University of Maryland (Dubayah *et al*., 1997). Over the mission's two year lifetime, VCL is planned to collect data over 3-5% of the Earth land area between 65° N and S latitude and will sample nearly all the major forest and woodland types. VCL laser footprints are 25 m wide and contiguous in the along track direction, while spaced 4 km apart across track.

Both small- and large-footprint lidars will contribute to a dramatic increase in lidar observations in the next decade, from which measurements such as earth surface topography, the vertical structure of vegetation, including sub-canopy, will be acquired with unprecedented accuracy. Small-footprint lidars are becoming widely available commercially, and new sensors meet the requirements of high sampling intensity and multiple return signals. The potential of forestry applications of small-footprint lidar is enriched by concurrent optical sensor images. Large-footprint lidar brings the advantages

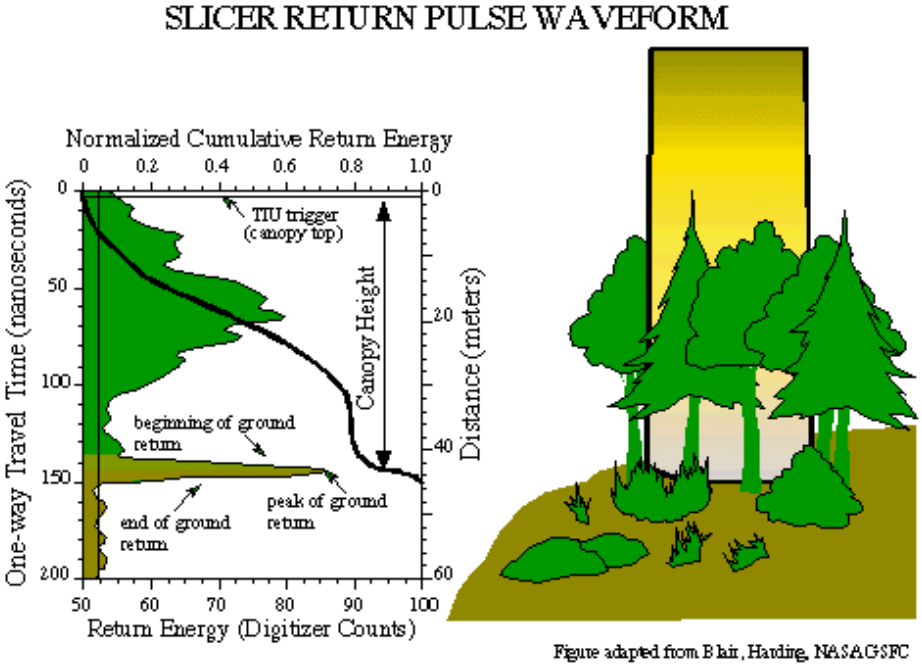


Figure 4. Lidar waveform collected by the SLICER instrument. (Lefsky, 1998)

of satellite observations with repeated and global coverage. The two types of lidar prove to be the foremost technology to study the earth surface and vegetation vertical structure and will complement needs for global lidar coverage and accurate fine-scale surface elevations.

3.2.3.3. Lidar and optical data fusion

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 digital photogrammetric studies 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). 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 is 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.

4. Materials and methods

4.1. Study site

The study area is located in the southeastern United States, in the Piedmont physiographic province of Virginia (Figure 1). It includes a portion of the Appomattox-Buckingham State Forest that is characterized by deciduous, coniferous, and mixed stands of varying age classes (37° 25' N, 78° 41' 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.

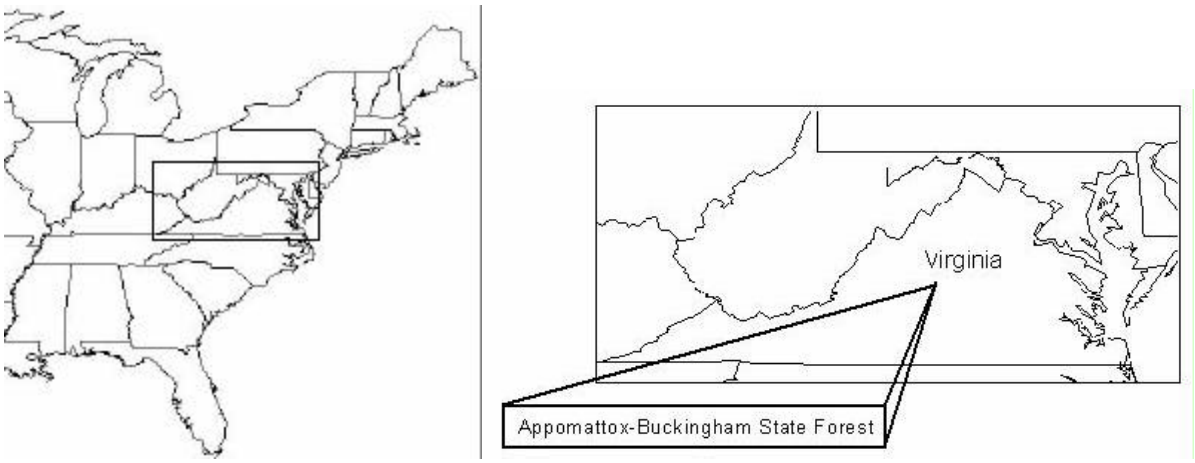


Figure 5. Map of eastern United States indicating the location of the study area

4.2. Ground reference data and modeled parameters

4.2.1. Ground inventory data

The ground truth data was 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 - 140 years for the upland-hardwood stands. Three stands of loblolly pine were exceptionally old, being of 60-65 years. The tree species found in the pine-hardwoods

stands were white oak (*Quercus alba* L.), chestnut oak (*Quercus prinus* L.), northern red oak (*Quercus rubra* L.), southern red oak *(Quercus falcata* Michx.), yellow poplar *(Liriodendron tulipifera* L.), red maple (*Acer rubrum* L.), and three species of pines – Virginia pine (*Pinus virginiana* Mill.), loblolly pine (*Pinus taeda* L.), and shortleaf pine (*Pinus echinata* Mill.). In addition to the hardwood species mentioned above, the following species were inventoried in the upland and bottomland hardwood stands: pignut hickory (*Carya glabra* (Mill.) Sweet), mockernut hickory (*Carya tomentosa* (Poir.) Nutt), scarlet oak (*Quercus coccinea* Muenchn.), black oak (*Quercus velutina* Lam.), blackgum (*Nyssa sylvatica* Marsh.), flowering dogwood (*Cornus florida* L.), sourwood (*Oxidendron arboreum* L.), Eastern redcedar (*Juniperus virginiana* L.), black cherry (*Prunus serotina* Ehrh.), hornbeam *(Carpinus caroliniana* Walt.), and American beech (*Fagus grandifolia* Ehrh.).

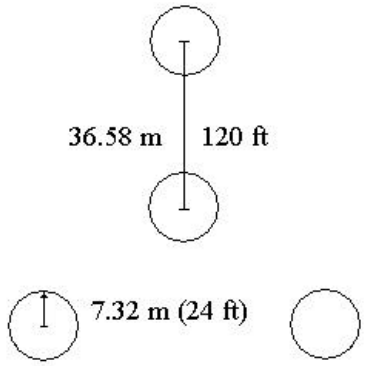


Figure 6. Layout of a single FIA plot with four subplots

The plot design followed the U.S. National Forest Inventory and Analysis (FIA) field data layout (Figure 6). 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) (U.S. Department of Agriculture, Forest Service, 2001, National Forest Inventory and Monitoring Core Field Guide). 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. The FIA program is in transition, changing in response to legislation and demands for increased consistency. Thus, the field data collection partly followed FIA standards and added new variables in an attempt to explore the integration of lidar measured information into the core FIA database and to allow a more detailed inventory of trees within the subplots. For the purpose of this 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

microplot with a radius of 2.07 m (6.8 ft) was located at the center of each subplot, to account for seedlings and saplings with a dbh above 2.54 cm (1 in), but less than 6.35 cm (2.5 in). A total of 16 plots were measured in the study area, each with 4 subplots. FIA plot centers (subplot 1 centers) were located systematically on a 200 x 200 m grid (656 x 656 ft), with rows oriented East-West, and columns oriented North-South (Figure 7). The origin of the grid relative to the map was randomly selected. The selection of plots location on the grid tried to cover all forest types present in the study area and, at the same time, to balance the ratio between plots with coniferous and deciduous species (Table 2).

Table 2.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

To simplify the analysis relative to tree species, subplots have been 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. Table 3 shows the number of FIA-type subplots per each category.

Table 3. Number of FIA-type subplots per each category

Category	**Hardwoods**	**Pines**
Number of FIA-type subplots	33	31

The centers of subplots 1, for most of the plots, were laid out in the field using a navigational GPS unit – PLGR (Rockwell Collins). 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, HP-GPS-L4 with a PC5-L data collector (Corvallis Microtechnology, Inc.). 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., 2001). Under forest canopy, GPS systems tend to yield from 1.5 - 3 times less accurate solutions (Craig Greenwald, 2001, Corvallis Microtechnology, Inc., Technical Support, personal communication). Therefore, we estimate sub-meter accuracy for locating the plot centers. Depending on the data availability, the following National Geodetic Survey continuously operating reference stations in Virginia, USA, were used for the differential correction: Blacksburg, Driver, Charlottesville, and Richmond, all within the base line distance of 300 km (187.5 miles) for this type of GPS receiver from the location of the study area. Subplot center coordinates are shown in Appendix 1.

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 3 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 tip of the apical meristem. The actual length was also measured for trees with a broken or missing top. For leaning trees, the height of the highest point above the ground, usually the tip, was recorded along with the bearing and distance from the base of the stem to the projected tip on the ground. 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). 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

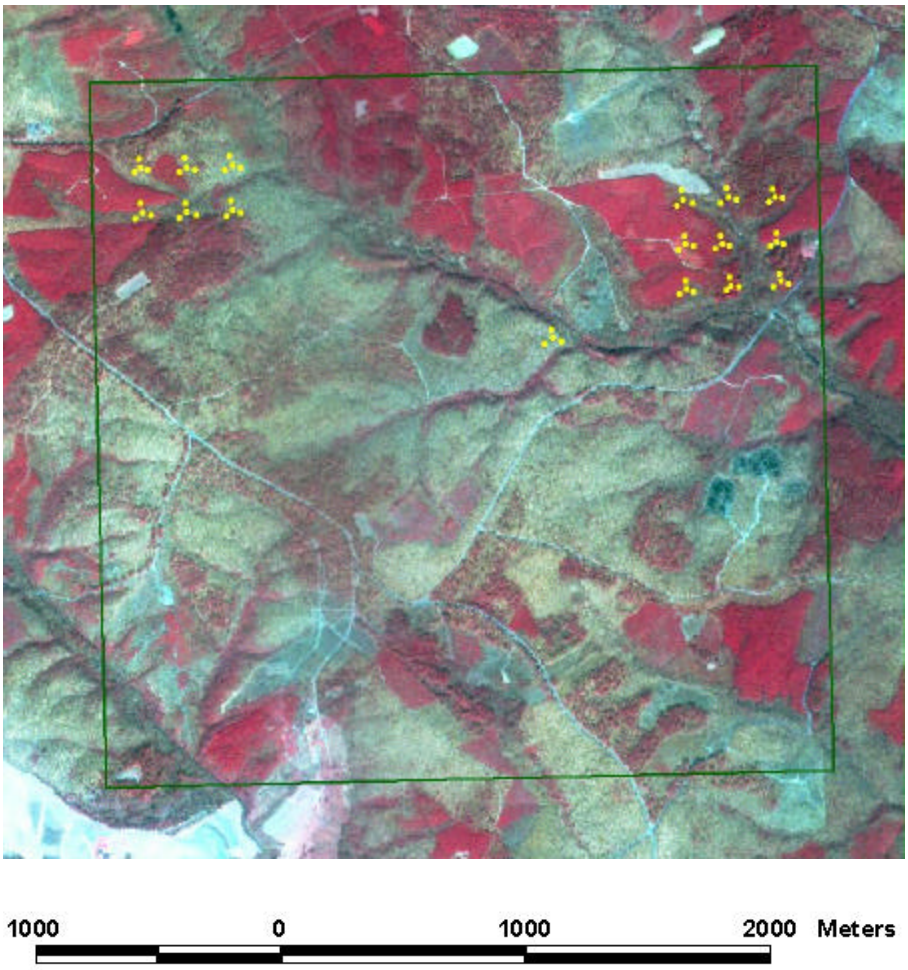


Figure 7. Location of study plots (yellow dots) on a leaf-off color infrared ATLAS image (NASA's Airborne Terrestrial Land Applications Scanner, 4 m resolution, 1998). The green square shows the lidar data coverage. (Copyright 2002, American Society for Photogrammetry and Remote Sensing, 2002 Annual Conference Proceedings)

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 0.01 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 (Appendix 2) 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 4.

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% of the total height; tree mapping: ± 3 degrees for azimuth, 0.3 m (1.0 ft) for distance; dbh: 0.25 cm (0.10 inch) for trees 50.8 cm (20 inches) or less, 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 at 0.6-0.9 m (2-3 ft).

Table 4: 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		
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 plots		
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

4.2.2. Forest biomass

The most common technique for deriving forest biomass is through the use of regression and destructive sampling. Sample trees are measured standing and then cut and weighed. The mass of the components of each tree is regressed to one or more dimensions of the standing tree. Equations of several different forms are used to predict biomass (y) from dbh or dbh and height (x). The most common forms are allometric ($y=ax^b$), exponential ($y=ae^{bx}$), and quadratic ($y=a+bx+cx^2$) (Tritton and Hornbeck, 1982). Biomass equations for individual trees for some of the species growing in the United States can be found in Hahn (1984), Smith (1985), Briggs *et al*. (1989), and Perala and Alban (1993). The tree is normally separated into three above ground components: (1) bole or main stem, (2) bole bark, and (3) crown (including branches and foliage). Some studies (e.g., Kurz *et al.*, 1996) also consider the below ground biomass incorporated in the stump and the major roots. All methods for estimating stand biomass

involve, at least in their developmental stages, a prediction of individual tree biomass (Parresol, 1999). It is customary to report biomass results as biomass density reported per unit area in tons per hectare (also expressed as Mg ha^{-1}) and thus, sample plot summations of individual trees biomass estimates are reported per unit area. The fresh biomass of an individual tree may be determined by weighing its components using field scales or by sampling. Oven-dry biomass is calculated by oven drying wood samples and determining the specific gravity. Green biomass is usually converted to dry biomass, because it can vary with environmental moisture conditions. For large trees, like the trees found in the Appomattox-Buckingham forest, such methods are time consuming and very laborious. As it is out of the reach of the current study to estimate biomass using destructive methods, the biomass will be estimated based on previously developed models for the southeastern forests in the United States. Schroeder *et al*. (1997) presented a general methodology for using FIA data to reliably estimate aboveground biomass density for temperate broadleaf forests in the United States and to develop expansion factors for converting volume directly to biomass from USDA FIA data. Forest volume inventories emphasize only the commercially valuable wood. Thus, it is necessary to develop biomass expansion factors that convert volume to biomass and account for noncommercial tree components, such as branches, bark, and foliage. Biomass density calculated using expansion factors is also useful for analyzing forest carbon budgets. The study of Schroeder *et al*. (1997) was based on data assembled from 12 published and unpublished studies of aboveground biomass of eastern U.S. forests composed of deciduous and coniferous species that are also common in the Appomattox-Buckingham forest. They analyzed 454 trees of 34 hardwood species, including maples and oaks. The conifer data contained 83 trees of 5 species, the majority being pine trees. In all studies, sample trees were selected from fully stocked stands to represent a wide range of tree diameters, from 1.3 to 85.1 cm for hardwood species, and 2.5 to 71.6 cm for conifer species. They used standard destructive sampling methods to determine aboveground biomass for each sample tree. All hardwood species were pooled together in one data set and the conifer species into a second data set, in an attempt to develop general equations that could be applied to forest inventory data classified by forest type. Although both height and diameter data were available for biomass prediction, their study showed that

height did not significantly improve models based on diameter alone. The same conclusion was reached by Crow (1971) for estimating biomass in natural stands of jack pine (*Pinus bancksiana* Lamb.), which found that the most reliable independent variable in regression equations is the diameter at breast height. The final models obtained by Schroeder *et al*. (1997) relating dbh (cm) to biomass (kg), with the smallest standard error and well-distributed residuals, together with their R^2 values are given below:

Deciduous: $\text{Biomass} = 0.5 + \frac{25000 dbh^{2.5}}{dbh^{2.5} + 246872}$ R^2=0.99 n=454 [1]

Coniferous: $\text{Biomass} = 0.5 + \frac{15000 dbh^{2.7}}{dbh^{2.7} + 364946}$ R^2=0.98 n=83 [2]

Equations [1] and [2] above were used to estimate individual trees aboveground biomass based on dbh measurements from ground inventory data. Tree biomass estimations were summed per FIA subplot and regressed against lidar measured tree dimensions. The errors associated with these estimates derive from the equations above, sampling errors and ground measurement errors. A summary of statistics for the field-calculated biomass is given in Table 6.

4.2.3. Stand volume

Total tree volume was calculated from the ground measurements for each inventoried tree and subplot volume was derived as the sum of individual trees within the plot. For all tree species, volume equations used to calculate total outside bark tree volume were of the form (Schumacher and Hall, 1933):

$$V_t = \beta D^{\gamma} H^{\delta} \qquad [3]$$

where: V_t = total outside bark volume;

b, g, d = parameters usually estimated from the data;

D = diameter at breast height (dbh); and

H = total tree height.

For loblolly pines, tree volume was calculated using equation [3] with the requirement that $\delta + \gamma = 3$. The dimensionless parameters were estimated by Sharma and

Oderwald (2001): $\beta = 0.83937$ and $\gamma = 2.18530$ (D and H are measured in the same units).

For hardwood species with dbh less than 28 cm (11 inches) and southern pines, including Virginia pines and shortleaf pines, with dbh less than 12.7 cm (5.0 inches), volume was calculated using the equation:

$$V_t = b(D^2H)^g \qquad [4]$$

For southern pines, parameter values for equation [4] can be found in Saucier and Clark (1985), while the same parameters for hardwoods species are given in Clark *et al.* (1986) (Table 5). Both references require dbh measured in inches and total height in feet.

For southern pine trees with dbh larger than 12.7 cm (5.0 inches) and hardwoods species with dbh larger than 28 cm (11 inches), the volume equation had the form:

$$V_t = b(D^2)^g H^d \qquad [5]$$

Parameters for equation [5] for pine and hardwood species were found in Saucier and Clark (1985) and Clark *et al.* (1986), respectively, and they are given in Table 5.

Table 5: Equation parameters for volume equations

Species	**Equation parameters**				
	Dbh less than: • 12.7 cm (pines) • 28 cm (hardwoods)		Dbh larger than: • 12.7 cm (pines) • 28 cm (hardwoods)		
	β	γ	β	γ	δ
Southern pines	0.00211	1.01241	0.00199	1.03101	1.01241
Red maple	0.00402	0.93484	0.00817	0.78674	0.93484
Sweetgum	0.00354	0.94353	0.00245	1.01987	0.94656
Yellow poplar	0.00430	0.93475	0.00347	0.97925	0.93475
Hickory species	0.00481	0.91795	0.00248	1.05655	0.91795
Chestnut oak	0.00301	0.96996	-	-	-
Southern red oak (including black oak and northern red oak)	0.00409	0.93293	0.00329	0.97797	0.93293
White oak	0.00544	0.90256	0.00293	1.03114	0.90256
Scarlet oak	0.00437	0.92917	0.00247	1.04824	0.92917
Other species (black gum, black cherry, sour wood, dogwood, etc.)	0.00392	0.94065	0.00278	1.00702	0.94065

In addition to volume, basal area was computed for each subplot as the sum of individual-tree basal area values. Descriptive statistics for basal area, volume, and biomass are given in Table 6. Given the dbh (cm), basal area (m^2) for each inventoried tree was calculated with the formula:

$$BA = \frac{pdbh^2}{4(10{,}000)} \qquad [6]$$

Table 6: Descriptive statistics for basal area, volume, and biomass

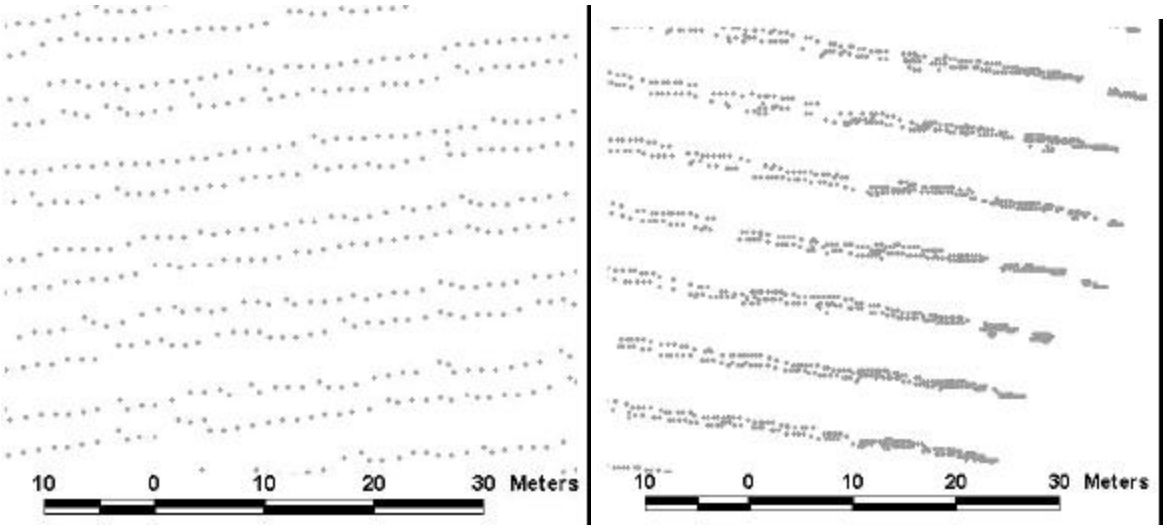
Statistic	Basal area (m^2/ha)	Volume (m^3/ha)	Biomass (Mg/ha)
		Pines	
Mean	20.06	122.88	79.62
Minimum	1.25	3.42	3.42
Maximum	57.19	571.99	314.75
Standard deviation	12.03	112.24	64.61
		Deciduous plots	
Mean	19.85	163.20	131.61
Minimum	2.76	17.51	15.13
Maximum	35.16	403.61	270.90
Standard deviation	8.15	82.90	62.06

4.3. Lidar data set

The lidar data were acquired on September 2nd, 1999, over an area of 1012 ha (2500 acres) located in the Appomattox-Buckingham (AB) State Forest in Virginia, USA. The lidar system (AeroScan, EarthData, Inc.) utilizes advanced technology in airborne positioning and orientation, enabling the collection of high-accuracy digital surface data. The aerial platform was a Piper Navajo Chieftain aircraft capable of carrying aerial cameras, airborne GPS, inertial measuring units, and the lidar sensor. 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 carrier airspeed was between 110-145 knots. The sensor uses a laser wavelength of 1064 μm with a pulse time width of 12 ns. The laser beam divergence was of 0.33 mrad and that

gave a footprint of 0.65 m from the flying height of 1980 m. Figure 8 shows the scanning pattern on the ground.

The AeroScan system is not capable of recording the intensity of the backscattered laser echo, but it is able of recording up to five returns for each laser pulse, depending on the ground cover. The laser point density on the ground, for one swath, is reported to be between 0.007 and 0.5 points/m^2 and for the laser data of AB forest was of 0.47 points/m^2 for the first return, 0.20 points/m^2 for the second return, 0.02 points/m^2 for the third return and 0.0001 points/m^2 for the fourth return. None of the pulses were able to produce a fifth return for the given ground and vegetation conditions. 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. For this study only the first and the last returns were used. The point density for the first or last return translates into an average point distance of 1.5 meters. The mission was designed with up to 70 % side overlap to increase the point density on the ground and to correct for the scanning pattern evident in Figure 8 (a) and (b). The resulting three-dimensional coordinates were compiled in an ASCII mass point file of x, y, z on the UTM projection, zone 17, North American Datum 83 (NAD 83), for each of the laser returns.



a) at the center of the swath b) at the edge of the swath

Figure 8. Lidar scanning pattern on the ground

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 9 FIA-type plots with a mixture of pines and hardwoods stands and is representative for the range of scanning patterns. The grid cell size was 1x1 m, therefore the statistical measures were reported directly per 1 m^2. The area included laser points from 7 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 results are summarized in Table 7. Figure 9 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 0.8 m.

Table 7: Basic statistical measures for the number of lidar points per 1m^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

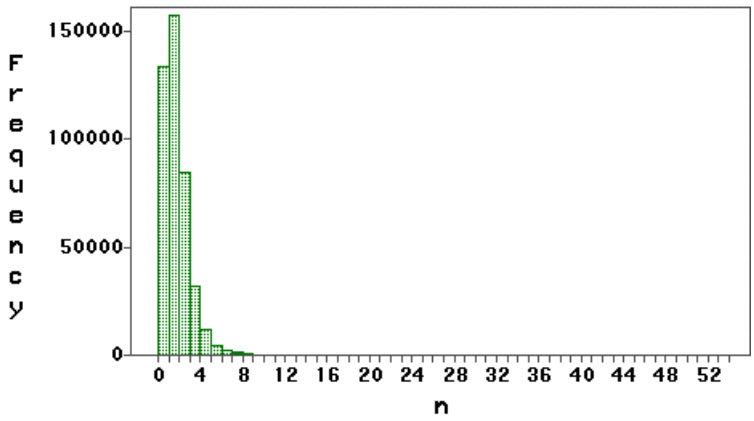


Figure 9: Frequency distribution of the number of lidar points per 1m^2.

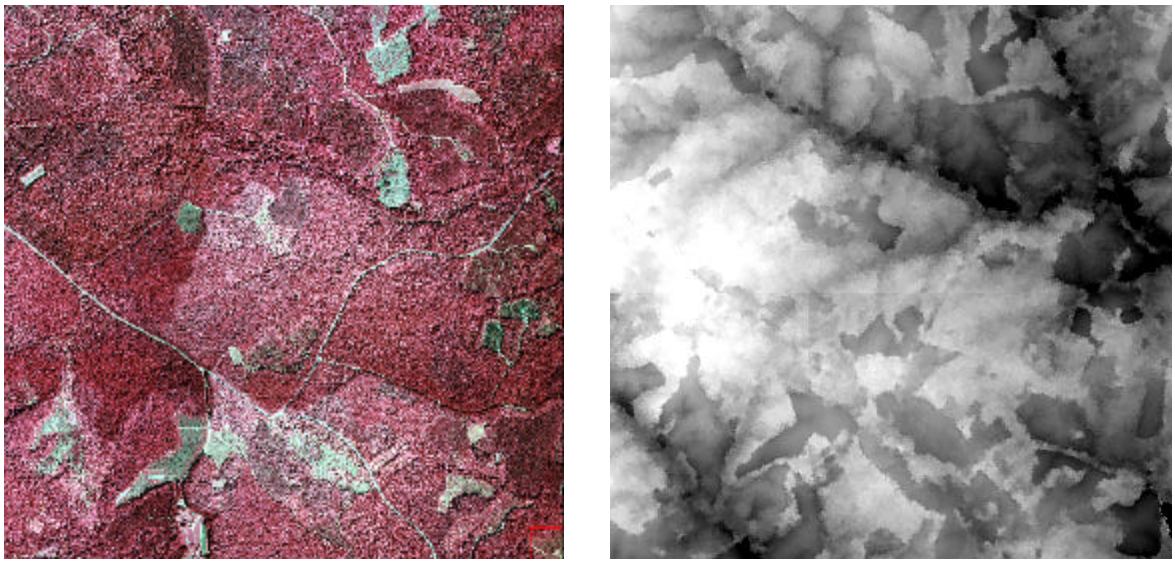
The provider performed an evaluation of the lidar data, including a comparison of the data from flight line to flight line. This comparison showed high relative accuracy and no anomalies in the data. All ranges were post-processed by EarthData, Inc., and

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 ±15 cm and an horizontal accuracy of ±25 cm (EarthData, Inc., 2001).

4.4. 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 March 17, 1998, at 2100 m AGL), shown in Figure 7, and a leaf-on ortho-image (Figure 10 (a)) provided by EarthData, Inc., derived from 1:13,000 color-infrared photography acquired by NASA in the fall of 1999 (0.5 m spatial resolution). A photogrammetrically derived DEM, also based on the NASA infrared photos of 1999, was provided by EarthData, Inc. The grid spacing of the photogrammetric DEM is 10 meters (Figure 10 (b)). The DEM was produced to enable production of the digital orthophoto and therefore, was not designed to be an accurate model of the terrain. A visual analysis of the DEM in Figure 10 (b) indicates that the DEM derived from CIR imagery actually models the top of the canopy and the bare ground elevation similar to the digital surface model of the first lidar return.



a) b)

Figure 10: Ortho-image (color-infrared) of the study area (a) and photogrammetrically-derived DEM (b)

4.5. Ground digital elevation model (DEM)

As previously stated, a basic task before attempting to determine the vegetation height is to characterize the ground elevation. An iterative approach was used to construct the terrain model from the raw lidar data points. The first step of the algorithm for constructing the terrain elevation model overlays a regular grid with a cell size of 10 m over the laser points and identifies the minimum laser point elevation in each cell. The terrain slopes are gentle and a smaller grid cell size is not justified, while a larger size would not adequately characterize the micro-relief. The number of lidar points in each of the 100 m^2 grid cells varies, depending on the spatial pattern of lidar points within each flight line and on the spatial overlay of adjacent swaths (Figure **11**).

For visualization purposes only, the minimum points in each cell were used to interpolate elevation values to a regular grid of 10-meter cell size shown in Figure 14 (a). The location of the DEM relative the whole study area is shown in Figure 15 (a). The interpolation technique was linear kriging implemented in Surfer (Version 7.02, Golden Software, Inc.). Popescu *et al.* (in press) investigated several interpolation techniques from raw lidar points to regular grids, such as kriging, inverse distance, and triangulation,

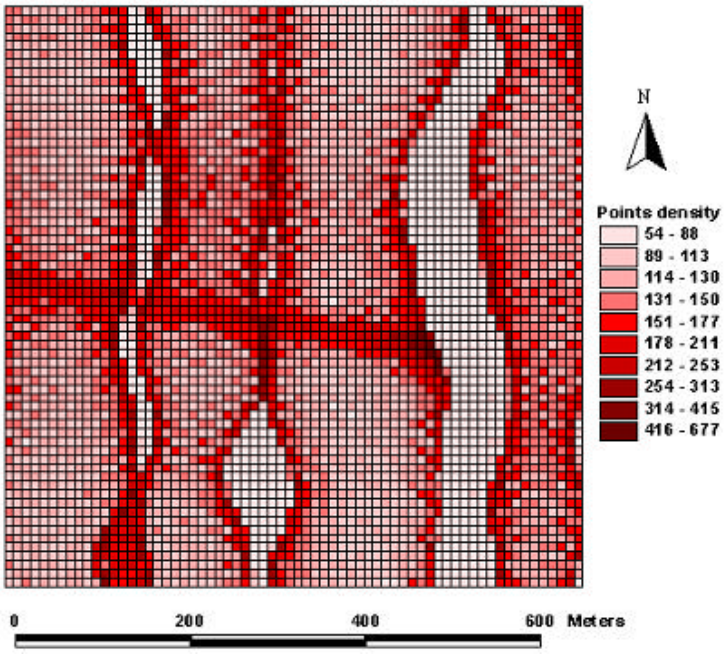


Figure 11: Lidar points density per 100 m^2 grid cells.

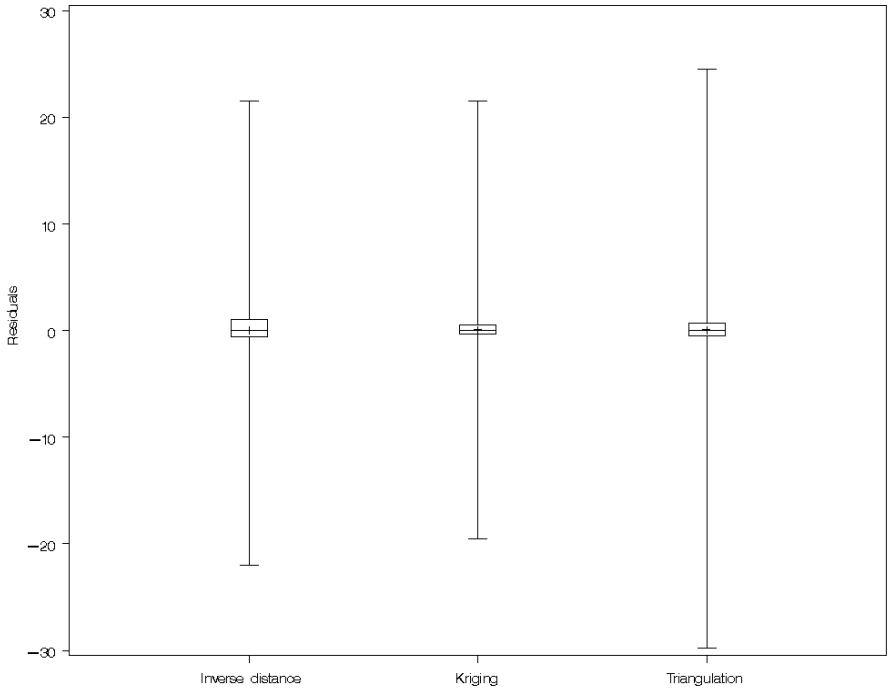


Figure 12: Boxplots of residuals for three interpolation techniques

and found that kriging gave the smallest residuals (Figure 12).

