An Object-Oriented Approach to Forest Volume and Aboveground Biomass Modeling using Small-Footprint Lidar Data for Segmentation, Estimation, and Classification

Jan A.N. van Aardt

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Forestry

Randolph H. Wynne, Chair
James B. Campbell
Ross F. Nelson
Richard G. Oderwald
Stephen P. Prisley
John R. Seiler

August 16, 2004
Blacksburg, Virginia

Keywords: Object-oriented, lidar distributions, forest volume and above-ground biomass, classification, multiresolution, hierarchical segmentation

An Object-Oriented Approach to Forest Volume- and Above-Ground Biomass-by-Type Modeling using Small-Footprint Lidar Data for Segmentation, Estimation, and Classification

Jan A.N. van Aardt

(ABSTRACT)

This study assessed the utility of an object-oriented approach to deciduous and coniferous forest volume and aboveground biomass estimation, based solely on small-footprint, multiple return lidar data. The study area is located in Appomattox Buckingham State Forest in the Piedmont physiographic province of Virginia, U.S.A, at 78°41' W, 37°25' N. Vegetation is composed of various coniferous, deciduous, and mixed forest stands. The eCognition segmentation algorithm was used to derive objects from a lidar-based canopy height model (CHM). New segment selection criteria, based on between- and within-segment CHM variance, and average field plot size, were developed. Horizontal point samples were used to measure in-field volume and biomass, for 2-class (deciduous-coniferous) and 3-class (deciduous-coniferous-mixed) forest schemes. Per-segment lidar distributional parameters, e.g., mean, range, and percentiles, were extracted from the lidar data and used as input to volume and biomass regression analysis. Discriminant classification was performed using lidar point height and CHM distributions. There was no evident difference between the two-class and three-class approaches, based on similar adjusted R^2 values. Two-class forest definition was preferred due to its simplicity. Two-class adjusted R^2 and root mean square error (RMSE) values for deciduous volume (0.59; 51.15 m^3/ha) and biomass (0.58; 37.41 Mg/ha) were improvements over those found in another plot-based study for the same study area. Although coniferous RMSE values for volume (38.03 m^3/ha) and biomass (17.15 Mg/ha) were comparable to published results, adjusted R^2 values (0.66 and 0.59) were lower. This was attributed to more variability and a narrower range (6.94 – 350.93 m^3/ha) in measured values. Classification accuracy for discriminant classification based on lidar point height distributions (89.2%) was a significant improvement over CHM-based classification (79%). A lack of modeling and classification differences between average segment sizes was attributed to the hierarchical nature of the segmentation algorithm. However,

segment-based modeling was distinctly better than modeling based on existing forest stands, with values of 0.42 and 62.36 m^3/ha (volume) and 0.46 and 41.18 Mg/ha (biomass) for adjusted R^2 and RMSE, respectively. Modeling results and classification accuracies indicated that an object-oriented approach, based solely on lidar data, has potential for full-scale forest inventory applications.

ACKNOWLEDGEMENTS

I came to the United States in 1998 as a raw South African and I am leaving a little more tempered. (That is good, for those that are wondering.) What a rewarding experience this was. The people of Blacksburg welcomed us in so many ways. These ranged from social right through to academic support. I owe special thanks to Dr. Harold Burkhart (department head) for his backing of graduate students and to my advisor, Dr. Randy Wynne. There is not much that I can say about "The Kernel" without sounding like a Hallmark card. Suffice it to say that a kernel can be interpreted as the hub of events, and it can be smoothing or sharpening. Randy was all that, with an open-door policy to boot. Thank you for the freedom you gave me to develop my career, to start a family, and to explore my academic and social interests. You will always be welcome to a mug of Rooibos tea around a campfire in Africa with me!

I would also like to extend my deepest appreciation to the other members of my graduate committee, Drs. Jim Campbell, Richard Oderwald, Steven Prisley, John Seiler (all from Virginia Tech), and Ross Nelson (NASA Goddard Space Flight Center). Each of these individuals contributed to my graduate career through their support and expertise in their respective fields of knowledge. Their doors were always open and without their input I would have been stuck up many trees in the forests of Appomattox.

To all my peers, I thank you from the bottom of my heart for the impact you had in my life and for the camaraderie we shared. We really were sounding boards to each other, in academic terms as well as outside the office. I would like to thank Sorin Popescu, Jared Wayman, Zack Bortolot, Christine Blinn, and Beccy Forest Musy for their friendship and support. It was a fantastic bunch to "grow-up" with in terms of my academic development. May the cords of friendship only stretch a little, until we meet again in some exotic location somewhere!

There were so many people that were integral to my studies, but Connie Noonkester and Sue Snow deserve special mention. As program support technicians, Connie and Sue were the people that helped to keep wheels turning, projects funded, and food in my fridge. I also would like to

thank Dr. John Scrivani at the Virginia Department of Forestry for his help in setting up field data collection and in subsequent research efforts.

I would like to thank Heidi and Wolfgang Glasser, who “adopted” us upon our arrival in Blacksburg. Thank you so much for you love - you truly were pillars of support during our stay in Blacksburg. I will never forget our long evenings of playing cards, discussing life, and sharing a glass of Cointreau. The Walter family, as our host family, also helped in making our family’s Blacksburg experience the joy that is was. What wonderful times we had together…

Thank you to my family if South Africa, who supported and missed us for the six years we were in the USA. Thanks dad Michiel and mom Sanet for the huge part you played in me being here.

Then there is my wife, Marleen. What can I say… We spent our first six years of marriage over here and I would not have wished to spend it anywhere else. Thank you for loving me, sometimes enduring me, and always being there for me. Thank you for raising Karla, our wonderful daughter, while I was glued to the computer. You are my best friend and loveliest critic!

My sincere thanks go to the William J. Fulbright scholarship program, American Society of Photogrammetry and Remote Sensing, Virginia Tech Department of Forestry, Graduate Student Assembly at Virginia Tech, NASA, and the McIntire-Stennis Research Program for financial support of this research.

I don’t believe in fate. My path to Blacksburg and beyond was and is set like a railroad track, and although I was given choices along the way, I was not the architect. I want to give honor and praise to my Lord and God, Jesus Christ, for all He has given me. Philippians 4:13 will forever hold true.

Soli Deo Gloria

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION AND OBJECTIVES

CHAPTER 3: DECIDUOUS AND CONIFEROUS FOREST VOLUME AND ABOVEGROUND BIOMASS ESTIMATION USING SMALL-FOOTPRINT LIDAR- DISTRIBUTIONAL PARAMETERS ON A PER-SEGMENT BASIS

CHAPTER 5: MULTIRESOLUTION, HIERARCHICAL SEGMENTATION OF SMALL-FOOTPRINT LIDAR DATA AS A FORESTRY INVENTORY PRECURSOR

LIST OF TABLES

CHAPTER 2: LITERATURE REVIEW

CHAPTER 3: DECIDUOUS AND CONIFEROUS FOREST VOLUME AND ABOVEGROUND BIOMASS ESTIMATION USING SMALL-FOOTPRINT LIDAR- DISTRIBUTIONAL PARAMETERS ON A PER-SEGMENT BASIS

CHAPTER 5: MULTIRESOLUTION, HIERARCHICAL SEGMENTATION OF SMALL-FOOTPRINT LIDAR DATA AS A FORESTRY INVENTORY PRECURSOR

CHAPTER 1

INTRODUCTION AND OBJECTIVES

1.1 Introduction

Accurate prediction of forest volume and biomass by type always has been an ideal of natural resource managers. Not only does such a prediction have far reaching financial implications for any forest company, but it also is of global ecological importance with regards to monitoring carbon sequestration. The traditional way to gauge forest resources involves large-scale, labor-intensive forest inventories, often incorporating intricate sampling schemes and extrapolation efforts (Avery and Burkhart, 1994; Shivers and Borders, 1996). Aerial photography, as an alternative approach, also has found application as a technique to estimate forest volume through stand-volume tables. However, existing methods are time consuming, subjective, and more applicable to smaller areas (Avery and Burkhart, 1994). Alternative approaches to forest inventory therefore have fostered intense research with the advent of new remote sensing technologies, specifically light detection and ranging (lidar) (Lefsky *et al*., 2002a). Lidar has enabled users to extract forest structural information, e.g., Means *et al.* (2000), Lefsky *et al.* (2002b), Naesset (2002), and Popescu *et al*. (2004). Remote sensing approaches, such as lidar, lend themselves to methods that are repeatable and objective, resulting in yield estimates that approach acceptable accuracy and precision.

Lidar involves the emission of a laser pulse from an airborne sensor, the measurement of the pulse's return-travel time from sensor to target, and the calculation of the distance traveled by the laser beam. The distance from the sensor to the target is often converted to target height above sea level, given that the sensor (flying) height is known (Baltsavias, 1999a). Applications in forestry have mainly focused on measurement of canopy height, sub-canopy topography, and the vertical distribution of intercepted surfaces in forested areas. These measured characteristics were extended to the modeling of above-ground biomass, stem counts, and crown widths (Dubayah and Drake, 2000; Lefsky *et al.*, 2002a). Lidar data successfully have been implemented by many groups to accurately gauge aboveground biomass of both temperate and

tropical forests, as well as relating biomass to merchantable volume estimates. Large footprint Scanning Lidar Imager of Canopies by Echo Recovery (SLICER) data have been used extensively to estimate biomass and basal area of yellow poplar (*Liriodendron tulipifera*), Douglas-fir (*Pseudotsuga menziesii),* and western hemlock (*Tsuga heterophylla*) stands (Lefsky *et al.*, 1999a; Lefsky *et al.*, 1999b; Means *et al,* 1999). Small footprint data, on the other hand, have the benefit of spatially concise and explicit returns that make a tree-specific approach possible. Forest structural information such as tree height, basal area, biomass, and volume have been modeled using small-footprint lidar data (Nilsson, 1996; Magnussen and Boudewyn, 1998; Magnussen *et al.*, 1999; Young *et al.*, 2000; Means et al., 2000; Næsset, 2002; Popescu 2003; Popescu, 2004), while forest type informational content can be derived from lidar canopy densities or distributions (Douglas *et al.,* 2003). Other forestry applications of lidar include estimation of forest fuel loads (Drake and Weishampel, 2001; Riaño *et al*., 2003) and derivation of DEMs applicable to forest management and site mapping (Popescu *et al*., 2002; Hodgson *et al*., 2003).

A lidar-based, type-specific volume and biomass assignment also has appeal as an elegant approach to using a single remote sensing data source. Biomass results further are of particular importance to the modeling of Net Primary Productivity (NPP), which addresses the question of how much carbon is sequestered where and by which forest type or species. One of twelve "National Priority" application areas in NASA's Earth Science program is "Sequestration capacity monitoring for carbon management." These application areas are deemed as being of "high national priority, have significant potential for increased socio-economic value from the application of Earth science, information and technology, and have an operational community that can utilize the Earth science inputs for generating assessments and decision support information" (NASA Webcast, Research Community Update, 2002). Socio-economic value and application feasibility are the first two high priority criteria that are mentioned as being important to Earth Science Enterprise (ESE) application (NASA ESE Strategic Plan, 2000; NASA Earth Science Enterprise Applications Strategy, 2002). Remotely sensed volume and biomass estimates have both regional (managerial) and global socio-economic value in terms of forest management and carbon modeling efforts, as well as being highly applicable in the forest management and

NPP monitoring efforts. Hence applicability of a seemingly forest inventory-only tool is extended to possible regional, national, or even global application.

Volume and biomass modeling at both local and larger scales are dependent on the initial measurement unit as the basis for extrapolation to more expansive areas. Such a measurement unit needs to be homogenous in order to derive accurate forest biophysical estimates (Avery and Burkhart, 1994; Shivers and Borders, 1996; Makela and Pekkarinen, 2001). One very likely approach is that of forest segmentation, or definition of a composite forest area by structurally uniform objects. Image segmentation is a technique well suited to the derivation of such objects in a forestry context (Nugroho *et al.,* 2002a; Engdahl *et al.,* 2003; Heyman *et al*., 2003), and has been implemented successfully for per-object parameter estimation (Woodcock *et al.*, 1994; Makela and Pekkarinen, 2001; Pekkarinen, 2002; Engdahl *et al.*, 2003; Kellndorfer and Ulaby, 2003). Makela and Pekkarinen (2001) stated that segmentation of forest areas into stratified units was very suitable for the estimation of forest variables, concluding that volume and biomass estimate errors could be minimized by using homogenous segments. Pekkarinen (2002) investigated stand-level errors and concluded that segmentation succeeded in delineating distinct forest areas for feature extraction. The author suggested that segment-level data could decrease associated segment-level errors. Per-segment forest parameter estimation has potential based on previous studies, but no inventory estimate is complete without assignment to specific forest types, thereby encompassing a complete inventory approach.

Forest type classification is especially important for vast tracts of land, where accessibility may be limited, while remote sensing products are available. Such a classification enables managers to derive type maps from remotely sensed imagery, or ecologists to attribute carbon stores to specific forest groups. Traditional approaches have been based on multispectral remote sensing inputs (Nelson *et al.*, 1984; Shen *et al.*, 1985; Franklin, 1994; White *et al.*, 1995), while type definition using lidar canopy distributions also has come to the fore (Douglas *et al*., 2003). Such approaches mainly have been pixel- or stand-based, while segment-based classification has gained popularity as an alternative method (Willhauck, 2000; Heyman *et al*., 2003).

A segment-based, object-oriented approach results in real-world object classification, with procedures applied to homogenous units. An object refers to a spatial entity that is homogenous in terms of a selected property, as opposed to the traditional, continuous fields approach found in spatial analysis (Burrough and McDonnel, 1998). Results also are representative of real-world objects and can be devoid of the salt-and-pepper appearance so often found with pixel-based classifiers. While accuracies between pixel-based and object oriented approached are similar in many cases, the visual, realistic classification rendering of the object-oriented approach is of importance. A comparison of standard maximum likelihood classifiers vs. object-oriented classification techniques in the Argentine Nothofagus forests showed that the object approach performed better in terms of classification accuracy (Willhauck, 2000). The author concluded that the object-oriented approach also resulted in a better visual result, while the maximum likelihood classification had a distinct salt-and-pepper appearance. Hill (1999) and Heyman *et al*. (2003) successfully applied per-segment classification approaches for aspen (*Populus tremuloides*) and swamp-forest, lower-, middle, and upper floodplain forests mapping. Accuracies as high as 91% highlighted the potential of object-oriented approaches to classification. Hill (1999) also stressed the usefulness of aggregation and disaggregation of segments at various scales to ecological managers. High accuracies, real-world object extraction, and hierarchical aggregation, important to scaling attempts, can therefore all be listed as advantages of object-oriented classification.

A complete, lidar-based, object-oriented approach to volume- and biomass-by-type estimation has potential as an encompassing forest inventory method. Volume- and biomass estimation based on lidar height distributions have proven useful in grid-cell studies (Means *et al.*, 2000; Næsset, 2002), but the extension of this idea to segments or objects is unique. Attribution of such per-segment estimates to specific types (Douglas *et al*., 2003) is the final step towards an almost stand-alone, remotely sensed inventory system, requiring only a limited amount of ground truth or calibration. However, lidar data segmentation, volume- and biomass modeling, and forest type assignment will require effective combination and application for the result to be acceptable to forest managers. Even though current inventory methods place a high burden on time and financial resources, they are often both unbiased and precise. The acceptance of remote sensing inventory methods by the forestry industry is therefore very dependent on these methods' ability

to either match or improve on current accuracy and precision standards, while being more cost effective in the long term.

1.2 Objectives

The overall objective of this study was to evaluate the application potential of a lidar-based, object-oriented approach to forest volume- and biomass-by-type estimation. Previous grid-cell, lidar distributional volume- and biomass modeling efforts (Means et al., 2000; Næsset, 2002) were extended to a segment-based approach. This was followed by object-oriented forest type classification, again based on lidar distributions (Douglas *et al*., 2003). Lidar data, as sole input data source to the objective, potentially could be used to derive objects, estimate volume and biomass, and classify objects as forest types. Considering the use of lidar as structural data source and the need for cost effective, unbiased, and precise results in an operational forestry context, the specific objectives of this study are as follows:

(i) Assess the utility of object-oriented, lidar distribution-based estimation of forest volume and biomass.

(ii) Determine the feasibility of object-oriented classification of lidar data for subsequent assignment of volume- and biomass estimates to forest types.

(iii) Develop a lidar-based, multiresolution, hierarchical approach to forest geographical differentiation whereby unique structural objects or segments can be used for subsequent forestry-related analyses. This primarily includes definition of a decision-tree to determine the optimum segmentation result, based on between- and within-segment variability, and validation of segmentation selection through volume- and biomass estimation and object-oriented classification.

If successful, such an integrated approach would build on existing work in the field of forest inventory, but incorporate a distinctly novel approach to parameter modeling using only lidar data. Figure 1.1 shows a flow chart detailing objectives and methods used to address each individual goal. Chapter 2 looks at prior work in the two broad areas of segmentation and associated object-oriented classification, and lidar forest volume- and biomass estimation. This is

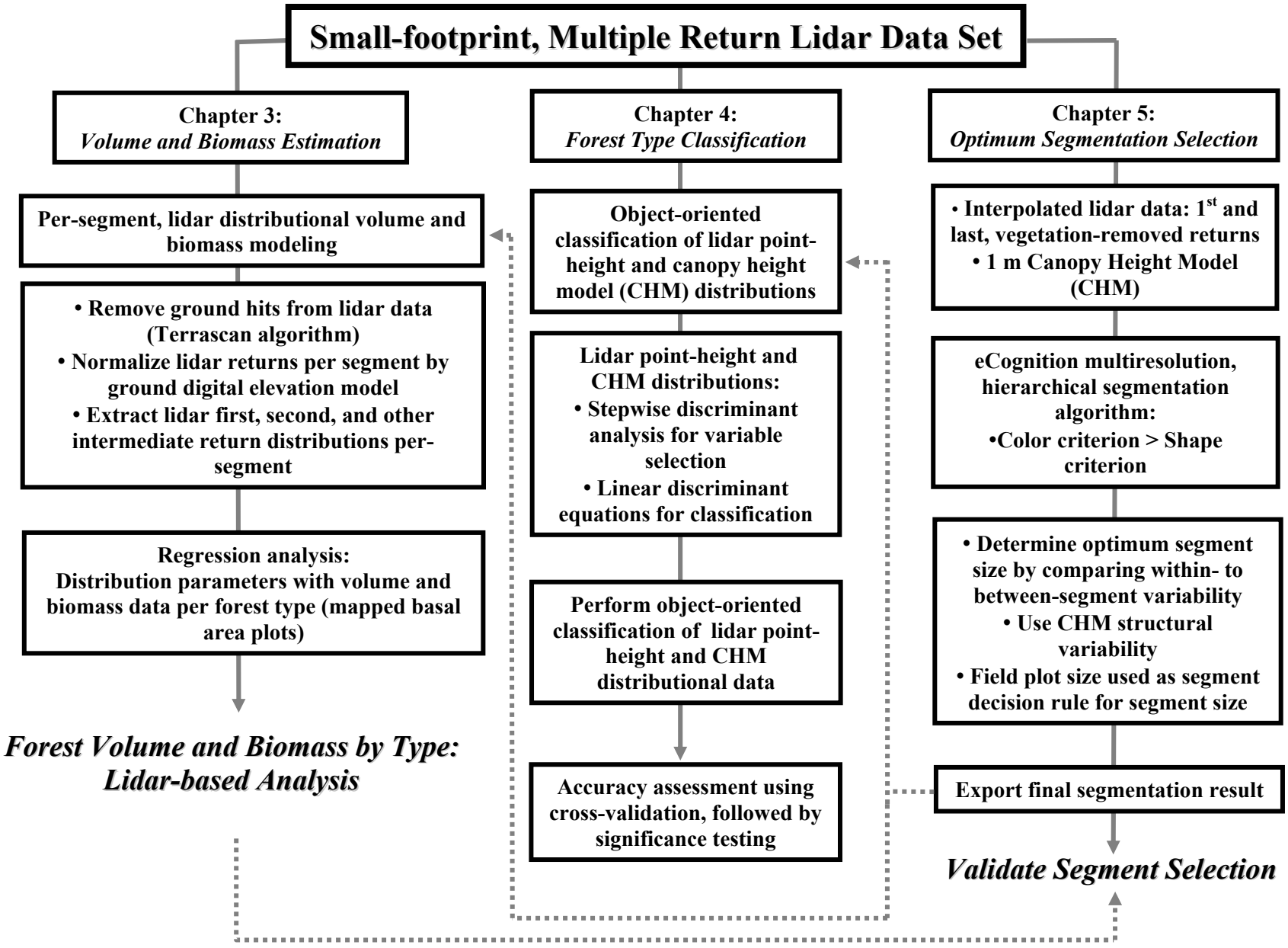


Figure 1.1 Flow chart of the lidar-based, object-oriented approach to forest volume- and biomass-by-type estimation

followed by per-segment, lidar-based forest volume and biomass modeling (Chapter 3), object-oriented segment assignment to distinct forest types (Chapter 4), and forest area segmentation and selection of an appropriate segmentation result (Chapter 5).

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Volume- and biomass-by-type estimation has three main research questions that have to be addressed. The first step is the derivation or choice of analysis units, based on segmentation of a study area. The second question is related to the development of per-segment and per-type volume and biomass models, while the third and last step involves assignment of per-segment estimates to forest types. Together these three components form a cohesive approach to volume-by-type assignment, while extension to biomass modeling is a logical addition. This latter aspect has far reaching implications that extend past the operational arena to the realm of ecological implementation in terms of carbon monitoring. Segmentation as it relates to the definition of analysis objects, associated object-oriented classification, and estimation of forest biophysical parameters using lidar data will be discussed next.

2.2 Forest Area Segmentation and Object-Oriented Classification

Image segmentation algorithms are defined as processes by which an original image description with specific gray levels is translated into a description of regions, with each region being distinctly different from the next. In remote sensing this is further refined as being the search for homogenous regions and later classification of these regions. In short, segmentation is the partitioning of an image into regions that share common properties. Segmentation needs to be both exhaustive and exclusive, while cells or points partitioned in a region share at least one common property. Therefore, regions have two key properties related to their (i) spectral, structural, or binary and (ii) spatial uniqueness. Such regions in images are important because they correspond to objects in a scene. For binary images segmentation is synonymous with thresholding, with either a 0 or 1 categorization (Wilson and Spann, 1988; Jain *et al.*, 1995; Russ, 1995; Darwish *et al.*, 2003).

Thresholding, statistical classification, edge detection, and region growing are defined as main types of image segmentation approaches. Each method has advantages and disadvantages, e.g., thresholding of image histograms is relatively straightforward, but disregards spatial information. On the other hand, statistical approaches take all image information into account, but ignore spatial explicitness inherent in remote sensing imagery. Edge detection and region growing share the common disadvantage of subjective user input, while spatial information recognition is an advantage in both cases (Wilson and Spann, 1988). Segmentation approaches include segmentation based on graph theory (Cheevasuvit, 1990), knowledge-based segmentation (Ton *et al.*, 1991), unsupervised segmentation using nonlinear regression (Acton, 1996), the Woodcock-Harward centroid-linkage algorithm (Shandley *et al.*, 1996), Markov random field model-based segmentation (Smits and Dellepiane, 1997; Sarkar *et al.*, 2002), a Hough transformed-based approach (Shankar *et al.*, 1998), watershed-based hierarchical segmentation (Li *et al.*, 1999), and iterative edge-region co-operation (Kermad and Chehdi, 2002).

Image segmentation itself is a topic that has fostered intense research in many fields over the past decades. Some recent, general natural resource studies will be looked at first to draw focus to this field, followed by a discussion of forest classification and volume estimation segmentation studies.

2.2.1 General Natural Resource Segmentation

Ryherd and Woodcock (1996) applied a multi-pass, pair-wise, region growing segmentation algorithm, also known as the Woodcock-Harward algorithm, to simulated conifer forest, natural vegetation, and mixed-use urban areas using Landsat TM simulation data. The algorithm generates spatially homogenous regions based on Euclidian distance in *n*-dimensional spectral space. Addition of textural data to spectral data generally increased segmentation accuracy, with no accuracy reduction with any of the data sets used. Lobo *et al.* (1998) cited improved discrimination among segments based on per-segment statistics vs. conventional per-pixel information as justification for a segmentation approach to grassland mapping in California. Color infrared photography with 13.5 cm resolution was used as input to an edge preserving algorithm (EPS). It was followed by application of a facet-merging segmentation algorithm

(IMORM). Image segmentation coupled with canonical discriminant analysis proved to be adequate for mapping patch dynamics at the required ecological scale. Abeyta and Franklin (1998) used the Woodcock-Harward algorithm to derive boundaries between desert shrub communities in the Anza-Borrego Desert State Park in California. In-field, line intercept sampling was used to verify results. Image region boundaries showed less than 10% omission error, but approximately 50% commission error when compared to field data. Principal component analysis (PCA) and texture did not increase boundary accuracy in this case. A region growing segmentation algorithm was applied to channels 1 and 2 of NOAA AVHRR data in Mato Grosso, Brazil. Eight classes, including dense forest, savanna, and floodplain vegetation, were mapped to a Kappa coefficient of 0.4. The authors concluded that segmentation, followed by supervised classification, is useful for vegetation mapping at a regional scale (Rodriguez-Yi *et al.*, 2000).

A comparison of standard maximum likelihood classifiers vs. object-oriented classification techniques in the Argentine Nothofagus forests showed the two approaches to be very similar in classification accuracy (93% and 96%, respectively) (Willhauck, 2000). The author concluded that the object-oriented approach resulted in a much better visual result, while the maximum likelihood classifier resulted in a distinct salt-and-pepper appearance. This is of importance, especially given that validation data consisted of single pixels. Pixel-based accuracy assessment on a pixel-based classification result could yield random results in the case of very fragmented segmentation outcomes. Ecological biomes usually are patch-like in nature. This makes the visual result of a study important for practical considerations. Schwarz *et al.* (2001) compared a pixel-based, parallelepiped supervised classification to eCognition's object-oriented approach for storm loss detection is Swiss alpine forests. The object-oriented approach resulted in better accuracies for IKONOS imagery with high spatial resolutions of 4 m (multispectral) and 1 m (panchromatic). However, similar accuracies were found when using SPOT imagery with a 10 m resolution. This result yet again highlights the possible shortcomings of pixel-based classification methods when high spatial resolution data are used.

Evans *et al.* (2002) used a canonically-guided region growing (CGRG) procedure for segmentation of multispectral Landsat TM data of western Australian farmland. Internal field

markers were generated from a local canonical Eigen value image. CGRG segmentation was found to be more accurate for field boundary positioning when compared to a similar region growing-spectral clustering algorithm. The multiresolution, hierarchical approach of the eCognition algorithm was used by Gomes and Marcal (2003) to segment ASTER data with 9 bands between 520 – 2430 nm (15 m spatial resolution) in northwest Portugal. After identification of 9 land cover classes the image was classified using both maximum likelihood and a fuzzy-logic approach. The maximum likelihood algorithm resulted in higher accuracies (71.5%) than the object-oriented, fuzzy-logic approach (46.3%). It was concluded that segmentation into objects should provide more realistic results as opposed to treating pixels as individual observations, independent of their neighborhood. Although segmentation results for this study seemed visually acceptable, the fuzzy-logic classification algorithm did not perform well in cases where objects had high spectral similarity (Gomes and Marcal, 2003). Some authors claim that object-based classifiers better exploit expert knowledge and contextual information than pixel-based classifications. Flanders *et al.* (2003) used traditional spectral information, polygon shape parameters, and context with other classes to classify forest cut blocks in British Columbia. The authors concluded that the object-based approach was more accurate due to the inherent shape characteristics of cut blocks, an aspect not considered by pixel-based methods. Results such as this lend credit to the claim that object-based approaches to segmentation and classification are more likely to keep spatial information intact than pixel-based methods.

2.2.2 Forest Area Segmentation and Classification

Segmentation approaches have been used extensively for forest inventory and object-based classification purposes. Häme and Tomppo (1987) used both Landsat TM and SPOT data and an edge-based segmentation algorithm to delineate forest stands in southern Finland. Segmentation formed part of a stand-based forest inventory scheme. The main concept was that forest stands are by definition homogenous units. Segmentation of images over forested areas delineates such homogenous units, thereby facilitating inventory and reducing error bounds. The SPOT images were deemed best for segmentation with their higher radiometric and spatial resolution when compared to Landsat. Applicability of SPOT data to natural resources management also was

investigated by Jaakkola (1989) who used an edge-preserving segmentation algorithm and subsequent classification of segments to delineate forest stands. The maximum likelihood classification of segments performed slightly worse than a contextual maximum likelihood classifier (77% vs. 88%), but executed up to 15 times faster. Even though the algorithm detected most boundaries, manual editing of final boundaries gave better results than when a developed expert system was used.

Woodcock *et al.* (1990) argued that an image segmentation algorithm could delineate forest stands based on spectral and texture data. They stated that three forest stand attributes needed differentiation, namely tree size, density, and type or species composition. The Woodcock-Harward region-growing algorithm was applied to Landsat TM data. Similar regions were merged on each pass based on a similarity criterion, which had to adhere to minimum and maximum size requirements. Preliminary requirements were set based on United States Forest Service (USFS) specifications. A minimum of 30 pixels (2.7 ha), a maximum of 200 pixels (18 ha), and a non-forest minimum of 22 pixels (2 ha) were set. Ryherd and Woodcock (1990) stated that addition of textural data to spectral bands increased stand boundary definition, especially in areas of changing tree size and in areas where a gradation between forest and non-forest land cover occurred. However, textural data could not be used as the only input to the segmentation algorithm or in a ratio of less than 1:3 with spectral bands, otherwise the texture band overwhelmed spectral information, which resulted in poorer stand delineation.

Woodcock *et al.* (1994) applied the Woodcock-Harward algorithm and unsupervised classification to estimate conifer forest stand attributes in the Stanislaus National Forest in California. The error for the estimate of total timber volume was 4.6% with the conclusion that such mapping projects were suited to large-scale mapping. In a study by Ryherd and Woodcock (1996), remotely sensed images were segmented using both spectral and texture data. The addition of texture data to spectral data helped to improve segmentation results and never decreased segmentation accuracies, even in a simulated coniferous stand. Textural data were especially useful in cases where the features of interest showed differences in local variances. The authors used the Woodcock-Harward approach with different combinations of bands and weighting of texture. They concluded that addition of texture data was the single most important

ingredient for successful segmentation of forested areas (Ryherd and Woodcock, 1996). Canters (1997) used pixel probability vectors to group similar pixels in distinct segments or fields. This was done under the assumption that attribute uncertainty is field-based and not pixel-based, thereby enabling the author to define classification uncertainty without added bias.

The idea of distinct forest units also came to the fore in a study by McCormick and Folving (1998), who used segmentation and fractal and dominance methods to estimate biodiversity per forest stand. Segmentation was done by following an edge-preserving smoothing algorithm. The segmentation algorithm was based on the linkage of each edge pixel in an image to its neighboring pixel with the minimum edge value. Data structures called "directed trees" were thus formed, with pixels belonging to the same tree forming a segment. A major advantage listed by the authors is that of spatial implicitness of the algorithm and its results. Heyman *et al*. (2003) used a per-segment approach to improve aspen (*Populus tremuloides*) mapping in Oregon, USA. They applied a histogram thresholding method based on hue and saturation values of high-resolution color-infrared photographs. The authors achieved 88% accuracy for mapping aspen segments into three broad categories (no aspens; 0-50% aspens; 50-100% aspens).

Current segmentation-based volume estimate errors are too high (≈ 59%) for use in forest management (Kilpeläinen and Tokola, 1999). The study was conducted in southern Finland on pine, birch, and other broadleaved species. The authors concluded that results were very dependent on the data and validation method used, but that segmentation could ultimately be used to stratify forests, thereby reducing variation and increasing sampling efficiency. With forest mapping and classification as main goals, Hill (1999) used a low-level edge detection and region growing segmentation approach in southeast Peru to delineate swamp-, lower-, middle, and upper floodplain forests. An accuracy of 91% was obtained when lower-level segments were aggregated into six meaningful forest classes. The author underlined the usefulness of aggregation and disaggregation of segments at various scales to ecological managers. A statistical segmentation approach was used by Bressers and Oevelen (1999) to identify erosion indicators in tropical forests. Two approaches, namely a statistical homogeneity approach and a successive edge detection and region growing algorithm, were used. These segmentation approaches were used in conjunction with texture analysis to identify clear cuts, followed by

radar backscatter measurements within identified regions. On the other hand, Abkar *et al.* (2000) used a likelihood-based segmentation approach, whereby posterior probabilities were calculated on a per-object basis, rather than per-pixel. This led to a reduction of spectral probabilities. This method was used to estimate deforestation extent and location in Thailand by comparing it to a per-pixel maximum likelihood classification of the same area. The segmentation approach improved accuracy by 10%, which was deemed significant.

Two of the more traditional segmentation approaches, edge detection and region growing, were used by Kermad and Chehdi (2002) to segment an agricultural landscape and aid in forest vegetation classification. The edge detection process was coupled to a region growing algorithm in an iterative manner. Over-segmented results were re-entered in the process while loosening constraints until outcomes converged and stable results were obtained. Hyyppä and Inkinen (1999) were able to derive accurate tree locations and crown areas from segmented lidar data when using a modified watershed segmentation procedure. This approach ultimately led to accurate height and volume estimations when crown area and stems-per-hectare derivations were used.

Of particular interest to this study are two Finish studies that attempted "segment-aided" timber volume estimation. The first by Makela and Pekkarinen (2001) attempted a Landsat TM plot-level volume-by-species estimation. The authors used a measurement space-guided clustering, defined as an ISODATA classification followed by connected component labeling (ISOCCL), which in turn was based on edge detection and linking. A directed trees approach, based on gradient analysis and edge detection, was applied as an alternative segmentation method. Spectral features used for volume estimation were extracted in two ways, namely (i) from a fixed window around the field sample plot, and (ii) from the pixels in the fixed window that belonged to the same segment as the sample plot. The ISOCCL approach yielded the most accurate volume estimation for pine and spruce species (*Pinus sylvestris* and *Picea abies*), as well as for the total volume. The directed trees algorithm yielded the best results for broad-leaf species (*Betula pendula*, *B. pubescens,* and *Populus tremula*). Improvements from the fixed window approach to segment-based approaches were relatively small (1%-11.3% in relative RMSEs) with RMSE values remaining high. The authors concluded that segmentation of forest areas was

very suitable for the estimation of forest variables in that errors could be minimized by extracting estimates from more homogenous segments (Makela and Pekkarinen, 2001).

Pekkarinen (2002) investigated large stand-level errors as a follow-up study. Departing from the premise that these large errors are related to the limited spatial resolution of sensors such as Landsat TM, the author investigated the use of very high resolution images for image-based multi-source forest inventory (MSFI). Segments again were created using the directed tree approach, followed by a region merging algorithm. Segment-based spectral features were compared to those derived from square shaped windows. Although the segmentation algorithm succeeded in delineating distinct forest areas for feature extraction, the segment-based approach was only marginally better than the reference. RMSEs of broad-leaved species decreased by a maximum of 10%, while only marginal decreases were seen for spruce species (≤ 8%). In general, plot-level errors remained high. The author suggested that these large errors could be due to the local nature of field data and that segment-level data could decrease associated segment-level errors (Pekkarinen, 2002).

Engdahl *et al.* (2003) incorporated the eCognition, multiresolution segmentation algorithm in a design aimed at estimating accurate stem volumes of Scots pine and Norwegian spruce in southern Finland. INSAR-based stem volume estimates were comparable to Finish National Forest Inventory (NFI) estimates, with RMSEs of 101 m^3/ha and 115 m^3/ha, and correlation values (r) of 0.79 and 0.69, respectively. The authors concluded that it is possible to produce an accurate, segmented land cover classification and stem volume estimates for forest segments. Kellndorfer and Ulaby (2003) used the eCognition segmentation algorithm for forest biomass inversion from SAR data. eCognition was found to generate accurate image objects which were spatially similar to existing stand boundaries and ecological units. Stand and object backscatter correlated well with ERS and JERS data ($R^2 = 0.89$). They concluded that image object biomass inversion can be trained on inventoried forest stands and applied to larger segment data, thus avoiding poor performing pixel-based inversion.

Kressler *et al.* (2003) used panchromatic KOMPSAT-1 and SPOT-5 data to classify basic land cover types in western Austria. Homogenous objects were derived with the multiresolution

eCognition algorithm based on scale, shape, color, smoothness, and compactness of segments. Overall accuracies of 89.9% (KOMPSAT) and 86.3% (SPOT) were obtained, with the largest confusion between agriculture and forest clearings. Kayitakire *et al.* (2002) used the same algorithm to map mixed oak, spruce, beech, and pine forest stands in Belgium. Overall accuracies of 88% (per-pixel clustering) and 83.3% (per-parcel derived map) were found. It should be noted that wrong classification of a parcel resulted in all the parcel pixels being misclassified, as opposed to singe pixel misclassification. Nugroho *et al.* (2002a) applied this multiresolution, hierarchical algorithm to SAR and Ikonos data to analyze forest spatial structure in Indonesia. They found that single tree crown shape and tree distribution defined forest spatial structure, with both characteristics having been measured from hierarchical segmentation results. These results were extended in a follow-up study. Tree objects were quantitatively clumped to various degrees and hierarchical levels, based on parameters such as distance to closest trees, tree height, and canopy structure (Nugroho *et al.,* 2002b). Many of these parameters were obtained using the fuzzy logic capabilities of eCognition software. De Kok *et al.* (1999) concluded that, due to eCognition's fuzzy logic capabilities, more advanced classification methods have been made available to users of data with increased spatial and spectral resolutions. The authors claimed that this algorithm is suited to forestry applications because of the ability to set the scale of segmentation, complemented by built-in hierarchical capabilities.

In conclusion, multiresolution, hierarchical segmentation, coupled with lidar-based structural input, has great potential for forest area segmentation. Sampling within such homogenous stands could provide the user with high degrees of accuracy with small associated errors (Makela and Pekkarinen, 2001; Pekkarinen, 2002), thereby contributing to the integration of remotely sensed data into forest resource inventory. A lidar-based approach should give researchers and field users an idea as to the degree that this technology is applicable to forest segmentation specifically, and eventual volume- and biomass-by-type modeling. Potential segment-based estimates also could be scaled, through recombination, to mimic a stand-based forest inventory.

2.3 Lidar Technology and its Forestry Applications

Estimation of accurate and unbiased forest biophysical attributes is a common goal shared by all natural resource managers. Traditional, empirical models that utilize growth-and-yield models are well established, but they do not necessarily add to our knowledge of natural forest processes. Recently, process models have come to the fore as attempts are being made to model natural ecosystem processes. These process models are crucial elements of many managerial systems. Examples include the use of process models in strict yield modeling and as part of defining carbon sequestration, and more importantly, potential effects of global warming. Carbon sequestration and global carbon budgets therefore have become topics that foster intense research efforts. Temperate forests in the USA provide a variety of ecosystem services. One of the most pronounced contributions is provision of negative feedback to the greenhouse gas accumulation, which subsequently retards global warming through carbon sequestration (Schlesinger, 1995). It is estimated that carbon stored on US timberland has increased by 38% to 8.8×10^{15} g from 1952 to 1993. This carbon was sequestered on 296×10^{6} ha of forests, which corresponds to 5% of the world's forested area. This sequestration accounts for as much as 21% of the possible carbon sink in temperate forests (Birdsey *et al.*, 1993). The US temperate forests could become even greater contributors to global carbon sequestration in decades to come, since net carbon accumulation of commercial forests in the mid-latitudes are below its biological capacity (Dixon *et al.*, 1994).

Regional carbon modeling, which extends to global initiatives, has become increasingly important internationally. The Kyoto Protocol, an international agreement seeking to reduce greenhouse gas concentrations, has been initiated with major relevance for US forests (Cathcart, 2000; Murray *et al.*, 2000). Relatively simple carbon sequestration estimation techniques for small forest areas have already been described (Hoover *et al.*, 2000; Haswell, 2000). These techniques enable land managers to determine the contribution of their forested land to the global carbon credit program. This quantification of carbon sequestration, as defined by net primary productivity (NPP), will play an even greater role in future natural resource management. Terrestrial contribution to the global carbon cycle is still not fully understood, with specifically forest vegetation being a major link in a complex system. Factors such as (a) the effect of forests

as either net carbon sinks or sources, (b) the temporal dependence of the sink or source status of forests, and (c) the effect of intensive management on this status, still have to be addressed in both the research and applied arenas. Since many NPP models have been established at a local scale, wider use of such models to calibrate and validate regional- and global-scale models is the next logical step. This model expansion process will be dependent on parameter supply from remote sensing platforms. Lidar technology offers a possible avenue for estimation of aboveground volume and biomass, which are derived from tree heights, stem counts, and crown diameters. Forest volume and biomass in turn are essential base parameters for many carbon models, as are variables such as stems-per-acre and a measure of canopy structure. The latter is related to leaf area index (LAI), with LAI aiding in the definition of photosynthetic ability (Savage, 1999). Combination of lidar data and hyperspectral data is likely to increase the effectiveness of current lidar algorithms when estimating forest biophysical parameters. Volume and biomass estimates by species or type could in turn result from use of hyperspectral data. Both empirical and process models may benefit from the scaling capabilities of remote sensing data.

Two main goals of lidar measurement of forest biophysical character are to extract volume and biomass. This will enhance forest managerial aspects and aid in the measurement of aboveground carbon allocation in forest ecosystems. Lefsky *et al.* (2002a) stated that lidar sensors are able to provide accurate and non-asymptotic estimates of various forest indices such as LAI and aboveground biomass. Laser scanning systems afford researchers an excellent opportunity to extract precise elevation points from an earth-bound surface, whether that entails a ground (digital elevation model) or vegetation (canopy height model) surface. In a forestry context the possibility of determining heights, stem counts, crown diameters, and gauging forest type and/or species through structural indices derived from lidar data is of extreme importance. What has become spectrally possible when using imagers, now is matched by structural possibilities when using lidar technology. But not unlike data from their spectral cousins, raw structural lidar data are a far cry from valuable information. In order to cover the broad range of lidar-related topics, lidar sensors and their working will therefore be discussed first, followed by an in-depth look at the application of lidar in forestry scenarios.

2.3.1 Airborne Laser Scanning using Lidar Sensors

The main purpose of all laser-scanning systems is to accurately measure the distance between target and sensor. The ranging unit of such a system includes both the emitting laser and the electro-optical receiver. The apertures of these two components are mounted such that the transmitting and receiving paths share the same optical path, thereby ensuring that laser-illuminated objects are in the field-of-view (FOV) of the optical receiver. Distance to the target is calculated by halving the time elapsed between emission and arrival of the reflected laser pulse at the receiver (Ackermann, 1999; Wehr and Lohr, 1999). Some benefits of lidar technology vs. traditional photogrammetric techniques include lidar's ability to operate during night time, its high automaticity, inherent structural information, canopy penetration capabilities, and high point density, while lack of full area coverage and cost can be listed as drawbacks (Ackermann, 1999; Baltsavias, 1999b). Future improvements may include pulse rate, resolution increases, and increased accuracy. Baltsavias (1999c) mentioned that airborne laser scanning (ALS) system manufacturers have increased from 1 to approximately 40 from 1996 to 1999.

Laser ranging systems can be divided into pulse and continuous wavelength (CW) lasers. Pulsed lasers are more prevalent in current systems than continuous wavelength or phase difference lasers. Pulsed laser altimeters can further be subdivided into discrete return or waveform-sampling sensors. Discrete return sensors measure single- or multiple return distances by evaluating the returned energy signal to find a peak or peaks that define discrete objects in the laser's path. Either the distance to peak-edge or maximum power of a peak is recorded (Wehr and Lohr, 1999). Multiple return systems are often used to extract both canopy and ground returns, assumed to be represented by first and last returns, respectively. Vegetation canopy height is then defined as the difference between the first and last returns (Lefsky *et al.*, 1999a).

Return waveform systems operate on the assumption that the shape of the waveform from the returned signal represents a vertical distribution of the intercepted surfaces within a given laser footprint. Such a waveform accounts for the spatial distribution of laser beam intensity along and across the laser beam's path. Discrete systems generally utilize a small footprint (< 1 m) sensor, as opposed to waveform systems with large footprint (5 - 15 m) sensors (Weishampel *et al.*,

1996; Blair and Hofton, 1999; Lefsky *et al.*, 1999a). Large footprint systems are useful for waveform sampling devices because recovery from tops of crowns and the ground are possible in the same waveform, while the resolution remains small enough to detect individual crown contribution to such a waveform (Lefsky *et al.*, 1999a).

Some considerations when using small-footprint lidar data in a forestry scenario include spatial distribution of tree stems, spacing or density of lidar returns, and the interpolation method used to derive a canopy surface. High return density lidar usually are costly to collect and expensive in terms of hardware requirements. Cost has to be offset with lidar point density in order to determine which cost-density setup would best fit a given stand's spatial distribution, as well as the user's pocket (Evans *et al.*, 2001).

2.3.2 Lidar Data Analysis: Algorithms and Processing Techniques

Some basic formulas that apply to laser scanning are listed in Table 2.1 (Baltsavias, 1999a). These formulas are helpful in planning and executing a lidar mission. They also are useful to determine properties of a lidar data set. Raw height data from lidar systems can be ordered based on provider specification ranges. Most data set attributes are dependent on the system used and conditions of the actual collection. These include flying height, airspeed, and overlap. Some of the attributes that may affect preprocessing algorithms are point density, registration of multiple returns, and amplitude registration. Point density can be varied through flying height and platform velocity, while multiple returns, which are used for vegetation and ground surface separation, is a function of both vegetation structure and the vertical resolution of the system. Removal of unwanted laser measurements, which may include noise, outliers, or gross errors, is referred to as "filtering". A basic approach to lidar processing can be described as one that (a) uses the original lidar data as long as possible, (b) separates surfaces and objects on the surfaces, and (c) develops algorithms that are based on applications for object classification and modeling (Axelsson, 1999; Petzold *et al.*, 1999). Although initial costs are high, laser scanning for derivation of DEMs was found to be 67% - 75% cheaper than photogrammetric compilation (Petzold *et al.*, 1999). High horizontal and vertical accuracies, as well as spatial resolution and coverage, also offset large initial investment in data acquisition.