A visual analysis of the DEM in Figure 14 (a) and its shaded relief representation in Figure 14 (c) reveals that some of the lidar points are vegetation points. These points are mainly located over the hardwood trees in the upper- and lower-left portions of the area shown in Figure 14 (b).

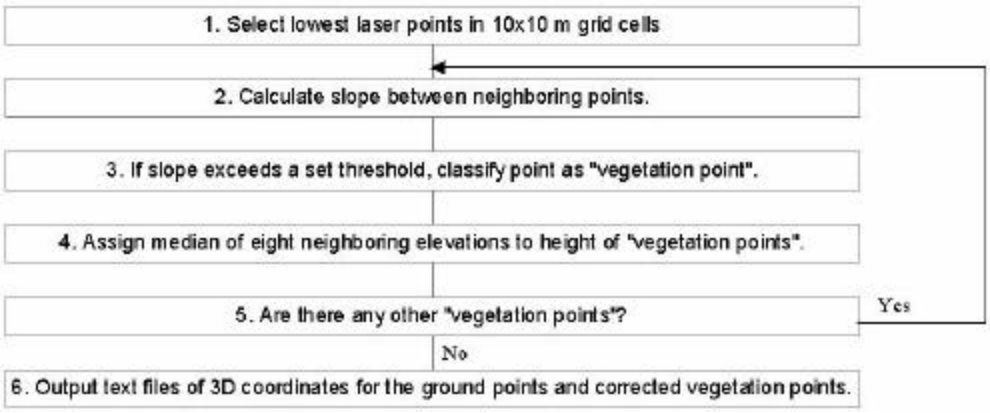


Figure 13: Flow chart of DEM algorithm.

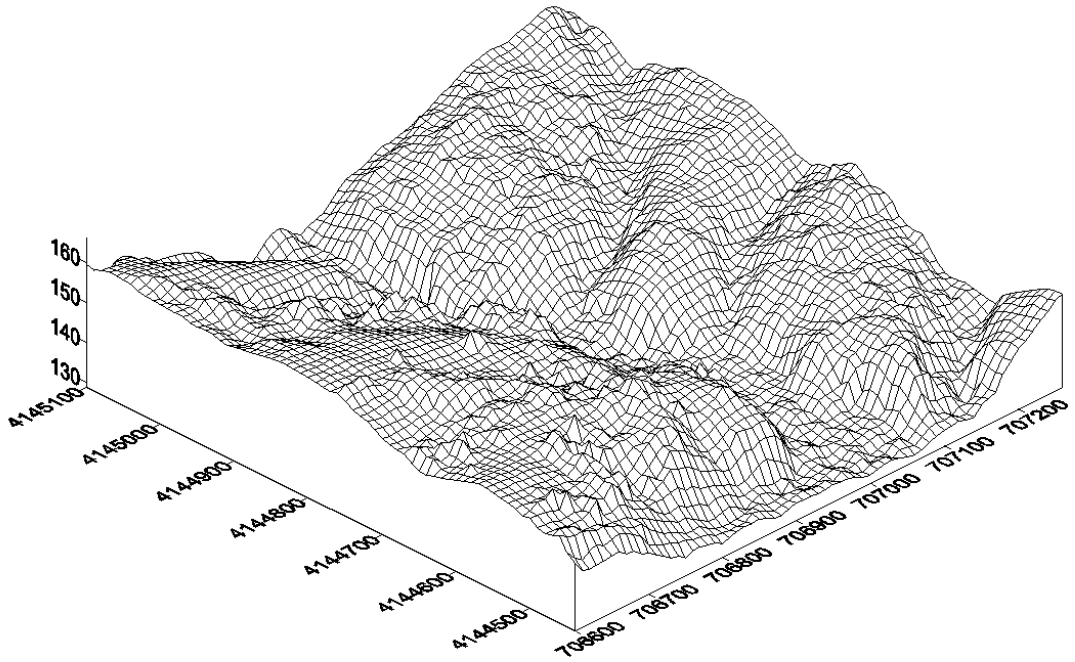
After the raw lidar points are filtered within 10x10 meters cells at the first step, the algorithm (Figure 13) proceeds iteratively and analyzes the slope between lidar points. The algorithm was implemented in IDL (version 5.5, Research Systems, Inc.). The basic assumption of this method is that a large slope between two nearby lidar points is unlikely to be attributed to a steep slope in the terrain, given the known ground conditions. More likely, the higher point is a vegetation hit and therefore needs to be removed. To define the filtering criteria, several methods can be used, such as incorporating simple knowledge about the terrain slope, using a training lidar ground sample, and calculating probabilities in order to minimize the number of classification errors (Vosselman, 2000).

For this study, in order to determine the nearby points that define the slope, the algorithm organizes points in the 10x10 meter grid that was used to initially filter the minimum points. For one lidar point in the central cell of a 3x3 neighborhood matrix, the slope is calculated relative to its eight direct neighbors. The raw lidar points have retained their three dimensional coordinates, therefore the slope is calculated based on the distance and height difference between two points with known coordinates. There is no interpolation involved up to this step, since by interpolating, some information in the original 3D lidar points is lost. It is recommended to use original lidar point values for as long as possible in the processing phase (Axelsson, 1999). When the slope values associated with one point exceed a slope threshold set when initializing the algorithm, the point is classified as vegetation point. Such a point is not entirely eliminated, but its elevation value is corrected. The new elevation value for a vegetation point is the median of the eight elevation values of its neighboring points. The algorithm runs iteratively until there are no more points classified as vegetation hits. For the area shown in Figure 14, the algorithm classified approximately 15 % of the raw lidar input points (minimum elevation points in 10x10 m grid) as vegetation points.

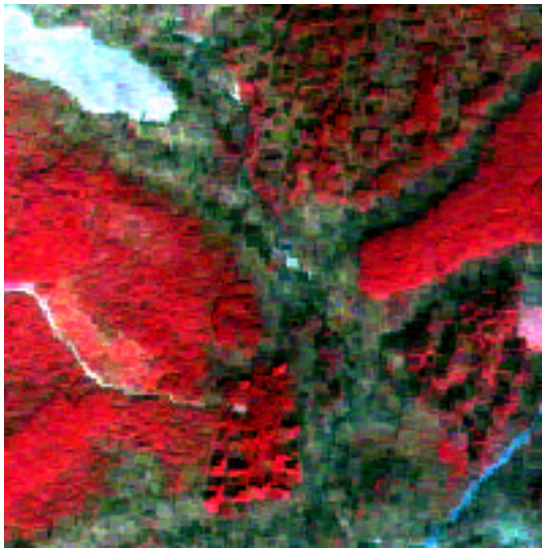
To initialize the slope threshold based on which a point is classified as either ground or vegetation hit, the user should have some prior knowledge of the terrain configuration. The U.S. Geological Survey (USGS) has developed a National Elevation Database (NED) that can be used to investigate the terrain slope. The NED DEMs offer a much-improved base of elevation data for calculating slope and hydrologic derivatives. A 7.5-

minute NED DEM of the study area with a grid spacing of 30 m (Figure 15 (a)) published by USGS in 1999 was used to find prior information about the slope of the terrain (Figure 15 (b)). The maximum slope (35%) was used as the slope threshold with the filtering algorithm.

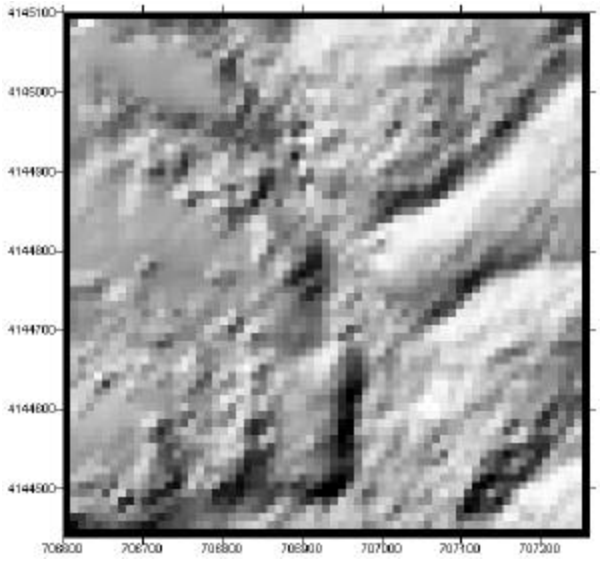
The resulting lidar-derived DEM shown in Figure 17 was interpolated using linear kriging from the lidar points output by the filtering algorithm. The lidar DEM was compared with four other sources of elevation data: the GPS points used to locate the FIA-type ground plots (Figure 7), the 7.5-minute USGS NED DEM of the study area (Figure 15 (a)), the 7.5-minute USGS digital raster graph (DRG), and the photogrammetrically-derived DEM provided by EarthData, Inc (Figure 10 (b)). The vertical datum for the GPS points, the lidar DEM, and the photogrammetric DEM is the World Geodetic System 1984 (WGS84), while for the NED DEM is the Geodetic Reference System (GRS) 80. WGS84 and GRS80 are considered equivalent for the study area. The horizontal datum is the North American Datum 1983 (NAD83), with the exception of the USGS DRG (NAD27). The USGS DRG has 10-feet elevation contours that are referenced to the National Geodetic Vertical Datum of 1929 (NGVD29 – sea level datum).



a)



b)



c)

Figure 14: Raw DEM (a) obtained from lidar points filtered at first step. Axes show coordinates in meters (UTM, zone 17, NAD83 datum). (b) ATLAS leaf-off image over the DEM area. (c) Shaded relief map of the raw DEM.

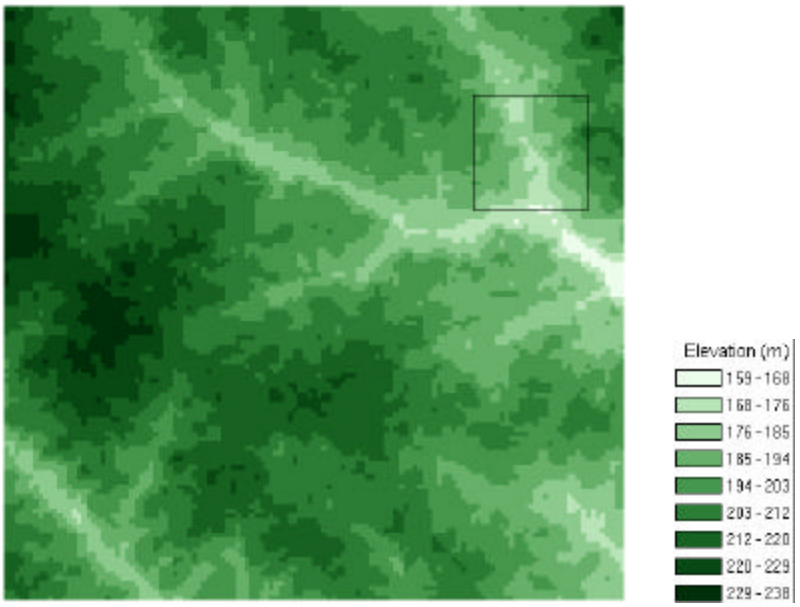
To compare the lidar DEM with the other sources, elevation differences were calculated at 78 point locations distributed over the study area. These points included the location of the 64 subplots and 14 additional points, which were collected following the same procedure (Appendix 1). Half of the 14 points were located with GPS in the open-ground area visible in the upper left corner on the ATLAS image in Figure 14 (a) and 7 points under the forest canopy. The elevation of the USGS DRG for the 78 points was interpolated from the contour lines. First, the DRG was projected to the NAD83 datum and the points were overlaid on the DRG (Figure 16 (a)). Interpolated elevation values were converted from NGVD29 to ellipsoid heights (WGS84) by using the VERTCON and GEOID99 programs of the National Geodetic Survey (NGS) Geodetic Tool Kit (National Geodetic Survey, 2001). Therefore, when calculating the elevation differences, each of the elevation sources was referenced in the same units (meters), horizontal and vertical datum, and projection. The root mean square error (RMSE) of the elevation differences was calculated using the following equation:

$$RMSE = (\sum (Z_l - Z_{es})^2 / n)^{0.5} \qquad [7]$$

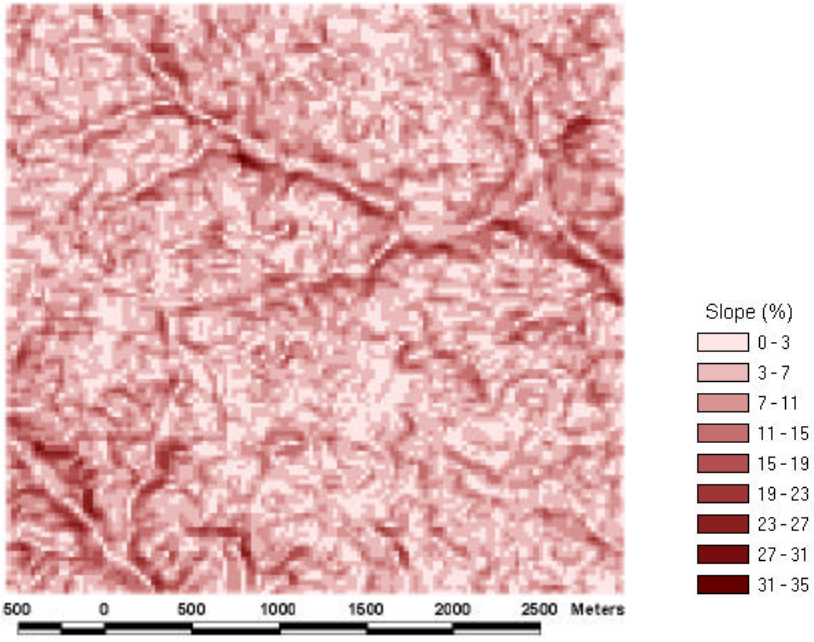
where: Z_l = the lidar elevation;

Z_{es} = the elevation of either GPS points, USGS DEM, or photogrammetric DEM;

n = the number of checkpoints.

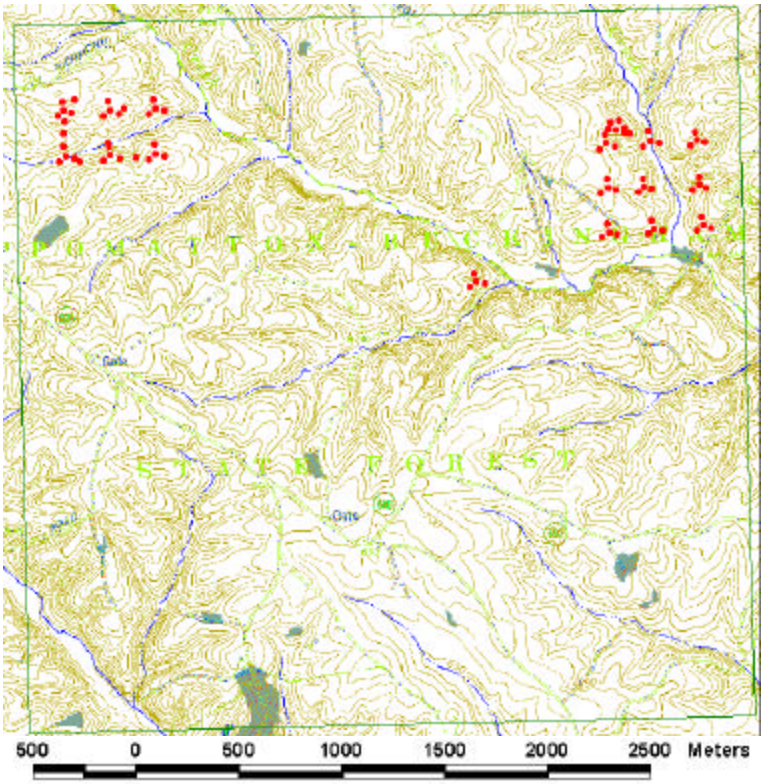


a)

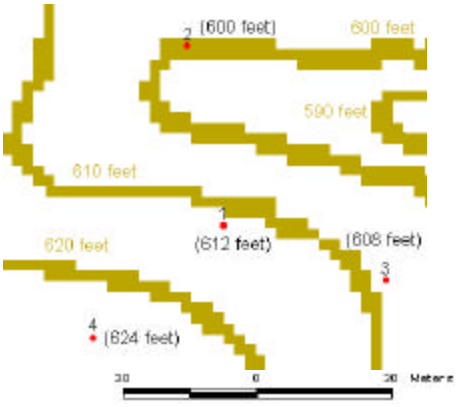


b)

Figure 15: (a) NED DEM of the whole study area; the square shows the location of the lidar –derived DEM in Figure 14; (b) slope image of the NED DEM.



a)



b)

Figure 16: (a) Portion of the Appomattox USGS DRG and the 78 GPS control points (shown in red); (b) interpolation of elevation values (shown in parenthesis) for 4 points from 10-feet contours.

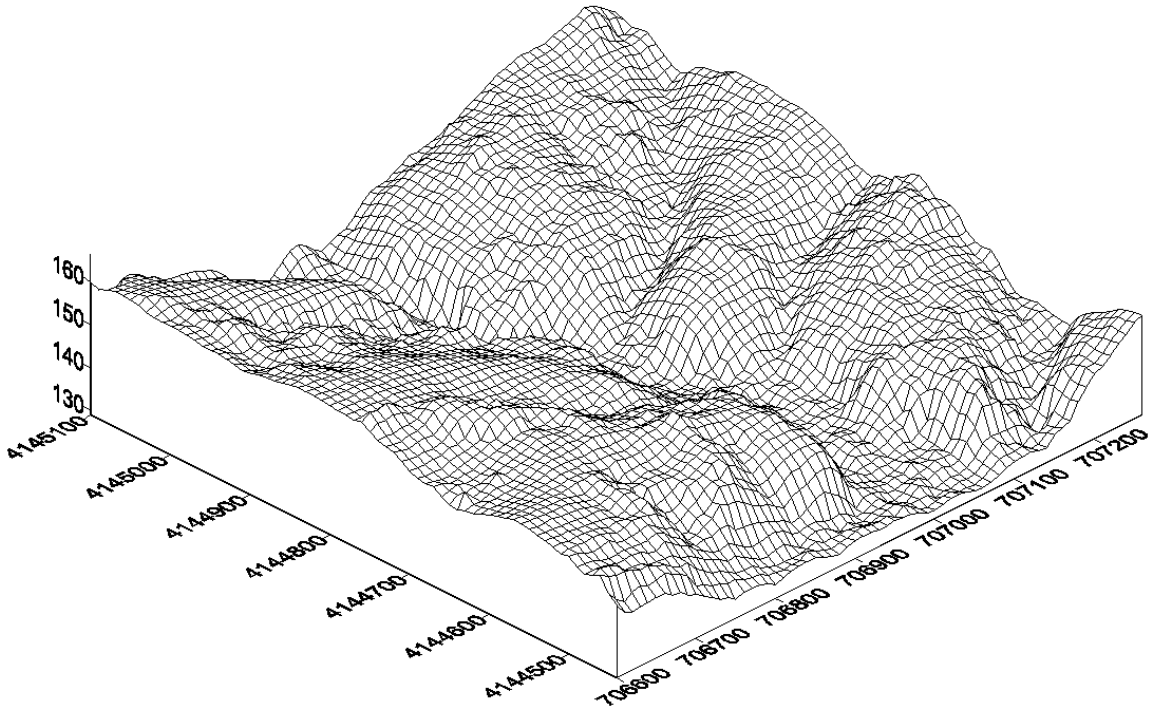


Figure 17: Final DEM (10x10 meters) of the area shown in Figure 14.

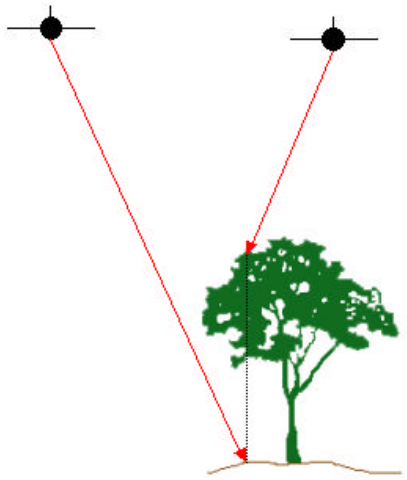


Figure 18: Difference in elevation over the same horizontal area due to combining data from adjacent flight lines.

4.6. Canopy Height Model (CHM)

The tree canopy height model was computed as the difference between tree canopy hits and the corresponding DEM values. Tree canopy hits or first-return lidar points are usually interpolated to a regular grid that corresponds to the digital surface model. 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 meters 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 from adjacent flight lines. Laser beams that fall over the same horizontally projected area from distinct flight lines have different incident angles and therefore, they could penetrate to various heights under the tree canopy (Figure 18).

When situations like the one depicted in Figure 18 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.

To obtain the tree canopy height model (CHM), the terrain elevation was subtracted from the top DSM. Figure 19 (a) shows a grayscale image of a portion (36.5 ha) of the CHM that includes pine plantations, shown in darker shades, and deciduous trees that are usually higher and are depicted with lighter shades of gray. The ground photo was taken in the leaf-off season and a hardwood stand is visible to the left of a fire line and a pine plantation to the right.

4.7. Tree dimensions

4.7.1. 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

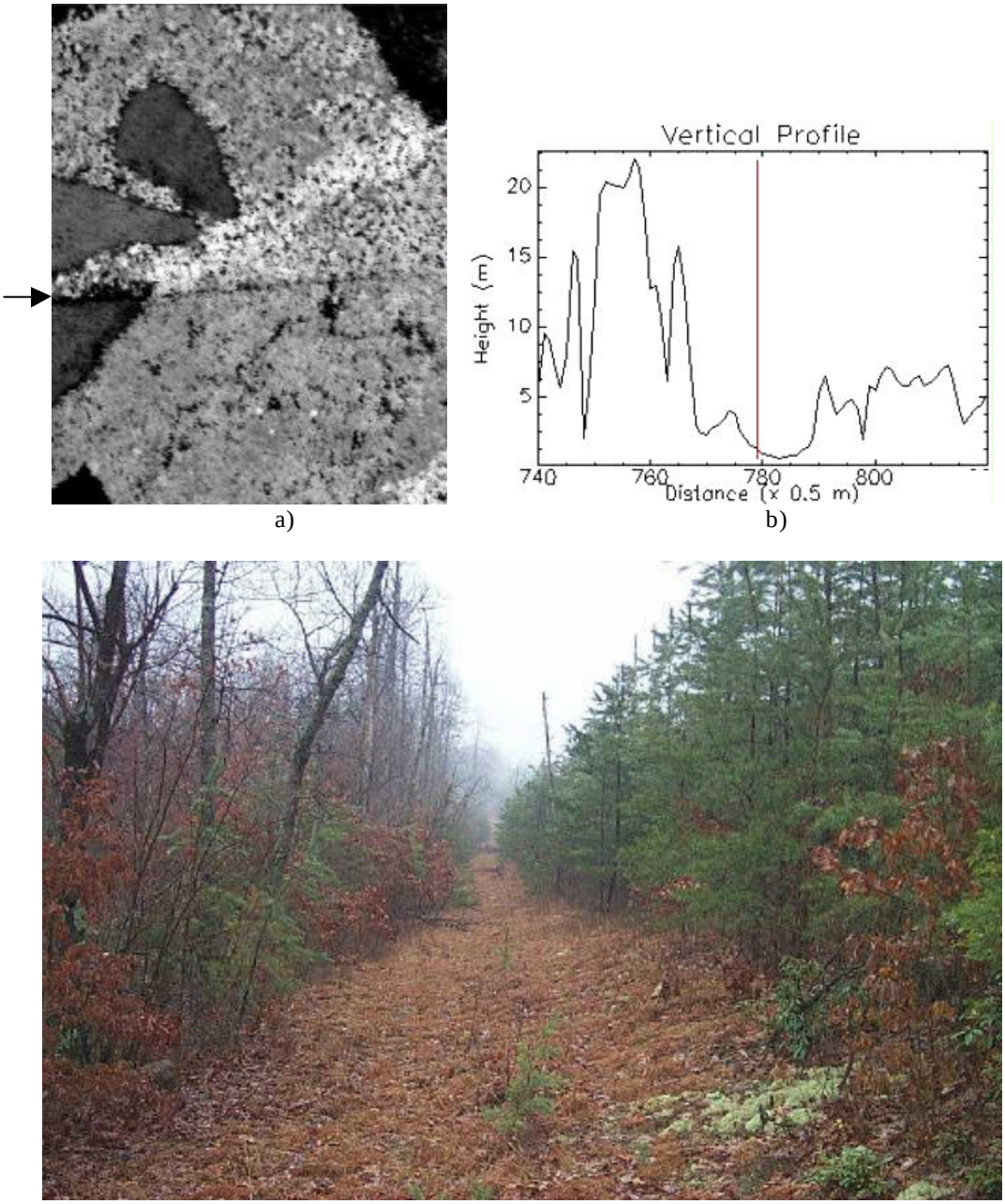


Figure 19. Portion of the canopy height model (a) and the vertical profile through the CHM (b). The ground photo (c) shows the location of the vertical profile through the CHM. The arrow to the left of the CHM image indicates the direction of sight.

diameter. Since 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 Figure 7 was used to differentiate between the two categories of species, deciduous and coniferous.

The multispectral ATLAS image was acquired in the leaf-off season of 1998. Only the first 8 bands covering the visible, near- and mid-infrared portion of the spectrum, were used in the classification process (Table 8).

Table 8: ATLAS bands used in the classification process

Band	Spectral coverage (mm)
1	0.45 – 0.52
2	0.52 – 0.60
3	0.60 – 0.63
4	0.63 – 0.69
5	0.69 – 0.76
6	0.76 – 0.90
7	1.55 – 1.75
8	2.08 – 2.35

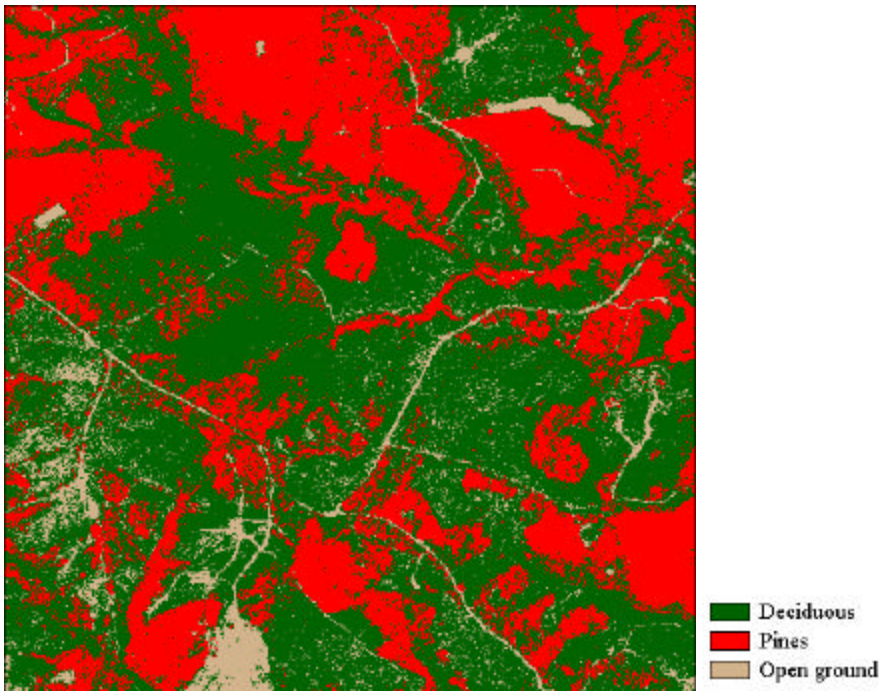


Figure 20: Classified ATLAS image

The image was classified into three classes – open ground, deciduous, and coniferous, using a supervised approach and the maximum likelihood classification decision rule (Figure 20). 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 co-dominant trees from the 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 trees dominant and co-dominant trees were selected for the accuracy assessment, including 133 deciduous and 292 pines. In addition, the seven GPS points collected in open ground were added to the reference data. Table 9 shows the error matrix associated with the classification process. The accuracy report is given in Table 10.

Table 9: 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

Table 10: 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						

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% is actually covered by pine species, while of the deciduous area, 72.39%. 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 (Campbell, 1996, p. 389). The 0.66 value for pines and 0.60 for deciduous species would imply that the classification process achieved an accuracy that is, respectively, 66% and 60%, better than it would be expected from chance assignment of pixels to the 3 categories.

The ATLAS image was georeferenced to the color-infrared ortho-image, and therefore spatially co-registered to the lidar data. The initial 4 meters 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 3 classes allows a pixel-level data fusion with the lidar surface. For a visual analysis of the co-registration between lidar data and the ATLAS image, the multiband image was draped over the three-dimensional CHM, as shown in Figure 21. Large individual deciduous crowns are visible in the foreground in Figure 21, 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.

4.7.1. Tree heights

Popescu *et al.* (in press) 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 that makes use of all laser height values within the subplots. For their study, Popescu *et al.* (in press) used a variable window of a square shape. A similar technique with variable window size and texture analysis was used by Daley *et al.* (1988) 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 3x3, 5x5, and 7x7 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 x n window and a circular window that is more appropriate to identify tree crowns. The algorithm (Appendix 3) was implemented in IDL Version 5.5

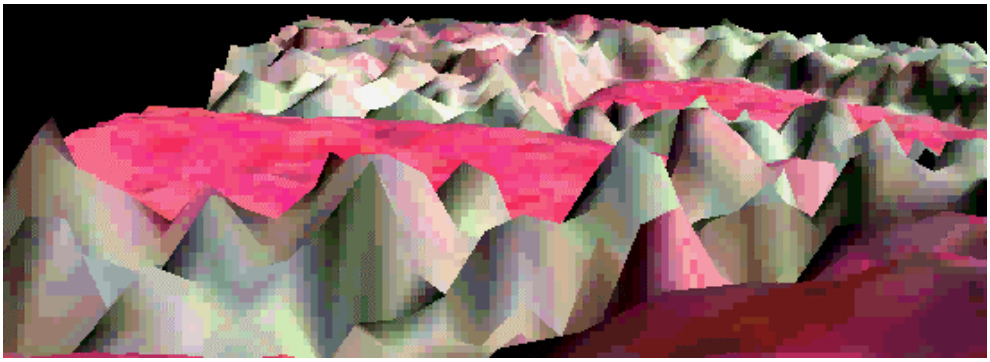


Figure 21: Multispectral ATLAS image draped over the CHM

(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 respectively, omission 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 to 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), 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 and nonlinear regression models were tested for deciduous trees, pines, and the combined data.

The relationship between crown width and tree height was investigated separately for the two species categories, pines and deciduous trees, and also for the combined data set. 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:

Deciduous: $Crown\ width = 3.09632 + 0.00895\ H^2$ ($R^2 = 0.54$, $s_{y.x} = 1.49$) [8]

Pines: $Crown\ width = 3.75105 - 0.17919\ H + 0.01241\ H^2$ ($R^2 = 0.58$, $s_{y.x} = 1.20$) [9]

Combined: $Crown\ width = 2.51503 + 0.00901\ H^2$ ($R^2 = 0.59$, $s_{y.x} = 1.45$) [10]

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, as height proved to be non-significant at 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 **[8]**, **[9]**, and **[10]** above, the window size varied between 3x3 and 31x31 pixels, which corresponds to crown sizes between 1.5 m to 15.5 m. The maximum crown diameter measured on the ground belonged to a white oak tree and was 13.8 m. In the case of the circular window for the LM filter (Figure 22), the window diameter varied between the same limits mentioned above for the size of the regular square windows.