Table 2.1 Basic lidar formulas (Baltsavias, 1999a)

Characteristic	Formula
Range and range resolution	$R = c\frac{t}{2};\ \Delta R = c\frac{\Delta t}{2}$
Vertical resolution (return separation)	$R_{\min} = c\frac{t_{\min}}{2}$
Swath width	$SW = 2h\tan\left(\frac{\theta}{2}\right) = ah$
Along track point spacing	$dx_{along} = \frac{v}{f_{sc}}$
Across track point spacing	$dx_{across} = \frac{SW}{N}$
Point density per unit area	$d = \frac{FnT_s}{A}$
R = Range (m); **c** = Speed of light (km/s); **t** = Time between sending and receiving a pulse (ns); **SW** = Swath width (m); **h** = Average flying height over ground (m); θ = Laser scanning angle (°; FOV); v = Laser frequency; $\mathbf{f_{sc}}$ = Scan rate (Hz; scan lines per second); dx_{along} = Average distance between scan lines, along track (m); dx_{across} = Average point spacing across track (m); **N** = Number of points per scan line; **d** = Average point density (points/m^2); **F** = Pulse rate (kHz); **n** = Number of flying strips to cover area; $\mathbf{T_s}$ = Flying time per strip (h); **A** = Covered area (km^2/h)	

Error checking is critical before processing lidar data. For this reason Petzold *et al.* (1999) recommend comparing lidar derived heights to known heights in relatively open, vegetation-free areas. Residual values of in a lidar data set can be used to determine the reliability of such a data set. Petzold *et al.* (1999) concluded that filtering techniques should be improved to take into account ground returns on steep slopes. Such returns would be discarded by many filtering algorithms since they might fall outside a specified threshold when compared to adjacent lidar points. Maas and Vosselman (1999) used invariant moments with closed solutions to determine parameters for simple building models. Precision ranged from 0.01 – 0.2 m for building dimensions and 1 - 2° for building orientation and roof slopes. The applicability of such an

approach could be worthwhile to investigate for coniferous, uniformly shaped crowns in even-aged plantations.

There are many established algorithms for the derivation of canopy heights from lidar data. Næsset and Bjerknes (2001) used a two-stage regression technique, first regressing mean height of dominant trees against laser-derived canopy heights and then using these equations to predict the mean height of selected spruce and pine stands. This method was further developed by Næsset (2002) who used stratum-specific regression equations to predict canopy parameters for reference stands. Næsset and Okland (2002) used a similar regression approach to predict canopy height, height to crown, and crown length as a proportion of tree height.

2.3.3 Estimating Forest Biophysical Parameters using Lidar Data

Photogrammetric techniques for the attempted measurement of tree heights, forest volume, and canopy density are well known (Gougeon, 1995; Brandtberg, 1997). Regression analysis of white spruce (*Picea glauca*) tree seedling silhouette area (derived from vertical photographs) and tree diameter and biomass were very promising, with R^2 values of up to 0.97 (Ter-Mikaelian and Parker, 2000). Crown projection area, percent exposed crown area, and relative height were used as independent variables in basal area growth equations for northern hardwood stands (*Acer saccharum, Fraxinus americana, Tilia americana*) in Wisconsin and had R^2 values ranging between 0.77 and 0.88 (Cole and Lorimer, 1994). Lidar makes the measurement of canopy profiles (crown dimensions) in 3-D spatial distributions possible, with an added benefit of ground-based elevations (DEMs) derived from "last-return" lidar signals (Weishampel *et al.*, 1996). Lidar systems, representing a unique approach, could negate the need for ground-based, small scale measurements of tree heights and/or canopy parameters, and provide more automation and positional accuracy than photogrammetric techniques.

2.3.3.1 Estimating Forest Biophysical Parameters using Large-footprint Lidar Data

Large-footprint lidar systems usually are based on waveform sampling techniques and are referred to as surface lidar systems. Waveform distributions define the strength or energy of the laser return from a given footprint at different time or distance intervals, thereby recording the vertical distribution of the returned laser illumination from all canopy elements and the ground (Lefsky *et al.*, 1999a). Waveform systems incorporate all vertical elements in a single footprint. They also directly relate waveform to biomass through lidar curve area. Disadvantages include inability to detect individual crowns and coarser ground spacing.

2.3.3.1.1 Plot-level Estimation of Forest Biophysical Parameters using Large-footprint Lidar Data

Lefsky *et al.* (1999b) used canopy structural indices obtained from SLICER data of Douglas-fir and western hemlock (*Pseudotsuga menziesii*, *Tsuga heterophylla*) to estimate species basal area and biomass. Plots were divided into very young ($n = 4$), young, ($n = 5$), mature ($n = 4$), and old-growth ($n = 9$) categories. Adjusted R^2 values for total biomass, total basal area, and number of stems (> 100 cm) were 0.91, 0.87, and 0.85, respectively. No root mean square errors (RMSE) were reported. The most significant independent variables were maximum canopy height, filled canopy volume, number of waveforms greater than 55 m, and closed and open gaps (Lefsky *et al.*, 1999b).

Blair and Hofton (1999) created pseudo-waveform distribution data from small-footprint lidar data in the tropical forests of Costa Rica. They used FLI-MAP data with a 10 cm diameter footprint and 30 cm spacing to simulate waveform return data using height distributions. These simulated data were correlated with actual waveform data that were derived from the LVIS system, which had a 25 m footprint of contiguous data. The LVIS sensor is the airborne simulator for NASA's Vegetation Canopy Lidar (VCL) spaceborne mission. It has proven useful for both topographic and canopy mapping in missions at the Sequoia National Forest and in Maryland (Blair *et al.*, 1999). Correlation between pseudo and recorded waveforms for simple, single mode waveforms in non-vegetated, flat areas was 0.99. High correlations also were found

in vegetated, more complex areas, but relative strengths of the two sensors at different canopy elevation levels differed. This was ascribed to a possible difference in tree cover conditions. The only systematic difference between pseudo and recorded data was consistently higher amplitude of the ground return in the recorded waveform. This was attributed to the single-return nature of the FLI-MAP system - a quality that permits only ground returns in 10 cm-wide gaps in the canopy. The authors concluded that vertical structure information for a return-waveform laser altimeter with a medium to large footprint, can be synthesized using a high resolution data set such as FLI-MAP. This will make algorithm testing on simulated data for large footprint, waveform lidar sensors such as the planned VCL satellite possible (Blair and Hofton, 1999).

Means *et al.* (1999) used metrics derived from large footprint SLICER data to estimate height, basal area, total biomass, and leaf biomass of coniferous forests (*Pseudotsuga menziesii* and *Tsuga heterophylla*) in the Pacific Northwest. R^2 values ranged from 0.84 (leaf biomass) to 0.96 (basal area; total biomass). RMSEs for predicted variables were 3.8 m (height), 9 m^2/ha (basal area), 88 Mg/ha (total biomass), and 2 Mg/ha (leaf biomass). Lidar height, canopy reflection sum, and quadratic mean height were the most useful independent variables. Canopy cover (0-1 range; $R^2 = 0.94$) was predicted using canopy closure as independent variable (Means *et al.*, 1999). Similarly high R^2 values were obtained by Sun and Ranson (2000) when using simulated lidar waveform models of trees and forest stands to derive forest vertical structure. R^2 values as high as 0.98 (height), 0.93 (crown width), and 0.94 (crown length) were found for Jack pine plots. RMSEs were not reported. The authors concluded that slower decay of airborne lidar waveforms could be explained by understory structure and scattering from the upper canopy. Although depiction of canopy structure when using SLICER data was highly reproducible, Harding *et al.* (2001) mentioned that differences among ground-based measures and lidar measures still exist in a mixed deciduous forest, which was dominated by yellow polar (*Liriodendron tulipifera*). Lidar-derived young and old growth maximum height was overestimated by 2 - 4 m. Mean height differences for old growth were overestimated by 2.2 m and those for young growth were underestimated by 1 m.

Drake *et al.* (2002a) used VCL simulation data from the Laser Vegetation Imaging Sensor (LVIS) to estimate quadratic mean stem diameter, basal area, and aboveground biomass at plot-

level. Plot sizes ranged between 0.25 and 0.5 ha for lowland primary and secondary tropical wet forests in Costa Rica. Independent variables included canopy height, the height of median energy, height/median energy ratio, and a canopy closure measure. The latter was calculated by dividing the ground- by canopy return bins. Results obtained included R^2 values of 0.93, 0.72, and 0.93, and RMSE values of 2 cm, 3 m^2/ha, and 18.39 Mg/ha, for quadratic mean stem diameter, basal area, and aboveground biomass, respectively. They concluded that large footprint lidar with waveform return can be used effectively to estimate forest structural parameters. Drake *et al.* (2002b) found correlations of up to 0.94 (RMSE = 16 Mg/ha) for large footprint lidar data and estimated aboveground biomass in primary and secondary neotropical rainforests. They concluded that the lidar data variables were highly correlated with aboveground biomass across varying forest types.

Lefsky *et al.* (2002b) extended application of lidar remote sensing of aboveground tree biomass to span three biomes. Biomes were defined by temperate coniferous (fir main species; *Pseudotsuga menziesii*), temperate deciduous (yellow poplar main species; *Liriodendron tulipifera*), and boreal coniferous forests (black spruce main species; *Picea mariana*). Twenty-one, 112, and 16 plots, were measured in each respective biome. Measured biomass ranged between 135.6 and 1329 Mg/ha, 11.4 and 716.3 Mg/ha, and 0 and 58.5 Mg/ha, respectively. Correlations of up to 0.90 were found for above-ground biomass and canopy indices, such as cover and mean canopy height squared, for all three biomes combined. An R^2 value of 0.84 was found for above-ground biomass in the case of all three biomes combined (standard error = 7.5 Mg/ha). Mean canopy height squared and the product of mean cover and mean canopy height were used as independent variables. The authors concluded that more research in this area is warranted, since successful application of a single biomass equation across three distinct biomes was not expected.

2.3.3.1.2 Stand-level Estimation of Forest Biophysical Parameters using Large-footprint Lidar Data

Lefsky *et al.* (1999a) used a large-footprint (10 m diameter; 10 m spacing) lidar system called SLICER (Scanning Lidar Imager of Canopies by Echo Recovery) to predict basal area and

biomass of tulip-poplar stands (*Liriodendron tulipifera*) on the coastal plain of Maryland. Canopy height profiles measured in-field were statistically indistinguishable from those derived from SLICER data. Four indices (maximum canopy height, mean canopy height, median canopy height, and quadratic mean canopy height) were chosen to relate field and remotely-sensed canopy height profile measurements to stand structure attributes. Linear regression techniques were used to relate these height indices to basal area and biomass. Quadratic mean height had the largest R^2 value ($R^2 = 0.7$) and smallest standard deviation of residuals (7.8 m^2/ha) for basal area, while both quadratic mean height and maximum canopy height had R^2 values of 0.8 for biomass prediction. Maximum canopy height had a smaller deviation of residuals (73.9 Mg/ha vs. 75.1 Mg/ha) (Lefsky *et al.*, 1999a).

The use of multifractal analysis to describe canopy height models that are extracted from lidar measurements also has been investigated. A multifractal can be defined as a measure of probability (or some physical quantity), distinguished from its geometric support, that has different fractal dimensions on different parts of the support. Drake and Weishampel (1998) used multifractals to describe canopy height models derived from SLICER data in Orlando, Florida. The landscape was characterized by a longleaf pine (*Pinus palustris*) overstory and saw palmetto (*Serenoa repens*), wiregrass (*Aristida stricta*), scattered gallberry (*Ilex glabra*), and oak (*Quercus myrtifolia, Q. chapmanii*) understory. The flown transects were similar, but multifractal spectra showed local differences within transects, thereby uncovering fine scale differences among transects (Drake and Weishampel, 1998).

Lidar data present inherent challenges, which are mainly due to their spatial nature and sensor characteristics. Some authors (Drake and Weishampel, 1998; Means *et al.*, 1999; Hofton *et al.*, 2000) mentioned co-registration problems even with large footprint data. Weishampel *et al.* (1996) concluded that although forest profile characterization using large footprint lidar data has benefits such as top-down structure, it might be difficult to delineate individual crowns when using such a large resolution. Knowledge of variables such as the laser scan center and the phase center of the GPS antenna, time delay associated with laser electronics, and an estimation of the timing correction over varying topography might aid positional calibration of lidar systems. The ultimate limitation to data precision is linked to the precise determination of airplane trajectory.

It was recommended that precise ephemeris data instead of broadcast data be used to correct differential GPS data (Hofton *et al.*, 2000).

Lefsky *et al.* (1999c) included spatial discontinuity of large waveform samplers and development of wide-range relationships to predict forest structure indices as potential hurdles to the application of waveform lidar. The authors believed that wall-to-wall mapping of biomass was possible when using integrated lidar, Landsat ETM+, and forest inventory data. An R^2 value of 0.73 was found when predicting aboveground biomass for Douglas-fir and western hemlock forests when using such an inventory model. Although costs were substantially higher, Lefsky *et al.* (2001) found that the SLICER system performed better than other remote sensing systems (Landsat TM, AVIRIS, and ADAR) in predicting forest structural attributes. R^2 values of 0.86 and 0.84 and standard errors of 24% and 21% were obtained for biomass and basal area when using SLICER data. Multi-temporal Landsat TM data performed second best with R^2 values of 0.60 and 0.62 and standard errors of 35% and 27%, for biomass and basal area, respectively. It was concluded that lidar offered substantial improvements over traditional sensors in forest structural attribute estimation.

2.3.3.2 Estimating Forest Biophysical Parameters using Small-footprint Lidar Data

Means (2000) listed the estimation of ground surface elevation and canopy top elevation as common goals for large- and small-footprint lidar systems. Differing horizontal resolution, with 0.5 to 3 m resolution for small-footprint lidar and 10 to 25 m resolution for large-footprint lidar, was mentioned as a conflicting characteristic. Small-footprint lidar therefore has great potential for estimating basic stand parameters such as height, crown diameter, and trees-per-acre. This makes derivation of timber volume, using these parameters, possible (Young *et al.*, 2000).

Nilsson (1996) used small-footprint lidar (0.75 – 3 m) to derive mean tree height for even-aged Scots pine (*Pinus sylvestris*) stands and found that laser mean height underestimated observed mean tree height by 2.1 to 3.7 m. This error could be compensated for by using field data for calibration (Nilsson, 1996). This common underestimation mainly is due to the fact that most lidar returns are from crown locations below tree tops (Magnussen and Boudewyn, 1998).

Loblolly pine (*Pinus taeda*) stands, ranging from 9 to 15 years of age, were measured using a small-footprint lidar system. Correlations as high as 0.9 for lidar trees-per-acre and field trees-per-acre and 0.95 for total lidar height vs. total field height were found (Young *et al.*, 2000).

2.3.3.2.1 Tree-level Estimation of Forest Biophysical Parameters using Small-footprint Lidar Data

Hyyppä and Inkinen (1999) were able to derive accurate tree locations, height (< 1 m standard error), crown area, basal area, and volume for Norway spruce and Scots pine using small-footprint data (10 hits/m^2). Relatively small errors were found when extrapolating single tree values to stand level (13.6% for height, 9.6% for basal area, and 9.5% for stem volume). Laser-derived tree height had an R^2 value of 0.97 and standard error of 2.3 m when compared to field measured heights. Associated statistical values for volume estimation were 0.88 (R^2) and 16.5 m^3/ha (standard error). Tree locations were determined using a modified watershed segmentation procedure, which also aided in the measurement of crown area.

Næsset and Okland (2002) found R^2 values of 0.75, 0.53, and 0.51 when laser derived canopy metrics (various quantiles, maximum and mean values, and coefficients of variation) were regressed against actual canopy height, height to crown, and relative crown length, respectively. Standard deviations associated with predicted values were 3.2 m, 2.2 m, and 10.5% for canopy height, height to crown, and relative crown length, respectively. They concluded that mean plot tree height, as opposed to individual tree heights, can be determined more accurately ($R^2 = 0.91$) when using laser data.

Brandtberg *et al.* (2003) used a fuzzy segmentation approach to first delineate individual tree crowns, followed by estimation of laser-based tree heights. The mean standard error for leaf-off individual trees was 1.1 m, with an overestimation bias for shorter trees and an underestimation bias for taller trees. Distinct differences were found among various tree species (oaks, red maple, and yellow poplar) for metrics such as mean, median, and mode of normalized height, maximum laser height, and maximum laser reflectance percentage.

2.3.3.2.2 Plot-level Estimation of Forest Biophysical Parameters using Small-footprint Lidar Data

Small-footprint lidar systems have been used for accurate forest merchantable volume estimation as early as 1986 (Maclean and Krabill, 1986). The authors used NASA's Airborne Oceanographic Lidar (AOL) system with a 0.7 m footprint to estimate volume in loblolly and mixed hardwood stands in Maryland. They effectively correlated lidar profile area with plot-measured volume and obtained a R^2 of 0.92 when using the natural logarithm of volume as dependent variable, and lidar profile area and predominant species as independent variables. No associated error was reported. Nelson *et al.* (1988) found site-specific biomass and volume estimates for coastal pine species (loblolly, shortleaf, slash, and longleaf pine) using a 0.75 m footprint lidar to be very variable, with only 20 to 30% of estimates within ± 10% of actual measured values. However, they concluded that logarithmic biomass and volume models could predict mean total tree volume to within 2.6% of the measured ground values, while biomass predictions were within 2% of ground values.

Weltz *et al.* (1994) used lidar technology to measure vegetation canopy height in desert shrub and semi-desert grasslands in Arizona. They found that lidar height measurements were not significantly different from field height measurements for vegetation taller than 0.3 m, in seven out of eight plots. However, canopy height was underestimated for vegetation shorter than 0.3 m and overestimated for vegetation taller than 0.5 m. Magnussen *et al.* (1999) eliminated a mean bias of -3 m by using two recovery models. In the first model crown height was sampled proportional to crown area, thereby eliminating tree laser hits and ground hits that were too close together in both time and space. The second model assumed a bias between a laser canopy hit and a hit at the top of a tree. The observed canopy heights could then be considered as the difference between the true tree height and this recognized bias. The authors found that model-based estimates of tree height were not significantly different from ground-measured values (Magnussen *et al.*, 1999).

Næsset (2002) refined this technique further by using stratum-specific, distribution-based regression equations to predict volume and crown parameters for Norway spruce and Scots pine

stands in Norway. Strata were defined by age, site index, and tree species. Various quantiles, maximum- and mean values, canopy density measures and coefficients of variation were used as independent variables. R^2 values for various strata ranged from 0.74 (mature forest, poor site quality), 0.85 (mature forest, good site quality), to 0.93 (young forest) for dominant height. R^2 values for volume ranged from 0.8 for mature forests, to 0.93 for young forests. Standard deviations of differences between predicted and measured values were 0.7 to 1.33 m and 18.3 to 31.9 m^3/ha for dominant height and volume, respectively. To achieve accurate results, optimal grid sizes for individual crown determination from interpolated canopy height models may need to be defined *a priori*.

Means *et al.* (2000) implemented a lidar distribution, grid cell based approach to estimate height and basal area for a variety of Douglas-fir stands. Measured heights and basal area were projected from 1996 to 1999 using appropriate growth models. A multiple return, small-footprint (0.6 m diameter; 0.6 – 3 m spacing) lidar system was used, with lidar returns extracted from 10 x 10 m grid cells within larger 50 x 50 m measured plots. Distribution, canopy cover percentiles, maximum height, elevation, average mean height, and average of the maximum heights were calculated for grid cells. Stepwise regression analysis was used to determine the relationships between ground data and lidar measurements, with dependent variables being height, basal area, and volume. R^2 values of 0.93 (RMSE = 3.4 m), 0.95, and 0.97 (no RMSEs for latter two values) were obtained for height, basal area, and volume, respectively. R^2 values for plots excluding old-growth plots were 0.98 (RMSE = 1.7 m), 0.94 (RMSE = 5.4 m^2/ha), and 0.95 (RMSE = 73 m^3/ha), for height, basal area, and volume, respectively. Various percentile variables, e.g., the 90^{th} height percentile and 20^{th} coverage percentile, were shown to be significant predictor variables. The authors concluded that regression estimates were acceptable, even with the relatively coarse 10 m cell extraction method used to derive the ground elevation model.

Popescu *et al.* (2002; 2003) developed a method of varying kernel size by return height to find local maxima, i.e., the height for an individual tree. The study focused on pine-hardwood and pine stands in the Virginia Piedmont. R^2 values of 0.85 - 0.90 and 0.84 - 0.85 were obtained for maximum and dominant, and co-dominant tree height (dbh > 12.7 cm; RMSEs not reported), respectively. Holmgren *et al.* (2003) obtained R^2 values of 0.90 and 0.82 and RMSEs of 37 m^3/ha and 43 m^3/ha for stem volume estimation. Independent variables for two volume models

included laser mean heights, crown coverage area, and laser derived stem number. Values were derived for 10 m radius plots in southern Sweden (Norway spruce, *Picea abies*; Scots pine, *Pinus sylvestris*; birch species, *Betula* spp.). Lidar measurements, used in conjunction with a field sample, were effective for estimation of tree plot height and volume (Holmgren *et al.,* 2003).

Hudak *et al.* (2002) combined Landsat TM data for horizontally generalized forest classes with small footprint lidar data (0.6 m footprint) to estimate canopy height. Regression, kriging, co-kriging, and kriging and co-kriging of the regression residuals were evaluated over a variety of sampling intensities that ranged from 250 - 2000 m for transects. Aspatial regression models kept vegetation distribution patterns intact. However, taller canopies were underestimated and shorter canopy heights were overestimated. The use of spatial models resulted in less biased results, but failed to keep vegetation distribution patterns intact, especially at coarser sampling resolutions of > 1000 m. Integrated models performed best overall ($r \sim 0.94$; standard deviation = 5.27 m), with accurate estimation and reliable vegetation structure results. The authors concluded that these methods would be best for estimation of canopy heights from Landsat data at locations unsampled by lidar (Hudak *et al.*, 2002).

2.3.3.2.3 Stand-level Estimation of Forest Biophysical Parameters using Small-footprint Lidar Data

Magnussen and Boudewyn (1998) showed that the distribution of canopy heights over a Douglas-fir (*Pseudotsuga menziesii*) stand was a function of vertical distribution of foliage area. The proportion of laser pulses returned from a given height was proportional to the fraction of leaf area above it. This relationship was used to estimate mean stand height and a strong correlation ($R^2 = 0.8$; standard deviation = 2.2 m) was found between laser and field estimates. Næsset and Bjerknes (2001) used regression equations to predict mean height of young (< 6 m) Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) stands using small-footprint lidar. They recorded a bias of 0.23 m between actual and predicted mean stand heights, with a standard deviation of 0.56 m for the residuals between predicted and ground-truth data.

Douglas *et al.* (2003) used lidar canopy density and reflectance, derived from small footprint lidar data, to classify mature pine (15 plots), immature pine (14 plots), and mature hardwood (15 plots) stands in Mississippi. Analyses were limited to the forest canopy by only using lidar returns in the upper 50% of the total tree height. The authors implemented a discriminant classification with the number of lidar hits per cubic meter and the variance of the intensity data within the canopy of each plot as variables. Accuracies of 100% (mature pine), 85.7% (immature pine), and 93.3% (mature hardwood) were achieved, with an overall accuracy of 86.4%. Overall accuracy was 65.9% when only using the number of hits per cubic meter as independent variable. Such an approach bodes well for the application of lidar data, a structural data source due to its height information, to forest classification.

Results from previous lidar-based, volume- and biomass estimation studies bode well for extension of methods to a lidar distribution-based, object-oriented modeling approach. Not only do segment-based approaches to volume- and biomass modeling have potential based on previous work (Makela and Pekkarinen, 2001; Pekkarinen, 2002), but lidar distributions also have been shown to be effective in modeling attempts (Means *et al*., 2000; Næsset, 2002) and forest type classification (Douglas *et al*., 2003). The added benefits of an object-oriented approach is that (i) the extraction of estimates can be done on homogenous units extracted from the data set itself, (ii) the volume estimates can possibly be assigned to forest types through object-oriented classification approaches, (iii) estimate-by-types can be scaled through recombination of the hierarchical segmentation results. From previous results it seems likely that high precision of estimates will remain as principal challenge. Extension of lidar-based forest volume- and biomass estimation to operational application is dependent on this factor, as well as the cost associated with large-scale lidar data acquisitions.

2.4 Literature Cited

Abeyta A.M., and J. Franklin, 1998. The accuracy of vegetation stand boundaries derived from image segmentation in a desert environment. *Photogrammetric Engineering & Remote Sensing* 64 (1): 59-66.

Abkar A., M.A. Sharifi, and N.J. Mulder, 2000. Likelihood-based image segmentation and classification: A framework for the integration of expert knowledge in image classification procedures. *Journal of Algebraic Geometry* 2 (2): 16 pp.

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

Acton, S.T., 1996. On supervised segmentation of remotely sensed imagery using nonlinear regression. *International Journal of Remote Sensing* 17 (7): 1407-1415.

Avery, T.E., and H.E. Burkhart, 1994. Forest Measurements. 4th Edition. McGraw-Hill, Boston, USA. 408 p.

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

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

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

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

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

Blair, J.B., and M.A. Hofton, 1999. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters* 26 (16): 2509-2512.

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

Brandtberg, T., T.A. Warner, R.E. Landenberger, and J.B. McGraw, 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment* 85 (2003): 290-303.

Bressers P.G., and P.J. Van Oevelen, 1998. Erosion indicators assessment using microwave remote sensing. Operational remote sensing for sustainable development: Proceedings of the 18th EARSeL Symposium on Operational Remote Sensing for Sustainable Development, Enschede, Netherlands, 11-14 May 1998.

Burrough, P.A., and R.A. McDonnel, 1998. Principles of geographical information systems. Oxford University Press. Oxford, England. 333 pp.

Canters F., 1997. Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. *Photogrammetric Engineering & Remote Sensing* 63 (4): 403-414.

Cathcart, J.F., 2000. Carbon sequestration. *Journal of Forestry* 98 (9): 32-37.

Cheevasuvit F., 1990. Image segmentation based on graph theory. Proceedings of the 23rd International Symposium on Remote Sensing of Environment, Bangkok, Thailand, April 18-25, 1990. pp. 1329-1334.

Cole, W.G., and C.G. Lorimer, 1994. Predicting tree growth from crown variables in managed northern hardwood stands. *Forest Ecology and Management* 67: 159-175.

Dixon, R.K., S. Brown, R.A. Houghton, A.M. Solomon, M.C. Trexler, and J. Wisniewski, 1994. Carbon pools and flux of global forest ecosystems. *Science* 263: 186-190.

Darwish, A., K. Leukert, and W. Reinhardt, 2003. Image segmentation for the purpose of object-based classification. Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse 3 pp.

De Kok, R., T. Schneider, and U. Ammer, 1999. Object based classification and applications in the Alpine forest environment. In: Proc. Joint ISPRS/EARSeL Workshop "Fusion of sensor data, knowledge sources and algorithms", Valladolid, Spain, June 3-4, 1999. 7 pp.

Douglas, T.E., D.L. Evans, K.L. Belli, and S.D. Roberts, 2003. Classification of pine and hardwood by the distribution and intensity of lidar returns. ISPRS "Three dimensional mapping workshop from InSAR and LIDAR", June 17-19, 2003, Portland, Oregon, USA. 5pp.

Drake, J.B., and J.F. Weishampel, 1998. Multifractal analysis of laser altimeter and ground-based canopy height measures of a longleaf pine savanna. First International Conference on Geospatial Information in Agriculture and Forestry, Lake Buena Vista, Florida, June 1-3, 1998.

Drake, J.B., and J.F. Weishampel, 2001. Simulating vertical and horizontal multifractal patterns of a longleaf pine savanna. *Ecological Modeling* 145 (2001): 129–142.

Drake, J.B., R.O. Dubayah, D.B. Clark, R.G. Knox, J.B. Blair, M.A. Hofton, R.L. Chazdon, J.F. Weishampel, and S.D. Prince, 2002a. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment* 79 (2002): 305-319.

Drake, J.B., R.O. Dubayah, R.G. Knox , D.B. Clark, and J.B. Blair, 2002b. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment* 81 (2002): 378-392.

Dubayah, R.O., and J.B. Drake, 2000. Lidar remote sensing for forestry. *Journal of Forestry* 98 (6): 44-46.

Engdahl, M.E., J. Pulliainen, and M. Hallikainen, 2003. Combined land-cover classification and stem volume estimation using multitemporal ERS tandem INSAR data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Evans, D.L., S.D. Roberts, J.W. McCombs, and R.L. Harrington, 2001. Detection of regularly spaced targets in small-footprint lidar data: Research issues for consideration. *Photogrammetric Engineering & Remote Sensing* 67 (10): 1133-1136.

Evans C., R. Jones, I. Svalbe, and M. Berman, 2002. Segmenting multispectral Landsat TM images into field units. *IEEE Transactions on Geoscience and Remote Sensing* 40 (5): 1054-1064.

Flanders D., M. Hall-Beyer, and J. Pereverzoff, 2003. Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. *Canadian Journal of Remote Sensing* 29 (4): 441-452.

Franklin, S.E., 1994. Discrimination of subalpine forest species and canopy density using digital CASI, SPOT PLA and Landsat TM data. *Photogrammetric Engineering & Remote Sensing* 60 (10): 1233-1241.

Gomes A., and A.R.S. Marcal, 2003. Land cover revision through object based supervised classification of ASTER data. ASPRS annual conference, May 2003, Anchorage, Alaska. 9 pp.

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

Häme T., and E. Tomppo, 1987. Stand based forest inventory from SPOT image – first experiments. SPOT first in-flight results.

Harding D.J., M.A. Lefsky, G.G. Parker, and J.B. Blair, 2001. Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment* 76 (2001): 283-297.

Haswell III, W.T., 2000. Techniques for estimating forest carbon, *Journal of Forestry* 98 (9): 1-6.

Heyman, O., G.G. Gaston, A.J. Kimerling, and J.T. Campbell, 2003. A per-segment approach to improving Aspen mapping from high-resolution remote sensing imagery. *Journal of Forestry* (June, 2003): 29-33.

Hill, R.A., 1999. Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of Remote Sensing* 20 (5): 1039-1044.

Hodgson, M.E., J.R. Jensen, L. Schmidt, S. Schill, and B. Davis, 2003. An evaluation of LIDAR- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs. *Remote Sensing of Environment* 84 (2003): 295–308.

Hofton, M.A., J.B. Blair, J.B. Minster, J.R. Ridgeway, N.P. Williams, J.L. Bufton, and D.L. Rabine, 2000. An airborne scanning laser altimetry survey of Long Valley, California. *International Journal of Remote Sensing* 21 (12): 2413-2437.

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

Hoover, C.M., R.A. Birdsey, L.S. Heath, and S.L. Stout, 2000. How to estimate carbon sequestration on small forest tracts. *Journal of Forestry* 98 (9): 13-19.

Hudak A.T., M.A. Lefsky, W.B. Cohen, and M. Berterretche, 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment* 82: 397-416.

Hyyppä J., and M. Inkinen, 1999. Detecting and estimating attributes for single trees using laser scanner. *Photogrammetric Journal of Finland* 16: 27-42.

Jaakkola S.P., 1989. Applicability of SPOT for forest management. *Advanced Space Research* 9 (1): 135-141.

Jain, R., K. Rangachar, and B.G. Schunck, 1995. Machine vision. MIT Press and McGraw-Hill, Inc., New York. 549 p.

Kayitakire, F., C. Farcy, and P. Defourny, 2002. Ikonos-2 imagery potential for forest stands mapping. ForestSAT Symposium, Heriot Watt University, Edinburgh, August 5-9, 2002. 11 pp.

Kellndorfer, J.M., and F.T. Ulaby, 2003. Forest biomass inversion from SAR using object oriented image analysis techniques. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Kermad C.D., and K. Chehdi, 2002. Automatic image segmentation system through iterative edge-region co-operation. *Image and Vision Computing* 20: 541-555.

Kilpeläinen, P., and T. Tokola, 1999. Gain to be achieved from stand delineation in Landsat TM image-based estimates of stand volume. *Forest Ecology and Management* 124 (1999): 105-111.

Kressler, F., Y. Kim, and K. Steinnocher, 2003. Object-oriented land cover classification of panchromatic KOMPSAT-1 and SPOT-5 data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 7 pp.

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

Lefsky, M.A., W.B. Cohen, S.A. Acker, G.G. Parker, T.A. Spies, and D. Harding, 1999b. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir and western hemlock forests. *Remote Sensing of Environment* 70: 339-361.

Lefsky M.A., W.B. Cohen, A. Hudak, S.A. Acker, and J.L. Ohman, 1999c. Integration of lidar, Landsat ETM+, and forest inventory data for regional forest mapping. ISPRS workshop: Mapping forest structure and topography by airborne and spaceborne lasers. November 9-11, La Jolla, CA, USA. 7 pp.

Lefsky, M.A., W.B. Cohen, and T.A. Spies, 2001. An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research* 31 (2001): 78-87.

Lefsky M.A., W.B. Cohen, G.G. Parker, and D.J. Harding, 2002a. Lidar remote sensing for ecosystem studies. *Bioscience* 52 (1): 19-30.

Lefsky M.A., W.B. Cohen, D.J. Harding, G.G. Parker, S.A. Acker, and S.T. Gower, 2002b. Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology & Biogeography* 11 (2002): 393-399.

Li W., G.B. Benie, D.C. He, S. Wang, D. Ziou, and Q.H.J. Gwyn, 1999. Watershed-based hierarchical SAR image segmentation. *International Journal of Remote Sensing* 20 (17): 3377-3390.

Lobo A., K. Moloney, and N. Chiariello, 1998. Fine-scale mapping of a grassland from digitized aerial photography: An approach using image segmentation and discriminant analysis. *International Journal of Remote Sensing 19* (1): 65-84.

Maas, H.G., and G. Vosselman, 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (1999): 153-163.

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

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

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

Makela H., and A. Pekkarinen, 2001. Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing of Environment* 77:66-75.

McCormick N., and S. Folving, 1998. Monitoring European forest biodiversity at regional scales using satellite remote sensing. Assessment of biodiversity for improved forest planning, 283-289, P. Bachman *et al.* (eds.). Kluwer Academic Publishers, Netherlands.

Means, J.E., S.A. Acker, D.J. Harding, J.B. Blair, M.A. Lefsky, W.B. Cohen, M.E. Harmon, and 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.

Means, J.E., 2000. Comparison of large-footprint and small-footprint lidar systems: Design, capabilities and uses. Second International Conference on Geospatial Information in Agriculture and Forestry, Lake Buena Vista, Florida, January 10-12, 2000.

Means J.E., S.A. Acker, B.J. Fitt, M. Renslow, L. Emerson, and C.J. Hendrix, 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering & Remote Sensing* 66 (11): 1367-1371.

Murray, B.C., S.P. Prisley, R.A. Birdsey, and R.N. Sampson, 2000. Carbon sinks in the Kyoto Protocol - potential relevance for US forests. *Journal of Forestry* 98 (9): 6-11.

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

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80 (2002):88-99.

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

NASA Earth Science Enterprise Applications Strategy, 2002. http://www.earth.nasa.gov/visions/appstrat2002.pdf. January 2002. 20 pp.

NASA Webcast, Research Community Update, 2002. http://gaia.hq.nasa.gov/rcuwebcasts/esfaq.cfm. July 16, 2002.

NASA ESE Strategic Plan, 2000. http://www.earth.nasa.gov/visions/stratplan/ese_strategic_plan.pdf. November 2000. NASA Headquarters, Washington D.C., 20546. 47 pp.

Nelson, R.F., R.S. Latty, and G. Mott, 1984. Classifying northern forests using Thematic Mapper simulator data. *Photogrammetric Engineering & Remote Sensing* 50 (5): 607-617.

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

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

Nugroho, M., D.H. Hoekman, and R. De Kok, 2002a. Analysis of the forests spatial structure using SAR and Ikonos data. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 10 pp.

Nugroho, M., D.H. Hoekman, and R. de Kok, 2002b. Analysis of forest spatial structure using spatial decision rule. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 8 pp.

Pekkarinen A., 2002. Image segment-based spectral features in the estimation of timber volume. *Remote Sensing of Environment* 82: 349-359.

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 and Remote Sensing* 54 (1999): 95-104.

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

Popescu, S.C., R.H. Wynne, and R.F. Nelson, 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing* 29 (5): 564–577.

Popescu, S.C., R.H. Wynne, and J.A. Scrivani, 2004. Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, U.S.A. *Forest Science* (In press)

Riaño, D., E. Meierc, B. Allgöwerc, E. Chuviecoa, and S.L. Ustin, 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment* 86 (2003): 177–186.

Rodriguez-Yi, J.L., Y.E. Shimabukuro, and B.T.F. Rudorff, 2000. Image segmentation for classification of vegetation using NOAA AVHRR data. *International Journal of Remote Sensing* 21 (1): 167-172.

Russ J.C., 1995. The image processing handbook. CRC Press, Ann Arbor, pp. 674.

Ryherd S., and C. Woodcock, 1990. The use of texture in image segmentation for the definition of forest stand boundaries. Proceedings of the 23rd International Symposium on Remote Sensing of Environment, Bangkok, Thailand, April 18-25, 1990. pp. 1209-1213.

Ryherd S., and C. Woodcock, 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering & Remote Sensing* 62 (2): 181-194.

Sarkar A., M.K. Biswas, B. Kartikeyan, V. Kumar, K.L. Majumder, and D.K. Pal, 2002. A MRF model-based segmentation approach to classification for multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 40 (5): 1102-1113.

Savage, N., 1999. Lidar sensor sees forest and the trees. *Laser Focus World* 35 (5): 71-72.

Schlesinger, W.H., 1995. An overview of the carbon cycle, soils and global change (R. Lal, J. Kimble, E. Levine, and B.A. Stewart, editors), Lewis

Schwarz, M., C.H. Steinmeier, and L. Waser, 2001. Detection of storm losses in alpine forest areas by different methodic approaches using high-resolution satellite data. Proceedings of the 21st EARSeL Symposium, Paris. 7pp.

Shandley, J., J. Franklin, and T. White, 1996. Testing the Woodcock-Harward image segmentation algorithm in an area of southern California chaparral and woodland vegetation. *International Journal of Remote Sensing* 17 (5): 983-1004.

Shankar B.U., C.A. Murthy, and S.K. Pal, 1998. A new gray level based Hough transform for region extraction: An application to IRS images. *Pattern Recognition Letters* 19 (1998): 197-204.

Shen, S.S., G.D. Badwar, and J.G. Carnes, 1985. Separability of boreal forest species in the Lake Jennette area, Minnesota. *Photogrammetric Engineering and Remote Sensing* 51 (11): 1775-1783.

Shivers, B.D., and B.E. Borders, 1996. Sampling techniques for forest resource inventory. John Wiley and Sons, Inc. New York, USA. 356 p.

Smits P.C., and S.G. Dellepaine, 1997. An irregular region label model for multi-channel image segmentation. *Patter Recognition Letters* 18 (1997): 1133-1142.

Sun G., and K.J. Ranson, 2000. Modeling lidar returns from forest canopies. *IEEE Transactions on Geoscience and Remote Sensing* 38 (6): 2617-2626.

Ter-Mikaelian, M.T. and W.C. Parker, 2000. Estimating biomass of white spruce seedlings with vertical photo imagery. *New Forests* 20: 145-162.

Ton J., J. Sticklen, and A.K. Jain, 1991. Knowledge-based segmentation of Landsat images. *IEEE Transactions on Geoscience and Remote Sensing* 29 (2): 222-232.

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

Weishampel, J.F., K.J. Ranson, and D.J. Harding, 1996. Remote sensing of forest canopies. *Selbyana* 17: 6-14.

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.

White, J.D., C.K. Glenn, and J.E. Pinder III, 1995. Forest mapping at Lassen Volcanic National Park, California, using Landsat TM data and a geographical information system. *Photogrammetric Engineering and Remote Sensing* 61 (3): 299-305.

Willhauck, G., 2000. Comparison of object oriented classification techniques and standard image analysis for the use of change detection between SPOT multispectral satellite images and aerial photos. ISPRS XXXIII conference, Amsterdam, Netherlands, 2000. 8 pp.

Wilson R., and M. Spann, 1988. Image segmentation and uncertainty. Research Studies Press Ltd., Letchworth, Hertfordshire, England, 180 pp.

Woodcock, C.E., V. Jakabhazy, S. Macomber, S. Ryherd, A.H. Strahler, and Y. Wu, 1990. Timber inventory using Landsat Thematic Mapper imagery and canopy reflectance models. Proceedings of the 23rd International Symposium on Remote Sensing of Environment, Bangkok, Thailand, April 18-25, 1990. pp. 937-943.

Woodcock, C.E., J.B. Collins, S. Gopal, V.D. Jakabhazy, X. Li, S. Macomber, S. Ryherd, V.J. Harward, J. Levitan, Y. Wu, and R. Warbington, 1994. Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment* 50: 240-254.

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

CHAPTER 3

DECIDUOUS AND CONIFEROUS FOREST VOLUME AND BIOMASS ESTIMATION USING SMALL-FOOTPRINT LIDAR-DISTRIBUTIONAL PARAMETERS ON A PER-SEGMENT BASIS

Abstract. This study assessed the utility of a lidar-based, object-oriented approach to deciduous and coniferous volume and aboveground biomass modeling. The study area is located in Appomattox Buckingham State Forest in the Piedmont physiographic province of Virginia, U.S.A, at 78°41' W, 37°25' N. Vegetation is composed of various coniferous, deciduous, and mixed forest stands. The eCognition segmentation algorithm was used to derive objects from a lidar-based canopy height model (CHM). Between- and within-segment variances of the CHM were used to select segments (0.035 ha/segment - 3.942 ha/segment) for subsequent modeling efforts. Existing stands and the segment size most closely related to stands also were used to evaluate segmentation results vs. results from operational stands. Horizontal point samples were used to calculate in-field volume and above-ground biomass, for 2-class (deciduous-coniferous) and 3-class (deciduous-coniferous-mixed) forest classification schemes. Per-segment lidar distributional parameters, e.g., mean, range, percentiles, were extracted from small-footprint, multiple return lidar data. These parameters were used as input to volume and biomass regression analysis. Adjusted R^2 and Mallow's Cp metrics used to select volume and biomass models for the range of segmentation results. There was no evident difference between the two-class and three-class approaches, based on similar adjusted R^2 values. Segment-based modeling (2-class overall adjusted $R^2 = 0.52 – 0.59$) resulted in a distinct improvement over stand-based attempts (2-class overall adjusted $R^2 = 0.42$). Two-class adjusted R^2 and root mean square error (RMSE) values for deciduous volume (0.59; 51.15 m^3/ha) and biomass (0.58; 37.41 Mg/ha) were better than those found in another, plot-based study for the same study area. Coniferous R^2 values for volume (0.66) and biomass (0.59) were lower than other published results. The lower adjusted R^2 values for conifers were attributed to more variability within a narrower measured range (6.94 – 350.93 m^3/ha). Although not conclusive, smaller segment sizes generally performed better than larger segments, due to the lower within-segment variability. The precision of volume and above-ground biomass estimates, as a percentage of the estimate, was lower in the

case of modeled (40% - 41% and 43% - 46%, respectively) versus field-measured (59% and 69%, respectively) values. Lidar-based object-oriented volume- and biomass modeling have significant potential, especially when the ease of scalability and the benefit of a single remote sensing data source, coupled with field data, are considered.

3.1 Introduction

Forest volume and biomass have long been estimated using extensive in-field inventory methods or aerial photography. Although field-based methods are typically unbiased, both approaches are time consuming and expensive. Digital, large-scale remote sensing could provide a cheaper option to estimation of forest biophysical parameters over large tracts, while potentially also providing accurate and unbiased estimates. The structural nature of lidar height data makes it especially suitable for estimation of forest volume and biomass. Lidar makes the measurement of canopy structure in 3-D spatial distributions possible, and has an added benefit of ground-based digital elevation models (DEMs) derived from "last-return" lidar signals (Weishampel *et al.*, 1996). Lefsky *et al.* (2002a) stated that lidar sensors are able to provide accurate and non-asymptotic estimates of various forest indices such as LAI and aboveground biomass. Lidar systems could negate the need for ground-based, small scale measurements of tree heights and/or canopy parameters, and provide more automation and positional accuracy than photogrammetric techniques. Lidar-based forest measurements also are of importance to general forest inventory and biomass modeling (Lefsky *et al.,* 2002a; Lefsky *et al*., 2002b; Næsset, 2002; Popescu *et al*., 2004), estimation of forest fuel loads (Riaño *et al*., 2003; Seielstad and Queen, 2003), and derivation of DEMs (Popescu *et al*., 2002; Hodgson *et al*., 2003), applicable to forest management and site mapping.

Lidar measurement of forest parameters have been attempted in many studies. These range from large-footprint to small-footprint lidar, deciduous to coniferous species, and height to volume measurements. A brief summary of pertinent studies is given in Table 3.1. The species variability, range of observed values, R^2, and root mean square error (RMSE) values have been included.