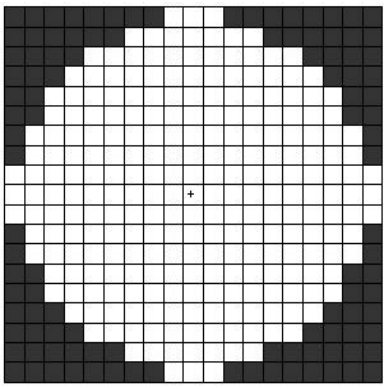


Figure 22: Circular window (white background) compared to a square window (19x19 pixels – 9.5 x 9.5 m).

The algorithm (Figure 23) reads the height value at each pixel and calculates the window size to search for the local maximum. If the current pixel corresponds to the local maximum, it is flagged as a tree top (Figure 25). Once the location of each identified tree crown has been established, the canopy 3-D surface of laser heights (CHM) is sampled only at the positions of the tree 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 concept of variable windows is illustrated in Figure 24, which shows a portion of the CHM with the filtering windows that identified tree tops.

The algorithm was run with both circular and square window LM filters, with and without data fusion with optical data. When no optical data was used to differentiate between deciduous and pines when calculating the width of search window, the filter size was calculated based on the relationship between height and crown size derived from all inventoried trees (Equation **[10]**).

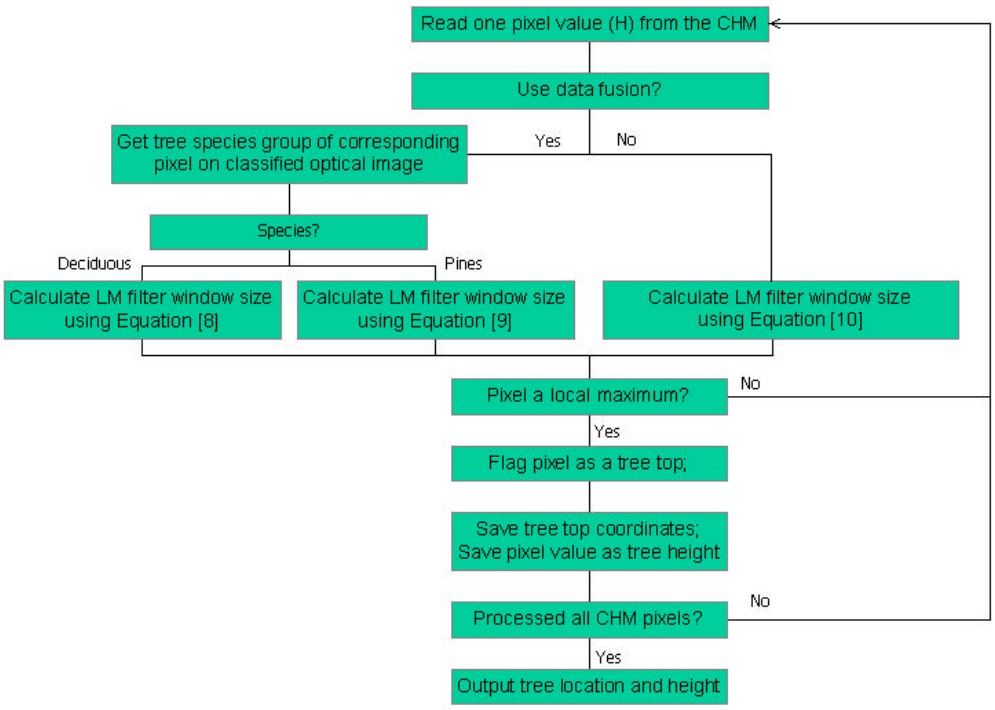
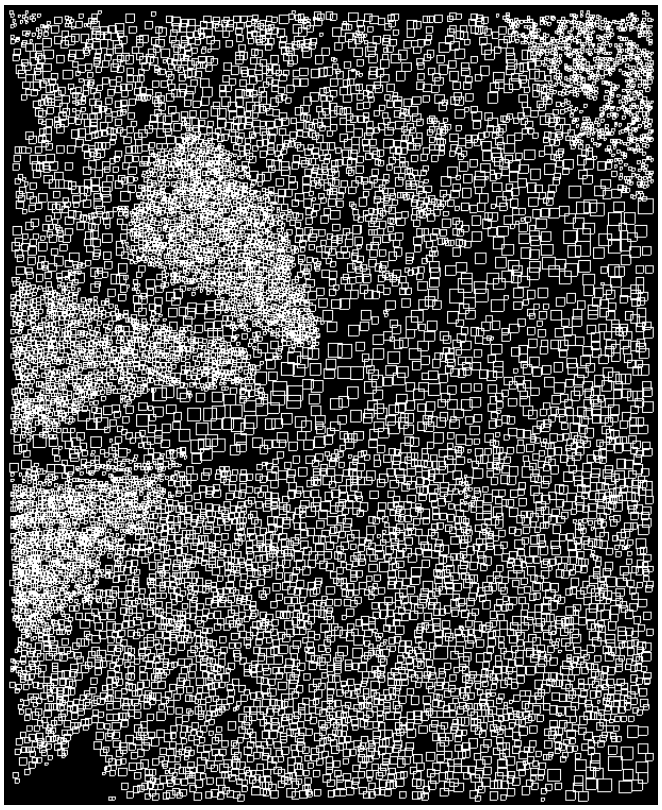
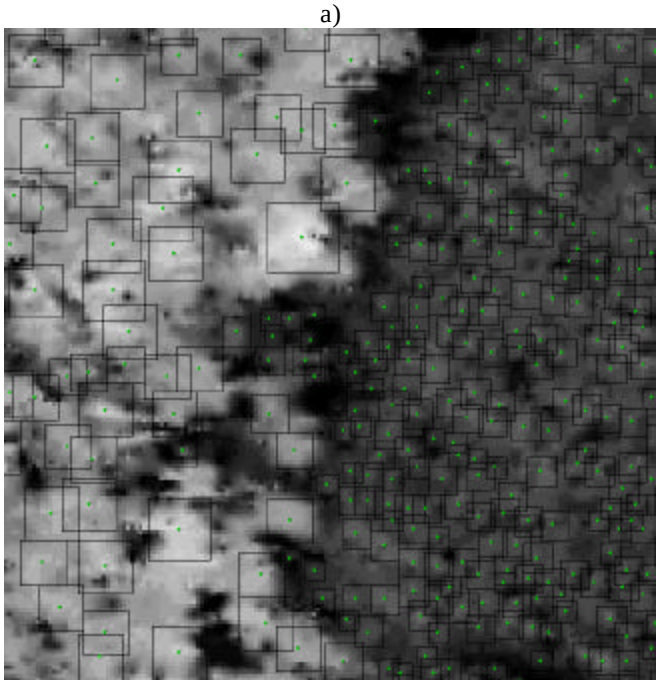


Figure 23: Flow chart of the algorithm for locating trees and measuring height



a)



b

Figure 24: Portion of the CHM variable windows (a) and tree tops (b)

4.7.2. Stand density

The variable window size LM technique that identifies tree tops was also used to estimate the number of trees per plot and, thus, the stand density. The total number of local maxima within one plot is an indicator of the number of stems per plot. Lidar estimated stand density was compared to the FIA field data at the subplot level.

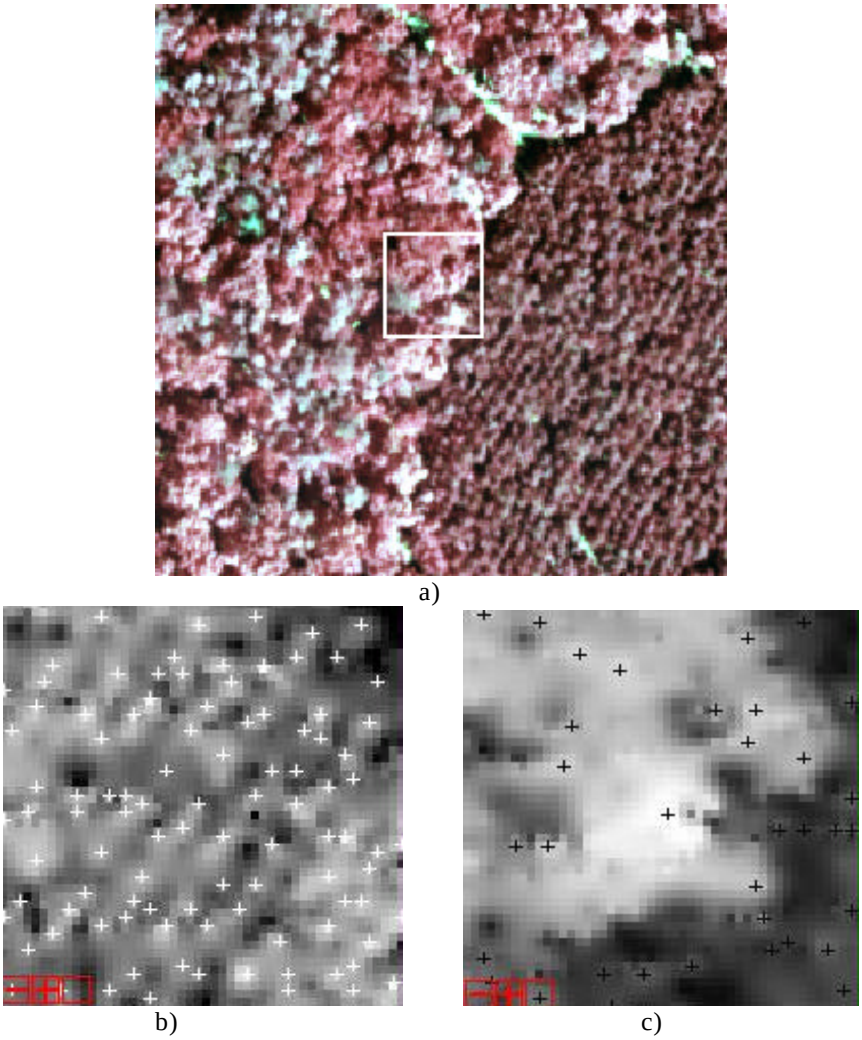


Figure 25. Ortho-image (a) and tree tops identified in the pine plantation (b) and the pine-hardwood mixed stand next to it (c). Rectangle on the ortho-image shows approximate location of zoom window c). Plantation row pattern oriented SW-NE is visible in a) and b). (Copyright 2001, American Society for Photogrammetry and Remote Sensing, 2001 Annual Conference Proceedings)

4.7.3. Crown width

In a simulation study, Popescu *et al.* (2000) derived the average crown width using the canopy closure and stand density. The average crown width can be estimated by dividing the canopy area, i.e., number of laser canopy hits multiplied by the area of one pixel of the interpolated canopy height model, to the number of stems. The average crown diameter for the whole stand is then computed assuming a circular crown with the area equal to average horizontal canopy area. Other approaches to the detection of individual tree crowns using high spatial resolution optical imagery are valley following (Gougeon, 1999), edge finding (Pinz, 1999), morphology (Barbezat and Jacot, 1999), semivariograms and slope breaks (Wulder *et al.*, 2000).

The algorithm developed for this study uses the location of individual trees identified with the LM filter. A 3x3 median filter is used with the CHM to avoid some of the noise in the highly complex surface representing the top of the canopy. The median filter was favored, since it is useful for noise suppression without affecting original values in the CHM. Also, it is an edge preserving filter (Erdas Imagine, 1997, p. 192), better suited for

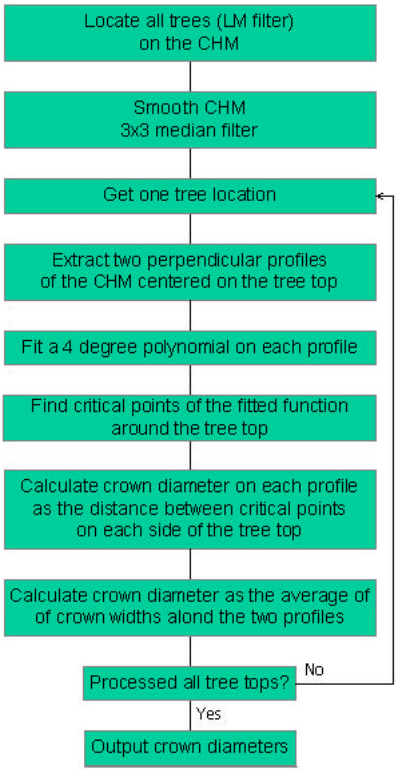


Figure 26: Flow chart of algorithm for measuring crown diameter

conserving the delineation between adjacent tree crowns.

The crown diameter is the average of two values measured along two perpendicular directions from the location of the tree top. To describe the crown profiles along the two directions on the CHM, the algorithm fits on both profiles a four-degree polynomial with least squares by use of the singular value decomposition (SVD) method (Press *et al.*, 1992, p. 676). The length of each of the two profiles is limited to twice the window size and is centered on the tree top. The four-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). The values of the independent variable at extreme functional values are known as the critical points. The independent variable in this case is the distance along the vertical profile through the CHM and the dependent variable is the CHM height. The sign of the first derivative indicates whether the graph of the fitted function is rising or falling. The first derivative is equal to zero at extreme values. 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 point is a local maximum or minimum. Points of inflection occur where the concavity of the fitted function changes. The algorithm (Figure 26) finds the critical points of the fitted function and analyzes the extreme values they yield, based on the first and second derivatives. Numerical differentiation with 3-point, Lagrangian interpolation is used to find the first and second derivatives in IDL.

The fitted function follows closely the vertical profile of a tree crown (Figure 27) and its graph has a maximum in the neighborhood of the tree top, where the first derivative equals 0 and the second derivative is negative. Points of inflection occur on the edges of a crown profile. When these conditions are met, i.e., the fitted function indicates a tree crown profile, the distance between critical points is used to calculate the crown diameter. The final value for a crown diameter is computed as the average of the crown diameters measured on the two perpendicular directions or profiles. Due to the complexity of the CHM, sometimes the first and second derivatives cannot provide real solutions and crown diameter cannot be measured. For 4.49 % of the trees identified on the three-dimensional lidar CHM in an area with both deciduous and pine trees, the

crown diameter could not be measured. As expected, in an area covered only by large deciduous trees with a complex spatial interaction between neighboring crowns, for 8.78 % of the tree tops identified by the LM filter the algorithm could not calculate the crown diameter. These trees that do not have a lidar measurement for crown diameter are ignored when computing average crown diameter per plot. This method seems appropriate to measure crown diameter for dominant and co-dominant trees that have individualized crowns on the CHM surface. This algorithm measures non-overlapping crown diameters, while the field measurements considered crowns to their full extent, therefore measured overlapping crown diameters.

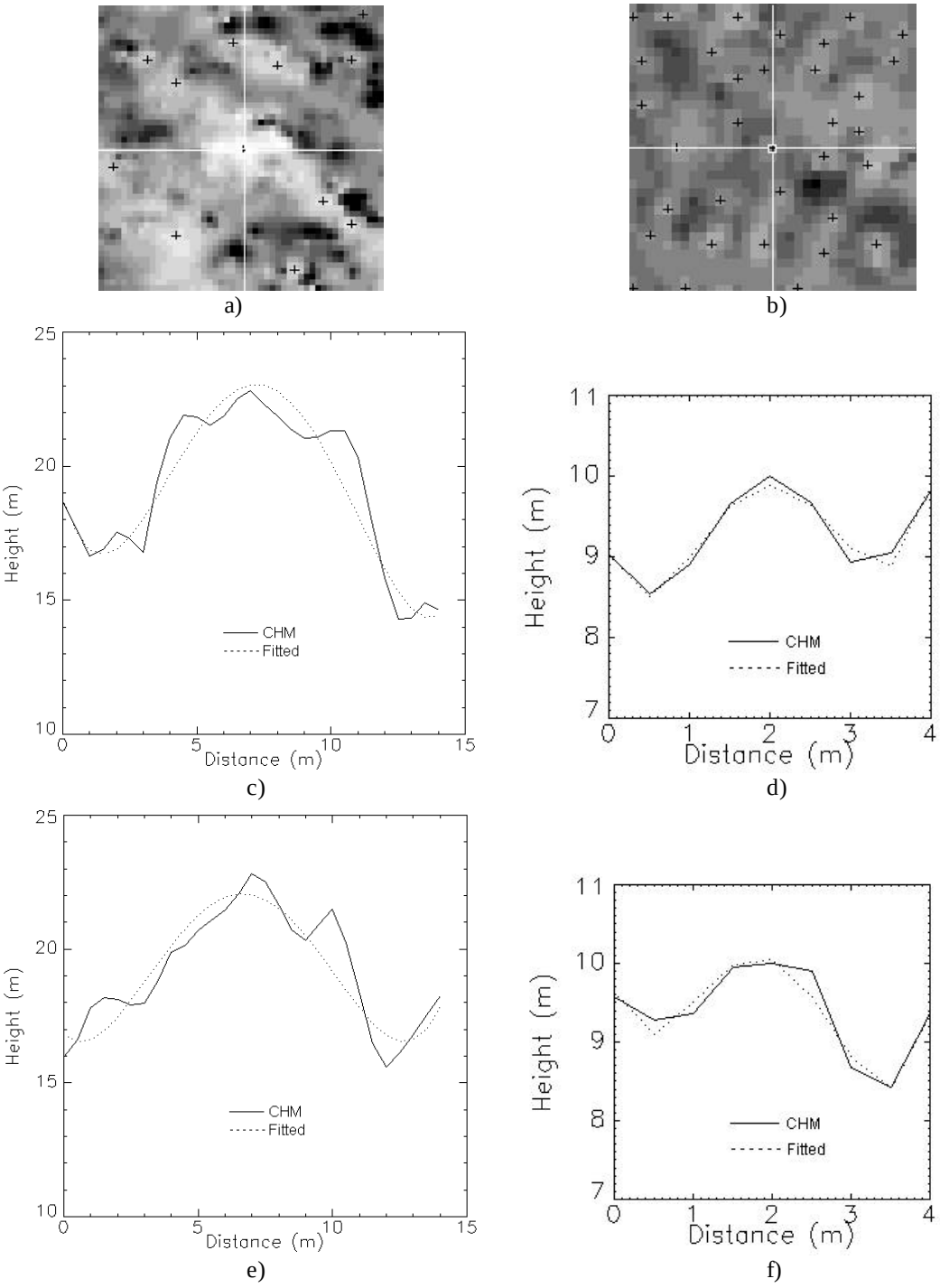


Figure 27: Vertical profiles through the CHM and the fitted polynomials for a deciduous tree and a pine located in the center of the CHM "image" (a) and (b), respectively; (c) and (d) show vertical profiles along the horizontal direction for the deciduous and the pine trees; (e) and (f) are vertical profiles along the vertical direction (deciduous and pine trees, respectively).

4.8. Regression analysis

Linear regression models (Appendix 4) were used to develop equations relating lidar-derived parameters, such as tree height, stand density, and crown width, with field inventory data and field-based estimates of volume and biomass 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 11) 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 volume, basal area, biomass, mean and maximum height, mean crown diameter, number of trees, mean and quadratic dbh (Table 11).

The study of Popescu *et al.* (in press) confirmed that lidar is better suited to measure trees in the upper layer of the canopy, mainly the dominant and co-dominant 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

Table 11. 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 Crown diameter • Average crown diameter / subplot • Minimum crown diameter • Maximum crown diameter • Standard deviation of individual tree crown diameters Number of trees	Tree height • Average tree height / subplot • Maximum height Crown diameter (average / subplot) Number of trees Dbh • Average / subplot • Quadratic mean dbh Basal area Volume Biomass

inch), and (3) dominant and co-dominant 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 (Equation [11]), 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 condition number (f) was calculated with the formula below:

$$f = \frac{l_{\max}}{l_{\min}} \qquad [11]$$

where $l_{\max}$ and $l_{\min}$ are respectively the largest and the smallest eigenvalues.

The use of regression analysis with subplot values raised concerns regarding the possible inflation of the explained variance due to the potential for spatial dependency between subplot values. A semivariogram plot can be used to examine spatial dependence of the ground-measured values of subplot parameters. Figure 28 shows the

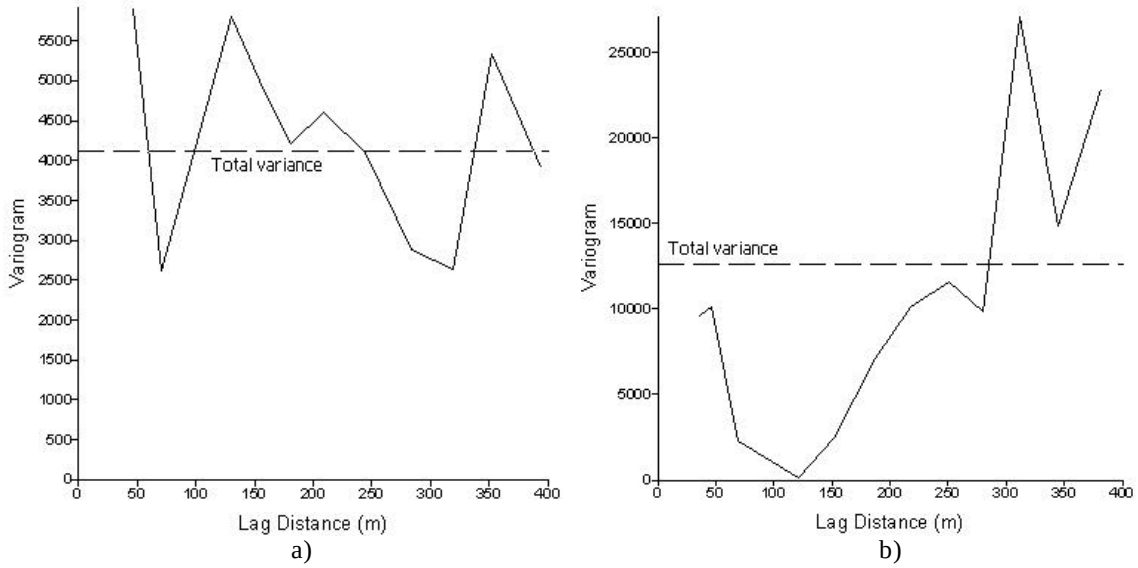


Figure 28. Semivariogram plots for subplot volume for hardwoods (a) and pines (b).

semivariogram plot for the ground-estimated volume for pine and hardwood plots obtained by using a lag distance of 35 m, while the minimum distance between subplots is 36.58 m. The sample semivariogram obtained for hardwood plots (Figure 28 (a)) shows that even at short distances there is still a high degree of variability. The virtually horizontal semivariogram displaying a pure nugget effect for hardwood plots indicates the absence of spatial dependency for volume. A different situation is shown in Figure 28 (b). The sample semivariogram for pine subplot volume indicates that spatial dependency exists at distances less than 300 m. After the dip in semivariance values between 100 and 150 m, the semivariogram increases gradually with distance, reaching the total variance between 250 and 300 m.

To investigate whether the use of subplot values artificially inflates the explained variance with the regression analysis, pine subplots belonging to the same plot were pooled together and plot values were used to regress volume. The independent variables in the regression analysis were the same variables proven significant when performing regression analysis at subplot level. Seven values of plot volume for pines were used to examine how well the lidar variables can predict the plot volume and results were compared with the outcome of the subplot-based analysis.

Since the ground-truth data was 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, was set aside from the sample, and the remaining observations were used to estimate the coefficients for a particular candidate model. The observation previously set aside is then replaced and another observation withheld with coefficients estimated again. 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 ($e_{i,-i}$, $i=1,\ldots n$). These residuals are true prediction errors, since one observation is not simultaneously used for fit and model assessment. The PRESS statistic is defined as:

$$\text{PRESS} = \sum_{i=1}^{n} (e_{i,-i})^2 \qquad [12]$$

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 standard deviation are reported for each model. For the choice of the best model, one might favor the model with the smallest PRESS.

4.8.1. 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 square regression model, not only for regressing maximum height, but also for the rest of the estimated parameters. Externally studentized residuals, also called R-Student, (Montgomery and Peck, 1992, p. 174-177; Neter et al., 1983, p. 406-407) were used for the outliers diagnostic. The R-Student residuals (t_i) are given by:

$$t_i = \frac{e_i}{\sqrt{s^2_{(i)}(1-h_{ii})}} \qquad [13]$$

Where $i = 1, 2, \ldots, n$; i-th observation out of n observations;

e_i = residuals;

$s_{(i)}^2$ = estimate of residuals variance; and

h_{ii} = leverage values or diagonal elements of the hat matrix that allows residuals to be expressed as a linear combination of the observations.

The R-student residuals follow the t distribution with $n - p - 1$ degrees of freedom, where 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, like the deciduous stands. Outliers were also investigated by analyzing the CHM and the ground data to gather nonstatistical evidence for discarding extreme values.

4.8.2. Investigating spatial autocorrelation

When reporting results for estimating forest biophysical parameters, most often through regression analysis, previous lidar studies neglected the investigation of spatial autocorrelation of residuals between lidar and ground-truth data. The errors of estimating ground plot values with lidar are georeferenced observations and measures of spatial dependencies can be used to investigate two aspects: (1) to see whether the spatial pattern displayed by these residuals is significant in some sense and therefore worth interpreting, and, in case it is, (2) obtain information on the factors that might affect it, such as the DEM, CHM, forest horizontal and vertical spatial structure, and anomalies in algorithm performance.

The spatial autocorrelation involves the correlation between values of the same variable at different spatial locations (Bailey and Gatrell, 1995, p. 269). One of the most widely used descriptive statistic for investigating spatial autocorrelation is Moran's I coefficient, defined as (Schabenberger and Pierce, 2001, p. 654):

$$I = \frac{n}{\sum_{i,j} w_{i,j}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} u_i u_j}{\sum_{i=1}^{n} u_i^2} \qquad [14]$$

where: n = number of sites with attribute $Z(s_i)$ (residuals) observed at site s_i (i =1,…,n);

$u_i = Z(s_i) - \overline{Z}$;

$w_{i,j}$ = neighborhood connectivity weights between sites s_i and s_j, with $w_{i,i} = 0$;

Moran's I calculation depends on the criteria for defining the neighborhood connectivity weights. Bailey and Gatrell (1995, p. 261) give several criteria for defining $w_{i,j}$, most of them based on a binary rule and a cutoff distance defining proximity. As a rule, the choice of criteria depends upon the type of data and the mechanism through which spatial dependence might arise. For a forestry situation, the inverse distance criteria shown below is appropriate to quantify proximity and was chosen to define neighborhood connectivity:

$$w_{i,j} = \begin{cases} d_{i,j}^{g} & \text{If the distance } i\text{-}j,\ d_{i,j} < D,\ \gamma < 0 \\ 0 & \text{Otherwise} \end{cases} \qquad [15]$$

The cutoff distance (D) was 300 m, for both pines and hardwoods subplots. This distance is equal to the range of the semivariogram shown in Figure 28 (b). The inverse distance formula ([15]) was calculated for $\gamma = -1$.

In the absence of spatial autocorrelation, I has expected value $E(I) = -1/(n\text{-}1)$. To determine whether I is statistically significant, the Z statistic can be used as follows:

$$Z_{obs} = \frac{|I - E(I)|}{s_I} \qquad [16]$$

Cliff and Ord (1981, Ch. 2.3) gives the formulas for the variance of I (s_I), with two approaches, Gaussian and randomization. In the Gaussian approach, the observed values at site i (i=1…n) are assumed to be independent drawings from a normal population. In the randomization approach, $Z(s_i)$ are considered fixed and are randomly permuted among the n sites. The Moran's I and its interpretation were performed using the SAS®

(version 8.02, SAS Institute, Inc.) macro provided by Schabenberger and Pierce (2001), which calculates I, Z_{obs}, and p-values under both the normality and randomization assumptions (Appendix 4). All these statistics were calculated for each of the biophysical parameters for the two species groups, in each case for only one model. The model was chosen based on the interpretation of the regression analysis and cross-validation. For each model, Moran's I was calculated using the inverse distance (Equation [15]) neighborhood connectivity matrix.

5. Results and discussion

5.1. 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

Table 12: Maximum height values and residuals, in meters, for lidar and ground measurements.

Obs.	Pines*					Deciduous*				
	Lidar	Ground	Fitted	Residual	R-Student	Lidar	Ground	Fitted	Residual	R-Student
1	**7.34**	**9.45**	**7.32**	**2.1318**	**2.4275**	16.73	23.47	18.51	4.9502	1.4167
2	11.42	11.28	11.08	0.1952	0.1975	10.18	17.37	14.82	2.5543	0.7928
3	10.86	10.36	10.56	-0.2078	-0.2107	25.74	22.86	23.62	-0.7550	-0.2039
4	17.16	16.15	16.38	-0.2340	-0.2356	24.61	22.56	22.97	-0.4160	-0.1119
5	14.99	15.24	14.38	0.8594	0.8757	30.63	26.21	26.38	-0.1704	-0.0475
6	22.67	20.73	21.47	-0.7408	-0.7687	20.29	20.42	20.53	-0.1130	-0.0306
7	14.24	14.33	13.68	0.6418	0.6502	24.26	28.65	22.78	5.8719	1.6490
8	**10.67**	**8.53**	**10.39**	**-1.8624**	**-2.0266**	22.53	22.56	21.80	0.7603	0.2046
9	10.58	9.45	10.31	-0.8593	-0.8840	**24.32**	**10.97**	**22.81**	**-11.8420**	**-3.9143**
10	11.41	9.75	11.08	-1.3256	-1.3873	12.44	14.94	16.09	-1.1537	-0.3391
11	7.27	6.10	7.25	-1.1535	-1.2232	**27.87**	**33.53**	**24.82**	**8.7104**	**2.6344**
12	13.08	12.50	12.62	-0.1173	-0.1182	30.70	23.77	26.42	-2.6500	-0.7466
13	10.97	9.75	10.67	-0.9194	-0.9463	31.59	28.96	26.92	2.0367	0.5776
14	13.53	12.19	13.03	-0.8428	-0.8595	26.84	24.99	24.24	0.7529	0.2042
15	10.79	10.06	10.50	-0.4432	-0.4508	21.10	19.51	20.99	-1.4811	-0.4005
16	16.07	15.85	15.38	0.4723	0.4767	21.03	19.20	20.95	-1.7515	-0.4742
17	8.74	9.75	8.61	1.1394	1.1959	22.00	20.42	21.50	-1.0800	-0.2910
18	9.89	9.75	9.67	0.0777	0.0790	18.02	16.15	19.25	-3.0993	-0.8598
19	10.99	10.67	10.69	-0.0178	-0.0181	22.31	22.56	21.68	0.8847	0.2382
20	**19.37**	**20.42**	**18.42**	**1.9958**	**2.1899**	26.66	25.30	24.14	1.1647	0.3159
21	12.68	12.80	12.24	0.5520	0.5594	17.60	15.85	19.01	-3.1618	-0.8805
22	**13.38**	**14.63**	**12.89**	**1.7357**	**1.8548**	27.87	24.69	24.82	-0.1296	-0.0353
23	31.02	30.18	29.18	1.0005	1.1465	29.03	25.91	25.48	0.4344	0.1194
24	30.64	28.35	28.83	-0.4787	-0.5353	27.59	28.04	24.66	3.3788	0.9326
25	30.42	27.74	28.63	-0.8856	-1.0001	24.16	25.30	22.72	2.5785	0.6989
26	23.53	21.95	22.26	-0.3147	-0.3255	24.52	22.25	22.93	-0.6751	-0.1817
27	21.14	20.73	20.06	0.6717	0.6900	23.93	20.42	22.59	-2.1715	-0.5871
28	11.43	11.28	11.09	0.1860	0.1882	**27.42**	**17.07**	**24.57**	**-7.4951**	**-2.1949**
29	14.69	13.41	14.10	-0.6937	-0.7035	20.68	24.99	20.75	4.2364	1.1699
30	20.82	19.20	19.76	-0.5629	-0.5758	22.97	20.73	22.05	-1.3186	-0.3551
31	-	-	-	-	-	22.81	23.77	21.96	1.8119	0.4890
32	-	-	-	-	-	22.48	22.25	21.77	0.4785	0.1287
33	-	-	-	-	-	19.42	18.90	20.04	-1.1410	-0.3104

* Observations in bold were identified as outliers

analysis. As an example, Table 11 shows the results obtained using the circular window filtering method with data fusion. The results obtained with the rest of the methods are shown in Appendix 6. Residuals and R-Student residuals in Table 11 were obtained 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 results for only 30 plots are shown in Table 11. The pine plot with no lidar trees has 5 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 [9]) 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 errors of locating the plot center could have led to this situation.

As explained in section 4.8.1., the R-Student residuals follow a *t* distribution, therefore, the *t* statistic can be used to ascertain outlying values. Tail areas of 5% 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 via hypothesis testing based on the *t* distribution, 4 of the pine plots and 3 of the deciduous plots were identified as outliers (shown in bold in Table 13). 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 13. The other three methods of processing the CHM gave similar results when comparing regression models with and without outliers (Appendix 7).