Table 3.1 Previous studies that are particularly pertinent to this research (lidar for forest volume and biomass estimation)

Study	Species	Observed range	R^2	RMSE
Lefsky *et al.* (1999) Large footprint	Douglas-fir and western hemlock (*Pseudotsuga menziesii*, *Tsuga heterophylla*)	22 plots (0.25 ha/plot); Biomass: 157 – 965.2 Mg/ha Basal area: 16.2 – 91.6 m^2/ha	Total biomass, total basal area: 0.91 and 0.87	None given
Means *et al.* (1999) Large footprint	Douglas-fir and western hemlock	26 plots (0.00785 ha/plot); Biomass: 2 – 1,000 Mg/ha Basal area: 0 – 92 m^2/ha	Total biomass and basal area: 0.96	88 Mg/ha (total biomass), and 9 m^2/ha (basal area)
Lefsky *et al.* (2002b) Large footprint	Temperate coniferous (Douglas fir main species), temperate deciduous (yellow poplar main species; *Liriodendron tulipifera*), and boreal coniferous forests (black spruce main species; *Picea mariana*)	21, 112, and 16 plots, respectively (0.0625 ha/plot) Biomass: 135.6 – 1329 Mg/ha, 11.4 – 716.3 Mg/ha, 0 – 58.5 Mg/ha, respectively	0.84 was found for above-ground biomass	standard error = 7.5 Mg/ha
Maclean and Krabill (1986) Small footprint	Loblolly pine (*Pinus taeda*) and mixed hardwood stands (*Quercus* spp., *Liriodendron tulipifera, Liquidambar styracifula, Nyssa sylvatica*)	Contiguous plots along track, number not given (0.08 ha/plot); Range not given	0.72 - 0.92 (natural logarithm of volume)	None given
Nelson *et al.* (1988) Small footprint	Coastal pine species (*Pinus taeda, P. elliottii, P. echinata, P. palustris*)	113 plots (± 0.0625 ha/plot); even-aged species plots	0.43 – 0.53 (total tree volume) 0.44 – 0.55 (total tree green weight)	69.2 – 76 m^3/ha standard deviation (volume) 67.8 – 75.5 Mg/ha (total tree green weight)
* Means *et al.* (2000) Small footprint	Douglas-fir	19 plots (0.01 ha/plot) (shrub – old-growth) Volume: 18 – 2,051 m^3/ha	0.97 (volume; all plots) 0.95 (volume; old-growth plots excluded)	None given for all plots 73 m^3/ha (old-growth plots excluded)
* Næsset (2002) Small footprint	Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*)	144 plots (0.02 ha/plot); 61 reference stands (56) Young forest = 41 – 498.2 m^3/ha (36) Mature forest (poor site) = 59 – 280.1 m^3/ha (52) Mature forest (good site) = 54 – 639.8 m^3/ha	0.91 (reference stands)	18.3 – 31.9 m^3/ha
Holmgren *et al.* (2003) Small footprint	Norway spruce, Scots pine, and birch species (*Betula* spp.)	65 plots (0.03 ha/plot); Volume: 14 – 366 m^3/ha	0.82 - 0.90 (two volume models)	37 - 43 m^3/ha

* Precursory lidar distributional studies that are important the approach taken in this study

Large footprint lidar sensors have been used extensively for forest volume and biomass estimation (Lefsky *et al.,* 1999; Means *et al.,* 1999; Lefsky *et al.,* 2002b). Most of these studies have focused on biomass estimation, as the waveform nature of large-footprint lidar is suited to detection of aboveground vegetative matter (Lefsky *et al.,* 1999). However, the spatially discontinuous nature of large footprint lidar sensors could limit their applicability at local scales. Small footprint, continuous lidar measurements enable users to measure volume and biomass even for small tracts of forest. Where necessary, such measurement could be scaled through sampling to regional or even global levels.

Small footprint lidar systems were used as early as 1986 to 1988 for forest volume estimation (Maclean and Krabill, 1986; Nelson *et al.,* 1988). Small footprint lidar-volume studies have used both plot- and stand-based approaches (Nilsson, 1996; 2003; Popescu *et al*., 2004). Lidar height distributional approaches also have come to the fore (Means *et al*., 2000; Næsset, 2002). Analyses in these cases were based on distributional height metrics, e.g., mean, range, skewness, and percentiles. A lidar distributional approach to forest volume- and biomass modeling lends itself to segment or stand level application, and also simulates a waveform-type return found in the case of large footprint lidar sensors. Such a pseudo-waveform has the benefit of vertical forest structure characterization, a feature which is potentially valuable for large scale, stand-level volume- and biomass estimation (Magnussen and Boudewyn, 1998; Means et al., 2000; Drake et al., 2002; Lefksy et al., 2002b; Næsset, 2002).

Magnussen and Boudewyn (1998) showed that the distribution of canopy heights over a Douglas-fir (*Pseudotsuga menziesii*) stand was a function of vertical distribution of foliage area. The proportion of laser pulses returned from a given height was proportional to the fraction of leaf area above it. This relationship was used to estimate mean stand height and a strong correlation (R^2 = 0.8; standard deviation = 2.2 m) was found between laser and field estimates. Such a result corroborated the usefulness of a distributional approach in characterizing vertical structure.

Means *et al.* (2000) implemented a lidar distributional approach to estimate height and basal area for Douglas-fir stands, ranging from shrub-like (18 m^3/ha) to old-growth (1313 – 2051 m^3/ha)

stands. Lidar returns were extracted from 10 x 10 m grid cells within larger 50 x 50 m measured plots. Distributional parameters, e.g., canopy cover percentiles, maximum height, elevation, average mean height, and average of the maximum heights were calculated for grid cells. Stepwise regression analysis was used to determine the relationships between ground data and lidar measurements, with dependent variables being height, basal area, and volume. R^2 values of 0.93 (RMSE = 3.4 m), 0.95, and 0.97 (no RMSEs for latter two values) were obtained for height, basal area, and volume, respectively. R^2 values for plots excluding old-growth plots were 0.98 (RMSE = 1.7 m), 0.94 (RMSE = 5.4 m^2/ha), and 0.95 (RMSE = 73 m^3/ha), for height, basal area, and volume, respectively. Various percentile variables, e.g., the 90^{th} height percentile and 20^{th} coverage percentile, were shown to be significant predictor variables. Næsset (2002) predicted volume and crown parameters for Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) stands in Norway, using a stratum-specific (young forests; old-growth, on poor and good sites) approach. Observed volume values ranged between 41 m^3/ha and 639.8 m^3/ha. Lidar first- and last pulse distribution based regression equations were used to model volume and crown parameters. Various quantiles, maximum- and mean values, canopy density measures and coefficients of variation were used as independent variables. R^2 values for 61 reference stands were 0.87 (dominant height) and 0.91 (volume). Standard deviations ranged between 0.7 – 1.33 m (dominant height) and 18.3 – 31.9 m^3/ha (volume).

Extension of distributional grid-cell approaches to object- or segment- and stand-level applications was a logical next step. An object refers to a spatial entity that is homogenous in terms of a selected property, as opposed to the traditional, continuous fields approach found in spatial analysis (Burrough and McDonnel, 1998). Segments can be treated as entities or objects, since each segment is homogenous in terms of a defined variable. Such an application required that distinct forest cover and structural types have different, unique lidar canopy densities or distributions (Douglas *et al.,* 2003), and that segment-level estimate errors could be minimized (Makela and Pekkarinen, 2001; Pekkarinen, 2002). Segment-based modeling not only extends a grid- or plot-level approach, but also is amenable to stand-level scaling, since segments can match existing structural boundaries in forests. However, scaling efforts assume that segments are hierarchical and topologically sound, i.e., smaller level segments are exact constituents of larger level segments.

Of particular interest to this study are two Finnish attempts at "segment-aided" timber volume estimation. The first by Makela and Pekkarinen (2001) attempted a Landsat TM plot-level volume-by-species estimation. The authors used a measurement space-guided clustering, defined as an ISODATA classification followed by connected component labeling (ISOCCL), which in turn was based on edge detection and linking. A directed trees approach, based on gradient analysis and edge detection, was applied as an alternative segmentation method. Spectral features used for volume estimation were extracted in two ways, namely (i) from a fixed window around the field sample plot, and (ii) from the pixels in the fixed window that belonged to the same segment as the sample plot. The ISOCCL approach yielded the most accurate volume estimation for pine and spruce species (*Pinus sylvestris* and *Picea abies*), as well as for the total volume. The directed trees algorithm had the best results for broad-leaf species (*Betula pendula*, *B. pubescens,* and *Populus tremula*). Improvements from the fixed window approach to segment-based approaches were relatively small (1%-11.3% in relative RMSEs), and the RMSE values remained high. The authors did conclude, however, that the segmentation of forest areas was suitable for the estimation of forest variables. Errors could be minimized by extracting estimates from more homogenous segments (Makela and Pekkarinen, 2001).

The second study by Pekkarinen (2002) was done as a follow-up to investigate the large stand-level errors. Departing from the premise that these large errors are related to the limited spatial resolution of sensors such as Landsat TM, the author investigated the use of very high resolution images for image-based multi-source forest inventory (MSFI). The author concluded that the segmentation algorithm succeeded in delineating distinct forest areas for feature extraction, and performed better than reference data, extracted from square-shaped windows. Segmentation resulted in a decrease of 10% in the case of broad-leaved species RMSEs, as well as decreases in the case of spruce species ($\leq$ 8%), while plot-level errors in general remained high. The author again suggested that these large errors could be due to the local nature of field data and that segment-level data could decrease associated segment-level errors (Pekkarinen, 2002). Such studies on segment-based estimates, coupled with grid-cell, lidar distributional volume- and biomass modeling, therefore hinted at the potential of a segment-based distributional approach.

The basic precept of this study was that extraction of lidar distributions on a grid-cell basis, used to model volume and biomass (Means *et al.*, 2000; Næsset, 2002), could be extended to estimation at the segment or forest stand level. The specific objective of this study therefore is to determine whether volume and aboveground biomass can be successfully estimated using object-oriented analysis of lidar distributions. The success Means *et al.* (2000) and Næsset (2002) had with their cell-distribution approaches bode well for the methodology in this study, since distributions were useful in the estimation of heights, basal area, and volume for coniferous species. Added benefits of the proposed methodology include extraction of estimates from homogenous, scalable, and operational units derived from the lidar data itself, and possible reductions in RMSE due to the use of segment-level data (Pekkarinen, 2002).

3.2 Methods

3.2.1 Study Area

The 946 ha (2,338 acres) study area is located in Appomattox Buckingham State Forest (Appomattox County) in the Piedmont physiographic province of Virginia, southeastern U.S.A at 78°41' W, 37°25' N (Figure 3.1). The mean elevation of the study area is 185 m (606 ft.), with minimum and maximum elevations of 133 m (436 ft.) and 225 m (738 ft.), respectively. Local topography can best be described as gentle rolling slopes and flat terrain. Vegetation is composed of various coniferous (*Pinus taeda, P. virginiana, P. echinata*, and *P. strobus*), deciduous (*Quercus coccinea, Q. alba,* and *Liriodendron tulipifera*), and mixed forest stands (a BAF plot-based description of the forest characteristics is given in Table 3.2).

3.2.2 Available Data

Lidar data were acquired by Spectrum Mapping, LLC using the DATIS II (small-footprint, high-density, multiple return) system. The lidar data were acquired on September 9, 2002, centered at 78°40'30" W, 37°25'9" N, and covered an area of approximately 958 ha (2,367 acres). Specifications of the lidar data set are given in Table 3.3.

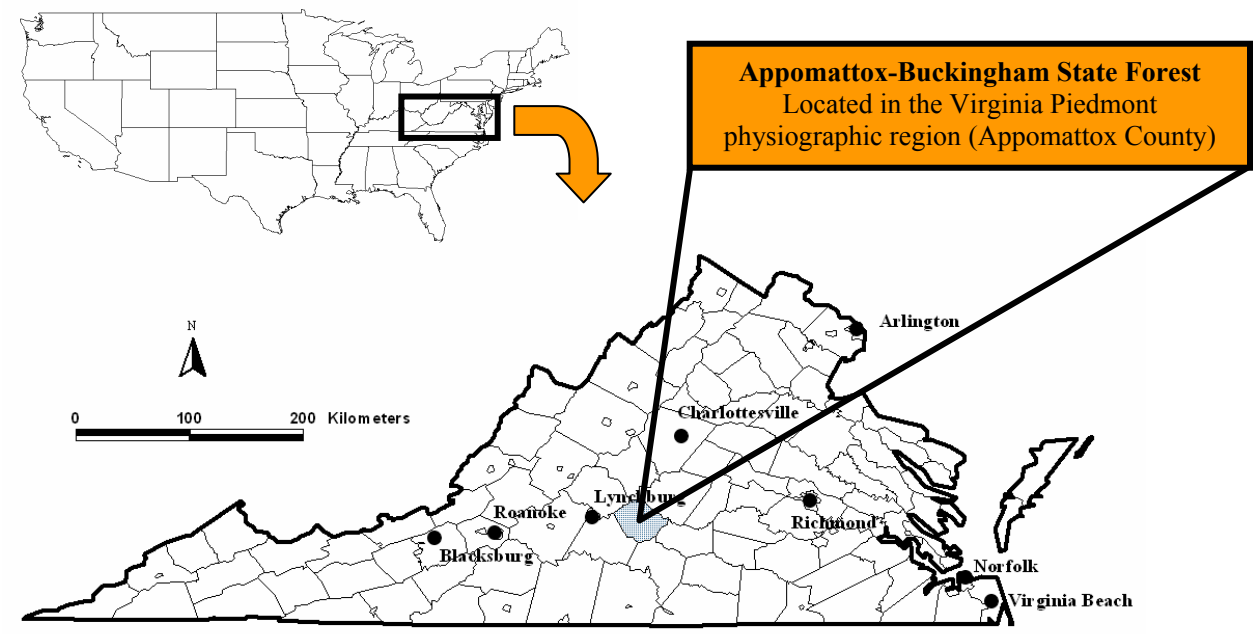


Figure 3.1 Study Area: Appomattox Buckingham State Forest

Table 3.2 General descriptive information for deciduous, coniferous, and mixed plots

Class	Type	Parameter	Minimum	Maximum	Average	σ
2-class	Deciduous plots (140)	Volume/ha (m^3/ha)	6.94	350.65	157.64	84.14
		Biomass/ha (Mg/ha)	11.11	269.01	113.60	58.60
		Basal area/ha (m^2/ha)	2.30	34.44	16.32	7.84
	Coniferous plots (79)	Volume/ha (m^3/ha)	8.32	350.93	114.49	75.44
		Biomass/ha (Mg/ha)	4.67	155.56	41.47	26.64
		Basal area/ha (m^2/ha)	2.30	36.73	14.24	7.91
3-class	Deciduous plots (112)	Volume/ha (m^3/ha)	6.94	350.65	156.16	89.32
		Biomass/ha (Mg/ha)	11.11	269.01	117.31	62.53
		Basal area/ha (m^2/ha)	2.30	34.44	15.97	8.21
	Coniferous plots (56)	Volume/ha (m^3/ha)	8.32	278.99	100.45	66.42
		Biomass/ha (Mg/ha)	4.67	81.65	33.66	19.95
		Basal area/ha (m^2/ha)	2.30	36.73	13.61	8.11
	Mixed plots (51)	Volume/ha (m^3/ha)	31.68	350.93	156.85	72.60
		Biomass/ha (Mg/ha)	20.06	175.75	81.49	38.93
		Basal area/ha (m^2/ha)	4.59	36.73	16.84	6.68

Table 3.3 DATIS II lidar data set characteristics

Characteristic	Specifications
Laser altitude	2,000 m (6,562 ft.) above ground level
Laser scan field-of-view	75° maximum
Swath width and centerline spacing	800 m (2,625 ft.) and 400 m (1,312 ft.)
Scan rate	25 Hz
Laser pulse rate	35 kHz
Scan angle	± 13.5°
Returns	≤ 5
Resolvable distance between returns	0.75 m
Footprint	0.46 m (1.51 ft.)
Spacing across / along track	1 m (3.3 ft.) / 2 m (6.6 ft.)
Accuracy (X,Y,Z)	X,Y: 0.5 m; Z: 0.15 m (X,Y: < 1.6 ft.; Z: < 0.49 ft.)
Post-processed GPS accuracy	< 0.05 m
Wavelength	1,064 nm

Field data consisted of 256 mapped basal area plots (BAF; basal area factor 10) on a 16 columns by 16 rows, 201.17 m (10 chains) grid. Field data were collected during the summer, fall, and winter months (May – December) of 2003. A Magellan Sportrak Pro GPS unit (WAAS enabled) was used to navigate to within 2 meters of each designated plot center. A Corvallis Microtechnologies, Inc. (CMT) March II GPS unit was used to accurately map the established plot center (120 second static point collection). All GPS plot center locations were differentially corrected using data from the National Geodetic Survey's Continually Operating Reference Stations (CORS, 2000) and Corvallis Microtechnologies, Inc. PC-GPS software (Version 3.7; Corvallis Microtechnologies, Inc.). The following reference stations from the CORS-network were used based on data availability:

- Richmond, VA (37° 32' 16.42936" N; 77° 25' 46.77568" W)
- Fan Mountain, VA (37° 52' 43.46536" N; 78° 41' 37.24955" W)
- Blacksburg, VA (37° 12' 21.63726" N; 80° 24' 52.27622" W)

For each sampling point, the following data were collected (**Appendix A**, an actual data sheet):

- Plot basal area ("in-tree" count); diameter at breast height (dbh) > 5 inches (12.7 cm) (10-factor prism)
- Dbh and height for all plot trees tallied (diameter tape and Vertex hypsometer)

- Azimuth and distance from plot center to each tallied tree (SUUNTO compass and Vertex hypsometer's range finding function)
- Species codes of tallied trees (**Appendix B**)
- Differentially corrected GPS point at plot center (CMT's March II GPS unit)

A total of 37 BAF non-forest plots had to be discarded due to their location on private land or having volume and biomass values of zero. Zero-value plots could not be used in subsequent model development since it was impossible to assign a forest type in such cases, with no trees being tallied. This left a total of 219 BAF plots (Figure 3.2) that were used in the statistical analysis (**Appendix C**).

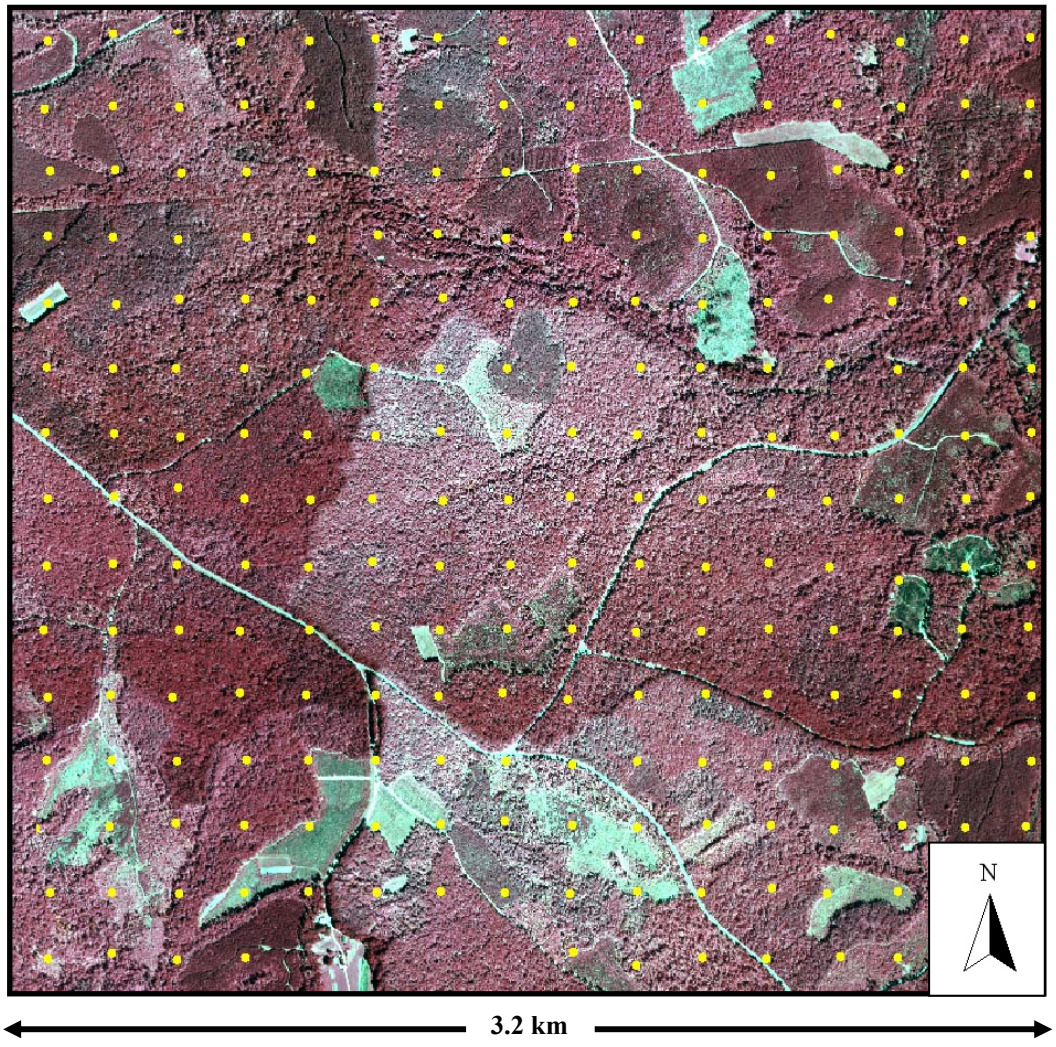


Figure 3.2 Mapped BAF plot locations on a 1999, leaf-on, color-infrared aerial photograph of the study area (bottom-middle plots missing plots due to locations on private land)

Each basal area plot was mapped based on plot center coordinates, azimuths, and distances to tallied trees. Plots were assigned to 2- and 3-class forest type schemes based on basal area percentages. "Deciduous" or "Coniferous" types were defined as plots that had 50% or greater basal area contribution from either deciduous or coniferous species, respectively. A "Mixed" class was added to the 3-class type designation for plots that had less than 90% basal area contribution for either deciduous or coniferous species. A 90% cut-off was based on sample numbers for the 2- and 3-class schemes. The 2-class analysis consisted of 140 deciduous and 79 coniferous plots, while the 3-class analysis consisted of 112 deciduous, 56 coniferous, and 51 mixed plots. This allowed for volume and biomass model development based on adequate plot samples (> 30) for both the 2- and 3-class analysis. Only 25 (11.4%) of the plots were mixed when a 75% cut-off was used, making this class redundant and too small for viable statistical analysis.