Table 13: Summary statistics for regression results with and without outliers

Statistics	**With outliers in**		**With outliers out**	
	Pines	Deciduous	Pines	Deciduous
R^2	0.9757	0.3626	0.9896	0.6094
RMSE (m)	0.99401	3.71714	0.68442	2.33173

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

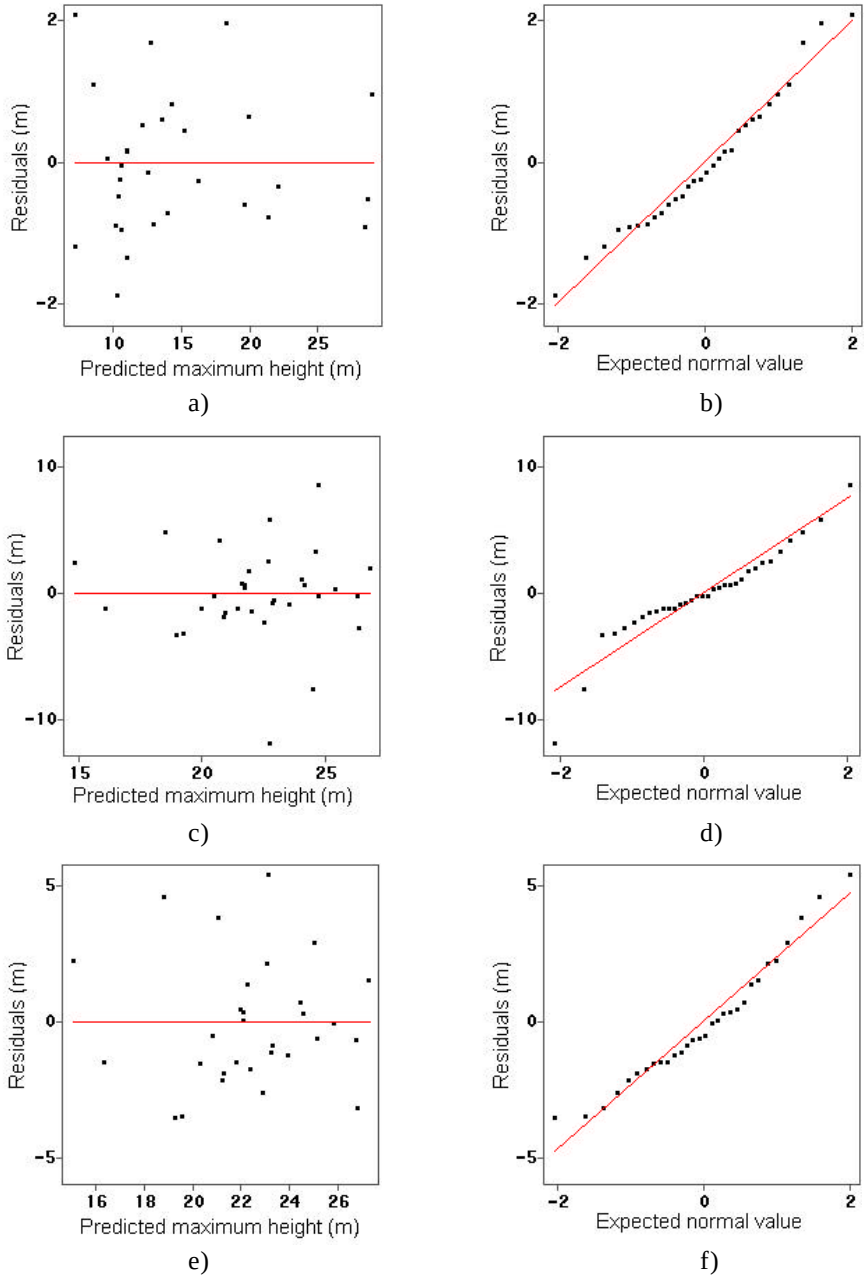


Figure 29: Plot of residuals versus fitted values (a) and normal probability plot (b) for pines, with outliers in; (c) plot of residuals versus fitted values and (d) normal probability plot for deciduous plots, with outliers in; (e) plot of residuals versus fitted values and (f) normal probability plot for deciduous plots, with outliers out;

significant increase in the R^2 value for the deciduous data set along with a substantial reduction in the RMSE. Figure 29 (a) and (b) shows the plot 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 in (Figure 29 (d)), 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 29 (f)) indicates a closer approximation of normality.

The two deciduous plots, 9 and 28, 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 plot 9 (ground maximum height 10.97 m) had a large hickory tree with a height of 18.90 m at 7.56 m from the plot center and a southern red oak of 27.74 m at 9.20 m from the plot center. A similar situation was found for plot 28 (ground maximum height of 17.07 m) that had a white oak of 25.30 m at 7.60 m from the plot center and chestnut oak 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. Plot 11 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 height of the dominant-codominant trees on this plot is 27.3 m, while the lidar maximum height is 27.87. The situation might be explained by inaccuracies in the lidar DEM. The vegetation profile of the plot estimated on the ground reveals a cover of 3.65 m high for 35% of the subplot area that, along with the dense overstory canopy, might prevent laser pulses to reach the ground.

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 dimensions, volume, basal area, and biomass 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.

For all the regression models for each of the biophysical parameters, for both pines and deciduous data, multicollinearity effects were investigated as explained in section 4.8. The highest condition number found was equal to 305.99, considerably lower than the value that raises concerns, i.e., 1000 (Myers, 1990, p. 370).

5.2. Comparison of the lidar DEM with other sources of elevation

The comparison of the lidar DEM with other sources of elevation, such as the 7.5-minute USGS NED DEM, the 7.5-minute USGS DRG, and the photogrammetrically-derived DEM provided by EarthData, Inc., showed that the lidar DEM is within error ranges currently found in other sources of elevation data. The elevation differences between the lidar-derived DEM and the other sources of elevation data for the 78 points are characterized in Table 14. Figure 30 shows the frequency distributions of the differences between lidar terrain elevations and the other sources, except the photogrammetric DEM.

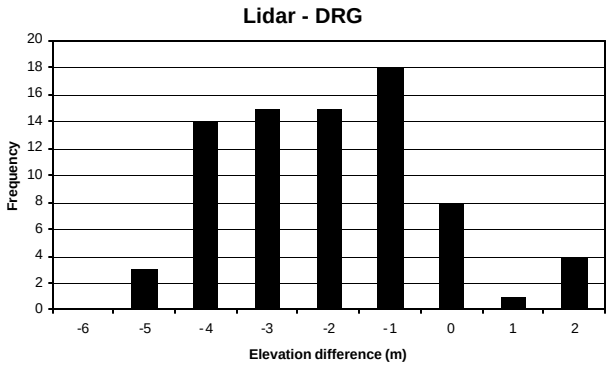
Table 14. Elevation differences (m)

Statistics \ Source (meters)	Lidar vs. GPS	Lidar vs. USGS DEM	Lidar vs. USGS DRG	Lidar vs. Photo DEM
Average difference	–3.53	–5.86	–2.45	–20.00
Standard deviation	2.97	3.84	1.67	8.16
Range Min	–10.89	–15.21	–5.90	–35.55
Max	3.02	5.90	1.88	–4.59
RMSE	4.60	6.99	2.96	21.58

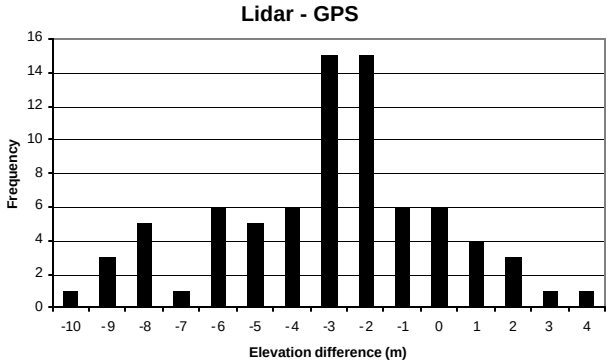
Table 14 shows that the smallest RMSE (2.96 m) was obtained when the lidar DEM was compared to the USGS DRG. As applied to the USGS 7.5-minute quadrangle topographic map, the vertical accuracy standard requires that the elevation of 90 percent of all points tested must be correct within half of the contour interval (U.S. Geological Survey, 2001). On a map with a contour interval of 10 feet, the map must correctly show 90 percent of all points tested within 5 feet (1.5 meters) of the actual elevation. The interpolation of the elevation for the 78 GPS points overlaid on the DRG is estimated to

be within 3 feet (approximately 1 m) when compared to the DRG contours (Figure 16 (b)).

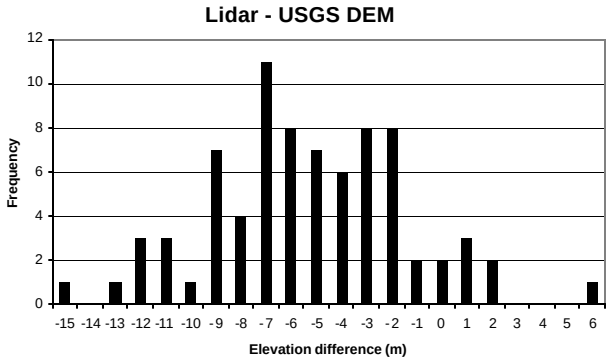
The GPS vertical accuracy is known to be smaller than the horizontal accuracy, due to the satellite configuration above the Earth (Oderwald and Boucher, 1997). A sub-meter horizontal accuracy was estimated for the GPS points locating the center of the ground plots (section 4.2.1). As expected, a large difference was observed between the lidar DEM and the photogrammetric DEM. The average difference of 20 m with a RMSE of 21.58 m certifies that the photogrammetric DEM actually models the top surface of vegetation and open ground. With respect to the USGS DEM, the lidar DEM was on average –5.86 meters lower than the 7.5-minute DEM with a RMSE of 6.99 meters. Reutebuch and McGaughey (1999) found a range of differences between lidar measurements (Saab TopEye scanning system) and a 7.5-minute USGS DEM of 54.3 to –50.3 meters in Washington State.



a)



b)



c)

Figure 30: Frequency distributions of the elevation differences between the lidar DEM and the USGS DRG (a), GPS (b), and USGS DEM (c).

It is very difficult to establish accurate ground truth for laser altimetry with commonly used techniques, such as carrier phase GPS, in particular for areas with dense

vegetation. Vosselman (2000) reached the same conclusion in his study. The RMSE of lidar vs. other sources of elevation only quantify differences in the data sets, but it gives no indication of which data set is most accurate. However, a comparison of the frequency distributions of elevation differences between the lidar DEM and the other sources of elevation data (Figure 30) indicates that there are considerable differences among the sources to which lidar is compared. This comparison illustrates that the lidar DEM is within error ranges currently found in other sources of elevation data. Still, there appears to be a systematic negative bias when comparing the lidar elevations with other sources. This situation is mitigated in the open ground.

Table 15 shows statistics on elevation differences for the seven points situated in open ground. The smallest RMSE (1.03 m) was obtained again when lidar elevations were compared to the USGS DRG, with a standard deviation of 0.33 m and an average of –0.98 m. The smallest average difference was obtained when the lidar DEM was compared to GPS-derived elevations (–0.17 m), with a range between –1.95 and 1.18 m. The comparison with the USGS DEM shows an average difference of –0.55 m and a range between –4.69 and 5.90 m. A small negative bias is expected in any situation on sloped terrain, since the first step of the algorithm for deriving terrain elevation from raw lidar points selects the lowest elevation in a 10 x 10 m cell. This algorithm also favors the selection of subsurface lidar points caused by lidar measurement errors. However, since the lidar DEM is further used to derive vegetation heights, the accuracy of the DEM is reflected indirectly in the accuracy of estimating tree heights.

Table 15. Elevation differences for open-ground points (m)

Statistics \ Source (meters)		Lidar vs. GPS	Lidar vs. USGS DEM	Lidar vs. USGS DRG	Lidar vs. Photo DEM
Average difference		–0.17	–0.55	–0.98	–6.30
Standard deviation		1.27	3.42	0.33	1.32
Range	Min	–1.95	–4.69	–1.29	–8.65
	Max	1.18	5.90	–0.41	–4.59
RMSE		1.19	3.22	1.03	6.42

Due to the high vertical accuracy (±15 cm) reported for the AeroScan lidar system flying at less than 2400 m, and the lidar ability to record multiple returns from a laser pulse and to penetrate to the forest floor through openings in the canopy, it is guite possible that the lidar DEM is more accurate than the other sources.

5.3. Investigating spatial autocorrelation

The semivariograms in Figure 28 indicate that no spatial dependency exists for ground-estimated volume for the deciduous plots, while for the pine plots, spatial dependency exists up to a range of about 300 m. The calculation of Moran's *I* coefficient for both pine and deciduous plots confirmed the same situation. Spatial autocorrelation of subplot averages for ground measurements was investigated for the height of dominant trees, volume, and biomass.

For the height of dominants on the pine plots, p-values for the significance of the *Z* statistic proved that spatial autocorrelation exists at 0.05 level (p-values for the randomization and normality assumptions were respectively 4.7086E-10 and 2.5316E-10). For the deciduous plots, p-values of 0.32, under both assumptions, indicate no spatial autocorrelation for the height of dominants.

Moran's *I* calculations for volume and biomass present the same situation as with the height of dominant trees. For the pine subplot volume, p-values under randomization and normality were respectively 0.002 and 0.003, indicating spatial autocorrelation exists at 0.05 level. For the deciduous plots, p-values for volume were not significant (0.76, for both randomization and normality). Spatial autocorrelation for ground-estimated pine biomass is only marginally significant (p-values of 0.04), while for the deciduous plots, it is not significant (p-values of 0.72).

Spatial autocorrelation for the dependent variables raised concerns about the possible inflation of the explained variance due to the potential for spatial dependency between subplot values. To examine whether the use of subplot values artificially inflates the explained variance with the regression analysis, pine subplots belonging to the same plot were pooled together into 7 plots and plot values were used as dependent variables. The analysis was done only for plot volume using the same significant variable, mean crown diameter. The R^2 value obtained using average plot values was 0.90 (RMSE 30.81 m^3/ha), while subplot analysis gave an R^2 of 0.83, with RMSE 47.90 m^3/ha. Moreover, confidence intervals for plot level parameter estimates include the parameter estimates obtained at subplot level. Therefore, the initial concerns regarding the possible inflation of variance when using subplot values for pines were not confirmed and the subsequent analysis was done at the subplot level.

5.4. 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 measures 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 (Table 16) and deciduous (data sets by obtaining better R^2 values for estimating the height of dominant and co-dominant 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 intercept 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, the co-location 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 almost 97% 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 18) 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 (Table 18). Naesset and Okland (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.

Table 16. 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	0.8963	$0.82647 + 0.80818H_{max}$
	SQF	H_{max}	1.87	0.8963	$0.82647 + 0.80818H_{max}$
	CW	H_{max}, CD_{ave}, CD_{std}	1.26	0.9563	$-0.77630 + 0.57257H_{max} + 0.91903CD_{ave} + 2.38845CD_{std}$
	CWF	H_{ave}, H_{min}, H_{std}, CD_{ave}	1.14	0.9656	$-0.21885 + 0.78538H_{ave} - 0.37937H_{min} + 0.50150H_{std} + 1.78613CD_{ave}$
All	SQ	H_{ave}, CD_{std}	1.50	0.7415	$4.52287 + 0.37786H_{ave} + 1.14151CD_{std}$
	SQF	H_{ave}, CD_{std}	1.48	0.7458	$4.31588 + 0.39436H_{ave} + 1.16062CD_{std}$
	CW	H_{ave}, CD_{std}	1.14	0.8493	$3.93120 + 0.39505H_{ave} + 1.64873CD_{std}$
	CWF	H_{ave}, CD_{std}	1.18	0.8394	$4.20656 + 0.37934H_{ave} + 1.48877CD_{std}$
FIA standard	SQ	H_{ave}, H_{max}, CD_{max}	1.18	0.9416	$3.31532 + 0.23535H_{ave} + 0.57874H_{max} - 0.54824CD_{max}$
	SQF	H_{ave}, H_{max}, CD_{max}	1.11	0.9481	$3.79115 + 0.22589\ H_{ave} + 0.60009H_{max} - 0.70397CD_{max}$
	CW	H_{max}, CD_{max}	1.09	0.9474	$1.08697 + 0.55224H_{max} + 0.69952CD_{max}$
	CWF	N, H_{max}, CD_{max}	1.07	0.9519	$3.52918 - 0.08886N + 0.48535H_{max} + 0.62258CD_{max}$
Deciduous					
Dominants	SQ	H_{max}, CD_{ave}	2.06	0.7637	$4.20097+ 0.83107H_{max} - 0.58420CD_{ave}$
	SQF	H_{max}, CD_{ave}	1.91	0.7912	$3.67791 + 0.82446\ H_{max} - 0.41608CD_{ave}$
	CW	H_{max}	2.08	0.7332	$4.22851 + 0.68477\ H_{max}$
	CWF	H_{max}	2.07	0.7354	$4.39941 + 0.67972\ H_{max}$
All	SQ	CD_{max}	1.22	0.4953	$8.20149 + 0.66389CD_{max}$
	SQF	H_{ave}	1.58	0.3972	$8.36490 + 0.22249H_{ave}$
	CW	H_{max}, CD_{ave}	1.60	0.3652	$8.24679 + 0.28781H_{max} - 0.33682CD_{ave}$
	CWF	CD_{max}	1.58	0.3578	$8.26377 + 0.66760CD_{max}$
FIA standard	SQ	H_{max}, CD_{max}	1.48	0.7533	$7.31106 + 0.63421H_{max} - 0.62351CD_{max}$
	SQF	H_{max}, CD_{max}	1.57	0.7357	$6.68390 + 0.64977H_{max} - 0.55472CD_{max}$
	CW	H_{max}, CD_{max}	1.57	0.7098	$7.41716 + 0.53921H_{max} - 0.33535CD_{max}$
	CWF	H_{max}	1.66	0.6628	$7.01565 + 0.45870H_{max}$

* 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}, minimu m 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).

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% improvement in R^2 values for the height of all trees measured on the plots (from 0.7458 to 0.8493) and 7% for the height of dominant trees (from 0.8963 to 0.9663). 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 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 they 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% 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 squared windows, with standard deviations of PRESS residuals between 1.30 and 2.20.

As explained when documenting outliers in the previous chapter, 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 17). This method explained 97 % of the variance with a sub-meter standard error of the estimate using only lidar maximum height as the independent variable. For deciduous plots, the best R^2 value was 0.6859 (RMSE 2.07 m).

Table 17. Regression results – dependent variable: maximum height (m) / subplot*

Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
All	SQ	H_{max}, CD_{max}	0.91	0.9804	$1.42214 + 1.02681H_{max} - 0.54600CD_{max}$
	SQF	H_{max}, CD_{max}	0.91	0.9804	$1.58459 + 1.02089H_{max} - 0.56401CD_{max}$
	CW	H_{max}, CD_{std}	0.96	0.9780	$0.41051 + 0.89765\ H_{max} + 0.60232CD_{std}$
	CWF	H_{max}	0.99	0.9757	$0.54189 + 0.92320\ H_{max}$
Deciduous					
All	SQ	H_{max}	2.24	0.6369	$9.35914 + 0.56096H_{max}$
	SQF	H_{max}	2.07	0.6859	$8.75699 + 0.59639H_{max}$
	CW	H_{max}	2.37	0.5965	$9.19615 + 0.57294H_{max}$
	CWF	H_{max}	2.33	0.6094	$9.21868 + 0.57395H_{max}$

* Method and variable abbreviations are the same as in Table 16

Table 18. 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);

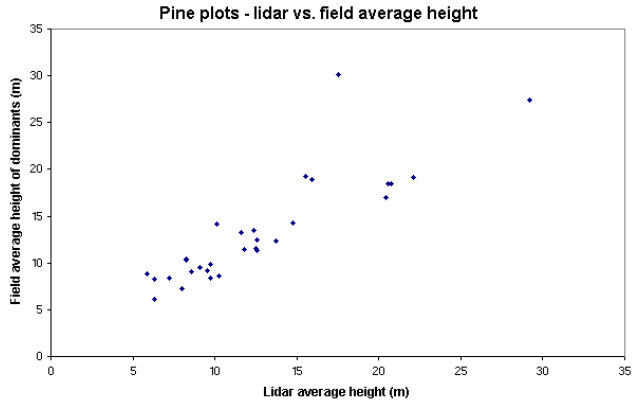
The investigation of spatial autocorrelation for residuals obtained when regressing height for the pine plots was done for the average height of dominant trees using LM filtering with circular windows and data fusion. P-values for the *Z* statistic under both assumptions, randomization and normality, were not significant when compared to the 0.05 significance level, being respectively 0.09 and 0.08. Therefore, the Moran's *I* coefficient indicated no spatial autocorrelation of residuals. For the deciduous plots, Moran's *I* coefficient was calculated for predicting the average height of all trees using the filtering techniques with square windows. The significance test for Moran's *I* coefficients under randomization and normality for binary and inverse distance connectivity matrices proved that no spatial autocorrelation exists at 0.05 level (p-values 0.56).

Figure 31 (a) and (b) shows scatterplots of lidar-measured vs. field-measured height and observed vs. predicted values for height, for the pine and deciduous models that were investigated for spatial autocorrelation of residuals.

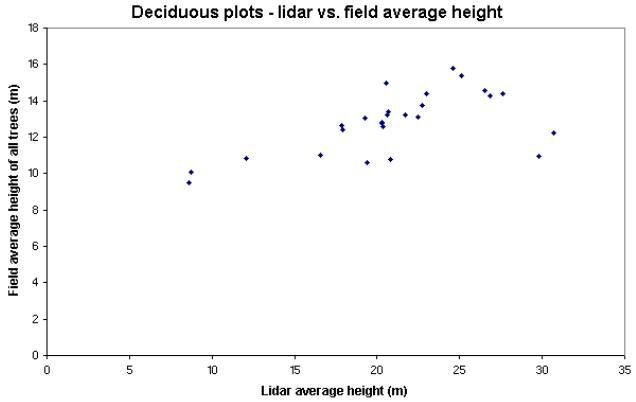
5.5. Crown diameter

As expected, the regression analysis indicated that the average crown diameter is less accurately estimated than the average height. Typically, crown diameter is not measured in forest inventories, but it is needed to calculate certain competition measures (Avery and Burkhart, 1994; Bigging and Dobbertin, 1995) and in determining canopy cover. When needed, it is usually subsampled and estimated from other tree measurements. The study of Gill *et al.* (2000) developed models of tree crown radius for several conifer species of California and obtained R^2 values in the range of 0.2691 to 0.6077 and RMSE values from 0.6081 to 1.48 m. The variable with the best prediction for their models was dbh.

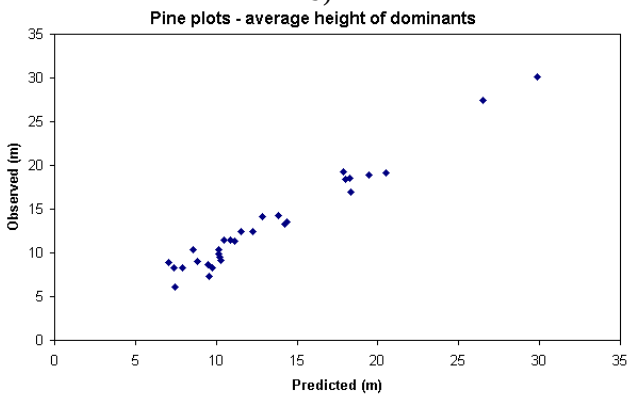
The regression analysis for the current study shows that results are similar for both pines and deciduous trees, with R^2 values of 0.62-0.63 for the dominant trees and a standard error of estimate of 1.36 – 1.41 m. For pines, data fusion improved the R^2 value by 8 to 10%, while for deciduous trees the boost was more significant, from 6 to 24%. The algorithm that gave best results for pines was the square LM filter with data fusion, while for deciduous trees was the circular LM filter with data fusion.



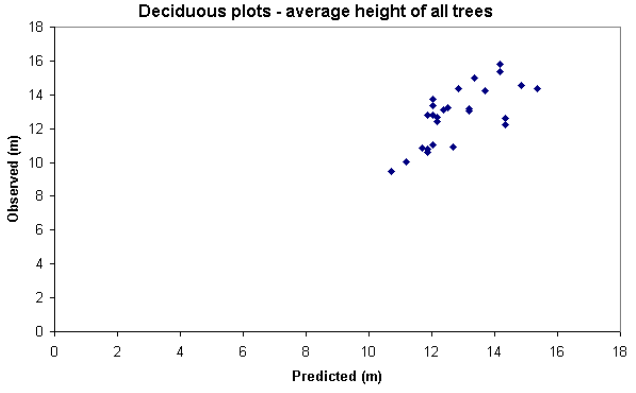
a)



b)



c)



d)

Figure 31: Scatterplots of predicted vs. observed and lidar vs. field height values

For the pine plots, the cross-validation (Table 20) also showed that using the optical data improved the prediction of crown diameter, revealing overall smaller PRESS statistics and standard deviations of PRESS residuals. With square windows filtering for pines, optical data makes a considerable difference in improving the estimation of crown diameter, as shown by both regression results and cross-validation. Though R^2 values are slightly better when filtering with square windows than with circular windows, prediction accuracy seems better for the latter.

For the deciduous plots, using the optical data consistently gave better results with both shapes of the filtering windows, as indicated by R^2 values and PRESS statistics. Despite the fact that for both pine and deciduous plots the R^2 values are higher when

Table 19. Regression results – dependent variable: average crown diameter (m) / subplot*

Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
Dominants	SQ	H_{max}, CD_{max}	1.49	0.5573	$1.73117 + 0.34275H_{max} - 0.64997CD_{max}$
	SQF	H_{max}, CD_{max}	1.36	0.6317	$2.53820 + 0.38918H_{max} - 0.99219CD_{max}$
	CW	H_{max}, CD_{std}	1.46	0.5767	$0.43083 + 0.16732H_{max} + 1.26712CD_{std}$
	CWF	H_{max}, CD_{max}	1.47	0.5687	$-0.51619 + 0.13903H_{max} + 0.53766CD_{max}$
All (FIA Standard)	SQ	H_{max}, CD_{max}	1.29	0.5095	$2.07945 + 0.27392H_{max} - 0.53800CD_{max}$
	SQF	H_{max}, CD_{max}	1.16	0.6057	$2.85880 + 0.32246H_{max} - 0.88105CD_{max}$
	CW	H_{max}, CD_{max}	1.32	0.4903	$0.41416 + 0.11895H_{max} + 0.35748CD_{max}$
	CWF	H_{max}, CD_{max}	1.30	0.5041	$0.33285 + 0.11296H_{max} + 0.39461CD_{max}$
Deciduous					
Dominants	SQ	H_{ave}	1.80	0.3337	$2.04905 + 0.22828H_{ave}$
	SQF	H_{max}	1.75	0.3513	$1.20716 + 0.25101H_{max}$
	CW	H_{ave}	1.71	0.3833	$1.93286 + 0.24868H_{ave}$
	CWF	H_{ave}, CD_{ave}, CD_{min}, CD_{std}	1.41	0.6236	$0.82883 + 0.42087H_{ave} + 1.45125CD_{ave} - 1.93922CD_{min} - 1.53391CD_{std}$
All (FIA Standard)	SQ	H_{ave}, CD_{ave}	0.99	0.4652	$2.59964 + 0.25530H_{ave} - 0.35977CD_{ave}$
	SQF	H_{ave}, CD_{ave}	0.98	0.4663	$2.82134 + 0.22236H_{ave} - 0.26108CD_{ave}$
	CW	H_{ave}	0.96	0.4360	$2.72765 + 0.15628H_{ave}$
	CWF	H_{ave}, CD_{min}	0.93	0.4930	$2.29544 + 0.21825H_{ave} - 0.19038CD_{min}$

* Method and variable abbreviations are the same as in Table 16

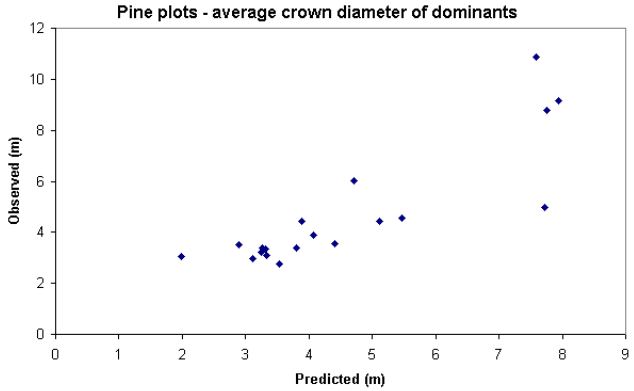
regressing the crown diameter of dominant trees, prediction accuracy is slightly better for the crown diameter of all tress measured by FIA standards. The smallest overall PRESS statistics and sub meter standard deviation of PRESS residuals were obtained for the average crown diameter of all FIA measured trees on the deciduous plots. Figure 32 (a) and b) shows scatterplots of lidar-measured vs. field-measured crown diameter and observed vs. predicted values for crown diameter, for the pine and deciduous models with the best fit and prediction.

For the pine plots, Moran's *I* coefficient was calculated for predicting the average crown diameter for the dominant trees using the filtering techniques with square windows and data fusion. The significance test for Moran's *I* coefficients under randomization and normality proved that no spatial autocorrelation exists at the 0.05 level (p-values 0.60).

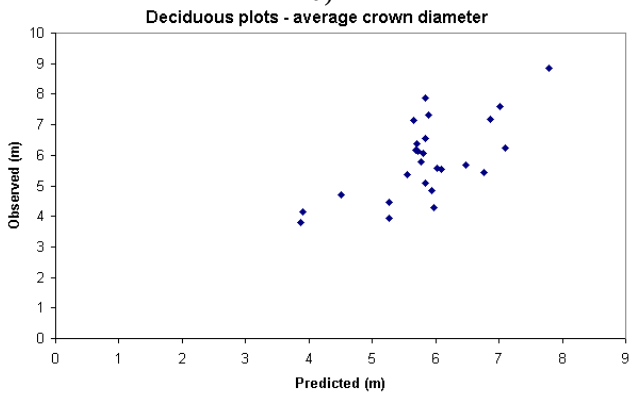
Table 20. PRESS statistics for predicting average crown diameter (m) / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
Dominants	SQ	88.56	-3.69	5.31	0.14	1.84
	SQF	59.39	-3.49	4.25	0.08	1.51
	CW	53.09	-2.71	3.13	0.02	1.43
	CWF	50.21	-2.73	3.01	0.04	1.39
All (FIA Standard)	SQ	67.37	-2.43	5.42	0.12	1.60
	SQF	45.72	-2.09	4.52	0.10	1.32
	CW	43.38	-2.31	3.78	0.05	1.29
	CWF	43.34	-2.24	3.64	0.08	1.29
Deciduous						
Dominants	SQ	91.32	-2.49	5.50	0.00	1.91
	SQF	91.08	-4.00	4.95	0.00	1.87
	CW	93.72	-2.79	5.26	0.00	1.80
	CWF	88.06	-2.55	3.86	0.10	1.74
All (FIA Standard)	SQ	27.64	-1.78	2.13	0.02	1.05
	SQF	27.80	-2.03	2.05	0.02	1.03
	CW	29.40	-1.58	2.06	-0.01	1.01
	CWF	28.15	-1.96	2.26	-0.01	0.98

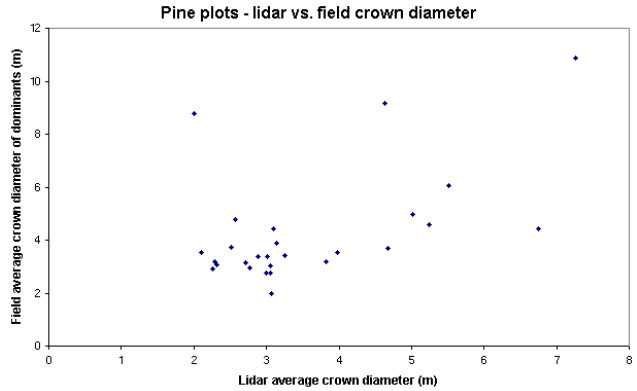
* Method refers to LM filtering technique: SQ (square window), SQF (square window with data fusion), CW (circular window), and CWF (circular window with data fusion);



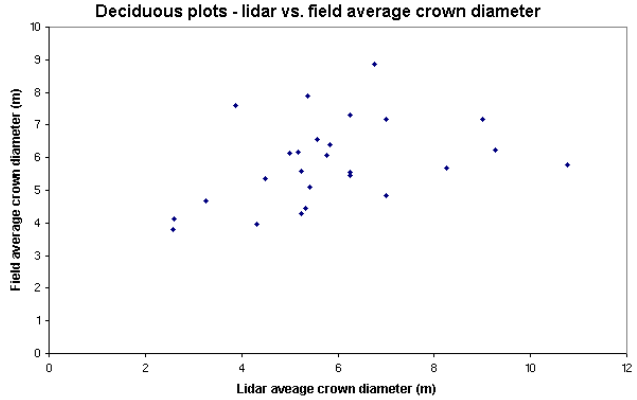
a)



b)



c)



d)

Figure 32: Scatterplots of predicted vs. observed and lidar vs. field crown diameter

The spatial autocorrelation of residuals for the deciduous plots was investigated for the average crown diameter of all trees measured by the FIA standard using LM filtering with square windows without data fusion. P-values for the *Z* statistic under both assumptions, randomization and normality were not significant when compared to the 0.05 significance level (0.21 for both assumptions). Therefore, the Moran's *I* coefficient indicated no spatial autocorrelation of residuals for estimating crown diameter.

The lidar-measured variables that proved significant for predicting crown diameter for the pine plots were maximum height and maximum crown diameter. For the deciduous plots, most frequently average height and average crown diameter appear as significant variables, which is a reasonable result from a biometrics standpoint. Despite the fact that maximum height can be accurately estimated with lidar, as shown in section 5.4., results for estimating crown diameter with lidar as not as good as for estimating height. Such a conclusion is not unexpected, but lidar-measured crown diameter still proves to be a significant variable for estimating other biophysical parameters. Appendix 8 shows the regression results obtained for estimating the ground-measured biophysical parameters, such as dbh, basal area, volume, and biomass, without including in the regression models the independent variables related to the lidar-measured crown diameter. The increase in R^2 values when using lidar-measured crown diameter variables was on average 0.07, with a maximum of 0.27 for regressing dbh.

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 CHM aimed at measuring the non-overlapping crown diameter, while the field measurements considered crowns to their full extent, therefore measured overlapping crown diameters. However, with an increased sampling intensity of lidar, the CHM should better portray the three-dimensional model of the tree crown and as a consequence, predicting crown diameter should become more accurate.

5.6. Diameter at breast height (dbh)

Diameter at breast height is, no doubt, the most frequent tree measurement made by foresters. Though dbh is a tree dimension that is not directly visible for airborne lidar sensors, lidar measurements correlated well with dbh for the pine plots. The algorithm that performed best for pines was the circular LM filter with data fusion (Table 21). The regression analysis for this method explained more than 86% of the variance associated with dbh for all three categories of trees measured on the ground, i.e., dominants, all trees, and FIA-standard trees. A very small standard error of the estimate (1.42 cm) was obtained for the average dbh of all trees with the circular LM filter and data fusion technique, with a 0.8976 R^2 value. Moreover, cross-validation (Table 22) proved that the same technique gave the overall smallest PRESS statistic and a very low standard deviation of PRESS residuals, of only 1.53 cm, which is approximately 11.5 % of the average dbh measured on the ground for all the pine plots (13.22 cm). As with the average height, data fusion and the circular LM filter consistently improved results for regressing the average dbh for pines. Among the two factors, the use of circular windows to filter for local maximum appears to be the most important to improve results for estimating dbh, as indicated by R^2 values and PRESS statistics. Results for deciduous plots were not as good as for pines when judged by the explained variance (highest R^2 value 0.5098), especially for the "all trees" category, which includes many overtopped trees. Nevertheless, the lowest standard deviation of PRESS residuals was 3.02 cm, for an average crown diameter of 17.18 cm observed on the ground. For the deciduous plots, LM filtering with square windows proved slightly better for predicting dbh.

Despite the fact that dbh is not directly imaged by lidar, results for estimating dbh for pines are not surprising. Prior studies have reported that dbh is best correlated with crown radius (e.g., Sprinz and Burkhart, 1987; Smith *et al*., 1992; Gill *et al.* 2000) and strongly correlated with stem height (Green, 1981; Arabatzis and Burkhart, 1992). Indeed, the variables that most often appeared significant in the regression models were maximum height and maximum crown diameter.

For the pine plots, the spatial autocorrelation of residuals was investigated for regressing average dbh of all trees using LM filtering with circular windows and data fusion. P-values for the Z statistic under both assumptions, randomization and normality,

proved not significant when compared to the 0.05 significance level (p-values 0.28). Therefore, the Moran's I coefficient indicated no spatial autocorrelation of residuals.

For the deciduous plots, Moran's *I* coefficient was calculated for predicting the

Table 21. Regression results – dependent variable: diameter at breast height (cm) / subplot*

Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
Dominants	SQ	H_{max}	6.14	0.6968	$-1.36696 + 1.36421H_{max}$
	SQF	H_{max}, CD_{max}	5.88	0.7318	$3.35988 + 1.81223H_{max} - 2.60656CD_{max}$
	CW	H_{max}, CD_{min}, CD_{std}	4.57	0.8437	$-3.17395 + 0.64622H_{max} + 2.23103CD_{min} + 9.11205CD_{std}$
	CWF	N, H_{std}, CD_{min}, CD_{max}	3.80	0.8963	$-10.76026 + 0.28287N + 1.65568H_{std} + 3.21778CD_{min} + 3.54253CD_{max}$
All	SQ	N, H_{max}	2.57	0.6367	$0.95133 + 0.27770N + 0.63073H_{max}$
	SQF	N, H_{max}	2.56	0.6382	$1.07482 + 0.27297N + 0.62887H_{max}$
	CW	N, H_{std}, CD_{ave}	1.63	0.8574	$-2.31428 + 0.27961N + 0.71110H_{std} + 3.04464CD_{ave}$
	CWF	N, H_{min}, H_{std}, CD_{ave}	1.42	0.8976	$2.86891 + 0.26064N - 0.23905H_{min} + 0.58301H_{std} - 4.08465CD_{ave}$
FIA standard	SQ	H_{max}, CD_{max}	4.35	0.7188	$6.05915 + 1.26971H_{max} - 1.55162CD_{max}$
	SQF	H_{max}, CD_{max}	3.99	0.7637	$8.51633 + 1.42548H_{max} - 2.63702CD_{max}$
	CW	H_{ave}, H_{max}, CD_{max}	3.95	0.7779	$-1.68235 - 0.56057H_{ave} + 0.96745H_{max} + 2.72069CD_{max}$
	CWF	H_{ave}, H_{max}, CD_{ave}, CD_{max}, CD_{std}	3.25	0.8617	$-5.65523 - 2.39290H_{ave} + 1.70061H_{max} + 4.72096CD_{ave} + 3.69662CD_{max} - 4.92991CD_{std}$
Deciduous					
Dominants	SQ	H_{max}	6.63	0.4089	$7.34141 + 1.04545H_{max}$
	SQF	H_{max}, CD_{max}	6.20	0.5098	$5.74215 + 1.65444H_{max} - 1.69833CD_{max}$
	CW	H_{max}	6.52	0.4415	$5.45407 + 1.15193H_{max}$
	CWF	H_{max}	6.56	0.4339	$6.00805 + 1.13187H_{max}$
All	SQ	CD_{max}	2.84	0.3107	$9.59958 + 1.04681CD_{max}$
	SQF	CD_{max}	3.55	0.2490	$10.15280 + 1.04358CD_{max}$
	CW	H_{max}, CD_{ave}	3.31	0.3399	$9.24804 + 0.56882H_{max} - 0.91448CD_{ave}$
	CWF	CD_{max}	3.49	0.2374	$9.51909 + 1.10275CD_{max}$
FIA standard	SQ	H_{max}	4.19	0.3483	$12.13503 + 0.58043H_{max}$
	SQF	H_{max}, CD_{max}	4.26	0.4236	$11.14339 + 0.98524H_{max} - 1.11980CD_{max}$
	CW	H_{max}, CD_{max}	4.01	0.4768	$12.15117 + 0.88036H_{max} - 0.95425CD_{max}$
	CWF	H_{max}	4.32	0.3716	$10.88770 + 0.65423H_{max}$

* Method and variable abbreviations are the same as in Table 16

average dbh of all trees using the filtering techniques with square windows. The significance test for Moran's *I* coefficients under randomization and normality proved that no spatial autocorrelation exists at 0.05 level (p-values 0.20 and 0.21, respectively).

For the models with the best fit and prediction, subplot values of predicted and observed measurements of dbh are plotted against each other in Figure 33. As these plots indicate, the variability of average dbh for the pine plots slightly decreases for larger trees, while for the deciduous plots, it increases with dbh.

Beside average dbh, the quadratic mean diameter could be of interest as it is related to basal area and the number of trees per unit area. Therefore, regression analysis was also performed with the quadratic mean diameter as the dependent variable. For the pine

Table 22. PRESS statistics for predicting average dbh (cm) / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
Dominants	SQ	1411.44	-17.47	15.73	0.08	6.98
	SQF	1534.21	-17.61	16.94	0.42	7.26
	CW	1007.36	-10.47	19.86	0.48	5.87
	CWF	606.84	-6.69	12.45	0.19	4.57
All	SQ	271.50	-10.71	7.39	0.05	3.06
	SQF	270.89	-10.82	7.34	0.05	3.06
	CW	109.26	-3.37	4.53	0.08	1.94
	CWF	67.78	-3.26	2.93	0.02	1.53
FIA standard	SQ	1030.70	-9.62	20.52	0.52	6.16
	SQF	668.72	-9.04	17.06	0.31	4.97
	CW	686.67	-7.99	18.67	0.25	5.04
	CWF	526.36	-8.31	14.39	0.24	4.41
Deciduous						
Dominants	SQ	1250.42	-14.32	15.02	-0.02	7.07
	SQF	1110.61	-10.44	15.16	0.12	6.53
	CW	1374.05	-15.55	14.20	-0.01	6.88
	CWF	1391.75	-15.49	14.13	-0.01	6.93
All	SQ	227.98	-8.09	6.52	-0.03	3.02
	SQF	368.42	-8.18	10.94	0.00	3.76
	CW	463.71	-7.55	10.72	0.15	4.00
	CWF	394.25	-9.64	10.98	-0.04	3.69
FIA standard	SQ	482.79	-6.44	13.38	-0.01	4.39
	SQF	518.02	-6.50	12.39	-0.01	4.46
	CW	502.82	-6.50	12.52	0.03	4.16
	CWF	588.68	-7.05	13.07	0.00	4.51

* 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);

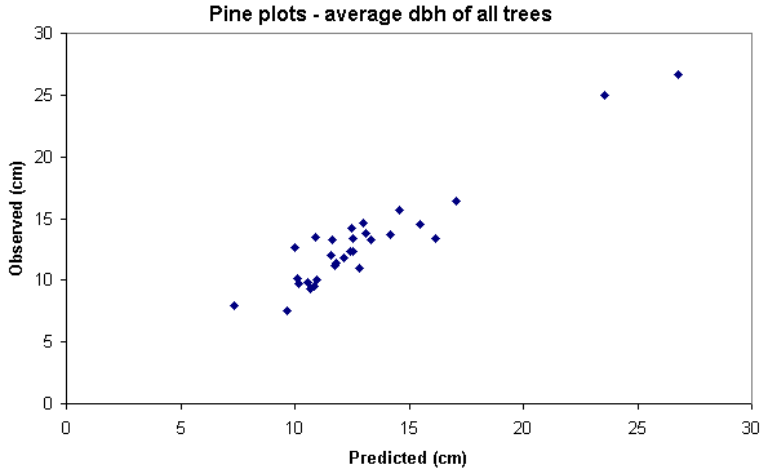
Table 23. Regression results – dependent variable: quadratic mean diameter (cm) / subplot*

Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
All	SQ	N, H_{max}	3.17	0.7067	$-1.77981 + 0.30835N + 0.87134H_{max}$
	SQF	N, H_{max}	3.17	0.7063	$-1.53675 + 0.29697N + 0.86571H_{max}$
	CW	N, H_{max}, H_{std}, CD_{min}, CD_{std}	2.17	0.8785	$-4.20248 + 0.35623N + 0.27050H_{max} + 0.41238H_{std} + 2.52264CD_{min} + 4.14874CD_{std}$
	CWF	N, H_{min}, H_{std}, CD_{ave}, CD_{std}	1.51	0.9409	$-7.33394 + 0.30877N - 0.56010H_{min} + 0.96391H_{std} + 6.91464CD_{ave} - 2.39785CD_{std}$
Deciduous					
All	SQ	H_{max}	3.39	0.3089	$9.88597 + 0.42941H_{max}$
	SQF	H_{max}	3.84	0.3343	$8.13267 + 0.53060H_{max}$
	CW	H_{max}, CD_{ave}	3.55	0.4361	$8.52424 + 0.74702H_{max} - 1.00961CD_{ave}$
	CWF	H_{max}	3.85	0.3094	$8.68092 - 0.50830H_{max}$

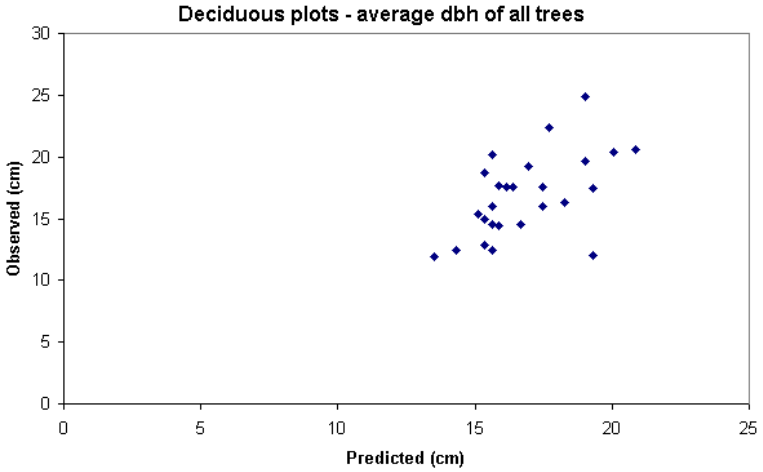
* Method and variable abbreviations are the same as in Table 16

plots, the circular LM filter with data fusion explained 94% of the variance associated with the quadratic mean diameter with a standard error of estimate of 1.51 cm. The circular window shape and data fusion considerably improved results for pines judged by the increase in R^2 and reduction of RMSE. For deciduous plots, model fit was not as good as for pines, with an R^2 of 0.4361 and an RMSE of 3.55 cm. Variables estimated with filtering with a circular shaped filter proved to give the best fit for deciduous plots.

When estimating pine quadratic mean diameter (Table 23), the circular windows and data fusion technique gave a high R^2 value (0.94) and a small RMSE (1.51 cm). The PRESS statistic (Table 24) for pines was also small (110.42).



a)



b)

Figure 33: Scatterplots of predicted vs. observed dbh for pine (a) and deciduous plots (b)

Table 24. PRESS statistics for predicting quadratic mean diameter / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	421.80	-12.00	9.18	0.11	3.81
	SQF	420.17	-12.10	9.15	0.10	3.81
	CW	274.44	-5.49	10.60	0.12	3.07
	CWF	110.41	-3.01	6.00	0.09	1.95
Deciduous						
All	SQ	328.42	-9.78	6.58	-0.08	3.62
	SQF	424.17	-9.04	-10.64	-0.05	4.04
	CW	490.25	-9.17	-9.21	0.12	4.11
	CWF	481.71	-10.93	-10.56	-0.06	4.08

* 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);

5.7. Number of trees

The number of trees per subplot proved to be one of the most difficult parameter estimated with airborne lidar. Results (Table 25) show similar R^2 values for pine and deciduous plots (0.2626 and 0.2562, respectively), while RMSE values are smaller for deciduous plots. As expected, the best results were obtained for dominant and co-dominant trees. The circular LM filter and data fusion gave best results for pines, while the square and circular LM filters gave similar results for deciduous plots. Part of the unexplained variance for the number of trees per FIA-type subplot can be attributed to the difficulty of an automated processing in counting trees in the forest canopy seen from above. The tree dominance is a key factor in the visibility of an individual crown from above. In addition, clumping of crowns strongly affects the number of trees and, unless, documented in the field, it is difficult to assess. A strict plot comparison of the number of trees is also affected by the tree position relative to the plot boundary, especially for trees near the plot periphery. The base of the stem can be mapped inside the plot whereas the top and most of its crown are projected outside the plot. Large trees near the edge of the plot, either inside or outside, are more liable to induce considerable differences. Leckie and Gougeon (1999) assessed the accuracy of both visual and automated tree counting with high resolution multispectral imagery in temperate forest stands of coniferous and deciduous tree species in eastern Ontario, Canada. They found that the automated detection of trees only had 40% direct one-to-one correspondence with ground truth trees. Their ground truth data consist of measurements of 600 trees of 17 species, including pines, spruces, and deciduous trees. They listed the tree species in order of both the visual and automated counting accuracy and found that deciduous species, such as maples and aspens, were contained in the bottom half of the list, while species of spruce made the top of the list.

The regression analysis for the pine subplots provided interesting results with respect to the variables that proved significant for estimating the number of trees. Every model included the standard deviation of individual tree heights on each subplot. In addition to this variable, the number of trees and the standard deviation of crown diameter were incorporated separately in three other models. In all regression models for pines, standard

error for height has a negative parameter estimate. This situation might indicate that the larger the variation of individual tree heights in a subplot, the smaller the number of trees. Also, the model that provided the highest R^2 value for pines (0.26) included the lidar-derived number of trees in addition to the standard deviation for height.

The number of trees estimated by lidar on the deciduous plots contributed significantly to the model fit. The models that included the number of trees as independent variables had the highest R^2 values (0.2562 and 0.2532). Other lidar estimated variables that most often remained in the models were maximum height and maximum crown diameter.

The calculation of Moran's *I* coefficient revealed no spatial autocorrelation for both pine and deciduous subplots. For the pine plots, the filtering technique with square windows and data fusion for predicting the number trees gave the smallest PRESS statistic (Table 26) and one of the highest R-square value. Therefore, it was chosen for investigating residuals spatial autocorrelation. P-values were not significant at 0.05 level for randomization and normality assumptions (p-values 0.26). For the deciduous plots, the spatial autocorrelation was examined with residuals obtained when using the circular windows filtering technique for counting the dominant trees. The *Z* test proved not significant at 0.05 level (0.09).

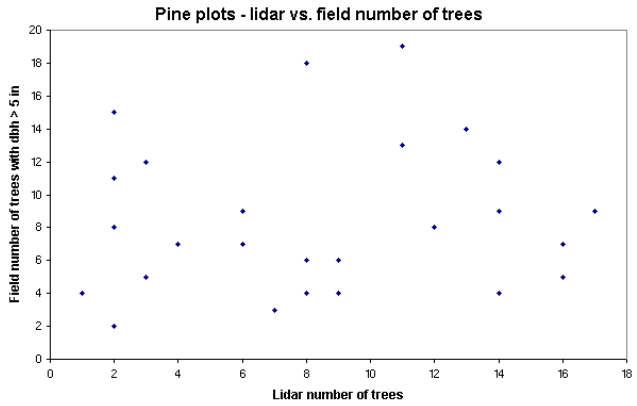
Scatterplots of field vs. lidar counted number of trees are presented in Figure 34 (a) and (b), for the pine and deciduous models with the best fit and prediction. Figure 34 (c) and (d) shows plots of the predicted vs. observed number of trees. As Figure 34 illustrates both lidar vs. ground counted trees and predicted vs. observed number of trees show high variability when compared on a subplot basis.

Table 25. Regression results – dependent variable: number of trees / subplot*

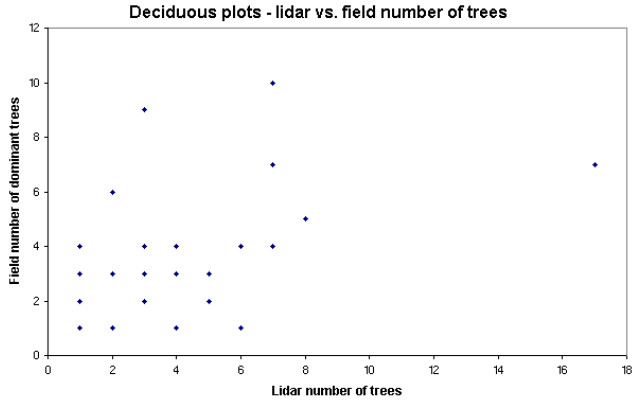
Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
Dominants	SQ	H_{std}	5.34	0.1993	$10.72023 - 0.62551H_{std}$
	SQF	H_{std}	5.39	0.1847	$10.62979 - 0.59909H_{std}$
	CW	H_{std}	5.34	0.1995	$10.82418 - 0.66831H_{std}$
	CWF	N, H_{std}	5.21	0.2626	$7.66785 + 0.23822N - 0.49468H_{std}$
All	SQ	H_{std}	12.7	0.1669	$25.24722 - 1.33610H_{std}$
	SQF	H_{std}	12.81	0.1544	$25.05097 - 1.27837H_{std}$
	CW	H_{std}	12.75	0.1617	$25.41249 - 1.40416H_{std}$
	CWF	H_{std}	12.76	0.1606	$25.39043 - 1.39835H_{std}$
FIA standard	SQ	H_{std}	4.28	0.1698	$9.25341 - 0.43968H_{std}$
	SQF	H_{std}, CD_{std}	4.18	0.2390	$7.33532 - 0.88842H_{std} + 3.75154CD_{std}$
	CW	H_{std}	4.32	0.1548	$9.27589 - 0.44848H_{std}$
	CWF	H_{std}, CD_{std}	4.15	0.2478	$7.28336 - 0.73702H_{std} + 2.89527CD_{std}$
Deciduous					
Dominants	SQ	N, H_{std}	2.09	0.2562	$3.07682 + 0.45336N - 0.28999H_{std}$
	SQF	H_{max}	2.16	0.1409	$7.69360 - 0.17051H_{max}$
	CW	N	2.06	0.1894	$2.42046 + 0.30644N$
	CWF	H_{max}	2.09	0.1695	$7.94221 - 0.18570H_{max}$
All	SQ	CD_{max}	4.75	0.1884	$20.78749 - 1.25903CD_{max}$
	SQF	H_{max}	4.78	0.1744	$21.55369 - 0.42896H_{max}$
	CW	N	4.57	0.2532	$8.07098 + 0.81861N$
	CWF	CD_{max}	4.64	0.2302	$21.80162 - 1.43735CD_{max}$
FIA standard	SQ	H_{max}	2.81	0.1012	$9.93287 - 0.17859H_{max}$
	SQF	-	2.85	0	5.85185
	CW	N	2.65	0.1410	$4.29000 + 0.33033N$
	CWF	H_{max}	2.69	0.1168	$10.06810 - 0.19262H_{max}$

* Method and variable abbreviations are the same as in Table 16

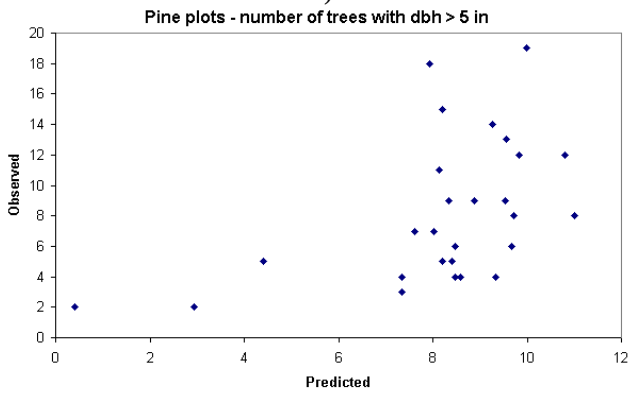
As anticipated, the local maximum filter has difficulties in identifying all the trees. Table 4 shows a large range for the number of pine trees per sublot (3 to 68) with standard deviation 13. The number of trees on deciduous plots varies between 3 and 22, with standard deviation 5. The variable window is nevertheless appropriate to vary the size based on tree height and forest type. However, errors of commission and omission can still occur, depending on the local forest structure and canopy closure. On small plots, like the FIA subplots, such errors can prove significant for estimating the number of trees.



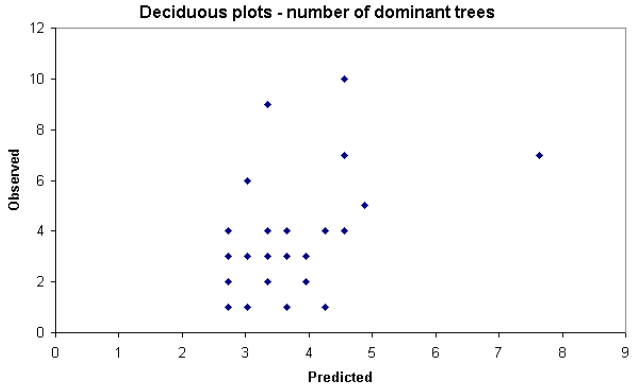
a)



b)



c)



d)

Figure 34: Scatterplots of lidar vs. field and predicted vs. observed number of trees

Table 26. PRESS statistics for predicting number of trees / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
Dominants	SQ	888.88	-9.52	16.20	0.12	5.54
	SQF	896.74	-9.45	16.28	0.09	5.56
	CW	931.26	-9.61	16.14	0.18	5.66
	CWF	874.31	-10.33	16.13	0.10	5.49
All	SQ	4961.54	-20.85	45.16	0.19	13.08
	SQF	5009.70	-20.69	45.32	0.13	13.14
	CW	5089.67	-21.00	45.10	0.28	13.24
	CWF	5089.03	-20.98	45.12	0.27	13.24
FIA standard	SQ	524.70	-5.32	10.48	0.04	4.41
	SQF	506.00	-5.58	10.64	-0.01	4.33
	CW	542.32	-5.33	10.45	0.07	4.48
	CWF	566.78	-5.79	9.74	0.12	4.58
Deciduous						
Dominants	SQ	174.31	-7.98	5.71	-0.23	2.63
	SQF	139.76	-5.66	5.73	-0.02	2.32
	CW	133.21	-3.42	5.88	-0.03	2.14
	CWF	144.79	-5.71	5.75	-0.02	2.23
All	SQ	631.07	-13.65	11.13	0.01	5.02
	SQF	711.63	-16.36	10.26	-0.07	2.23
	CW	686.82	-8.63	11.92	-0.13	4.86
	CWF	701.68	-14.00	9.93	0.01	4.92
FIA standard	SQ	233.64	-8.34	7.28	-0.08	3.06
	SQF	No variable met the 0.15 significance level to entry into the model				
	CW	260.74	-7.24	8.02	-0.13	3.00
	CWF	244.63	-8.22	7.43	-0.06	2.90

* 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);

5.8. Tree volume

The tree volume of a forest is one of the most important characteristics in forest management. The individual tree volume is usually considered to be a function of tree dbh, tree height, and an expression of tree form (Clutter *et al.*, 1983), but most practitioners prefer to use volume equations that involve only dbh and height.

Regression analysis for average plot height, crown diameter, and dbh demonstrated a strong correlation between lidar estimates and ground measurements, especially for the pine plots, hence high coefficients of determination (R^2) were expected for volume per plot. Indeed, for the pine plots, the circular LM filter with data fusion gave the best results, with an R^2 value of 0.8297 and a standard error of estimate of 47.90 m^3/ha (Table 27). The same method had the lowest PRESS statistic and standard deviation of PRESS residuals (Table 28). However, all four methods gave good results for plot volume for pines (all R^2 above 0.7578), with the average crown diameter and average height as significant variables in the models.

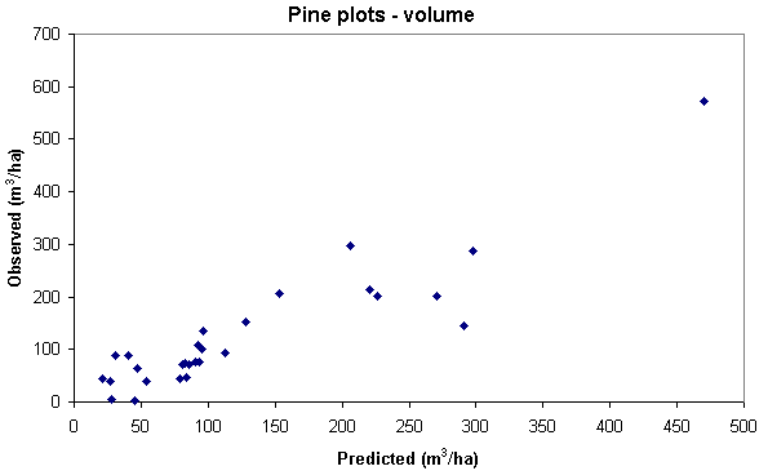
For deciduous plots, the circular LM filter explained 38.84% of the variance associated with plot volume, with RMSE of 52.84 m^3/ha. The range of ground-estimated volume was 3.42 - 571.99 m^3/ha and 17.51 - 323.68 m^3/ha for pines and deciduous plots, respectively (Table 6).

Table 27. Regression results – dependent variable: volume (m^3 / ha) / subplot*

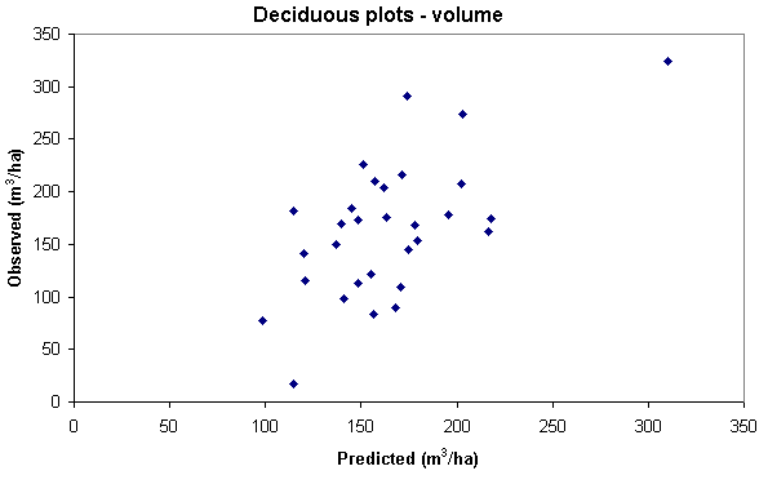
Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
All	SQ	H_{ave}	57.13	0.7578	$-101.41407 + 17.19689H_{ave}$
	SQF	N, H_{ave}, CD_{max}	50.74	0.8225	$-123.43376 + 4.38645N + 24.58183H_{ave} - 25.15525CD_{max}$
	CW	CD_{ave}	50.68	0.8094	$-140.01291 + 77.14394CD_{ave}$
	CWF	CD_{ave}	47.90	0.8297	$-149.51800 + 79.96584\ CD_{ave}$
Deciduous					
All	SQ	H_{ave}, CD_{min}	54.93	0.2530	$46.46559 + 8.14499H_{ave} - 11.39907CD_{min}$
	SQF	H_{ave}	60.12	0.2361	46.23773 + 5.79718Have
	CW	H_{max}, CD_{ave}	52.84	0.3884	$53.44176 + 9.62684H_{max} - 20.82779CD_{ave}$
	CWF	H_{max}, CD_{ave}	59.21	0.2321	$45.75842 + 9.51363H_{max} - 16.92139CD_{ave}$

* Method and variable abbreviations are the same as in Table 16

Most of the regression models included lidar estimates of the average crown diameter. For the pines plots, crown diameter alone was able to explain up to almost 83% of the variance associated with total volume. For the deciduous plots, maximum height and average crown diameter provided the best fit. The significant impact of crown diameter in explaining the variation noted in ground-measured tree volume is of high interest for practical volume estimation using remote sensing techniques with optical and lidar data. Perfecting algorithms for measuring crown diameter on remote sensing images and increasing the lidar sampling density could significantly improve volume estimates in operational forest inventories.



a)



b)

Figure 35: Scatterplots of predicted vs. observed volume for pine (a) and deciduous plots (b)

Cross-validation and percent variance explained show that the circular window shape for LM filtering improved results for both pines and deciduous plots. The use of optical data to differentiate between pine and deciduous species provided a better fit only for the pines plots. The relationship between predicted and measured volume for the models with the best fit and prediction is shown in Figure 35.

Residuals obtained when estimating volume for pine and deciduous plots are not spatially autocorrelated. The models that were chosen for investigating residuals spatial autocorrelation were obtained, for pines, when filtering with circular windows and data fusion, and for deciduous plots, with the same window shape, but without using optical data. P-values for the significance of Moran's I statistics for pines and deciduous data, were respectively 0.12 and 0.76. Normality and randomization assumptions provided the same conclusions.