Since field plot data intended for model development and validation were used on a per-segment basis, the BAF plots were expanded to a per-hectare basis for each segment. This was done using standard BAF expansion equations. Basal area plot estimates also were tallied to derive total volume and biomass for the entire study area:

$$\text{BA/hectare} = \frac{\sum Trees\ Tallied}{\sum Samples} * BAF * Metric\ Conversion\ Factor_1 \quad \ldots[1]$$

$$\text{Volume/hectare (or Biomass)} = \frac{\sum (VBAR * 10)}{\sum Samples} * Metric\ Conversion\ Factor_2 \quad \ldots[2]$$

$$\text{Total Volume } or \text{ Biomass} = \frac{\sum_1^n Volume / Hectare}{\sum Samples} * \text{Hectares} \quad \ldots[3]$$

where

BA/hectare = Basal area per hectare for each segment ($m^2.ha^{-1}$)

BAF = Basal Area Factor (10, in this case)

$$\text{Metric Conversion Factor}_1 = \frac{0.092903 \longrightarrow (ft^2 \text{ to } m^2)}{0.4046856 \longrightarrow (\text{acre to hectare})}$$

Volume or Biomass/hectare = Volume ($m^3.ha^{-1}$) or Biomass ($kg.ha^{-1}$) per hectare for each segment

VBAR = Volume-Basal-Area Ratio

$$= \frac{Volume}{\pi\left(\frac{D}{2}\right)^2 * \frac{1}{144}}$$ (dbh in inches; substitute Biomass for Volume for Biomass per acre calculations)

Metric Conversion Factor $_2$ $= \frac{1}{0.4046856} \longrightarrow$ (acre to hectare)

Total Volume *or* Biomass = Total for volume (m^3) or biomass (kg) for the study area

n = Number of segments

(Avery and Burkhart, 1994)

The same volume and biomass equations (Saucier and Clark, 1985; Clark *et al*., 1986; Schroeder *et al*., 1997; Sharma and Oderwald, 2001) for per-tree calculations found in Popescu *et al.* (2004) were used in this study and are shown in **Appendix D**. Popescu *et al.* (2004) and this study were situated within the same geographical boundaries, with the same species being studied. Specific volume and biomass equations were used for loblolly and other southern pines, as well as for hardwoods. Volume and biomass were calculated on an individual tree basis for each plot and expanded to per-hectare values for each segment using equation [2]. Plots were assigned to the segment in which they were located. This was done through post-stratification for selected segmentation results. BAF plot values were averaged on a per-segment basis in the case of larger segments that contained more than one field plot. Segments without BAF plots were excluded from the model development process, but not from the prediction part of this study. Descriptive statistics for all basal area plots are given in Table 3.2.

3.2.3 Lidar Data Processing

A canopy height model (CHM) was needed for segment derivation as a precursor to per-segment volume- and biomass modeling. First and last (vegetation-removed) returns from the lidar data set were extracted and corrected for possible errors (suspect low and high, or "bird" hits).

Peripheral outlier height values with a low frequency and a distinct difference (> 6 m) from the next smallest or largest value were removed as outliers. This resulted in the removal of one return smaller than –75 m and six returns larger than 31 m. It was decided that large values likely represented "bird-hits", while outlying small values were due to possible lidar error. The first returns were median-filtered by 1 m grid cells in order to remove per-cell values that were redundant to subsequent interpolation procedures. First and last returns were interpolated to a 1 m spatial resolution grid using regular Kriging, since Popescu *et al.* (2002) found this to be the most accurate interpolation technique using similar data over the same study area. This approach effectively addressed instances where a 1 m grid cell lacked an original input value. The resultant 1 m resolution was detailed enough to detect road and stand breaks in the segmentation process. It had the additional benefits of requiring less computing power, as opposed to a 0.5 m grid, and likely produced a smoother canopy digital terrain model (DTM) and ground digital elevation model (DEM). Interpolation was performed using Surfer 7.0 software (Golden Software, Inc.). It was assumed that first return data is representative of top-most canopy heights, while last, vegetation–removed returns were attributed to ground hits. The differenced first- and last return surface (CHM) was used as input to the eCognition segmentation algorithm. This allowed for extraction of forest segments based on height homogeneity and distinct stand breaks, e.g., roads and slope breaks.

The distributional modeling approach, based on height distributional parameters, required that lidar data be processed on a per-return basis in order to retain information related to the return hierarchy. Peripheral outlier height values again were removed for all return data sets, based on the same approach as in the case of the CHM. Ground hits were removed using Terrascan V. 003.002 (Terrasolid, Inc.) and MicroStation V. 08.00.04.01 (Bentley Systems, Inc.) software. This algorithm identifies ground hits based on iterative slope analysis of lidar returns. *Grid cell size* and *maximum slope of the area* are required input parameters. *Grid cell size* is the smallest cell size for which a ground return can be extracted. A cell size of 10 m was used in order to extract a maximum number of ground returns for the first (31,294,660), second (11,101,215) and third (2,121,989) returns. Grid cell sizes of 39 m and 119 m were used for the fourth (175,093) and fifth (5,379) returns, respectively. Larger grid cell sizes were required for the last two categories due to the small number of returns in each case. Each of these two cell sizes resulted

from cases where the number of ground hits reached a maximum for the fourth and fifth returns, based on the assumption that most of the hits from these categories would be ground hits due their ranking in the return hierarchy. A slope percentage parameter of 35% was used as a maximum for the area, obtained from a USGS DEM. Ground returns constitute an important component of overall lidar distributional patterns and were retained as data sets on a per-return basis.

Non-ground hits, designated as vegetation hits, were normalized for terrain by calculating the actual return height above a lidar-derived 1 m digital elevation model (DEM) of the study area. The actual height of each vegetation hit was calculated as the difference between the vegetation hit and the bilinear interpolated height of the four corner cells of the DEM cell directly beneath each hit. This was done using Surfer V. 8.1 software (Golden Software, Inc.). This process normalized all vegetation hits for varying terrain elevations, thereby enabling volume and biomass models to incorporate actual lidar point heights (Means *et al*., 2000).

3.2.4 Segmentation of the Study area

Segmentation was performed using a multiresolution, hierarchical algorithm (eCognition) applied to the lidar-derived CHM of the study area. Lidar data were considered a structural component, ideally suited to defining unique structural segments. The eCognition algorithm required Color:Shape and Smoothness:Compactness ratios as input parameters. The Color:Shape ratio was set at 0.8:0.2, based on the recommendation of the developers (Baatz and Schäpe, 2000; eCognition, 2003) and evaluation of alternative parameter inputs. Smoothness of shape was considered more important than shape in a forestry context, since smooth, boundary-following segments are preferable to compact, blocky segments. The Smoothness:Compactness weight combination therefore was set at 0.8:0.2.

Although eCognition was chosen as the preferred segmentation approach, one could argue that the segmentation method is subordinate in importance to the utility that resultant objects have to analyses. Even though a multitude of segmentation approaches exist in literature, e.g., the Woodcock-Harward (centroid linkage) algorithm (Shandley *et al.*, 1996), a Hough transform-

based approach (Shankar *et al.*, 1998), and watershed-based hierarchical segmentation (Li *et al.*, 1999), it is ultimately of great importance that segmentation results are robust. Other important factors are ease of operational use, widespread availability, and adequate software support. eCognition also was the preferred approach to segmentation because of its hierarchical nature, correspondence to input data, and because results from this algorithm have been proven in the natural resources context (Kayitakire *et al.*, 2002; Nugroho *et al.*, 2002a; Nugroho *et al.* 2002b; Engdahl *et al.*, 2003; Kellndorfer and Ulaby, 2003; Kressler *et al.*, 2003).

The decision of which segmentation results to use for model development was based on between- and within-segment variability of the CHM. Models were fitted to segmentation results where within-segment variability was smaller than between-segment variability, so as to minimize within-segment variability. The smallest selected segment size corresponded to the circular plot area, with radii defined by average tallied tree distance from field-collected BAF plot centers, plus one and two standard deviations. This ensured that segments were representative of plot-level field data, based on corresponding areas. Ten average segment sizes, ranging from 0.035 ha/segment to 3.942 ha/segment, were chosen for subsequent volume and biomass model development. These selections corresponded to segment sizes where within-segment variance was smaller than between-segment variance of the CHM heights. This was done in order to evaluate model performance across a range of average segment sizes. The current Appomattox stand map (167 segments; 5.666 ha/stand) also was selected, as well as the segmentation result that corresponded to the number of operational stands (168 segments; 5.632 ha/segment). Operational stands were used in order to compare segmentation-based modeling to stand-based modeling. Figure 3.3 shows the segmentation result of 6,687 segments (average distance from plot center plus 2 standard deviations) overlaid on the CHM of the study area. Vegetation and ground lidar data sets were extracted on a per-segment basis for all segmentation results using ARCGIS V. 8.3 software (ESRI). Resultant data sets were exported to SAS V. 8.02 software (Level 02M0; SAS, Inc.) for subsequent regression analysis.

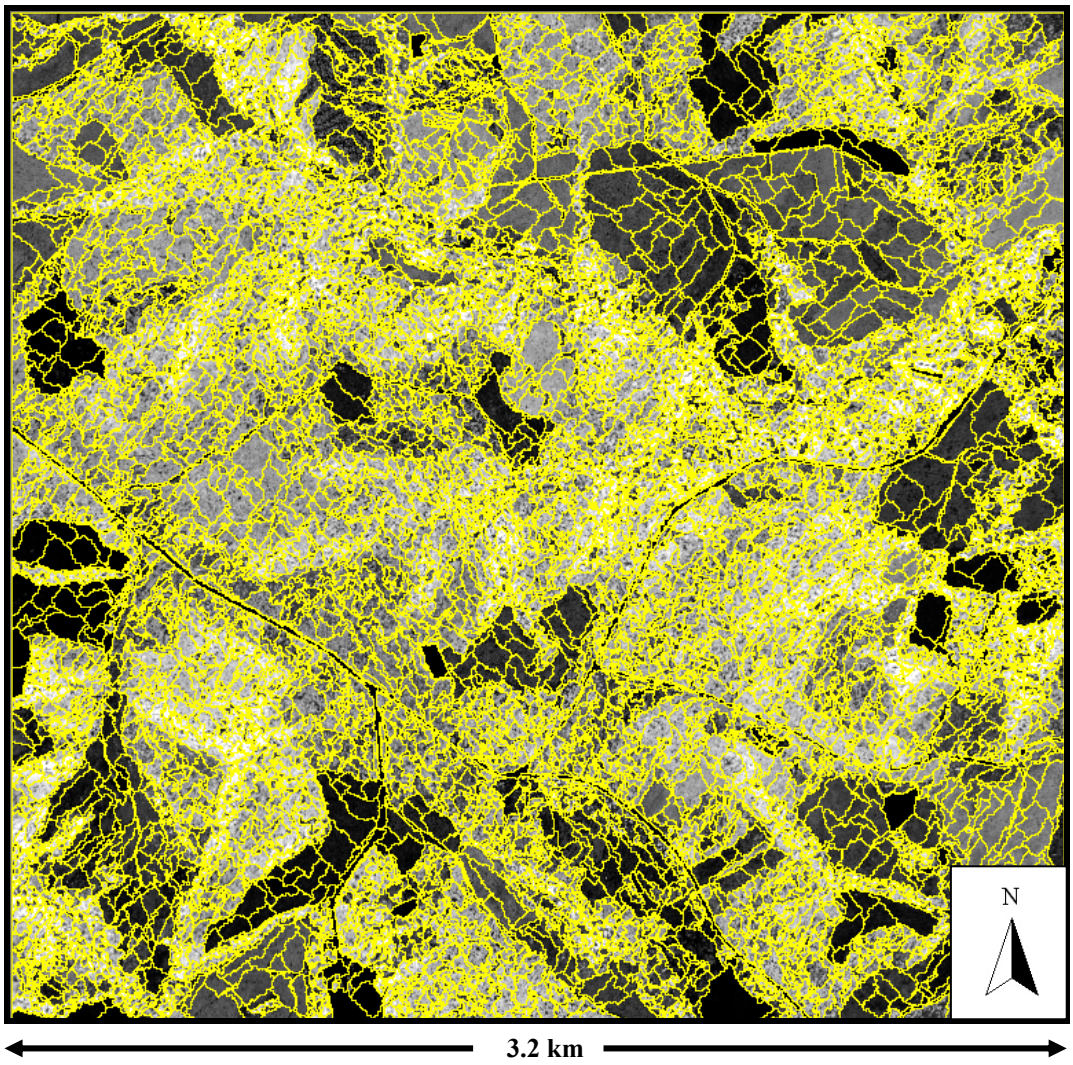


Figure 3.3 Segmentation results for 6,687 segments (0.141 ha/segment) overlaid on the canopy height model of the study area (946 ha)

3.2.5 Regression Analysis

The intent of this study was to extend lidar distributional, grid-cell forest volume and biomass modeling (Means *et al.*, 2000; Næsset, 2002) to estimation at the segment or forest stand level. It was assumed that distinct forest cover and structural types have different, unique canopy densities or distributions (Douglas *et al.*, 2003). These distributions could be characterized through construction of height distributions for vegetation returns per-segment using small-footprint DATIS II lidar data (Table 3.3). Lidar distributions should be representative of stand

structural characteristics such as canopy closure and stand height distribution. Theoretically, distributions from whole segments should approximate waveform lidar data for that segment, which in turn gives an indication of vertical vegetation distribution, a property closely related to biomass (Magnussen and Boudewyn, 1998; Means *et al.*, 2000; Næsset, 2002). Intermediate return distributions could also be useful, since these returns represent forest structure. A multi-tiered forest structure will theoretically have many intermediate returns, while an even-aged, weed-controlled, and thinned pine stand might exhibit a majority of first and last returns, with few intermediate hits.

Distributional parameters were derived for vegetation return data sets. Only first and second return variables from vegetation return data sets were used because of segments with missing values for the third through fifth returns. Distributional parameters included the mean, coefficient of variation, kurtosis, maximum, minimum, mode, range, standard error of the mean, skewness, standard deviation, number of observations, height percentile points at 10% intervals of height values, and canopy cover percentiles. Canopy cover percentiles were based on the proportion of first returns smaller than a given percentage of maximum height. The ratio of the number of vegetation or ground hits and the total number of lidar hits per segment also was calculated. This was done for second, and third through fifth group vegetation hits, as well as first, second, and third through fifth group ground hits. The vegetation ratio for each segment was calculated as the ratio of the number of vegetation hits per segment and the total hits for that segment. These distribution metrics have been shown to be useful descriptors of tree volume for 10x10 m grid cells in Douglas-fir, western Oregon stands (Means *et al.*, 2000) and 200 m^2 sample plots in Norway spruce and Scots pine stands in southeast Norway (Næsset, 2002). Lidar intensity (Figure 3.4) distributional parameters values for the first and second returns included the intensity mean, median, coefficient of variation, maximum, minimum, range, standard error of the mean, and standard deviation.

Linear regression analysis was performed using volume and biomass as dependent variables. Independent variables were reduced by using a forward selection process with α-values set between 0.075 and 0.350 as significance levels for remaining in the model. The goal was to

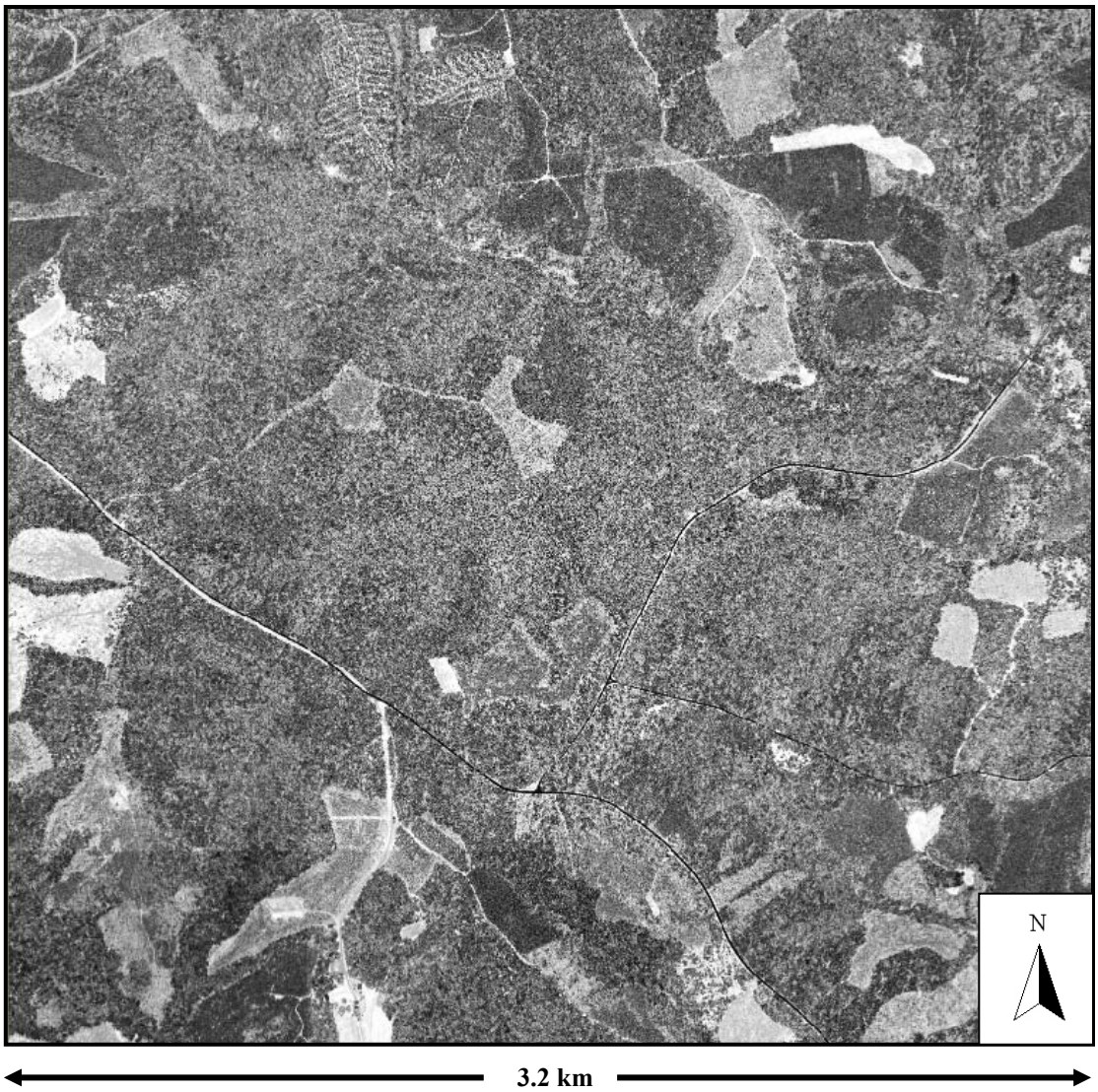


Figure 3.4 Lidar 1st return intensity image. Brighter tones are indicative of higher intensities

reduce independent variables from 75 initial variables to fewer than 10 variables in all cases. Forward selection was chosen over the stepwise selection used by Means *et al*. (2000), because forward selection retains all significant predictor variables, whereas stepwise selection discards variables that become less significant as more variables are added. This allowed for more user flexibility in the selection of final significant variables. Variables were validated through evaluation of Pearson's correlation coefficients between independent and dependent variables. All variables with correlations of 0.8 or lower were retained. However, only the variable with the highest correlation to the dependent variable was retained in cases where independent variable correlations were higher than 0.8. A value of 0.8 was chosen based on data characteristics, with

the knowledge that all lidar-derived variables are height-related, and hence some correlation was to be expected. This value resulted in adequate variables remaining in volume and biomass models as predictors, while high correlations ($r > 0.9$) were eliminated. These methods were crucial in order to avoid overfitting and invalid models in the final regression step, linear regression using Mallow's Cp and adjusted R^2 as selection criteria.

Mallow's Cp selection takes all combinations of independent variables into account, while calculating a value related to the mean square error of a fitted value for all models (Draper and Smith, 1981; Montgomery *et al*., 2001). Approximately ten or fewer candidate models from many recombination possibilities were selected for each model based on Mallow's Cp and adjusted R^2 values, as well as number of independent variables. A natural break in Cp values was found in many cases, while adjusted R^2 values and variable numbers were used to define a set of candidate models in other instances. Candidate models were evaluated based on the minimum Cp or where the Cp value equaled the number of initial independent variables that were entered into the Cp selection process (Draper and Smith, 1981; Montgomery *et al*., 2001). Candidate models also were restricted to the higher adjusted R^2 values. Selected candidate models ultimately were very similar, due to the recombination of variables in the Mallow's Cp regression procedure. A variety of fit criteria were used to select the best option from a set of candidate models. Although Mallow's Cp and adjusted R^2 values alone were valid fit criteria, root mean square errors (where applicable), model simplicity, and model validity also were considered. A compromise ultimately was required in the case of all selection criteria. While Mallow's Cp and adjusted R^2 values gave a good indication of "best-fit" models, model simplicity also was considered very important. Cases with a very slight increase in Cp values (< 1 unit) and decrease in adjusted R^2 values (~ 0.01), with the benefit of one or two fewer independent variables, were considered simpler with marginal sacrifice in fit statistics. Lastly, models with less abstract independent variables, e.g., range, mean, and max values, were favored over models with variables related to standard error of the mean, coefficient of variation, standard deviation, and the like.

Regression analyses were performed for segmentation results of 0.035 ha/segment, 0.091 ha/segment, 0.141 ha/segment, 0.318 ha/segment, 0.642 ha/segment, 0.964 ha/segment, 1.263 ha/segment, 1.885 ha/segment, 2.53 ha/segment, 3.942 ha/segment, 5.632 ha/segment, and the Appomattox Forest stands (5.666 ha/segment). Analyses were applied to 2- and 3-class schemes,

as well as for all segments combined. In the first case, models were fitted to "Deciduous" and "Coniferous" groups. Deciduous segments numbered from 61 to 140 segments, while coniferous segments ranged between 34 and 79 segments, depending on the average segment size and number of BAF plots that were averaged for larger segment sizes. In the second case, analyses were performed on "Deciduous" (43 - 112 segments), "Coniferous" (22 - 56 segments), and "Mixed" (30 - 51 segments) classes. Three-class analysis was based on segments that had less than 90% basal area representation by both deciduous and coniferous species, with such plots assigned to a "Mixed" class. Adjusted R^2 and RMSE values were calculated for all segmentation results in order to evaluate model performance across a range of average segment sizes. Regression analyses were limited to segments with non-missing values for distributional parameters included in Mallow's Cp regression selection. The only case with missing distributional parameter values were for 27,050 segments (0.035 ha/segment; 2-class: 124/140 deciduous and 70/79 coniferous segments; 3-class: 98/112 deciduous and 50/51 mixed segments).

Combined-class volume and biomass models were applied to the 27,050 segments (0.035ha/segment) and 6,687 segments (0.141 ha/segment) to estimate standing volume and biomass for all segments in the study area. Model estimates were adjusted for cases where segments had missing independent variables, and hence missing volume or biomass estimates. Such cases mainly could be attributed to small segments, or outlier segments with non-representative height distributions. Areas of segments with missing variables were tallied and the model totals adjusted using average volume and biomass per hectare values from all modeled segments. These results were compared to estimates obtained from the BAF field plots to gauge applicability of the developed models in operational conditions. Estimates derived from the two different methods were quantitatively compared using each estimate and its associated precision metric. RMSE values were used in the case of model outcomes, while standard deviation was used for the BAF plot estimate. The SAS program code for forward variable selection, correlation analysis, and Mallow's Cp are shown in **Appendix E**.