Results obtained for this study correspond to and exceed some of the previous findings that typically used homogeneous coniferous data sets. Nelson *et al*. (1988b) successfully predicted the volume for southern pine forests using a profiling lidar. By deriving several estimates of canopy height and cover, they explained between 43 and 53% of the variance associated with field-measured volume. A later study by Nelson *et al*. (1997) in tropical forests obtained similar results for predicting volume (R^2 between 0.64 and 0.66), basal area, and biomass. Næsset (1997b) reported R^2 values for volume in the range of 0.45 to 0.88 for the use of small-footprint scanning lidar data acquired over stands of Norway spruce (*Picea abies* Karst) and Scots pine (*Pinus sylvestris* L.). Nilsson (1996) obtained an R^2 value of 0.78 by using estimates of waveform area and laser height from a small-footprint waveform laser scanner flown over a coastal even-aged stand of Scots pine. Hyyppä *et al*. (2001) reported an RMSE of 51.72 m^3/ha for estimating volume for stands of Norway spruce and Scots pine, without mentioning the coefficient of determination. They used a segmentation-based approach to retrieve individual stems and calculate height, basal area, and volume estimates.

Table 28. PRESS statistics for predicting volume / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	133071	-160.88	239.24	1.83	67.71
	SQF	114929	-125.06	202.76	4.36	62.80
	CW	109398	-158.95	208.96	1.39	61.40
	CWF	95474	-166.64	175.68	1.72	57.35
Deciduous						
All	SQ	98733	-131.46	129.22	-0.01	62.84
	SQF	110172	-136.89	133.17	-0.59	65.09
	CW	93722	-120.30	121.12	-0.54	56.85
	CWF	124961	-123.97	174.71	-1.04	65.63

* 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);

5.9. Basal area

Basal area is of interest for forest inventories since it is highly correlated with the volume and growth of forest stands. In addition, many silvicultural and forest management considerations, such as thinning intensity, are based on basal area ground measurements. Regression models (Table 29) for pines perform similarly for the four methods of processing the CHM, with R^2 values in the range of 0.736 to 0.765 and RMSE values between 5.23 and 5.44 m^2/ha. However, the cross-correlation indicated a better prediction when using optical data in conjunction with circular window filtering. The standard deviation of PRESS residuals in this case was 5.85 m^2 / ha. For deciduous plots, the circular LM filter obtained a model fit with an R^2 of 0.249. While the PRESS statistic (Table 30) and the standard error of PRESS residuals for deciduous plots are comparable with the values obtained for the pine plots, only one lidar processing method, filtering with circular windows, was able to estimate variables that met the 0.15 significance level to entry into the regression model. Lidar measurements of crown diameter and tree height were the variables that contributed significantly to model fit for both pine and coniferous plots. The relationship between lidar-predicted and field-measured basal area is shown in Figure 36 for the best pine and deciduous models.

Residuals of basal area revealed no spatial autocorrelation. For both species groups, pine and deciduous trees, Moran's *I* was calculated for models referring to filtering with

circular windows, using optical data only for the pines. P-values for the significance of the *Z*-test under normality and randomization were not significant (0.23 and 0.59, for pine and deciduous plots, respectively).

Table 29. Regression results – dependent variable: basal area (m^2 / ha) / subplot*

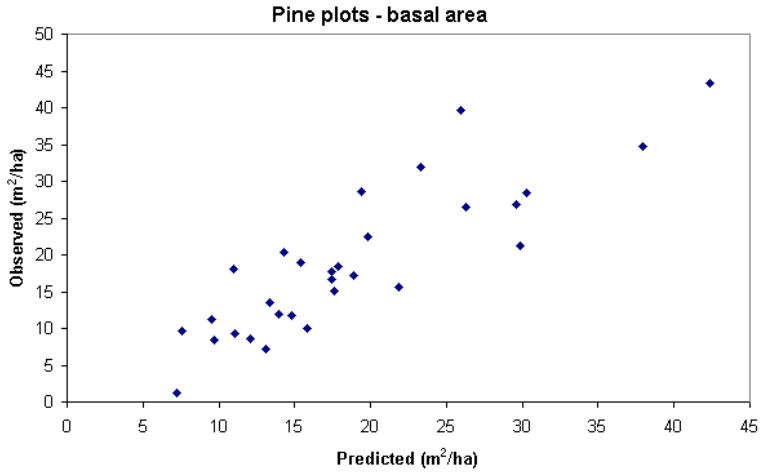
Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
All	SQ	H_{ave}, H_{min}, CD_{min}, CD_{std}	5.23	0.7656	$2.77019 + 1.87501H_{ave} + 0.90246H_{min} - 5.46788CD_{min} - 5.45925CD_{std}$
	SQF	H_{ave}, H_{max}, CD_{min}, CD_{std}	5.28	0.7617	$4.22093 + 3.38990H_{ave} - 0.91157H_{max} - 4.85450CD_{min} - 4.71406CD_{std}$
	CW	H_{min}, CD_{min}, CD_{max}	5.44	0.7368	$-4.82335 + 1.58962H_{min} - 4.07682CD_{min} + 3.65456CD_{max}$
	CWF	H_{min}, CD_{min}, CD_{max}	5.36	0.7439	$-3.61459 + 1.52023H_{min} - 3.98634CD_{min} + 3.44638CD_{max}$
Deciduous					
All	SQ	-	6.83	0	20.32692
	SQF	-	7.20	0	20.83370
	CW	H_{min}, CD_{min}	6.31	0.2498	$17.74306 + 0.57719H_{min} - 1.98161CD_{min}$
	CWF	-	7.03	0	20.33933

* Method and variable abbreviations are the same as in Table 16

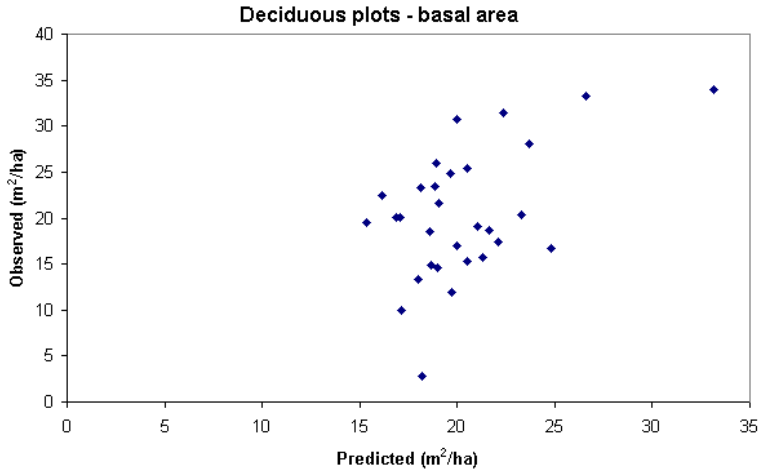
Table 30. PRESS statistics for predicting basal area (m^2 / ha) / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	1492.98	-16.39	19.59	0.48	7.16
	SQF	1431.85	-14.50	19.02	0.34	7.02
	CW	1044.69	-1147	15.94	-0.10	6.00
	CWF	991.82	-10.37	16.55	-0.10	5.85
Deciduous						
All	SQ	No variable met the 0.15 significance level to entry into the model				
	SQF	No variable met the 0.15 significance level to entry into the model				
	CW	1380.42	-18.70	12.35	-0.04	6.90
	CWF	No variable met the 0.15 significance level to entry into the model				

* 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);



a)



b)

Figure 36: Scatterplots of predicted vs. observed basal area for pine (a) and deciduous plots (b)

Results of the current study are comparable to the findings of previous studies that attempted to assess basal area using either lidar (Nelson *et al.*, 1997; Lefsky *et al.*, 1999) or high resolution imagery (e.g., Wulder *et al.*, 2000). In tropical forests, Nelson *et al.* (1997) explained between 16 and 55% of the field variance for basal area, mainly depending on the on the length of the lidar sampling segments. Lefsky *et al.* (1999) used various canopy height indices derived with the SLICER instrument to estimate basal area and biomass in deciduous forests of eastern Maryland, USA. Their fitted regression models explained between 60 and 70% of the variance associated with basal area. When using a validation data set, their models explained between 3 and 37% of the variance associated with basal area. The reason for lower R^2 for the validation data set was

attributed to the differences in forest conditions found in each data set. Standard deviation of residuals ranged between 44 and 8.9 m^2/ha, for both fitted and validation models. Wulder *et al.* (2000) attempted to assess basal area in stands of Douglas fir and western red cedar using a summary field sample to obtain dbh and a remotely estimated number of stems from high resolution imagery. Their results ranged from underestimating ground-measured basal area (67%) to overestimating it by 158%, depending on the window size and stand height.

5.10. Biomass

Biomass estimated from ground measurements was obtained using dbh only, since previous studies (Crow, 1971; Schroeder *et al*., 1997) proved that dbh is the most reliable variable for biomass estimation. Nevertheless, the ability of lidar estimated parameters to predict biomass was expected to be strong, since regression analysis demonstrated a good fit for predicting height and dbh, which best explain biomass. For pines, the differences between the four processing methods were small, all models explaining more than 78% of the variance (Table 31). Data fusion increased the fit for both the square and circular LM filters, with R^2 of 0.8193 and 0.8183 and RMSE of 29.48 and 29.00 Mg/ha, respectively. For deciduous plots, the explanatory power of the lidar-derived metrics for predicting biomass is lower than for pines, with the highest R^2 of 0.3276 (RMSE of 44.41 Mg/ha) obtained with the model for the circular LM filter. PRESS statistics (Table 32) revealed a standard error of residuals for the best prediction models of 34.37 Mg/ha for the pine plots and 48.49 Mg/ha for deciduous plots. Ground estimated biomass ranged for the pine plots from 3.42 to 314.75, with mean 79.62 Mg/ha and standard deviation 64.61 Mg/ha (Table 6), and for deciduous plots between 15.13 to 270.90 Mg/ha, with mean 131.61 Mg/ha and standard deviation 62.06 Mg/ha. Best fit and prediction was obtained when using the LM filtering technique with circular windows, for both pine and deciduous plots. For the pines, results were improved when using optical data and therefore differentiating between the two species groups.

Table 31. Regression results – dependent variable: biomass (Mg/ha) / subplot*

Trees	Method	Significant independent variables	$S_{y.x}$	R^2	Model
Pines					
All	SQ	N, H_{ave}, CD_{max}	32.09	0.7859	$-59.67489 + 2.74671N + 15.40265H_{ave} - 18.95494\ CD_{max}$
	SQF	N, H_{ave}, CD_{max}	29.48	0.8193	$-40.11945 + 2.44191N + 15.20587H_{ave} - 22.36038CD_{max}$
	CW	CD_{ave}	31.28	0.7809	$-69.00725 + 43.62412CD_{ave}$
	CWF	CD_{ave}, CD_{max}	29.00	0.8183	$-88.20307 + 29.48658CD_{ave} + 14.71462CD_{max}$
Deciduous					
All	SQ	H_{ave}	47.78	0.0924	$74.82695 + 2.72609H_{ave}$
	SQF	H_{ave}	50.68	0.1646	$55.34649 + 3.90188H_{ave}$
	CW	H_{max}, CD_{ave}	44.41	0.3276	$64.34655 + 6.82296H_{max} - 16.39028CD_{ave}$
	CWF	H_{max}	50.60	0.0947	$60.06451 + 3.22552H_{max}$

* Method and variable abbreviations are the same as in Table 16

All the regression models for pines and the model with the highest R^2 value for deciduous plots included lidar estimates of the average crown diameter. For the pine plots, the average crown diameter estimated with lidar when filtering with circular windows explained 78% of the variance associated with biomass. The filtering window shape had a small influence for the fit of regression models for pine plots, but for deciduous plots, a considerable gain in explaining the variance associated with biomass was obtained when filtering with circular windows. Figure 37 shows the 1:1 relationship

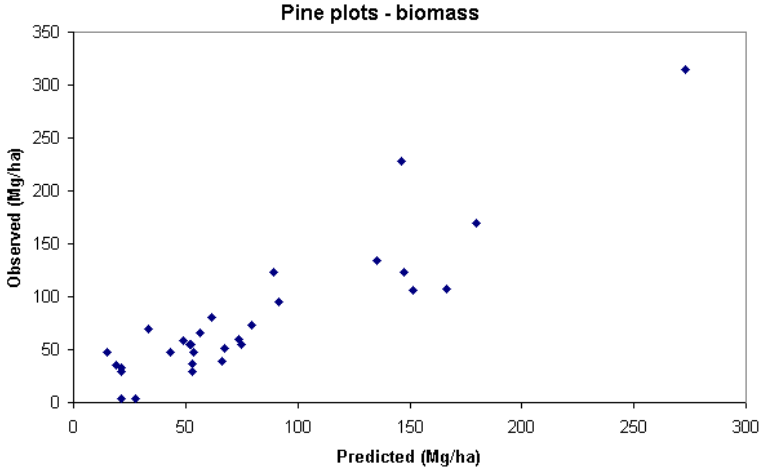
Table 32. PRESS statistics for predicting biomass (Mg/ha) / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	53526	-81.88	125.15	3.84	42.78
	SQF	39270	-72.81	96.88	2.84	36.68
	CW	36346	-72.73	108.78	0.48	35.40
	CWF	34280	-69.60	103.08	0.78	34.37
Deciduous						
All	SQ	71845	-116.93	115.54	-1.47	53.59
	SQF	79626	-113.73	119.85	-0.62	55.34
	CW	68188	-109.03	90.96	-0.39	48.49
	CWF	86103	-111.01	123.21	-1.25	54.47

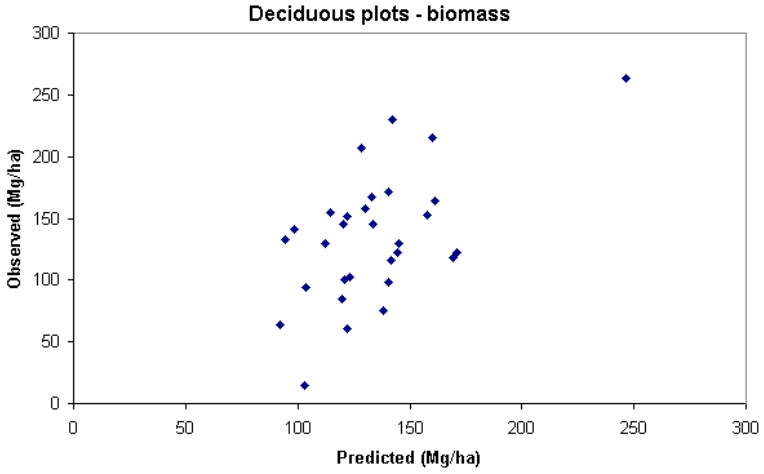
* Method refers to LM filtering technique: SQ (square window), SQF (square window with data fusion), CW (circular window), and CWF (circular window with data fusion);

between lidar-predicted and field-measured biomass for pine and deciduous plots, for the best models only.

The coefficients of determination for total above ground biomass are situated in the range reported for other studies, though it is difficult a make a direct comparison due to differences in lidar sensors, forest types, and ground truth data collection. Nelson *et al.* (1988b and 1997) reported R^2 values in the range of 0.40 to 0.65 for predicting biomass of tropical forests with laser profiling data. Studies using large footprint lidar data (SLICER) reported R^2 values in the range of 0.90 to 0.96 for stands of Douglas fir (*Pseudotsuga menziesii* Franco) and western hemlock (*Tsuga heterophylla* Sarg.) in the Pacific Northwest (Means *et al.*, 1999), and in the range of 0.70 to 0.80 for deciduous



a)



b)

Figure 37: Scatterplots of predicted vs. observed biomass for pine (a) and deciduous plots (b)

forests of Eastern Maryland, USA (Lefsky *et al.*, 1999). When using a validation data set, Lefsky *et al*. (1999), obtained lower R^2 (0 to 36% variance explained), due to the wider stand conditions found in the fitted data set. They found a standard deviation of residuals between 45.8 and 68.1 Mg/ha, approximately between 19 and 28.5% of the average above ground biomass measured on the ground). Biomass residuals exhibit no spatial autocorrelation. For both species groups, pine and deciduous trees, Moran's *I* was calculated for models referring to filtering with circular windows, using optical data only for the pines. P-values for the significance of the *Z*-test under normality and randomization were not significant (0.61 and 0.73, for pine and deciduous plots, respectively).

5.11. Comparison between processing techniques

Processing techniques were compared based on the variance of field-based estimates that each regression model was able to explain. Table 34 indicates the lidar processing technique that led to the regression model with the highest R^2 value for each of the biophysical parameters measured on the field and estimated with lidar. A similar comparison was done based on the prediction ability of each model when judged by the PRESS statistic. Table 33 indicates the lidar processing technique that was associated with the regression model with the lowest PRESS statistic for each of the field-measured forest parameters.

The conclusions that can be drawn from Table 34 are not surprising. Overall filtering for local maximum with circular windows gives better fitting models, since one would expect that a circular window shape is more appropriate for identifying individual tree crowns. Local maximum filtering worked best for both pines and deciduous plots, as judged by the R^2 values of the regression models.

All pine regression models, with only one exception, proved to explain a higher percentage of the variance associated with field-measured parameters when the size of the filtering windows was calibrated for the tree species groups, i.e., when using data fusion in conjunction with the lidar processing techniques. For deciduous plots, the majority of the regression models for estimating field-based estimates (4 out of 7) had a better fit without using optical data. Still, the optical data (ATLAS multispectral imagery)

Table 33. Comparison between processing techniques based on model prediction (PRESS statistic)

Estimated parameter	Species group	Filtering window shape*		Data fusion*	
		Square windows	Circular windows	No	Yes
Height	Pines		•	•	
	Deciduous	•		•	
Crown diameter	Pines		•		•
	Deciduous	•		•	
Dbh	Pines		•		•
	Deciduous	•		•	
Number of trees	Pines	•			•
	Deciduous		•	•	
Volume	Pines		•		•
	Deciduous		•	•	
Basal area	Pines		•		•
	Deciduous		•	•	
Biomass	Pines		•		•
	Deciduous		•	•	
Total		4 (1 pine + 3 deciduous)	10 (6 pines + 4 deciduous)	8 (1 pine + 7 deciduous)	6 (6 pines + 0 deciduous)

* "•" indicates method with lowest PRESS statistic

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; Naesset 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 is 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. All pine models, with one exception, and the majority of the deciduous models provided smaller PRESS residuals when the lidar estimates were obtained by identifying individual trees with circular search windows. While pine models

work better when using data fusion, all deciduous models better predicted field estimates without using optical data.

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. However, both model fit and prediction for deciduous plots indicated that the circular window shape is more appropriate to locate individual trees on the lidar three-dimensional depiction of the complex canopy surface.

Table 34. Comparison between processing techniques based on model fit (R^2 values)

Estimated parameter	**Species group**	**Filtering window shape***		**Data fusion***	
		Square windows	Circular windows	No	Yes
Height	Pines		•		•
	Deciduous	•			•
Crown diameter	Pines	•			•
	Deciduous		•		•
Dbh	Pines		•		•
	Deciduous	•			•
Number of trees	Pines		•		•
	Deciduous	•		•	
Volume	Pines		•		•
	Deciduous		•	•	
Basal area	Pines	•		•	
	Deciduous		•	•	
Biomass	Pines		•		•
	Deciduous		•	•	
Total		5 (2 pines + 3 deciduous)	9 (5 pines + 4 deciduous)	5 (1 pine + 4 deciduous)	9 (6 pines + 3 deciduous)

* "•" indicates method with highest R^2 value

6. Conclusions

The results of the current study show that lidar data could be used to accurately estimate biophysical parameters of forest stands by focusing at 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 co-dominant 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 forest biophysical parameters 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 forest inventory parameters by focusing at the individual tree level. Plot level tree height and crown diameter calculated from individual tree lidar measurements were particularly important in contributing to model fit and prediction of most of the forest parameters. As expected, among the biophysical parameters measured with lidar, tree height was most accurately estimated. Moreover, 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, since 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.

Lidar has proved that it is the most suitable technology for the derivation of high-resolution ground DEMs, despite the fact that vegetation removal is still a challenge for the automated processing of lidar data. However, the current study proved that the lidar

DEM is within error ranges currently found in other sources of elevation data. The accuracy of the lidar ground elevations is indirectly reflected in the results obtained for estimating forest biophysical parameters, especially the tree height. Lately, due to the great interest for creating accurate lidar DEMs for a variety of applications and the availability of lidar processing software for vegetation removal, it is expected that less effort will be allocated to the derivation of the terrain DEM, for the benefit of an increased focus on forestry measurements on lidar surfaces.

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 forest parameters, 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. With the availability of high-density lidar data, with pulse repetition rates that exceed 50,000 points per second, with multiple returns and intensity registration for each pulse, it is expected that lidar will prove even more accurate for estimating forest biophysical parameters of interest. It is therefore expected, that the transition from research to practical applications and operational use of lidar in forestry will accelerate.

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 provides the tools to reliably measure not only tree height, but also crown dimensions, and forest volume and biomass.

7. **References**

Ackermann, F., 1999. Airborne laser scanning – present status and future expectations. *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):64-67.

Arabatzis, A.A. and H.E. Burkhart, 1992. An evaluation of sampling methods and model forms for estimating height-diameter relationships in loblolly pine plantations. *Forest Science* **38**(1): 192-198.

Avery, T.E. and H.E. Burkhart, 1994. *Forest Measurements*. 4th edition, McGraw-Hill, Inc.

Axelsson, P., 1999. Processing of laser scanner data – algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):138-147.

Bailey, T.C. and A.C. Gatrell, 1995. *Interactive Spatial Data Analysis*. Longman Group, Ltd.

Baltsavias, E.P., 1999a. A comparisons between photogrammetry and laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):83-94.

Baltsavias, E.P., 1999b. Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):199-214.

Baltsavias, E.P., 1999c. Airborne laser scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):164-198.

Barbezat, V. and J. Jacot, 1999. The CLAPA project: automated classification of forest with aerial photographs. In *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, Victoria, BC, Feb. 10-12, 1998, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Center, 345-356.

Biging, G.S. and M. Dobbertin, 1995. Evaluation of competition indices in individual tree growth models. Forest Science **41**(2): 360-377.

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

Birdsey, R.A., A.J. Plantiga, and L.S. Heath, 1993. Past and prospective carbon storage in United States forests. *Forest Ecology and Management* **58**(1-2): 33-40.

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.

Briggs, R.D., J.H. Porter, and E.H. White, 1989. Comp onent biomass equations for *Acer rubrum* and *Fagus grandifolia*. *Faculty of Forestry Technical Publication Number 4 ESF 89-005*, SUNY College of Environmental Science and Forestry, Syracuse, New York. 14 p.

Brown, S.L., P. Schroeder, and J.S. Kern, 1999. Spatial distribution of biomass in the eastern USA. *Forest Ecology and Management* **123**: 81-90.

Bufton, J.L., J.B. Garvin, J.F. Cavanaugh. L. Ramos-Izquierdo, T.D. Clem, and W.B. Krabill, 1991. Airborne lidar for profiling of surface topography. *Optical Engineering* **30**(1): 72-78.

Campbell, J.B., 1996. *Introduction to Remote Sensing*. Second edition. The Guilford Press. 622 p.

Ciais, P., P.P. Trolier, J.W.C. White, and R.J. Francey, 1995. A large Northern Hemisphere terrestrial CO_2 sink indicated by the 13C/12C ratio of atmospheric CO2. Science 269: 1098-1102.

Clark III, A., D.R. Phillips, and D.J. Frederick, 1986. Weight, Volume, and Physical Properties of Major Hardwood Species in the Piedmont. USDA Forest Service, Southeastern Forest Experiment Station Research Paper SE-255, June 1986.

Cliff, A.D. and J.K. Ord, 1981. *Spatial Processes – Models and Applications.* Pion, Ltd.

Clutter, J.L., J.C. Fortson, L.V. Pienaar, G.H. Brister and R.L. Bailey (1983). *Timber Management: a Quantitative Approach*. John Wiley and Sons.

Corvallis Microtechnology, Inc., (2001). CMT Products - HP-GPS-L4. 2000. Available http://www.cmtinc.com/nav/frprod.html [October 30, 2001].

Crow, T.R., 1971. Estimation of biomass in an even-aged stand - Regression and "mean tree" techniques. In *Forest Biomass Studies*, 25th IUFRO Congress, Gainesville, Florida, USA: 35 – 48.

Daley, N.M.A, C.N. Burnett, M. Wulder, L.O. Niemann, and D.G. Goodenough, "Comparison of fixed-size and variable-sized windows for the estimation of tree crown position," *IEEE Transactions on Geoscience and Remote Sensing*, 1998.

Detwiler, R.O., and C.A.S. Hall, 1988. Tropical forests and the global carbon cycle. *Science* **239**:42-47.

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.

Dubayah, R., J.B. Blair, J.L. Bufton, D.B. Clarke, J. Jaja, R. Knox, S.B. Luthcke, S. Prince, and J. Weishampel, 1997. The Vegetation Canopy Lidar Mission. In *Land Satellite Information in the Next Decade II: Sources and Applications*, ASPRS Proceedings, Washington, DC.

Erdas Imagine, 1997. *Erdas Field Guide*. Fourth Edition. Erdas, Inc., Atlanta, Georgia.

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

Flood, M. and B. Gutelius, 1997. Commercial Implications of Topographic Terrain Mapping Using Scanning Airborne Laser Radar. *Photogrammetric Engineering and Remote Sensing* **63**(4): 327-366.

Gill, S.J., G.S. Biging, and E.C. Murphy, 2000. Modeling conifer tree crown radius and estimating canopy cover. *Forest Ecology and Management* 126: 405-416.

Gillett, P., 1984. *Calculus and Analytic Geometry*. Second edition, D.C Heath and Company.

Gomes Pereira, L.M., Wicherson, R.J., 1999. Suitability of laser data for deriving geographical information - a case study in the context of management of fluvial zones. . *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):105-114.

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., 1999. Automatic individual tree crown delineation using a valley-following algorithm and a rule-based system. In *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, Victoria, BC, Feb. 10-12, 1998, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Center, 11-23.

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*, Sainte-Foy, Québec, Canada, August 21-24, 2001.

Green, E.J., 1981. Models of stand basal area distributions, individual tree basal area growth, and height-diameter relationships for loblolly pine. Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

Hahn, J.T., 1984. Tree volume and biomass equations for the Lake States. Research Paper NC-250. St. Paul, MN: USDA Forest Service, North Central Experiment Station. 10 p.

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*. ESA Scientific & Technical Pub. Noordwijk, The Netherlands, pp. 1251-1253.

Hill, J.M., L.A. Graham, and R.J. Henry, 2000. Wide-area topographic mapping and applications using airborne light detection and ranging (LIDAR) technology. *Photogrammetric Engineering & Remote Sensing* **66**(8): 908-914.

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 Transaction on Geoscience and Remote Sensing* **39**(5): 969-975.

Jaafar, J., G. Priestnall, P.M. Mather, and C.A. Vieira, 1999. Construction of DEM from Lidar DSM using morphological filtering, conventional statistical approaches and Artificial Neural Networks. *In Earth Observation: From Data to Information*, Proceedings of the 25th Annual Conference and Exhibition of the Remote Sensing Society, University of Wales at Cardiff and Swansea, 8-10 September, 1999: 299-306.

Krabill, W.B., J.G. Collins, L.E. Link, R.N. Swift, and M.L. Butler, 1984. Airborne laser topographic mapping results. *Photogrammetric Engineering & Remote Sensing* **50**(6): 685-694.

Kraus, K., and N. Pfeifer, 1998. Determination of terrain models in wooded areas with airborne scanner data, *ISPRS Journal of Photogrammetry & Remote Sensing, **53***: 193-203.

Kurz, W.A., S.J. Beukema, and M.J. Apps, 1996. Estimation of root biomass and dynamics for the carbon budget model of the Canadian forest sector. *Canadian Journal of Forest Research* **26**: 1973-1979.

Lam, N.S., (1983). Spatial interpolation methods: A review. *The American Cartographer*, **10**(2): 129-149.

Leckie,, D.G. and F. Gougeon, 1999. An assessment of both visual and automated tree counting and species identification with high spatial resolution multispectral imagery. In *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, Victoria, BC, Feb. 10-12, 1998, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Center, 141-142.

Lefsky, M.A, (1998, August 27). Lidar Remote Sensing of Forest Structure. Available http://www.fsl.orst.edu/~lefsky/slicerpage.html [October 5, 2000]

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. In *Natural Resources Management Using Remote Sensing and GIS*. Greer, J.D., Ed., ASPRS, Washington, D.C.: 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 G.L. Martin, 1984. Merchantable timber volume estimation using cross-sectional photogrammetric and densitometric methods. *Canadian Journal of Forest Research* **14**(6): 803-810.

Maclean, G.A. andx 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. 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.

Means, J.E., 2000. Comparison of large-footprint and small-footprint lidar systems: design, capabilities, and uses. In Proceedings: *Second International Conference on Geospatial Information in Agriculture and Foretry*, Lake Buena Vista, Florida, 10-12 January 2000: I-185-192.

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.

Myers, R.H., 1990. *Classical and Modern Regression with Applications.* The Duxbury advanced series in statistics and decision sciences. PWS-Kent, Boston.

National Geodetic Survey (2001, March 5). NGS Geodetic Tool Kit. Available http://www.ngs.noaa.gov/TOOLS/ [October 30, 2001].

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.

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

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., 1994. The use of airborne laser altimetry to estimate tropical forest basal area, volume, and biomass. Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, 145 p.

Nelson, R.F., 1997. Modeling forest canopy heights: the effects of canopy shape. *Remote sensing of environment* **60**: 327-334.

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.

Nelson, R.F., T.G. Gregoire, and R.G. Oderwald, 1998. The effects of fixed-area plot width on forest canopy height simulation. *Forest Science* **44**(3): 438-444.

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

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

Oderwald, R.G. and B.A. Boucher, 1997. *Where in the World and What? An Introduction to Global Positioning Systems.* Kendall/Hunt, Iowa.

Pachepsky, Y.A., J.C. Ritchie, and D. Gimenez. 1997. Fractal modeling of airborne laser altimetry data. *Remote Sensing of Environment* **61**: 150-161.

Parresol, B.R., 1999. Assessing tree and stand biomass: a review with examples and critical comparisons. *Forest Science* **45**(4): 573-593.

Perala, D.A. and D.H. Alban, 1993. Allometric biomass estimators for aspen-dominated ecosystems in the upper Great Lakes. *Research Paper NC-314. St. Paul, MN: USDA Forest Service, North Central Experiment Station. 38 p.*

Petzold, B., P. Reiss, and W. Stössel, 1999. Laser scanning - surveying and mapping agencies are using a new technique for the derivation of digital terrain models. *ISPRS Journal of Photogrammetry & Remote Sensing* **54**:95-104.

Pinz, A., 1999. Tree isolation and species classification. In *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, Victoria, BC, Feb. 10-12, 1998, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Center, 127-139.

Popescu, S.C., R.H. Wynne and R.H. Nelson, 2000. Estimating forest vegetation biomass using airborne lidar measurements. In Proceedings: *Second International Conference on Geospatial Information in Agriculture and Foretry*, Lake Buena Vista, Florida, 10-12 January 2000: II 346-353.

Popescu, S.C., R.H. Wynne and R.H. Nelson, In press. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture*. Elsevier Science.

Powell, D.S., J.L. Faulkner, D.R. Darr, Z. Zhou, and D.W. MacCleerry, 1993. The forest resources in the United States, 1992. *USDA Forest Service General Technical Report*, RM-234. 132 p.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, 1992. *Numerical recipes in C: The art of scientific computing*. Second edition. Cambridge University Press.

Renslow, M.S., 1999. Applications of advanced lidar for DEM applications. In *From Image to Information*, ASPRS Proceedings, Portland, Oregon.

Reutebuch, S.E. and R.J. McGaughey, 1999. Comparison of an airborne laser terrain model to USGS digital elevation models. In *From Image to Information*, ASPRS Proceedings, Portland, Oregon.

Ritchie, J.C, 1995. Airborne laser altimeter measurements of landscape topography. *Remote Sensing of Environment* **53**: 91-96.

Ritchie, J.C., D.L. Evans, D. Jacobs, J.H. Everitt, and M.A. Weltz, 1993. Measuring canopy structure with an airborne laser altimeter. *Transaction of the ASAE* **36**(4): 1235-1238.

Saucier, R.J. and A. Clark III, 1985. *Tables for Estimating Total Tree and Product Weight and Volume of Major Southern Tree Species and Species Groups.* Southwide Energy Committee, American Pulpwood Association, Inc.

Schabenberger, O. and F.J. Pierce, 2001. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press.

Schlesinger, W.H., 1995. An overview of the carbon cycle. *Soils and Global Change* (R. Lala, J. Kimble, E. Levine, and B.A. Stewart, editors). Lewis.

Schreier, H., J. Lougheed, C. Tucker, and D. Leckie, 1985. Automated measurements of terrain reflection and height variations using and airborne infrared laser system. *International Journal of Remote Sensing* **6**(1): 101-113.