3.3 Results and Discussion

Variable reduction when using forward selection was successful in reducing independent variables from the original 75 variables to fewer than 10 in each case. Variables shown to be significant by forward selection are shown in **Appendix F**. Variables were well distributed across the entire range of possible selections, with no clear trends in variable selection that were evident. Distributional variables were present for vegetation hits from the 1st and 2nd returns, while reflectance variables, percentiles, canopy cover percentiles, and ratio variables all were represented across forest types and segmentation treatments. Reflectance mean, maximum, and range variables of both first and second return vegetation hits were especially well represented. This indicated that reflectance values are of significance in the modeling of forest biophysical parameters (Means *et al*., 1999; Brandtberg *et al*., 2003). Kurtosis and skewness variables were prevalent in deciduous and coniferous volume and biomass variable sets. Percentile variables also were well represented, even in the case of second return vegetation percentiles. Although vegetation or canopy cover percentiles often were well represented, only the percentile most highly correlated to the dependent variable ultimately was selected. Specific significant variables, related to a forest type, were frequently present in both the volume and biomass models for that type. This was not unexpected due to volume and biomass being highly correlated metrics ($r \approx 0.87$). Strong representation from a wide range of distributional parameters indicated that simple metrics such as mean and extreme values were supplemented by parameters such as skewness, kurtosis, percentiles, and canopy percentiles. Results such as these build a strong case for the use of multiple return lidar data, and even associated reflectance-per-hit, for the modeling of forest biophysical parameters. This might be especially critical in areas that contain forests with high variability in site, growth, and composition. The inclusion of second return variables indicates that forest structure is an important aspect in volume and biomass modeling approaches. Second return variables, by definition, contribute to defining height levels other than the topmost canopy, describing aspects of forest vertical structure besides canopy height.

Lidar height distributions were found to be representative of BAF plot measurements. Figure 3.5 shows the lidar first return, vegetation distributions for randomly selected deciduous and

coniferous segments across a range of volume-per-hectare field measurements. Segments with lower measured volume-per-hectare exhibited either fewer hits at taller tree heights, or generally shorter trees than segments with higher volume-per-hectare measurements. Changes in distribution types also were evident with distributions for lower volume-per-hectare segments being skewed to the right, and *vice versa*. Distributions for intermediate volume-per-hectare segments resembled approximate normal-type distributions. The selection of percentile, skewness, canopy cover, and kurtosis variables are evident when distributions are evaluated. Factors such as skewness and percentage hits below percentiles were logical selections for distinguishing between different volume-per-hectare levels, based on distribution shapes and height frequencies. Metrics such as minimum and mean values also played a role due to the number of returns at the lower and upper limits of each distribution, and their contribution to the measured minimum and average value.

Figure 3.6 shows an example of deciduous and coniferous segments with similar volume-per-hectare values across increasing average segment sizes. Distributions for each type visually remain similar in shape as average segment size increases. Coniferous distributions, for example, remain relatively similar in shape as averages segment size increases, but the number of returns increases, resulting in a smoother curve. There are no distinct trends visible when deciduous are compared to coniferous distributions, although there appear to be more distinct upper- and lower tail values in the case of deciduous segments. This can be attributed to trees of above-average height and undergrowth, respectively, as commonly found in an uneven-aged stand.

Correlation analysis formed a critical component of the pre-processing analysis by removing unwanted high correlations, reducing independent variables even further, and ensuring viable, valid models. Correlated variables were intuitive in most cases, e.g., 1st return vegetation percentiles 20, 30, and 50. However, thorough evaluation of variable correlations was required to identify other highly correlated variables, e.g., the 70th percentile of the first returns and the 30th canopy cover percentile. Percentile metrics especially were problematic in terms of correlations. Two-thirds of percentile parameters often had to be removed due to high inter-variable correlations. Final variable sets for all forest types and segmentation treatments that

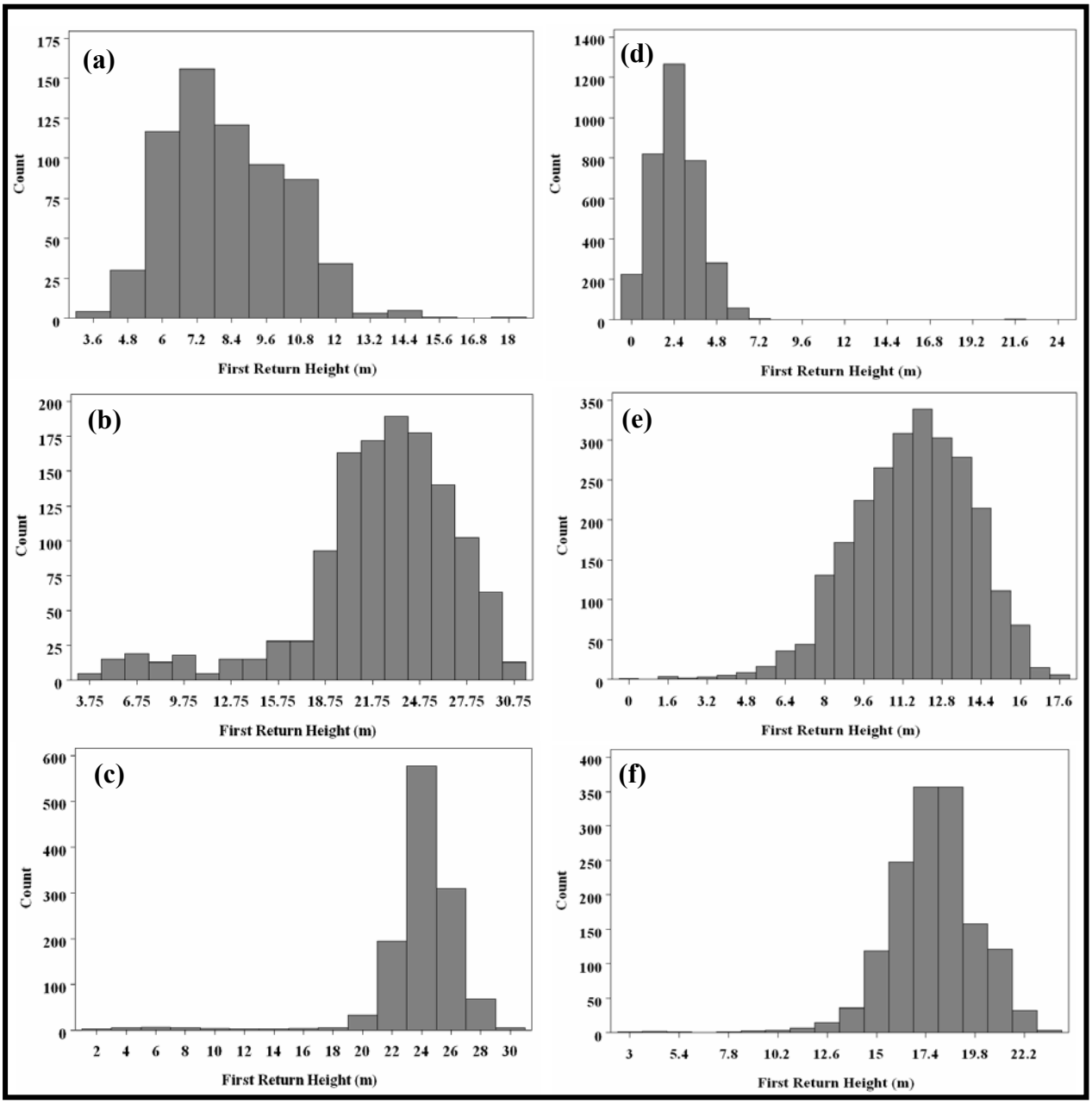


Figure 3.5 Per-segment (0.035 ha/segment) histogram plots for lidar first return vegetation hits across a range of field-measured volume-per-hectare. Deciduous segments are shown in (a) 10.45 m^3/ha, (b) 151.20 m^3/ha, and (c) 350.65 m^3/ha. Coniferous segments are shown in (d) 10.16 m^3/ha, (e) 154.76 m^3/ha, and (f) 350.93 m^3/ha

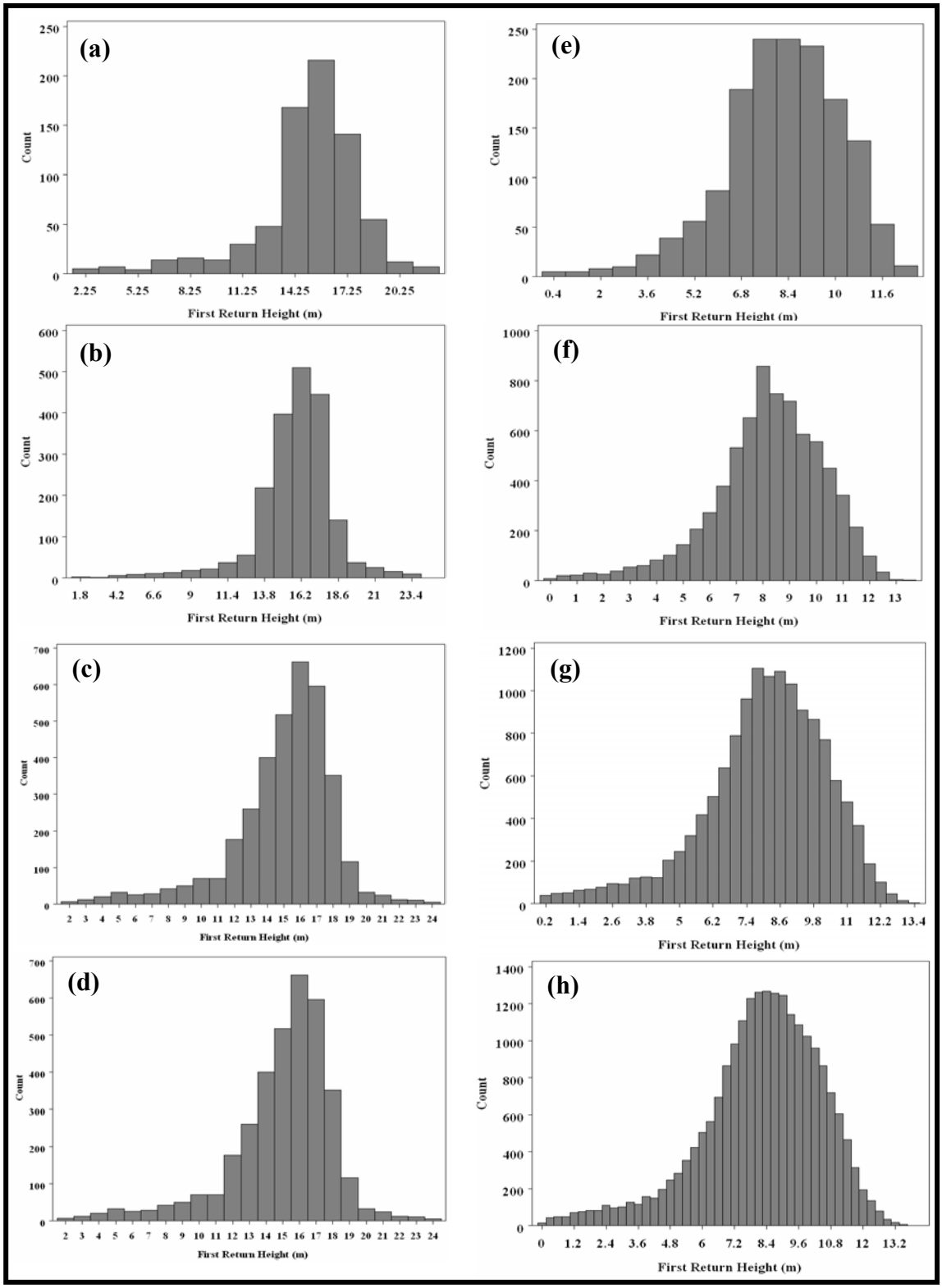


Figure 3.6 First return, vegetation height distributions for a deciduous (a – d; 153.02 m^3/ha) and coniferous (e – h; 159.50 m^3/ha) BAF plot for increasing segment sizes 0.035 ha/segment, 0.091 ha/segment, 0.141 ha/segment, and 0.318 ha/segment, respectively

were used as input to Mallow's Cp selection, after significance and correlation reduction, are listed in **Appendix F**. Variables were again distributed across the entire spectrum of possibilities, with canopy cover and regular percentiles well represented, and even reflectance values still being present. Mallow's Cp selection, which followed variable reduction, resulted in adequate combination of independent variables, from fewer variables with small Cp and large adjusted R^2 values, to more variables and poor metrics. Candidate volume and biomass models for 2- and 3-class schemes across segmentation treatments are shown in **Appendix G**.

Tables 3.4 - 3.15 list the selected models and associated descriptive statistics for 2- and 3-class models and each forest type, for all eleven selected segment sizes (0.035 – 5.632 ha/segment) and the existing Appomattox stands (5.666 ha/segment). Final selection was based on Cp, adjusted R^2, RMSE, and especially number of independent variables. The latter criterion was crucial due to the close performance of candidate models based on other fit statistics. Except for coniferous volume and biomass models, most models could be limited to 5 or fewer independent variables without appreciable loss in goodness-of-fit (adjusted R^2, RMSE, Cp). This could be attributed to a large amount of variation that had to be explained in the case of coniferous segments. Pure forest stands in public ownership are relatively limited in the Virginia Piedmont, resulting in highly variable stands with basal area contribution from both deciduous and coniferous species. While mixed deciduous stands have similar characteristics, a deciduous-coniferous mix likely added more variability to the coniferous group.

Models listed in Tables 3.4 - 3.15 resulted in the highest adjusted R^2 and RMSE values for a given number of variables, kept to a minimum, from all candidate models. Adjusted R^2 values for coniferous species volume were lower than those found in two comparable studies by Means *et al*. (2000; adjusted R^2) and Næsset (2002; R^2), both of which used a grid-cell based lidar distribution approach to volume modeling. R^2 values for these two studies ranged from 0.91 to 0.97, while species were limited to Douglas-fir (*Pseudotsuga menziesii*) (Means *et al*., 2000) and Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) (Næsset, 2002). Stands also varied from shrub-like (18 m^3/ha) to old-growth (2051 m^3/ha) (Means *et al*., 2000) and young forest (41 m^3/ha) to mature forest (639.8 m^3/ha) in the case of Næsset (2002). The range of forest volume and growth-types, low single species variability, averaging effect of plot-based

Table 3.4 Selected lidar distributional volume and biomass models for 27,050 segments (0.035 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	-0.60777 + 9.77043 P_Veg1_60 -203.58544 MinVeg2	0.51	0.51	5.22	59.04
	C	437.02548 + 10.40478 MeanVeg1 -2.33549 KurtosisVeg1 + 42.24961 SkewnessVeg1 -0.11889 MedianRef1 -7.01933 ModeVeg2 -176.08474 Canopy70P	0.65	0.62	7.95	45.43
	A	75.38293 + 9.8795 P_Veg1_60 -0.09684 MedianRef1 -4.86239 P_Veg2_10 + 0.14098 RangeRef2 -0.20538 StdRef2 + 73.76854 Canopy30P	0.60 (0.58)*	0.58 (0.56)*	10.00	52.87 (55.20)*
Biomass	D	-142834 + -146838 StdMeanVeg1+ 6893.76747 P_Veg1_70 + 157.11125 StdRef1 + 609.67491 CVVeg2 + 1874.25982 ModeVeg2	0.56	0.54	5.31	38.33
	C	-370563 + 3631.11161 MeanVeg1 + 6567.61127 SkewnessVeg1 -3515.45559 ModeVeg2 + 3098.78132 P_Veg2_30 + 178.00093 RangeRef2 -39262 Canopy70P	0.61	0.57	8.59	17.56
	A	-48312 + 8399.53101P_Veg1_70 + 49098 Canopy30P	0.60	0.60	2.13	38.36
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or kg/ha**
Volume	D	-29.47558 + 11.38110 P_Veg1_70 + 3.27119 ModeVeg2 -13.11890 P_Veg2_20	0.61	0.59	3.78	55.39
	C	-2.23294 + 14.04162 P_Veg1_30 -67.10545 StdMeanVeg2 + 127.74123 Canopy10P	0.50	0.47	4.91	48.23
	M	-2439.40986 + 0.76326 MinRef1 -0.16578 MedianRef1 -518.99961 MinVeg2 + 11.15849 P_Veg2_60 + 1.16738 RangeRef2	0.60	0.56	6.07	48.27
Biomass	D	-62235 + 7903.39544 P_Veg1_70 + 482.86523 CVVeg2 -175778 MinVeg2 + 1762.93269 ModeVeg2	0.58	0.56	4.83	39.95
	C	2979.68239 + 4298.51676 P_Veg1_30 -36826 StdMeanVeg2 + 114802 Grnd1ratio	0.53	0.50	4.00	14.13
	M	14871 + 6032.01411 MeanVeg1 -553.57582 StdMeanRef1 -359910 MinVeg2	0.51	0.48	2.55	28.20

* Revised values based on the application of a reduced set of independent variables to segments. In some cases, a full set of independent variables resulted in segments with missing values, while a reduced variable set could be fitted to more segments

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.5 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 10,352 segments (0.091 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	2456.94201 + 10.23106 P_Veg1_60 -0.94477 MaxRef1 + 0.74295 CVVeg2 -333.14409 MinVeg2	0.56	0.55	6.96	56.44
	C	676.82516 + 8.17600 MeanVeg1 + 19.83171 MinVeg1 -0.09257 MedianRef1 -0.23512 StdRef2 -261.01640 Vegratio	0.67	0.64	5.25	45.07
	A	309.84855 + 0.29731 CVVeg1 + 13.66277 MinVeg1 + 11.12989 P_Veg1_50 -0.14246 MedianRef1 -432.13149 MinVeg2 + 55.39894 Canopy30P	0.60	0.59	11.60	53.75
Biomass	D	1226311 + 7168.34654 P_Veg1_60 -476.28380 MaxRef1 + 514.89198 CVVeg2	0.52	0.51	5.45	40.95
	C	185653 + 2262.86568 P_Veg1_20 -29.74409 MedianRef1 + 3533.08872 P_Veg2_40 -91682 Vegratio -20694 Canopy70P	0.61	0.59	9.31	17.15
	A	1719863 -590.42558 MaxRef1 + 23259 StdVeg2 -427.47707 MinRef2	0.60	0.59	1.58	38.69
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	2235.65821 + 11.05422 P_Veg1_70 -0.87761 MaxRef1 + 0.92661 CVVeg2	0.62	0.60	6.85	56.18
	C	-9.43622 + 18.20808 P_Veg1_20 + 21.54957 P_Veg2_20 -7.69948 P_Veg2_80 + 324.55966 Grnd1ratio	0.58	0.54	5.51	44.82
	M	-2762.51289 + 22.40821MinVeg1 + 13.52503 P_Veg1_10 -0.35097 MeanRef1 -2.49465 StdMeanRef1 -42.54848 SkewnessVeg2 -37.85424 P_Veg2_10 + 1.33287 MaxRef2 + 248.54668 Canopy40P	0.69	0.63	9.56	44.29
Biomass	D	1612071 + 7448.2498 P_Veg1_70 -625.40333 MaxRef1 + 594.78959 CVVeg2	0.59	0.57	5.17	40.77
	C	55120 + 664.18556 MaxVeg1 + 202433 StdMeanVeg1 + 5840.05250 P_Veg1_20 -92789 StdMeanVeg2 -3307.55529 P_Veg2_80 -74.11209 StdRef2 + 120805 Grnd1ratio	0.67	0.62	8.00	12.29
	M	2557988 + 282757 StdMeanVeg1 + 4915.20326 P_Veg1_20 -938.51666 MaxRef1 -17615 P_Veg2_20 + 5558.84449 P_Veg2_40 -1888.17315 StdMeanRef2 + 119143 Canopy20P	0.61	0.55	6.74	26.18

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.6 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 6,687 segments (0.141 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	-68.70843 + 11.38628 P_Veg1_60 + 0.51347 CVVeg2 + 6.91620 ModeVeg2 -13.92585 P_Veg2_10	0.53	0.52	7.13	58.33
	C	-750.87351 + 16.09367 MeanVeg1 -18.84639 StdVeg1 -0.11117 MedianRef1 + 20.14318 P_Veg2_20 + + 0.60177 RangeRef2 -425.15619 Vegratio + 64.51644 Canopy40P	0.65	0.62	8.82	46.75
	A	1959.75744 + 0.43495 CVVeg1 + 10.71208 MedianVeg1 -0.66120 MaxRef1 -0.08310 MedianRef1 + 5.97669 ModeVeg2 -18.15179 P_Veg2_10	0.57	0.56	10.81	55.44
Biomass	D	-36094 + 7839.97744 P_Veg1_60 + 24.11436 MedianRef2	0.49	0.48	3.09	42.32
	C	-20274 + 5244.59189 P_Veg1_40 -344.24577 MaxRef1+ 2554.41273 ModeVeg2 + 218.15360 RangeRef2 + 127355 Grnd1ratio + 434629 Canopy90P	0.60	0.57	8.02	17.44
	A	-40509 + 538.48029 CVVeg2 + 9614.75760 P_Veg2_80 - 113772 Grnd2ratio	0.59	0.58	4.49	38.99
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	264.78319 + 8.67042 P_Veg1_70 + 4.59012 ModeVeg2 + 307.02298 StdMeanVeg2 -0.85096 MinRef2	0.59	0.58	6.22	58.04
	C	202.84813 + 12.58907 P_Veg1_20 -0.08711 MedianRef1 + 16.01838 ModeVeg2 + 383.85188 Grnd1ratio	0.59	0.56	5.75	44.09
	M	-1852.31116 + 11.81371 P_Veg1_10 -1.32651 MaxRef1 -0.88835 MinRef1 -0.23104 MedianRef1 + 7.39724 P_Veg2_40 + 2.34851 MaxRef2 + 120.87184 Canopy40P	0.67	0.62	8.00	44.91
Biomass	D	-17080 + 178071 StdMeanVeg1 + 6547.19639 P_Veg1_80	0.53	0.52	5.08	43.27
	C	111868 + 5370.99706 P_Veg1_20 + 5177.01988 ModeVeg2 -1965.57531 P_Veg2_80 -114770 Vegratio	0.62	0.59	6.51	12.84
	M	-1137236 + 2430.87976 5MaxVeg1 + 3325.16987 P_Veg1_20 -872.16181 MaxRef1 + 1307.94943 MaxRef2 - 1043.44207 StdMeanRef2	0.62	0.58	4.48	25.14