Schroeder, P., S. Brown, J. Mo, R. Birdsey, and C. Cieszewski, 1997. Biomass estimation for temperate broadleaf forests of the United States using inventory data. *Forest Science* **43**: 424 – 434.

Schumacher, F.X. and F.D.S. Hall, 1933. Logarithmic expression of timber-tree volume. *Journal of Agriculture Research* **47**: 719-734.

Sharma, M. and R.G. Oderwald, 2001. Dimensionally compatible volume and taper equations. Canadian *Journal of Forest Research* **31**(5): 797-803.

Smith, W. B., 1985. Factors and equations to estimate forest biomass in the North Central region. Research Paper NC-268. St. Paul, MN: USDA Forest Service, North Central Forest Experiment Station. 6 p.

Smith, W.R., R.M. Farrar, Jr., P.A. Murphy, Y.L. Yeiser, R.S. Meldahl, and J.S. Kush, 1992. Crown and basal relationships of open-growth southern pines for modeling competition and growth. *Canadian Journal of Forest Research* 22: 341-347.

Sprinz, P.T. and H.E. Burkhart, 1987. Relationships between tree crown, stem, and stand characteristics in unthinned loblolly pine plantations. *Canadian Journal of Forest Research* 17: 534-538.

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.

Tans, P.P., I.Y. Fung, and T. Takahashi, 1990. Observational constraints on the global atmospheric CO_2 budget. *Science* **247**:1431-1438.

Toth, C.K., S. Berning, J. Leonard, and D.A. Grejner-Brzezinska, 2001. Integration of lidar data with simultaneously acquired digital imagery. In Proceedings: *ASPRS 2001: Gateway to the New Millennium,* April 23-27, St. Louis, Missouri.

Tritton, L.M. and J.W. Hornbeck, 1982. Biomass equations for major tree species of the Northeast. Broomall, PA: Northeast Forest Experiment Station; *USDA Forest Service General Technical Report NE-69*. 46p.

U.S. Department of Agriculture, 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, D.C., 20250.

U.S. Geological Survey (August 13, 2001). Map accuracy standards. Available http://mac.usgs.gov/mac/isb/pubs/factsheets/fs17199.html [October 30, 2001].

Vosselman, G., 2000. Slope base filtering of laser altimetry data. *International Archives of Photogrammetry and Remote Sensing* **33**: 935-942.

Wehr, A. and U. Lohr, 1999. Airborne laser scanning – an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):68-82.

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 of SPIE*: Advances in Laser Remote Sensing for Terrestrial and Oceanographic Applications* **3059**: 184-189.

Weltz, M.A., J.C. Ritchie, and H.D. Fox, 1994. Comparison of laser and field measurements of vegetation height and canopy cover. *Water Resources Research* **30**(5): 1311-1319.

Wu, Y. and A.H. Strahler, 1994. Remote estimation of crown size, stand density, and biomass on the Oregon transect. *Ecological Applications* **4**(2): 299-312.

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 the Environment* **73**: 103-114.

Yegorov, A.D. and I.A. Potapova, 2000. In Proceedings: *Second International Conference on Geospatial Information in Agriculture and Foretry,* Lake Buena Vista, Florida, 10-12 January 2000: I 478-480.

Young, B., D.L. Evans, and R.C. Parker, 2000. Methods for comparison of lidar and field measurements of loblolly pine. In Proceedings: *Second International Conference on Geospatial Information in Agriculture and Foretry,* Lake Buena Vista, Florida, 10-12 January 2000: I-193-199.

8. **Appendices**

Appendix 1. Location of FIA subplots and GPS points used for DEM assessment

Point type	UTM coordinates, zone 17, NAD83		
	X	Y	Z
Pine subplots			
1	706700.4395	4144922.6180	156.83
2	706670.4357	4144896.8660	157.70
3	706933.2529	4144923.1060	145.50
4	707102.6020	4144936.5000	156.99
5	707094.4832	4144973.0980	154.90
6	707136.0937	4144927.7740	150.04
7	707076.1704	4144912.2490	157.41
8	706707.3403	4144727.9690	160.53
9	706702.3575	4144766.1160	158.63
10	706744.6964	4144715.5020	156.85
11	706676.4897	4144702.5190	157.13
12	706869.4423	4144729.4630	151.08
13	706861.4283	4144765.4760	151.55
14	706903.0114	4144720.7030	146.21
15	706841.3859	4144706.0470	150.84
16	707110.8193	4144776.0650	157.46
17	707143.1561	4144726.3880	150.28
18	706711.9331	4144530.5390	149.42
19	706706.4035	4144566.2720	152.06
20	706681.3444	4144508.9260	152.65
21	706911.7860	4144546.1090	149.21
22	706903.5218	4144582.5270	152.71
23	706886.3382	4144524.1410	144.08
24	707129.8272	4144562.9310	144.11
25	707123.5577	4144596.6520	150.76
26	707164.7546	4144550.0490	148.12
27	704290.9920	4145066.7428	179.18
28	704286.0608	4145105.0467	187.73
29	704325.0063	4145057.8256	176.86
30	704290.0871	4144909.8506	177.53
31	704672.1557	4144852.6515	183.30
Deciduous subplots			
1	706692.1091	4144963.9670	149.17
2	706737.9841	4144909.9790	155.21
3	706896.0417	4144935.3610	141.05
4	706889.1317	4144975.7380	139.46
5	706864.6541	4144909.1210	141.08
6	707110.9073	4144738.6960	149.18
7	707083.3936	4144716.0230	146.17
8	706746.2304	4144517.2220	145.73
9	706947.6675	4144537.3710	144.76
10	707102.0046	4144537.8390	148.13
11	706121.7928	4144313.0024	158.41
12	706113.9834	4144347.2063	156.13
13	706156.1488	4144301.5959	145.45

14	706093.4493	4144288.2754	153.71
15	704265.7052	4145045.3591	176.03
16	704499.4489	4145068.8960	180.29
17	704489.8331	4145111.4129	179.48
18	704538.4765	4145055.0895	177.49
19	704469.7449	4145043.4761	172.22
20	704697.9733	4145078.9802	165.14
21	704691.0696	4145117.7774	171.21
22	704736.0816	4145065.9882	163.27
23	704670.5252	4145055.0373	168.78
24	704300.7672	4144867.0131	170.46
25	704338.5198	4144854.5235	168.70
26	704269.7683	4144839.1685	171.56
27	704497.0261	4144872.3246	176.08
28	704491.8135	4144910.9068	168.15
29	704538.8005	4144857.6128	177.83
30	704468.4358	4144844.7159	174.98
31	704700.3829	4144877.4241	182.56
32	704692.6830	4144918.1714	171.82
33	704736.0535	4144866.1298	178.72
GPS points under the forest canopy			
1	704361.3096	4144840.889	170.60
2	704612.8891	4144859.151	172.12
3	704488.6336	4144922.756	164.48
4	704290.2516	4144966.554	174.43
5	704297.707	4145021.264	171.69
6	704341.0454	4145115.111	181.22
7	704560.6689	4145078.385	174.19
Open ground GPS points			
1	706707.5687	4144984.0004	153.41
2	706742.2257	4144981.7086	155.36
3	706780.6479	4144972.6630	153.83
4	706806.0233	4144967.2300	149.87
5	706783.0328	4144991.5357	152.10
6	706758.7578	4145025.8660	152.75
7	706717.2078	4145013.1848	154.60

Appendix 2. Field data measurements for the pine subplots (average values / subplot)*

	$H_{2.5}$ m	H_5 m	H_{dom} m	H_{max} m	H_{min} m	$N_{2.5}$	N_5	N_{dom}	$Dbh_{2.5}$ cm	Dbh_5 cm	Dbh_{dom} cm	Dbh_q cm	CD_5 m	CD_{dom} m	Volume m^3/ha	Biomass Mg/ha	BA m^2/ha
Pines																	
1	8.05	8.58	8.89	9.45	6.40	12	7	6	12.70	14.99	15.28	13.15	3.09	2.92	40.24	32.81	9.69
2	8.86	10.40	10.40	11.28	5.49	15	9	8	13.38	16.14	16.51	13.90	3.64	3.73	64.46	47.86	13.54
3	7.28	-	10.36	10.36	5.64	5	-	1	7.57	-	10.41	7.71	-	-	5.36	4.19	8.54
4	11.68	13.28	13.28	16.15	4.27	13	7	7	14.17	18.14	18.14	14.84	3.13	3.13	75.33	54.69	11.72
5	10.45	13.45	13.49	15.24	5.64	23	9	8	12.29	17.66	18.14	13.35	no data	no data	101.90	73.49	18.39
6	13.30	16.94	16.94	20.73	6.10	28	12	12	13.66	19.66	19.66	14.49	3.67	3.67	201.07	133.98	28.45
7	10.41	12.90	14.17	14.33	4.27	7	3	2	13.32	19.64	22.86	14.77	4.48	4.77	39.18	29.62	7.13
8	6.67	7.35	7.28	8.53	5.49	18	8	9	10.94	13.65	13.52	11.28	1.97	1.97	43.15	38.45	9.95
9	8.08	8.72	8.34	9.45	6.70	18	4	10	11.80	13.56	12.72	11.89	2.77	2.77	47.74	36.20	11.89
10	8.30	8.97	9.02	9.75	5.49	23	14	12	13.44	15.42	15.77	13.77	3.13	3.19	88.57	69.90	20.37
11	5.03	-	6.10	6.10	4.27	4	-	1	7.94	-	11.43	8.19	-	-	3.42	3.42	1.25
12	10.85	11.89	11.50	12.50	9.75	47	6	14	9.75	15.62	13.08	10.12	4.43	4.43	107.14	66.22	22.47
13	8.07	8.52	8.60	9.75	6.09	30	9	16	11.35	14.14	13.30	11.62	3.36	3.36	73.08	58.22	18.94
14	10.42	11.63	11.31	12.19	7.01	30	6	11	10.19	14.35	13.23	10.56	3.90	3.90	71.61	47.40	15.64
15	8.49	9.14	9.14	10.06	5.49	20	12	11	13.25	15.20	15.42	13.55	3.01	3.04	75.76	59.26	17.17
16	12.65	14.07	14.30	15.85	7.31	29	18	12	14.63	17.74	19.35	15.37	3.25	3.54	207.29	122.99	31.99
17	16.09	24.31	29.26	34.75	5.49	19	8	5	19.50	33.72	45.01	25.38	6.11	7.61	148.80	93.29	57.19
18	7.84	8.72	8.26	9.75	6.10	25	5	11	9.51	13.26	12.08	9.83	3.52	3.52	45.17	35.77	11.28
19	8.09	9.14	8.35	9.75	5.79	15	5	13	11.19	14.53	11.92	11.56	3.08	3.08	38.66	29.11	9.36
20	8.57	9.61	9.84	10.67	6.10	23	13	7	12.02	14.38	15.31	12.43	2.95	2.96	74.15	54.27	16.60
21	11.11	17.80	19.28	20.42	5.79	28	5	4	13.43	34.85	40.07	17.40	7.78	8.79	296.92	227.92	39.59
22	10.87	12.19	11.47	12.80	7.62	68	4	26	9.27	14.54	11.34	9.50	2.74	2.74	134.59	79.88	28.66
23	11.83	12.65	12.50	14.63	9.14	38	4	19	9.87	13.53	11.59	10.10	3.40	3.40	93.56	51.52	17.75
24	17.37	23.01	30.18	30.18	6.10	3	2	1	26.67	36.20	52.83	32.82	6.87	9.16	213.07	105.85	15.10
25	15.61	25.53	27.43	28.35	6.40	9	4	3	24.98	46.36	51.73	32.11	10.12	10.89	571.99	314.75	43.36
26	9.75	18.90	18.90	27.74	5.79	7	2	2	12.37	26.29	26.29	16.02	4.97	4.97	89.14	47.02	8.39
27	14.80	17.90	19.09	21.95	7.32	23	15	11	16.45	20.95	23.09	17.96	1.98	2.21	288.08	169.49	34.65
28	14.12	17.40	18.42	20.73	5.79	19	11	7	15.73	20.78	23.91	17.28	4.32	4.57	200.84	123.59	26.52
29	8.33	10.15	9.47	11.28	6.71	35	7	13	10.06	14.88	13.31	10.53	3.37	3.37	70.83	54.59	18.14
30	11.49	12.18	12.38	13.41	7.62	29	19	16	13.75	15.60	16.00	14.09	3.12	3.20	153.05	95.23	26.88
31	12.97	15.96	18.49	19.20	7.92	16	8	3	14.57	20.89	29.29	16.84	5.02	6.04	145.23	107.39	21.20

* $H_{2.5}$ – average height of all measured trees; H_5 – average height of trees with dbh larger than 12.7 cm (5 in); H_{dom} – average height of dominant trees; H_{max} – maximum height; H_{min} - minimum height; $N_{2.5}$ – number of trees; N_5 – number of trees with dbh larger than 12.7 cm (5 in); N_{dom} – number of dominant and co-dominant trees; $Dbh_{2.5}$ – average dbh of all trees; Dbh_5 – average dbh of trees with dbh larger than 12.7 cm (5in); Dbh_{dom} – average dbh of dominant trees; Dbh_q – quadratic mean diameter; CD_5 – average crown diameter of all trees with dbh larger than 12.7 cm (5 in); CD_{dom} – average crown diameter of dominant trees; BA – basal area.

Appendix 2 (cont). Field data measurements for the deciduous subplots (average values / subplot)*

	$H_{2.5}$ m	H_5 m	H_{dom} m	H_{max} m	H_{min} m	$N_{2.5}$	N_5	N_{dom}	$Dbh_{2.5}$ cm	Dbh_5 cm	Dbh_{dom} cm	Dbh_q cm	CD_5 m	CD_{dom} m	Volume m^3/ha	Biomass Mg/ha	BA m^2/ha
1	10.83	13.85	16.84	23.47	6.10	19	9	4	15.40	22.75	31.62	18.60	4.69	5.83	226.36	206.65	30.71
2	9.50	10.80	10.80	17.37	6.71	19	7	7	11.87	16.73	16.73	12.85	3.79	3.79	76.92	63.27	14.66
3	12.59	18.90	21.03	22.86	4.57	10	5	4	17.48	25.45	28.38	20.01	6.38	6.79	153.20	122.42	18.71
4	15.80	15.80	21.79	22.56	10.97	6	6	2	24.93	24.93	37.47	26.74	7.16	8.17	141.54	140.93	20.05
5	15.37	20.73	26.21	26.21	7.01	7	4	2	19.70	27.94	37.59	23.10	5.44	5.44	174.60	122.75	17.45
6	12.77	19.71	19.71	20.42	6.40	9	3	3	18.71	36.66	36.66	22.68	7.89	7.89	149.62	154.29	21.62
7	14.91	19.75	23.37	28.65	7.01	11	5	3	16.21	24.94	32.68	19.51	5.35	7.21	182.11	132.44	19.56
8	13.22	20.12	21.34	22.56	5.49	13	6	3	17.60	28.45	31.07	20.46	7.31	8.94	203.59	166.96	25.43
9**	8.58	10.97	10.29	10.97	4.57	13	1	4	8.42	14.22	11.11	8.70	6.20	6.20	20.76	18.64	4.60
10	10.06	14.94	14.94	14.94	7.01	3	1	1	12.45	21.59	21.59	14.03	4.13	4.13	17.51	15.13	2.76
11	10.61	16.23	16.23	19.51	5.18	16	4	4	12.83	24.83	24.83	14.93	6.18	6.18	109.87	98.27	16.67
12	12.80	16.40	17.85	19.20	4.88	17	10	7	15.97	21.67	24.35	17.87	5.09	5.40	184.59	144.80	25.98
13	13.20	15.08	16.59	20.42	4.57	22	13	9	16.03	20.53	23.14	17.48	4.27	4.61	216.34	171.40	31.39
14	12.43	14.19	15.42	16.15	6.40	17	11	5	14.40	17.09	20.47	15.25	4.45	5.30	116.05	94.21	18.46
15	10.79	17.07	18.29	22.56	3.96	15	5	4	14.90	27.33	30.86	18.25	5.57	5.75	172.82	151.33	23.34
16	18.64	21.64	22.86	25.29	8.23	7	5	4	28.88	35.92	40.96	32.25	6.96	7.58	323.68	262.90	34.01
17	11.03	12.86	12.86	15.85	5.49	21	10	10	12.45	16.41	16.41	13.13	3.95	4.11	97.87	84.22	16.92
18	12.42	17.83	24.69	24.69	5.49	4	2	1	21.34	34.42	47.24	26.65	7.28	9.88	121.49	102.65	13.28
19	14.56	22.17	24.28	25.91	6.40	9	4	3	20.40	35.62	42.50	25.83	7.19	8.05	273.35	215.36	28.05
20	14.39	17.58	19.96	28.04	6.40	9	6	4	20.63	26.46	31.69	23.61	5.77	6.69	207.10	164.00	23.45
21	15.00	16.99	25.30	25.30	7.01	5	4	1	22.35	25.97	46.99	25.95	6.55	12.02	145.30	116.12	15.73
22	12.64	17.58	19.58	22.25	4.57	11	6	4	17.71	25.61	30.73	20.89	5.37	5.41	178.05	152.38	22.42
23	14.22	16.38	20.27	20.42	9.14	6	4	2	20.66	25.65	34.16	23.00	6.77	9.39	113.50	100.39	14.83
24**	10.17	12.34	11.95	17.07	5.18	8	4	5	11.21	14.03	13.61	11.68	4.64	4.64	27.14	23.65	5.10
25	13.39	17.42	20.93	24.99	6.71	17	7	3	14.58	22.64	34.71	17.68	6.06	7.83	209.94	157.96	24.83
26	14.39	17.37	20.12	20.73	6.10	9	6	2	19.25	24.00	35.56	22.01	4.85	7.48	168.26	129.85	20.37
27	13.72	18.44	21.87	23.77	4.88	12	6	4	20.17	30.56	37.34	24.36	7.60	8.15	290.69	229.68	33.27
28	13.08	18.64	18.64	22.25	5.49	12	6	6	17.57	26.67	26.67	20.06	5.54	5.54	175.25	145.39	19.03
29	13.06	16.76	17.98	18.90	7.62	6	3	2	17.53	25.99	32.00	20.60	6.13	6.45	89.07	75.64	11.90
30**	17.01	22.00	27.28	33.53	7.62	10	6	4	22.48	31.54	40.26	27.43	7.96	9.46	403.61	270.90	35.16
31	12.25	21.64	21.64	23.77	6.71	10	2	2	12.07	27.05	27.05	14.61	6.22	7.22	83.46	60.85	9.97
32	10.93	23.01	28.96	28.96	3.96	8	2	1	14.57	32.89	51.05	20.22	8.85	12.80	162.53	117.70	15.29
33	14.25	17.42	21.13	24.99	5.49	12	7	3	16.34	21.77	29.55	18.91	5.68	7.74	169.54	129.93	20.04

* $H_{2.5}$ – average height of all measured trees; H_5 – average height of trees with dbh larger than 12.7 cm (5 in); H_{dom} – average height of dominant trees; H_{max} – maximum height; H_{min} - minimum height; $N_{2.5}$ – number of trees; N_5 – number of trees with dbh larger than 12.7 cm (5 in); N_{dom} – number of dominant and co-dominant trees; $Dbh_{2.5}$ – average dbh of all trees; Dbh_5 – average dbh of trees with dbh larger than 12.7 cm (5in); Dbh_{dom} – average dbh of dominant trees; Dbh_q – quadratic mean diameter; CD_5 – average crown diameter of all trees with dbh larger than 12.7 cm (5 in); CD_{dom} – average crown diameter of dominant trees; BA – basal area.

** observations not used in analysis (outliers, see section 4.8.1. and 5.1.).

Appendix 3. IDL program to implement local filtering with variable size windows for measuring tree height and crown diameter

```
;  Author:  Sorin Popescu  5/31/01       last revised:3/12/02
;  Program to identify treetops on lidar array using a variable window size,
;  and measure tree height and crown diameter
;
;*** define minimum height for forest vegetation
minH = 3.9 ; (m) based on height values for the FIA plots

;*** define LIDAR "image" resolution
res=0.50 ; (m)
minwsize=1000
maxwsize=1
;meters to ft conversion
m2ft = 3.28084
;define I/O files
  inFileName = 'CHM_grid_file' ;input file name
  outFileName = 'output_file_name' ;output file name with location of each tree, crown radius, and height
;open I/O files
  openr, 1, inFileName
  openw, 2, outFileName
;define variables to hold header information
LetterCode = '' ;define variable to hold letter code that ids the ASCII grid file
dimensions = intarr(2)        ;array to hold grid dimensions
Xdim = dblarr(2) ; array to hold min and max X coords of the grid
Ydim = dblarr(2) ; array to hold min and max Y coords of the grid;
;read file header
READF, 1, LetterCode      ;read first line
READF, 1, dimensions ;read dimensions
ncol = dimensions (0)
nrow = dimensions (1)
READF, 1, Xdim;read x min-max coords
xmin = Xdim(0)
xmax = Xdim(1)
READF, 1, Ydim;read ymin-max coords
ymin = Ydim(0)
ymax = Ydim(1)
;close file after reading header info
close, 1
chmgr = fltarr(ncol, nrow); variable to hold CHM grid, imported from ENVI (must be opened)
stndht=fltarr(ncol,nrow)     ; array to hold height info
treetop=intarr(ncol, nrow)  ; binary array to hold treetop info
crownDiam = fltarr(ncol, nrow, 3); array to hold crown diameter
wsmatrix = intarr(ncol, nrow) ; holds window size
radii=fltarr(4) ; vector to hold crown radius measured in 4 directions: R, L, Up, and Dw
stndht = chmgr
print, "Now getting treetops..."
;*** get the tree tops
for j=1, ncol-2 do $
  for i=1, nrow-2 do begin
            h=stndht(j,i)
            skip = 1
            cw = 2.51503 + 0.00901*(h)^2 ;crown RADIUS as a funtion of height
            wsize = round(cw/res) ; round wsize to an integer, IN PIXELS
```

```
        if (wsize LT 3) then wsize = 3 ; 3 is minimum for window size
        if ((wsize mod 2) EQ 0) then wsize = wsize - 1 ; if wsize is even, subtract 1
   if (wsize GT 31) then wsize = 31 ; max cw on the ground was 14 m= 28 pixels
   winmat=fltarr(wsize,wsize)
        ; calculate window indeces
        colmin = j-(wsize-1)/2 ;left col
          if colmin LT 0 then skip = 0 ; skip point if index not good
        colmax = j+(wsize-1)/2 ;right col
          if colmax GT ncol-1 then skip = 0
        rowmin = i-(wsize-1)/2 ;upper row
          if rowmin LT 0 then skip = 0
        rowmax = i+(wsize-1)/2 ;lower row
          if rowmax GT nrow-1 then skip = 0
        if (skip EQ 1) then begin
          winmat=stndht(colmin:colmax, rowmin : rowmax)
           if ((max(winmat) EQ h) AND (h GE minH)) then begin  ;tree height should be greater than
minH
                    treetop(j,i)=1       ; there's a tree top
             wsmatrix(j,i)=wsize; record the window size that identified the tree top
           endif
   endif
  endfor
print, "Found ", total(treetop), " trees"
print, "Now getting diameters..."
print, "Smoothing array first...median 3x3..."
stndhtMedian = median (stndht, 3)
for j = 0, ncol-1 do $
  for i = 0, nrow-1 do begin
    if treetop(j,i) EQ 1 then begin
    wsize=wsmatrix(j,i)
    ;__________ columns
          maxicol = j+wsize-1
          minicol = j-wsize+1
          reduceColDim=0
          if (maxicol GE ncol) then begin
                reduceColDim = maxicol-ncol
                maxicol = ncol-1
                minicol = minicol+reduceColDim ;to keep the array symetric around the treetop
          endif
          if (minicol LE 0) then begin
                reduceColDim = 0-minicol
                minicol = 0
                maxicol = maxicol-reduceColDim ;to keep the array symetric around the treetop
          endif
          HColvector = fltarr(maxicol-minicol+1); linear array to hold height values
          HColvector = stndhtMedian(minicol:maxicol, i)
          xx=findgen(maxicol-minicol+1) ; define vector to hold
          xx=xx*res
          lungime = maxicol-minicol+1
                crownDiam (j,i, 0) = cdflex4deriv (xx, HColvector, lungime, res)
                ;rows
                maxirow = i+wsize-1
          minirow = i-wsize+1
          reduceDim=0
          if (maxirow GE nrow) then begin
                reduceDim = maxirow-nrow
```

```
                maxirow = nrow-1
                minirow = minirow+reduceDim ;to keep the array symetric around the treetop
            endif
            if (minirow LE 0) then begin
                reduceDim = 0-minirow
                minirow = 0
                maxirow = maxirow-reduceDim ;to keep the array symetric around the treetop
            endif
            Hvector = fltarr(maxirow-minirow+1); linear array to hold height values
            Hvector = stndhtMedian(j, minirow:maxirow)
            xx=findgen(maxirow-minirow+1)
            xx=xx*res
            lungime = maxirow-minirow+1
            Hvector = rotate(Hvector, 3)
                crownDiam (j,i, 1) = cdflex4deriv (xx, Hvector, lungime, res)

            if (crownDiam(j,i,0) EQ 0 and crownDiam(j,i,1) NE 0) then $
                        crownDiam(j,i,2) =  crownDiam(j,i,1) $
            else if (crownDiam(j,i,0) NE 0 and crownDiam(j,i,1) EQ 0) then $
                        crownDiam(j,i,2) =  crownDiam(j,i,0) $
                 else crownDiam(j,i,2) = mean([crownDiam(j,i,0), crownDiam(j,i,1)])
    endif
  endfor
 nrtrees =long(Total(treetop))
 Print, 'Tree tops : ',nrtrees
;calculate coordinates of trees
nrtree = 1
for j = 0, ncol-1 do $
  for i = 0, nrow-1 do begin
    if treetop(j,i) EQ 1 then begin
      xtree = xmin + j*res + res/2 ;add 0.25m or 1/2 of resolution to center tree in the pixel
      ytree = ymin + i*res - res/2 ;subtract for the same reason as above
      ws = wsmatrix(j,i)
                TopHeight = stndht(j,i)
      printf, 2, format = '($, i6,",",f12.2,",", f12.2,",", f7.2, ",", f7.2, /)', nrtree, xtree, ytree,
crownDiam(j,i,2)/2, TopHeight
      nrtree = nrtree + 1
    endif
  endfor
printf, 2, 'end'
;*** close file units
close, /all
print, "The end"
end
```

Appendix 4. Regression analysis implemented in SAS

```
/* File name:  regression_analysis.sas */
title 'Regression analysis';
data all;
infile 'z:\appomattoxlidar\FieldData\FIA\SASdatafiles\cw_no.txt'
firstobs = 2;
input Pl Sp Nl MeanH MinH MaxLH StdH MeanCD MinCD MaxCD StdCD Hq CDq Nq
      Avdbhq Vol Biomass BA Hmax H25in H5in N25in N5in CD5 DBH25in
      DBH5in DBHq;
run;

data pines; set all; if Sp = 1;
data hardw; set all; if Sp = 2;

title1 'Pine plots data analysis';
proc reg data = pines;
model Biomass = Nl MeanH MinH MaxLH StdH MeanCD MinCD MaxCD StdCD /
      selection =stepwise vif collinoint influence PRESS;
output  r=res p=pred rstudent= Rstudent press = press;
run;

proc univariate; var press; run;

title1 'Deciduous plots data analysis';
proc reg data = hardw;
model Biomass = Nl MeanH MinH MaxLH StdH MeanCD MinCD MaxCD StdCD /
      selection =stepwise vif collinoint influence PRESS;
output  r=res p=pred rstudent= Rstudent press = press;
run;

proc univariate; var press; run;
```

Appendix 5. SAS program to calculate Moran's I coefficient (shown here for residuals of pine volume)

```
/*_____________________________________________________________*/
/* *** Author: Sorin Popescu                                   */
/* *** It uses the MoranI macro in Schabenberger and Pierce (2001)  */
/*     Macro to calculate Moran's statistic for residuals between   */
/*     lidar predicted volume and field-measured volume             */
/*     Connectivity weights are based on inverse distance           */
/*_____________________________________________________________*/

data coords;
 input x y;
 datalines;
706700.4395 4144922.6180
706670.4357 4144896.8660
706933.2529 4144923.1060
707102.6020 4144936.5000
707094.4832 4144973.0980
707136.0937 4144927.7740
707076.1704 4144912.2490
706707.3403 4144727.9690
706702.3575 4144766.1160
706744.6964 4144715.5020
706676.4897 4144702.5190
706869.4423 4144729.4630
706861.4283 4144765.4760
706903.0114 4144720.7030
706841.3859 4144706.0470
707110.8193 4144776.0650
706711.9331 4144530.5390
706706.4035 4144566.2720
706681.3444 4144508.9260
706911.7860 4144546.1090
706903.5218 4144582.5270
706886.3382 4144524.1410
707129.8272 4144562.9310
707123.5577 4144596.6520
707164.7546 4144550.0490
704290.9920 4145066.7428
704286.0608 4145105.0467
704325.0063 4145057.8256
704290.0871 4144909.8506
704672.1557 4144852.6515
::
run;
data z;
input z;
datalines;
13.033
17.262
-22.646
-18.248
6.723
-25.251
-15.215
-36.034
-36.242
48.569
-42.179
14.362
-10.103
-13.972
-14.620
54.537
24.361
11.453
-8.233
90.590
37.813
-19.210
-7.654
101.773
58.735
-9.411
-69.463
-9.954
25.087
-145.864
::
run;
data volume; merge coords z; run;
proc print; run;
data weights;
 input w1-w30;
 datalines;
0.000000 0.025291 0.004295 0.000000 0.000000 0.000000 0.000000 0.005134 0.006389 0.004722 0.004517 0.003896 0.004445
0.003496 0.003870 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.025291 0.000000 0.003786 0.000000 0.000000 0.000000 0.000000 0.005784 0.007430 0.005103 0.005143 0.003845 0.004314
0.003427 0.003903 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
```

```
0.004295 0.003786 0.000000 0.005887 0.005924 0.004929 0.006977 0.003350 0.003582 0.003566 0.000000 0.004905 0.005773
0.004886 0.004243 0.004338 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.005887 0.000000 0.026675 0.028894 0.027878 0.000000 0.000000 0.000000 0.000000 0.000000 0.003382
0.003402 0.000000 0.006225 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.005924 0.026675 0.000000 0.016253 0.015737 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.005058 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.004929 0.028894 0.016253 0.000000 0.016155 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.006502 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.006977 0.027878 0.015737 0.016155 0.000000 0.000000 0.000000 0.000000 0.000000 0.003624 0.003845
0.003873 0.000000 0.007116 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.005134 0.005784 0.003350 0.000000 0.000000 0.000000 0.000000 0.000000 0.025994 0.025393 0.025004 0.006169 0.006306
0.005107 0.007362 0.000000 0.005064 0.006184 0.004533 0.003655 0.004095 0.003686 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.006389 0.007430 0.003582 0.000000 0.000000 0.000000 0.000000 0.025994 0.000000 0.015154 0.014565 0.005846 0.006286
0.004861 0.006603 0.000000 0.004241 0.005003 0.003875 0.000000 0.003672 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.004722 0.005103 0.003566 0.000000 0.000000 0.000000 0.000000 0.025393 0.015154 0.000000 0.014403 0.007967 0.007875
0.006313 0.010293 0.000000 0.005324 0.006491 0.004628 0.004203 0.004828 0.004200 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.004517 0.005143 0.000000 0.000000 0.000000 0.000000 0.000000 0.025004 0.014565 0.014403 0.000000 0.005133 0.005119
0.004400 0.006063 0.000000 0.005695 0.007169 0.005164 0.003539 0.003894 0.003631 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.003896 0.003845 0.004905 0.000000 0.000000 0.000000 0.003624 0.006169 0.005846 0.007967 0.005133 0.000000 0.027105
0.028824 0.027364 0.004068 0.003941 0.004335 0.003450 0.005314 0.006630 0.004854 0.000000 0.003488 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.004445 0.004314 0.005773 0.003382 0.000000 0.000000 0.003845 0.006306 0.006286 0.007875 0.005119 0.027105 0.000000
0.016365 0.015944 0.004006 0.003591 0.003962 0.000000 0.004443 0.005327 0.004122 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.003496 0.003427 0.004886 0.003402 0.000000 0.000000 0.003873 0.005107 0.004861 0.006313 0.004400 0.028824 0.016365
0.000000 0.015787 0.004650 0.003709 0.004000 0.000000 0.005720 0.007237 0.005069 0.003619 0.003952 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.003870 0.003903 0.004243 0.000000 0.000000 0.000000 0.000000 0.007362 0.006603 0.010293 0.006063 0.027364 0.015944
0.015787 0.000000 0.003592 0.004585 0.005146 0.003938 0.005723 0.007232 0.005337 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.004338 0.006225 0.005058 0.006502 0.007116 0.000000 0.000000 0.000000 0.000000 0.004068 0.004006
0.004650 0.003592 0.000000 0.000000 0.000000 0.000000 0.000000 0.003526 0.000000 0.004673 0.005560 0.004304 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.005064 0.004241 0.005324 0.005695 0.003941 0.003591
0.003709 0.004585 0.000000 0.000000 0.027656 0.026700 0.004989 0.005037 0.005730 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.006184 0.005003 0.006491 0.007169 0.004335 0.003962
0.004000 0.005146 0.000000 0.027656 0.000000 0.015979 0.004846 0.005056 0.005411 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.004533 0.003875 0.004628 0.005164 0.003450 0.000000
0.000000 0.003938 0.000000 0.026700 0.015979 0.000000 0.004284 0.004273 0.004865 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.003655 0.000000 0.004203 0.003539 0.005314 0.004443
0.005720 0.005723 0.000000 0.004989 0.004846 0.004284 0.000000 0.026778 0.029746 0.004573 0.004593 0.003953 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.004095 0.003672 0.004828 0.003894 0.006630 0.005327
0.007237 0.007232 0.003526 0.005037 0.005056 0.004273 0.026778 0.000000 0.016431 0.004402 0.004535 0.003799 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.003686 0.000000 0.004200 0.003631 0.004854 0.004122
0.005069 0.005337 0.000000 0.005730 0.005411 0.004865 0.029746 0.016431 0.000000 0.004056 0.004031 0.003576 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.003619 0.000000 0.004673 0.000000 0.000000 0.000000 0.004573 0.004402 0.004056 0.000000 0.029155 0.026862 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.003488 0.000000
0.003952 0.000000 0.005560 0.000000 0.000000 0.000000 0.004593 0.004535 0.004031 0.029155 0.000000 0.016077 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.004304 0.000000 0.000000 0.000000 0.003953 0.003799 0.003576 0.026862 0.016077 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.025893 0.028438 0.006374 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.025893
0.000000 0.016337 0.005122 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.028438
0.016337 0.000000 0.006577 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.006374
0.005122 0.006577 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000

::
run;

%MoranI(data=volume,row=x,col=y,y=z,w_data=weights,print=1,sort=0,local=0);
```

MoranI macro from Schabenberger and Pierce (2001):

```
/* DISCLAIMER
 THIS INFORMATION IS PROVIDED BY O. SCHABENBERGER "AS IS".
 THERE ARE NO WARRANTIES, EXPRESSED OR IMPLIED, AS TO
 MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE REGARDING
 THE ACCURACY OF THE MATERIALS OR CODE CONTAINED HEREIN.   */
/* ------------------------------------------------- */
/* --- Author: Oliver Schabenberger                  */
/* --- Macro to calculate Moran's global I statistic */
/*     for attribute variable y= in data set data=   */
/*     Connectivity weights are in data set w_data=  */
/* ------------------------------------------------- */
/* If the data set does not contain row and column   */
/* variables set sort=0                              */
/* If local=1 then local Moran's I statistics are    */
/* calculated.                                       */

%macro MoranI(data=_last_,row=gxc,col=gyc,y=y,w_data=weights,print=1,sort=1,local=0);
  %if &sort=1 %then %do;
    proc sort data=&data out=moranI; by &row &col; run;
  %end;
  proc iml; reset noprint fw=7;
    use &data;  read all var {&y} into z;
                read all var {&row &col} into locs; close &data;
    use &w_data; read all         into C; close &w_data;
          print z;
    n = nrow(z);
    u = z - z[+]/n;      /* center the observations */
    W = C / c[+,+] * n;  /* standardize the weights so that 1'W1 = n  and I = (u'Wu)/(u'u) */
    I  = (u`*W*u)/(u`*u);
    result=J(2,7,0);
    EI = - 1/ (n-1);
    S0 = w[+,+];
          *print W; /***************/
    S1 = 0.5 * ((w+w`)##2)[+,+];
          *print S1: /***************/
    S2 = ((w[+,]` + w[,+])##2)[+];
          *print S2: /***************/
    b  = n * ((z##4)[+]) / ((z`*z)##2);
          *print b; /***************/
    /* Variance of Moran's I under randomization */
    VarI   = n * ((n**2 -3*n + 3)*S1 - n*S2 + 3*(S0**2)) - b*((n**2-n)*S1 - 2*n*S2 + 6*(S0**2));
    VarI   = VarI / (n-1)/(n-2)/(n-3)/(S0**2);
    VarI   = VarI - 1/((n-1)**2);
    StdI   = Sqrt(VarI);
    Zobs   = (I - EI)/StdI:
    pright = 1-Probnorm(Zobs);
    result[1,] = 1 || i || Ei || VarI || StdI || Zobs || pright;

    /* Variance of Moran's I under Gaussianity */
    VarI = (n*n*S1 - n*S2 + 3*S0**2)/(S0**2 * (n*n - 1));
    VarI = VarI - 1/((n-1)**2);
    StdI = Sqrt(VarI);
    Zobs   = (I - EI)/StdI:
    pright = 1-Probnorm(Zobs);
    result[2, ] = 2 || i || Ei || VarI || StdI || Zobs || pright;
    name = {"Assump","Iobs","EI","VarI","StdI","Zobs","PRight"};
    create _moranI from result[colname=name]; append from result; close _MoranI;
    /* Is the local version of Moran's I desired ? */
    %if &local=1 %then %do;
      LocalI = J(n,4,.);
      sumui2 = (u`*u);
      do i = 1 to n;
        locMoran = (n*u[i])#(W[i,]*u);
        Locali[i,1] = Locs[i,1];
        Locali[i,2] = Locs[i,2];
        LocalI[i,3] = locMoran/Sumui2;
        LocalI[i,4] = (-(W[i,][+])/(n-1));
      end;
      name = {"&row","&col","LocalI","ELocalI"};
      create _localI from Locali[colname=name]; append from Locali; close _localI;
    %end;
   quit;
   data _moranI; set _moranI; length _Type_ $13;
     if assump=1 then _type_ = "Randomization";
     if assump=2 then _type_ = "Gaussianity";
     drop assump;
   run;
   %if &print=1 %then %do;
     proc print data=_MoranI label noobs;
        label Iobs = 'Observed I'
              EI   = 'E[I]'
              StdI = 'SE[I]'
              Zobs = 'Zobs'
              Pright = 'Pr(Z > Zobs';
       var _type_ Iobs EI StdI Zobs Pright;
      run:
      %if &local=1 %then %do;
        title 'First observations of local I (WORK._LOCALI)';
        proc print data=_localI(obs=50); run;
      %end;
   %end:
%mend MoranI;
```

Appendix 6. Output for identifying outlying observations by analyzing maximum height

Identifying outlying observations for pines - filtering method: square windows

Obs	Lidar	Ground	Fitted	Residuals	R-Student
1	7.34	9.45	7.3209	2.12907	2.71767
2	11.42	11.28	10.9643	0.31568	0.35007
3	10.86	10.36	10.9353	-0.57530	-0.64839
4	16.14	16.15	15.8109	0.33914	0.37710
5	14.99	15.24	14.0840	1.15597	1.32334
6	22.67	20.73	21.5550	-0.82497	-0.94087
7	14.24	14.33	13.5869	0.74308	0.82772
8	10.67	8.53	9.9212	-1.39121	-1.65765
9	10.58	9.45	10.2328	-0.78284	-0.88058
10	11.41	9.75	11.2270	-1.47705	-1.72388
11	7.27	6.10	7.2491	-1.14905	-1.33891
12	13.08	12.50	12.7999	-0.29986	-0.33110
13	10.97	9.75	10.7753	-1.02525	-1.16344
14	13.53	12.19	13.2619	-1.07193	-1.21507
15	10.79	10.06	9.7714	0.28857	0.33839
16	16.07	15.85	15.4660	0.38402	0.42301
17	8.74	9.75	8.7585	0.99154	1.13612
18	9.89	9.75	9.9393	-0.18930	-0.21152
19	10.99	10.67	10.6538	0.01617	0.01789
20	19.37	20.42	20.2195	0.20054	0.33687
21	12.68	12.80	12.5311	0.26890	0.29815
22	13.38	14.63	12.9769	1.65314	1.94488
23	31.02	30.18	29.4518	0.72820	0.90991
24	30.64	28.35	28.2426	0.10739	0.13673
25	30.42	27.74	28.8357	-1.09571	-1.37989
26	23.53	21.95	21.7610	0.18901	0.21952
27	21.14	20.73	20.1259	0.60409	0.67778
28	11.43	11.28	11.1056	0.17438	0.19266
29	14.69	13.41	13.6340	-0.22403	-0.25191
30	20.82	19.20	19.3824	-0.18237	-0.20600

Identifying outlying observations for pines - filtering method: square windows and data fusion

Obs	Lidar	Ground	Fitted	Residuals	R-Student
1	7.34	9.45	7.3859	2.06413	2.60964
2	11.42	11.28	10.9871	0.29292	0.32421
3	10.86	10.36	10.9794	-0.61939	-0.70047
4	16.14	16.15	15.8057	0.34433	0.38242
5	14.99	15.24	14.0676	1.17237	1.34353
6	22.67	20.73	21.4794	-0.74940	-0.85114
7	14.24	14.33	13.5840	0.74603	0.83030
8	10.67	8.53	9.7928	-1.26276	-1.51265
9	10.58	9.45	10.2649	-0.81489	-0.91633
10	11.41	9.75	11.2589	-1.50887	-1.76505
11	7.27	6.10	7.3144	-1.21441	-1.41931
12	13.08	12.50	12.8171	-0.31711	-0.35006
13	10.97	9.75	10.8097	-1.05968	-1.20414
14	13.53	12.19	13.2765	-1.08651	-1.23200
15	10.79	10.06	9.7799	0.28009	0.32779
16	16.07	15.85	15.4522	0.39780	0.43781
17	8.74	9.75	8.8151	0.93489	1.06894
18	9.89	9.75	9.9891	-0.23913	-0.26775
19	10.99	10.67	10.6835	-0.01346	-0.01487
20	19.37	20.42	20.2312	0.18884	0.32135
21	12.68	12.80	12.2734	0.52661	0.58265
22	13.38	14.63	12.9880	1.64198	1.92790
23	31.02	30.18	29.3044	0.87557	1.09480
24	30.64	28.35	28.7699	-0.41985	-0.51311
25	30.42	27.74	28.6919	-0.95190	-1.18310
26	23.53	21.95	21.6580	0.29201	0.34363
27	21.14	20.73	20.0641	0.66592	0.74780
28	11.43	11.28	11.1326	0.14735	0.16260
29	14.69	13.41	13.6147	-0.20472	-0.23040
30	20.82	19.20	19.3088	-0.10875	-0.12346

Identifying outlying observations for pines - filtering method: circular windows

Obs	Lidar	Ground	Fitted	Residuals	R-Student
1	7.34	9.45	7.2502	2.19982	2.63482
2	11.42	11.28	11.1616	0.11842	0.12356
3	10.86	10.36	10.5050	-0.14500	-0.15156
4	17.16	16.15	16.5736	-0.42356	-0.44396
5	14.99	15.24	14.7573	0.48267	0.51729
6	22.67	20.73	21.3178	-0.58776	-0.62955
7	14.24	14.33	13.7725	0.55755	0.58214
8	10.67	8.53	10.4342	-1.90420	-2.16124
9	10.58	9.45	10.2184	-0.76843	-0.81455
10	11.41	9.75	11.1119	-1.36195	-1.47783
11	7.27	6.10	7.1961	-1.09607	-1.19831
12	13.08	12.50	12.5400	-0.04000	-0.04156
13	10.97	9.75	10.5569	-0.80694	-0.85601
14	13.53	12.19	12.9930	-0.80298	-0.84410
15	10.79	10.06	10.7539	-0.69387	-0.74172
16	16.07	15.85	15.2842	0.56580	0.59086
17	8.74	9.75	8.5745	1.17548	1.27760
18	9.89	9.75	9.6369	0.11306	0.11855
19	10.99	10.67	10.7380	-0.06801	-0.07100
20	19.37	20.42	18.8905	1.52949	1.76646
21	12.68	12.80	12.0999	0.70013	0.73785
22	13.38	14.63	12.6859	1.94412	2.21806
23	31.02	30.18	30.2744	-0.09436	-0.17157
24	30.64	28.35	28.4361	-0.08606	-0.10265
25	30.42	27.74	27.9299	-0.18994	-0.24875
26	23.53	21.95	21.9922	-0.04222	-0.04564
27	21.14	20.73	19.8937	0.83630	0.89588
28	11.43	11.28	11.0370	0.24304	0.25369
29	14.69	13.41	14.2261	-0.81609	-0.85961
30	20.82	19.20	19.7384	-0.53843	-0.56764

Identifying outlying observations for deciduous plots - filtering method: square windows

Obs	Lidar	Ground	Fitted	Residuals	R-Student
1	16.83	23.47	18.4800	4.9900	1.39568
2	10.18	17.37	14.8189	2.5511	0.77582
3	25.74	22.86	23.3854	-0.5254	-0.13847
4	24.61	22.56	22.7633	-0.2033	-0.05338
5	30.63	26.21	26.0776	0.1324	0.03622
6	20.29	20.42	20.3849	0.0351	0.00924
7	22.53	22.56	21.6182	0.9418	0.24717
8	24.32	10.97	22.6037	-11.6337	-3.81052
9	12.44	14.94	16.0631	-1.1231	-0.32242
10	27.87	33.53	24.5581	8.9719	2.70782
11	30.70	23.77	26.1162	-2.3462	-0.64769
12	31.59	28.96	26.6062	2.3538	0.65746
13	26.84	24.99	23.9910	0.9990	0.26489
14	21.10	19.51	20.8309	-1.3209	-0.34795
15	21.03	19.20	20.7923	-1.5923	-0.41999
16	22.00	20.42	21.3264	-0.9064	-0.23799
17	18.02	16.15	19.1352	-2.9852	-0.80709
18	21.26	22.56	20.9190	1.6410	0.43264
19	17.60	15.85	18.9039	-3.0539	-0.82899
20	29.03	25.91	25.1968	0.7132	0.19206
21	27.59	28.04	24.4040	3.6360	0.98522
22	24.16	25.30	22.5156	2.7844	0.73797
23	24.52	22.25	22.7138	-0.4638	-0.12176
24	27.42	17.07	24.3104	-7.2404	-2.07764
25	20.68	24.99	20.5996	4.3904	1.18623
26	22.97	20.73	21.8604	-1.1304	-0.29676
27	22.81	23.77	21.7723	1.9977	0.52635
28	22.48	22.25	21.5906	0.6594	0.17294
29	21.72	18.90	21.1722	-2.2722	-0.60040

Identifying outlying observations for deciduous plots - filtering method: square windows and data fusion

Obs	Lidar	Ground	Fitted	Residuals	R-Student
1	16.73	23.47	18.2862	5.1838	1.55490
2	10.18	17.37	14.3942	2.9758	0.96323
3	25.74	22.86	23.6398	-0.7798	-0.21914
4	24.61	22.56	22.9684	-0.4084	-0.11430
5	30.63	26.21	26.5454	-0.3354	-0.09772
6	20.29	20.42	20.4015	0.0185	0.00520
7	22.53	22.56	21.7325	0.8275	0.23141
8	24.32	10.97	22.7961	-11.8261	-4.28766
9	12.44	14.94	15.7371	-0.7971	-0.24262
10	30.70	33.53	26.5870	6.9430	2.19768
11	27.87	23.77	24.9055	-1.1355	-0.32309
12	31.59	28.96	27.1158	1.8442	0.54662
13	26.84	24.99	24.2934	0.6966	0.19671
14	21.10	19.51	20.8828	-1.3728	-0.38550
15	21.03	19.20	20.8412	-1.6412	-0.46150
16	22.00	20.42	21.4176	-0.9976	-0.27925
17	18.02	16.15	19.0527	-2.9027	-0.83591
18	21.26	22.56	20.9779	1.5821	0.44450
19	26.66	25.30	24.1865	1.1135	0.31450
20	17.60	15.85	18.8031	-2.9531	-0.85360
21	29.03	25.91	25.5947	0.3153	0.09038
22	27.59	28.04	24.7391	3.3009	0.95123
23	24.16	25.30	22.7010	2.5990	0.73381
24	22.41	22.25	21.6612	0.5888	0.16460
25	27.42	17.07	24.6381	-7.5681	-2.35227
26	20.68	24.99	20.6332	4.3568	1.25685
27	22.97	20.73	21.9939	-1.2639	-0.35386
28	22.81	23.77	21.8989	1.8711	0.52533
29	22.48	22.25	21.7028	0.5472	0.15295
30	19.08	18.90	19.6825	-0.7825	-0.22118

Identifying outlying observations for deciduous plots - filtering method: circular windows

Obs	Lidar	Ground	Fitted	Residuals	R-Student
1	16.83	23.47	18.5312	4.9388	1.40666
2	10.18	17.37	14.7666	2.6034	0.80831
3	25.74	22.86	23.5753	-0.7153	-0.19217
4	24.61	22.56	22.9356	-0.3756	-0.10059
5	30.63	26.21	26.3436	-0.1336	-0.03709
6	20.29	20.42	20.4900	-0.0700	-0.01885
7	24.26	28.65	22.7375	5.9125	1.65281
8	22.53	22.56	21.7581	0.8019	0.21479
9	24.32	10.97	22.7714	-11.8014	-3.86617
10	12.44	14.94	16.0460	-1.1060	-0.32455
11	27.87	33.53	24.7811	8.7489	2.63285
12	30.70	23.77	26.3832	-2.6132	-0.73255
13	31.59	28.96	26.8871	2.0729	0.58526
14	26.84	24.99	24.1980	0.7920	0.21374
15	21.10	19.51	20.9485	-1.4385	-0.38719
16	21.03	19.20	20.9089	-1.7089	-0.46055
17	22.00	20.42	21.4580	-1.0380	-0.27841
18	18.02	16.15	19.2049	-3.0549	-0.84393
19	22.31	22.56	21.6335	0.9265	0.24828
20	26.66	25.30	24.0961	1.2039	0.32496
21	17.60	15.85	18.9672	-3.1172	-0.86454
22	27.87	24.69	24.7811	-0.0911	-0.02472
23	29.03	25.91	25.4378	0.4722	0.12918
24	27.59	28.04	24.6226	3.4174	0.93880
25	24.16	25.30	22.6808	2.6192	0.70668
26	24.52	22.25	22.8847	-0.6347	-0.16999
27	23.93	20.42	22.5506	-2.1306	-0.57315
28	27.42	17.07	24.5264	-7.4564	-2.16938
29	20.68	24.99	20.7108	4.2792	1.17687
30	22.97	20.73	22.0072	-1.2772	-0.34235
31	22.81	23.77	21.9166	1.8534	0.49794
32	22.48	22.25	21.7298	0.5202	0.13928
33	21.72	18.90	21.2995	-2.3995	-0.64761

Appendix 7. Comparison of regression models with and without outliers

Summary statistics for models with and without outliers – square windows

Statistics	With outliers in		With outliers out	
	Pines	Deciduous	Pines	Deciduous
R^2	0.9804	0.3564	0.9891	0.6369
RMSE (m)	0.90962	3.81064	0.70105	2.23573

Summary statistics for models with and without outliers – square windows with data fusion

Statistics	With outliers in		With outliers out	
	Pines	Deciduous	Pines	Deciduous
R^2	0.9804	0.4229	0.9903	0.6859
RMSE (m)	0.91055	3.57571	0.67863	2.06801

Summary statistics for models with and without outliers – circular windows

Statistics	With outliers in		With outliers out	
	Pines	Deciduous	Pines	Deciduous
R^2	0.9780	0.3566	0.9905	0.5965
RMSE (m)	0.96401	3.73466	0.65614	2.36983

Appendix 8. Regression results for estimating ground-measured forest biophysical parameters without using independent variables related to lidar-estimated crown diameter

The following tables have the same number as the corresponding tables in chapter 5 (Results and discussion), that show results for estimating the same dependent variables. The character "a" follows the table number in this appendix to differentiate between the corresponding tables in chapter 5.

Table 20a. Regression results – dependent variable: diameter at breast height (cm) / subplot*

Trees	Method	Significant independent variables**	$S_{y.x}$	R^2	Model
Pines					
Dominants	SQ	H_{max}	6.14	0.6968	$-1.36696 + 1.36421H_{max}$
	SQF	H_{max}	6.14	0.6968	$-1.36696 + 1.36421H_{max}$
	CW	H_{max}	6.16	0.6943	$-1.35411 + 1.36037H_{max}$
	CWF	H_{max}	6.16	0.6943	$-1.35411 + 1.36037H_{max}$
All	SQ	N, H_{max}	2.57	0.6367	$0.95133 + 0.27770N + 0.63073H_{max}$
	SQF	N, H_{max}	2.56	0.6382	$1.07482 + 0.27297N + 0.62887H_{max}$
	CW	H_{max}	2.68	0.5883	$5.77563 + 0.46987H_{max}$
	CWF	N, H_{max}	2.61	0.6257	$1.53907 + 0.18034N + 0.61075H_{max}$
FIA standard	SQ	H_{max}	4.47	0.6918	$3.52989 + 0.98074H_{max}$
	SQF	H_{max}	4.47	0.6918	$3.52989 + 0.98074H_{max}$
	CW	H_{max}	4.48	0.6898	$3.52761 + 0.97863H_{max}$
	CWF	H_{max}	4.48	0.6898	$3.52761 + 0.97863H_{max}$
Deciduous					
Dominants	SQ	H_{max}	6.63	0.4089	$7.34141 + 1.04545H_{max}$
	SQF	H_{max}	6.56	0.4339	$6.00805 + 1.13187H_{max}$
	CW	H_{max}	6.52	0.4415	$5.45407 + 1.15193H_{max}$
	CWF	H_{max}	6.56	0.4339	$6.00805 + 1.13187H_{max}$
All	SQ	H_{max}	3.06	0.1953	$10.35924 + 0.28599H_{max}$
	SQF	H_{ave}	3.59	0.2337	$10.11436 + 0.34351H_{ave}$
	CW	H_{max}	3.56	0.2067	$9.15253 + 0.36127H_{max}$
	CWF	H_{max}	3.57	0.2033	$9.32384 + 0.35509H_{max}$
FIA standard	SQ	H_{max}	4.19	0.3483	$12.13503 + 0.58043H_{max}$
	SQF	H_{max}	4.39	0.3638	$11.10618 + 0.64709H_{max}$
	CW	H_{max}	4.30	0.3779	$10.57118 + 0.66566H_{max}$
	CWF	H_{max}	4.32	0.3716	$10.88770 + 0.65423H_{max}$

* Method and variable abbreviations are the same as in Table 15

** Independent variables (lidar-measured): number of trees, average height, minimum height, maximum height, and standard deviation of tree height values

Table 21a. PRESS statistics for predicting average dbh (cm) / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
Dominants	SQ	1411.44	-17.47	15.73	0.08	6.98
	SQF	1411.44	-17.47	15.73	0.08	6.98
	CW	1419.59	-17.31	15.79	0.08	7.00
	CWF	1419.59	-17.31	15.79	0.08	7.00
All	SQ	271.50	-10.71	7.39	0.05	3.06
	SQF	270.89	-10.82	7.34	0.05	3.06
	CW	285.58	-9.70	8.11	0.03	3.14
	CWF	277.27	-10.70	7.60	0.05	3.09
FIA standard	SQ	693.98	-8.98	16.33	0.06	5.07
	SQF	693.98	-8.98	16.33	0.06	5.07
	CW	697.00	-8.88	16.39	0.07	5.08
	CWF	697.00	-8.88	16.39	0.07	5.08
Deciduous						
Dominants	SQ	1250.42	-14.32	15.02	-0.02	7.07
	SQF	1391.75	-15.49	14.13	-0.01	6.93
	CW	1374.05	-15.55	14.20	-0.01	6.88
	CWF	1391.75	-15.49	14.13	-0.01	6.93
All	SQ	269.89	-8.17	7.88	-0.08	3.28
	SQF	378.88	-8.39	10.41	-0.03	3.82
	CW	411.34	-9.23	10.64	-0.07	3.77
	CWF	413.31	-9.22	10.64	-0.07	3.77
FIA standard	SQ	482.79	-6.44	13.38	-0.01	4.39
	SQF	546.65	-7.17	13.01	0.00	4.59
	CW	583.47	-7.04	13.17	0.00	4.49
	CWF	588.68	-7.05	13.07	0.00	4.51

* 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 22a. Regression results – dependent variable: quadratic mean diameter (cm) / subplot*

Trees	Method	Significant independent variables**	$S_{y.x}$	R^2	Model
Pines					
All	SQ	N, H_{max}	3.17	0.7067	$-1.77981 + 0.30835N + 0.87134H_{max}$
	SQF	N, H_{max}	3.17	0.7063	$-1.53675 + 0.29697N + 0.86571H_{max}$
	CW	H_{max}	3.28	0.6740	$3.58466 + 0.69185H_{max}$
	CWF	H_{max}	3.28	0.6740	$3.58466 + 0.69185H_{max}$
Deciduous					
All	SQ	H_{max}	3.39	0.3089	$9.88597 + 0.42941H_{max}$
	SQF	H_{max}	3.84	0.3343	$8.13267 + 0.53060H_{max}$
	CW	H_{max}	3.84	0.3155	$8.41879 + 0.51788H_{max}$
	CWF	H_{max}	3.85	0.3094	$8.68092 - 0.50830H_{max}$

* Method and variable abbreviations are the same as in Table 15

** Independent variables (lidar-measured): number of trees, average height, minimum height, maximum height, and standard deviation of tree height values

Table 23a. PRESS statistics for predicting quadratic mean diameter / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	421.80	-12.00	9.18	0.11	3.81
	SQF	420.17	-12.10	9.15	0.10	3.81
	CW	431.89	-10.85	9.97	0.05	3.86
	CWF	431.89	-10.85	9.97	0.05	3.86
Deciduous						
All	SQ	328.42	-9.78	6.58	-0.08	3.62
	SQF	424.17	-9.04	-10.64	-0.05	4.04
	CW	477.36	-10.97	10.57	-0.06	4.06
	CWF	481.71	-10.93	-10.56	-0.06	4.08

Table 26a. Regression results – dependent variable: volume (m^3 / ha) / subplot*

Trees	Method	Significant independent variables**	$S_{y.x}$	R^2	Model
Pines					
All	SQ	H_{ave}	57.13	0.7578	-101.41407 + 17.19689H_{ave}
	SQF	H_{ave}	55.88	0.7682	-100.36769 + 17.03419H_{ave}
	CW	N, H_{ave}	53.26	0.7969	-187.65347 + 3.32079N + 21.36929H_{ave}
	CWF	N, H_{ave}	53.31	0.7966	-195.05828 + 3.73659N + 21.58452H_{ave}
Deciduous					
All	SQ	H_{ave}	57.24	0.1534	69.94108 + 4.35691H_{ave}
	SQF	H_{ave}	60.12	0.2361	46.23773 + 5.79718Have
	CW	H_{max}	61.60	0.1379	51.26639 + 4.89988H_{max}
	CWF	H_{max}	61.21	0.1489	48.30938 + 5.04534H_{max}

* Method and variable abbreviations are the same as in Table 15

** Independent variables (lidar-measured): number of trees, average height, minimum height, maximum height, and standard deviation of tree height values

Table 27a. PRESS statistics for predicting volume / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	133071	-160.88	239.24	1.83	67.71
	SQF	127425	-158.14	228.94	1.92	66.25
	CW	121930	-127.82	214.33	3.13	64.76
	CWF	122358	-125.09	214.03	3.15	64.88
Deciduous						
All	SQ	100702	-143.95	127.13	-1.58	63.45
	SQF	110172	-136.89	133.17	-0.59	65.09
	CW	125393	-133.57	149.43	-1.28	65.74
	CWF	123847	-135.31	148.54	-1.27	65.34

* 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 28a. Regression results – dependent variable: basal area (m^2 / ha) / subplot*

Trees	Method	Significant independent variables**	$S_{y.x}$	R^2	Model
Pines					
All	SQ	H_{min}	6.11	0.6422	$5.23877 + 1.30487H_{min}$
	SQF	H_{min}	6.17	0.6348	$5.62464 + 1.25383H_{min}$
	CW	H_{min}	6.40	0.6071	$5.60872 + 1.32244H_{min}$
	CWF	H_{min}	6.45	0.6016	$5.76974 + 1.30254H_{min}$
Deciduous					
All	SQ	-	6.83	0	20.32692
	SQF	-	7.20	0	20.83370
	CW	-	7.03	0	20.33933
	CWF	-	7.03	0	20.33933

* Method and variable abbreviations are the same as in Table 15

** Independent variables (lidar-measured): number of trees, average height, minimum height, maximum height, and standard deviation of tree height values

Table 29a. PRESS statistics for predicting basal area (m^2 / ha) / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	1168.34	-12.03	17.72	-0.05	6.35
	SQF	1186.51	-11.98	18.01	-0.06	6.40
	CW	1286.84	-13.10	20.38	-0.06	6.66
	CWF	1306.33	-12.77	20.43	-0.05	6.71
Deciduous						
All	SQ	No variable met the 0.15 significance level to entry into the model				
	SQF	No variable met the 0.15 significance level to entry into the model				
	CW	No variable met the 0.15 significance level to entry into the model				
	CWF	No variable met the 0.15 significance level to entry into the model				

* 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 30a. Regression results – dependent variable: biomass (Mg/ha) / subplot*

Trees	Method	Significant independent variables**	$S_{y.x}$	R^2	Model
Pines					
All	SQ	H_{ave}, H_{std}	34.74	0.7393	$-44.20629 + 10.03814H_{ave} - 2.89935H_{std}$
	SQF	H_{ave}, H_{std}	34.17	0.7479	$-42.39288 + 9.84044H_{ave} - 2.89452H_{std}$
	CW	H_{ave}	35.18	0.7228	$-49.63329 + 10.15500H_{ave}$
	CWF	H_{ave}	35.49	0.7180	$-48.62698 + 10.05874H_{ave}$
Deciduous					
All	SQ	H_{ave}	47.78	0.0924	$74.82695 + 2.72609H_{ave}$
	SQF	H_{ave}	50.68	0.1646	$55.34649 + 3.90188H_{ave}$
	CW	H_{max}	50.84	0.0861	$62.63466 + 3.10311H_{max}$
	CWF	H_{max}	50.60	0.0947	$60.06451 + 3.22552H_{max}$

* Method and variable abbreviations are the same as in Table 15

** Independent variables (lidar-measured): number of trees, average height, minimum height, maximum height, and standard deviation of tree height values

Table 31a. PRESS statistics for predicting biomass (Mg/ha) / subplot

Trees	Method*	PRESS	Range of PRESS residuals		Mean of PRESS residuals	Standard deviation of PRESS residuals
			Min	Max		
Pines						
All	SQ	51670	-97.74	116.05	1.01	42.20
	SQF	49912	-94.98	117.58	1.13	41.47
	CW	44293	-67.99	125.27	0.89	39.07
	CWF	45131	-67.42	125.75	0.88	39.44
Deciduous						
All	SQ	71845	-116.93	115.54	-1.47	53.59
	SQF	79626	-113.73	119.85	-0.62	55.34
	CW	86980	-109.64	123.88	-1.27	54.75
	CWF	86103	-111.01	123.21	-1.25	54.47

* 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);

VITA

Name: Sorin C. Popescu

Research experience: - Remote sensing and GIS integration and applications in forestry, biometrics, natural resources management, growth and yield, and land cover change
- New optical and lidar remote sensors, multisensor data fusion, algorithm development for automated image processing, DEM generation, vegetation extraction and assessment
- Spatial statistics, sources of spatial data in natural resources, spatial analysis techniques

Positions held: AUGUST 1997 – MAY 2002 Graduate Research and Teaching Assistant, Department of Forestry, Virginia Tech, USA

JUNE-AUGUST 1997 GIS Analyst
Canadian Geomatic Solutions Ltd., Calgary, Alberta, Canada

JUNE - AUGUST 1996 Research Assistant
Dept. of Forest Biometrics, University of Freiburg, Germany

SEPTEMBER 1992 - MAY 1997 Assistant Lecturer
"Transilvania" University of Brasov, Dept. of Forest Management, Romania

Education: - Ph.D. (completion May 2002), Forestry, Remote Sensing, Dept. of Forestry, Virginia Tech
Course concentrations: GIS, remote sensing, spatial data analysis, forest biometrics; GPA: 3.98 / 4.00

- Diploma degree (1992), Forest Engineer "Transilvania" University of Brasov, Faculty of Forestry, Brasov, Romania

_____
Sorin C. Popescu