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.7 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 2,972 segments (0.318 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	-564.81439 + 16.10186 MeanVeg1 + 8.98072 ModeVeg2 -36.31668 P_Veg2_10 + 93.22649 Canopy20P + 490.97586 Canopy90P	0.57	0.55	9.13	56.49
	C	574.97402 + 10.61338 P_Veg1_40 -0.28438 StdRef2 -350.63049 Vegratio	0.56	0.54	5.09	51.29
	A	491.92959 + 230.28927 StdMeanVeg1 + 9.03057 P_Veg1_40 + 7.79524 ModeVeg2 -27.39259 P_Veg2_10 - 0.30435 StdRef2 - 232.81179 Vegratio	0.57	0.56	8.95	55.74
Biomass	D	-2780.49794 + 232267 StdMeanVeg1 + 7026.53626 P_Veg1_50 + 7034.25843 ModeVeg2 -21897 P_Veg2_10	0.52	0.51	4.17	41.17
	C	1485291 + 3515.19422 P_Veg1_40 -473.87803 MaxRef1 -254.91609 MinRef2 -121676 Vegratio	0.54	0.51	4.61	18.58
	A	-244074 + 9601.46203 MeanVeg1 + 88.69273 MedianRef1 + 4011.26767 ModeVeg2 -6433.95950 P_Veg2_30	0.60	0.60	6.99	38.33
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	14.70720 + 0.86668 MedianVeg1	0.53	0.53	4.34	61.21
	C	394.66046 + 17.83889 MeanVeg1 -16.35526 StdVeg2 -0.34355 StdRef2 -500.19809 Vegratio + 355.86955 Canopy80P	0.64	0.61	5.07	41.53
	M	430.85696 + 10.70375 P_Veg1_10 + 4.11804 StdMeanRef1 -1.10940 MinRef2	0.47	0.43	4.91	54.63
Biomass	D	5293.50302 + 7134.35335 P_Veg1_60 + 6556.06088 ModeVeg2 -19342 P_Veg2_10	0.55	0.53	2.77	42.73
	C	218012 + 3417.21375 P_Veg1_30 -90749 StdMeanVeg2 -81.01839 StdRef2 -148364 Vegratio	0.61	0.58	6.99	12.91
	M	-1100788 + 5080.87525 P_Veg2_60 + -152838 Canopy70P + 1272185 Canopy90P	0.49	0.46	4.14	28.65

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.8 Selected lidar distributional volume and biomass models for 1,473 segments (0.642 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	-465.37444 + 10.30145 P_Veg1_40 + 1.08466 MaxRef1 + 3.80564 RangeVeg2 -23.23921 P_Veg2_10 -0.94386 MaxRef2	0.54	0.52	8.80	58.35
	C	284.8419 + 11.08849 P_Veg1_40 -0.11948 MedianRef1 + 561.49193 ZeroNgrnd1ratio	0.53	0.51	3.11	52.84
	A	287.76520 + 486.37259 StdMeanVeg1 + 9.84571 P_Veg1_40 -26.50995 P_Veg2_10 -0.33749 StdRef2 + 555.58684 ZeroNgrnd1ratio	0.55	0.54	9.04	56.73
Biomass	D	-27169 + 8411.9874 P_Veg1_50 -12428 P_Veg2_10 + 1087591 ZeroNgrnd1ratio	0.49	0.48	5.06	42.36
	C	103555 + 4078.69040 P_Veg1_40 -106610 Vegratio	0.47	0.46	3.68	19.62
	A	-233237 + 259.95190 CVVeg1 + 8485.39537 MedianVeg + 80.07989 MedianRef1 -3486665 MinVeg2	0.59	0.58	5.78	39.01
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or kg/ha**
Volume	D	866.66394 + 8.57492 MedianVeg1 + 0.87639 RangeRef1 + 4.48030 RangeVeg2 -1.13409 MaxRef2	0.54	0.52	5.41	61.72
	C	303.72815 + 15.71060 P_Veg1_30 -1.78646 StdMeanRef2 -0.59669 StdRef2 + 0.06230 MedianRef2 + 737.63803 ZeroNgrnd1ratio + 146.83730 Canopy70P	0.70	0.67	7.00	38.24
	M	5554.20294 -1.69275 MinRef1 + 13.97249 StdMeanRef1 -1.68066 MaxRef2 -0.86973 MinRef2 -223.08804 Canopy60P	0.67	0.63	4.56	44.22
Biomass	D	-70748 + 8362.88108 P_Veg1_60 + 548.50455 CVVeg2	0.51	0.50	2.88	44.21
	C	148826 + 5417.28312 P_Veg1_25 -163577 StdMeanVeg2 -1033.93279 CVRef2 -126.02192 StdRef2 + 249279 ZeroNgrnd1ratio + 31849 Canopy70P	0.66	0.62	7.00	12.35
	M	-365004 + 75501 ZeroNVeg3_5ratio -140881 Canopy60P + 484061 Canopy90P	0.53	0.50	5.90	27.63

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.9 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 981 segments (0.964 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	-3736.42054 + 512.04312 StdMeanVeg1 + 4.45043 P_Veg1_10 + 10.85680 P_Veg1_40 + 1.35914 MaxRef1 + 889.88518 ZeroNgrnd1ratio	0.56	0.54	7.32	57.39
	C	429.26030 + 15.45378 MeanVeg1 + 6.8230 P_Veg1_10 -0.19552 MedianRef1 -4.56810 MaxVeg2 + 445.05874 ZeroNgrnd1ratio + 108.96666 Canopy30P	0.59	0.55	6.25	50.30
	A	246.52778 + 11.22933 P_Veg1_40 -0.10604 MedianRef1 + 398.51814 ZeroNgrnd1ratio	0.54	0.54	6.61	57.25
Biomass	D	-1935102 + 9713.75052 P_Veg1_40 + 699.30933 MaxRef1 + 915946 ZeroNgrnd1ratio	0.53	0.52	6.54	40.83
	C	9305.68329 + 3329.81470 MeanVeg1 + 3111.7526 P_Veg1_10 -972.10810 MaxVeg2 -7423115 MinVeg2 + 131907 ZeroNgrnd1ratio	0.51	0.47	7.11	19.70
	A	138729 + 462.34099 CVVeg1 + 683280 StdMeanVeg + 9404.78486 P_Veg1_40 -2094.90407 CVRef1 -4994001 MinVeg2 -417.35164 MinRef2	0.63	0.62	9.33	37.71
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	-62.89592 + 13.06912 P_Veg1_40 + 1048.43400 ZeroNgrnd1ratio	0.55	0.54	4.53	60.80
	C	599.31733 + 16.10637 MeanVeg1 -20.60615 StdVeg2 -0.43003 StdRef2 -571.66877 Vegratio + 314.71837 Canopy80P	0.63	0.59	6.00	43.89
	M	647.33429 -0.20329 MeanRef1 + 5.4947 P_Veg2_60 -176.87863 Canopy60P	0.51	0.48	3.75	51.96
Biomass	D	-49500 -26739 MinVeg1 + 9371.92384 MedianVeg1 + 780467 ZeroNgrnd1ratio	0.55	0.54	2.00	42.71
	C	273252 + 4041.65160 P_Veg1_30 -2234.92703 P_Veg2_80 -123.80839 StdRef2 -165789 Vegratio	0.60	0.56	5.00	13.55
	M	385335 + 3212.53150 P_Veg2_60 -601.99729 MinRef2 -145583 Canopy70P	0.58	0.55	3.02	26.54

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.10 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 749 segments (1.263 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	-3088.40721 + 13.35664 P_Veg1_40 + 1.12535 MaxRef1 + 2.10464 StdMeanRef2	0.54	0.53	6.19	58.28
	C	434.11729 + 6.46116 MeanVeg1 + 9.03449 P_Veg1_10 -0.18789 MedianRef1 + 535.30959 ZeroNgrnd1ratio	0.57	0.55	5.00	51.07
	A	348.17050 + 12.30712 -0.15941 -6563.10706 + 2.30318 + 296.34639	0.56	0.55	6.00	56.85
Biomass	D	-2292797 + 10386 P_Veg1_40 + 913.24284 MaxRef1 -2900.52040 CVRef2	0.54	0.53	5.10	40.73
	C	150465 + 4368.91277 P_Veg1_10 -9417359 MinVeg2 -113895 Vegratio -35928 Canopy50P	0.59	0.56	4.23	18.06
	A	-2411604 + 10042 P_Veg1_40 + 872.82293 MaxRef1 -4867899 MinVeg2 + 2356.74161 StdMeanRef2	0.62	0.61	7.65	38.03
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	-2502.31959 + 0.77043 CVVeg1 + 14.52804 P_Veg1_40 + 0.89607 MaxRef1	0.57	0.55	7.10	60.37
	C	-17.67697 + 19.94468 MeanVeg1 -23.11140 StdVeg1 -20786 MinVeg2 + 609.04443 ZeroNgrnd1ratio	0.62	0.59	5.00	44.46
	M	842.91045 -0.23230 MedianRef1 + 145.97738 ModeVeg2 -230.58477 Canopy60P	0.57	0.54	3.56	48.74
Biomass	D	-2476904 + 540.49099 CVVeg1 + 10299 P_Veg1_40 + 898.58313 MaxRef1	0.56	0.55	4.91	42.36
	C	217932 + 3526.66163 P_Veg1_20 -7617358 MinVeg2 -90.26707 StdRef2 -139121 Vegratio	0.61	0.57	4.60	13.58
	M	154386 + 67175 ModeVeg2 -130353 Canopy60P	0.56	0.54	1.81	26.11

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.11 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 502 segments (1.885 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	-1884.83701 + 12.28622 P_Veg1_30 + 1.39228 MinRef1 + 0.54302 MaxRef2	0.57	0.56	4.10	56.62
	C	604.75316 + 7.81933 P_Veg1_10 -0.24832 MeanRef1	0.47	0.46	3.00	52.34
	A	20.19012 + 12.12259 P_Veg1_40 + 1.02845 MinRef1 -0.14781 MedianRef1 + 428.44724 ZeroNgrnd1ratio	0.58	0.57	4.15	54.79
Biomass	D	-373365 + 10197 P_Veg1_40 + 1081.74271 MinRef1 -394893 StdMeanVeg2 + 969199 ZeroNgrnd1ratio	0.58	0.56	6.99	39.61
	C	324634 -104.22724 MedianRef1 + 5114.15139 MedianVeg2 -75580 ZeroNVeg2ratio + 35214 Canopy30P - 68583 Canopy50P + 48540 Canopy70P	0.61	0.57	6.73	15.64
	A	-1633282 + 380.67463 CVVeg1 + 10088 P_Veg1_40 + 50.36197 MeanRef1 + 585.8662 MinRef1 + 480.55129 MaxRef2	0.64	0.63	8.73	37.46
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	-301.48364 + 12.42276 P_Veg1_30 + 0.96603 MinRef1	0.59	0.59	4.15	58.10
	C	207.91562 + 22.8444 MeanVeg1 -31.16739 StdVeg2 + 1.21929 MinRef2 -0.49008 StdRef2 -527.20557 Vegratio + 118.38753 Canopy30P -178.00186 Canopy50P + 358.16686 Canopy70P	0.72	0.66	9.00	40.13
	M	3343.74474 + 1.92861 MinRef1 -0.17179 MedianRef1 -0.90853 MinRef2 -146.77500 Canopy50P -2999.47600 Canopy90P	0.54	0.49	6.32	50.05
Biomass	D	-242861 + 9808.94384 P_Veg1_40 + 639.32025 MinRef1 + 734940 ZeroNgrnd1ratio	0.59	0.58	6.06	41.48
	C	252508 + 2898.64810 P_Veg1_30 -129.65326 StdRef2 -141167 Vegratio	0.57	0.54	4.64	13.96
	M	314557 -437.05370 MinRef2 -981664 ZeroNgrnd1ratio -93250 Canopy60P	0.50	0.46	4.45	28.42

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.12 Selected lidar distributional volume and biomass models for 374 segments (2.530 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	2394.28642 + 10.80258 P_Veg1_50 -2411.49284 Canopy90P	0.53	0.52	4.75	56.92
	C	5316.89010 + 12.81651 P_Veg1_40 -0.18392 MeanRef1 + 1.25638 MaxRef1 -5.16699 RangeVeg2 -0.37447 StdRef2 + 120.34206 Canopy30P -7947.60598 Canopy90P	0.58	0.53	9.23	47.92
	A	3421.47769 + 0.32953 CVVeg1 + 10.72745 P_Veg1_40 -0.11546 MedianRef1 -3151.45656 Canopy90P	0.55	0.54	5.04	55.26
Biomass	D	1698988 + 8805.95521 MedianVeg1 -2255.26886 CVRef2 -1567288 Canopy90P	0.52	0.51	6.02	41.34
	C	-209372 + 4566.23638 P_Veg1_40 + 220.40006 RangeRef1 -6270.53711 StdVeg2 -298007 Canopy80P	0.44	0.40	3.48	18.01
	A	-55961 + 203.67372 CVVeg1 + 9360.78776 P_Veg1_50 + 108.81192 MeanRef1 -232322 Canopy80P	0.63	0.62	6.17	37.40
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or kg/ha**
Volume	D	1376.37784 + 13.97668 P_Veg1_50 + 0.84046 CVVeg2 -0.56206 MaxRef2	0.57	0.55	5.80	60.45
	C	2054.20958 -3.76694 KurtosisVeg1 + 27.57012 P_Veg1_25 -0.91041 RangeRef1 -6.00273 P_Veg2_70 + 832.37060 ZeroNgrnd1ratio + 95.00570 Canopy30P	0.67	0.62	6.18	40.45
	M	6982.20079 -34.71224 SkewnessVeg2 -78.94625 Canopy60P -6768.80680 Canopy90P	0.55	0.53	4.00	42.60
Biomass	D	-94995 + 10460 MedianVeg1 + 57.27534 MedianRef2	0.53	0.52	5.93	44.62
	C	-1215.60779 + 6357.87095 P_Veg1_25 -17827381 MinVeg2 -2642.10039 P_Veg2_70 + 143313 ZeroNgrnd1ratio	0.58	0.54	6.37	12.98
	M	394817 -22263 SkewnessVeg1 + 822.14771 MinRef1 -587313 Canopy80P	0.68	0.66	5.29	20.12

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.13 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 240 segments (3.942 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	5859.62024 + 14.39002 P_Veg1_60 -16.34060 StdMeanRef1 -0.24402 RangeRef2 -451.84584 ZeroNVeg2ratio -4949.51387 Canopy90P	0.58	0.55	5.76	55.31
	C	9758.28548 + 1.72784 RangeVeg1 + 11.66538 P_Veg1_40 -8.28570 RangeVeg2 -9614.11196 Canopy90P	0.58	0.55	4.76	48.33
	A	4250.50221 + 10.27277 P_Veg1_40 -0.12528 MedianRef1 + 370.68205 ZeroNgrnd1ratio -3947.55809 Canopy90P	0.53	0.52	6.03	56.23
Biomass	D	3668508 + 8566.23369 P_Veg1_60 + 3257.35068 MaxVeg2 -418143 ZeroNVeg2ratio -3434640 Canopy90P	0.57	0.55	6.20	39.70
	C	-507343 + 4485.89116 P_Veg1_30 + 226.38774 RangeRef1 -1187.0808 RangeVeg2	0.49	0.46	5.36	17.83
	A	3005737 -8000.67916 StdMeanRef1 + 431.22358 CVVeg2 + 14386 P_Veg2_75 + 130097 ZeroNgrnd3_5ratio -3138008 Canopy90P	0.62	0.61	8.14	37.13
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	-146.89327 + 14.85985 P_Veg1_60 + 444.41586 ZeroNgrnd2ratio	0.61	0.60	2.17	58.79
	C	-121.50767 + 24.71668 P_Veg1_25 + 211.66780 ZeroNgrnd2ratio + 49.93875 Canopy30P	0.68	0.65	5.52	41.37
	M	11639 -959.87957 MinVeg1 -7.38541 KurtosisVeg2 -199.31365 Canopy10P -11499 Canopy90P	0.58	0.54	3.66	43.87
Biomass	D	250556 + 10749 P_Veg1_60 -358486 ZeroNVeg2ratio	0.59	0.57	5.77	43.81
	C	-84498 + 6498.76606 P_Veg1_25 + 44.56384 MedianRef1 -81.96051 StdRef2 + 78560 ZeroNgrnd3_5ratio	0.67	0.63	5.70	12.06
	M	1188812 + 3454.89453 MeanVeg1 -560331 MinVeg1 -416.30166 RangeRef1 -171320 Canopy70P	0.65	0.62	6.81	22.17

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.14 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 168 segments (5.632 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	262.37385 + 19.92957 P_Veg2_70 + 208.14833 ZeroNgrnd3_5ratio -387.67008 Canopy80P	0.61	0.59	2.62	51.15
	C	458.52340 -4.18276 ModeVeg1 + 15.43186 P_Veg1_40 -5.59238 RangeVeg2 -0.36692 StdRef2	0.69	0.66	6.02	38.03
	A	1.04499 + 10.98685 P_Veg1_40 + 396.51430 StdMeanVeg2	0.55	0.54	2.96	52.16
Biomass	D	271644 + 13993 P_Veg2_75 -286090 ZeroNVeg2ratio -75577 Canopy70P	0.60	0.58	5.96	37.41
	C	37468 + 4618.23146 P_Veg1_40 -1469.46246 MaxVeg2	0.55	0.52	1.93	15.98
	A	343583 -1370.95705 CVRef1 + 316.85290 CVVeg2 + 15082 P_Veg2_75 -132911 ZeroNVeg3_5ratio -314776 Canopy80P	0.68	0.66	7.57	33.14
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	-31.77814 + 19.67658 P_Veg2_70	0.63	0.62	2.55	55.98
	C	-98.73963 + 16.18796 P_Veg1_40 + 0.68551 CVVeg2	0.52	0.47	1.82	52.13
	M	255.71328 -3.17225 ModeVeg1 + 1.54155 MinRef1 -5.84654 StdMeanRef2 -444.06932 ZeroNgrnd1ratio -111.50951 Canopy10P -145.92581 Canopy50P -413.21393 Canopy80P	0.78	0.74	9.35	28.02
Biomass	D	-134083 + 16205 P_Veg2_75 + 2460.48834 StdMeanRef2 + 159205 ZeroNgrnd3_5ratio	0.65	0.62	5.32	39.48
	C	50377 -1125.64714 ModeVeg1 + 477.72255 RangeVeg1 + 4851.42013 P_Veg1_30 -2390.38376 RangeVeg2	0.61	0.52	5.00	14.16
	M	1493940 -601912 MinVeg1 + 4984.27818 P_Veg1_10 -370.80540 RangeRef1 + 84546 ZeroNVeg3_5ratio -612016 Canopy80P	0.81	0.79	8.83	16.32

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 3.15 Selected lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 167 Appomattox forest stands (5.666 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M; all segments/types = A*)

2-class Model		Variables (# variables in parenthesis)	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	3867.04600 + 9.85476 MedianVeg1 -1.10198 MaxRef1 + 7.62947 RangeVeg2 -685.46642 StdMeanVeg2 -407.38497 ZeroNVeg2ratio + 186.14739 Canopy10P -751.23148 Canopy80P	0.51	0.44	9.82	63.56
	C	9338.59799 -2.95046 MaxRef1 -0.31137 MedianRef1 -1.44601 MinRef2	0.53	0.48	1.87	55.61
	A	2359.40526 + 9.19486 MedianVeg1 -6.53562 CVRef1 -0.54508 RangeRef1 -0.38331 MedianRef1 + 7.80485 P_Veg2_90 -3.85625 StdMeanRef2 + 687.76857 ZeroNgrnd1ratio	0.46	0.42	10.12	62.36
Biomass	D	272817 + 10550 MedianVeg1 + 53982 P_Veg2_10 -406842 ZeroNVeg2ratio + 157228 Canopy10P	0.46	0.43	5.00	45.84
	C	265637 + 4242.61474 P_Veg1_30 -45.81002 MedianRef1 -160253 Vegratio	0.45	0.40	4.00	19.45
	A	7863210 -392.16539 RangeRef1 -373.92441 StdRef1 -174.12588 MedianRef1 -1958636 StdMeanVeg2 + 19131 P_Veg2_75 + 66.96329 MedianRef2 + 361589 ZeroNgrnd1ratio -6383782 Canopy90P	0.51	0.46	8.66	41.18
3-class Model		**Variables (# variables in parenthesis)**	**R^2**	**Adjusted R^2**	**Cp**	**RMSE m^3/ha or Mg/ha**
Volume	D	-235.12102 + 15.25761 P_Veg1_50 + 100.81556 P_Veg2_10 + 818.29503 ZeroNgrnd2ratio + 211.14342 Canopy10P	0.51	0.46	5.96	68.16
	C	2645.64234 + 15.14541 P_Veg1_25 -1.67213 MaxRef1 -5.59027 MaxVeg2 + 2051.86395 Canopy80P	0.78	0.73	5.00	40.08
	M	8211.65768 + 7.95393 MedianVeg1 -2.68983 MinRef1 -7385.42988 Canopy90P	0.61	0.57	2.08	46.68
Biomass	D	-167585 + 10763 MedianVeg1 + 73900 P_Veg2_10 + 644151 ZeroNgrnd2ratio + 138488 Canopy10P	0.51	0.46	5.00	48.61
	C	3804460 -1369.71403 MaxRef1 -7927.01472 StdVeg2 -4581.48960 P_Veg2_30 -3555.09352 StdMeanRef2	0.75	0.70	3.36	12.56
	M	5424440 + 7871.71191 MeanVeg1 -22716 StdMeanRef1 + 121654102 MinVeg2 -18719 P_Veg2_20 -5414494 Canopy90P	0.73	0.68	6.00	20.29

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

measurements, and fixed plot measurements that directly corresponded with lidar plot boundaries could have contributed to higher R^2 values in these two studies. However, RMSE values were comparable to those found by Means *et al.* (2000; 73 m^3/ha, old-growth plots excluded) and Næsset (2002; 18.3 – 31.9 m^3/ha), indicating that a segment-based approach has potential for extension to operational application.

Coniferous adjusted R^2 values for volume in this study ranged from 0.46 (2-class; 1.885 ha/segment) to as high as 0.67 (3-class; 0.642 ha/segment. A narrower range in volume and biomass-per-hectare values found in this study (6.94 – 350 m^3/ha; 4.67 – 269.01 Mg/ha) was one possible reason for lower adjusted R^2 values. This was due to more intrinsic variability found in this narrower range, while an increased observed range with lower variability likely will result in better model fit statistics. Plot sampling technique also was a potential source of variability. Unlike the complete grid-cell inventory by Means *et al*. (2000), not every tree within a segment was measured in this approach. Although BAF plot measurement is an established forestry inventory technique, it does not account for all trees on a given plot. Each segment was assumed to be represented by its enclosed BAF plot. This assumption also could have impacted the results.

Coniferous RMSE values for this study (38.03 – 56.73 m^3/ha; 12.06 – 19.70 Mg/ha) compared favorably with those found by Means *et al*. (2000; 73 m^3/ha) and Nasesset (2002; 18.3 – 31.9 m^3/ha). This is of practical importance, since model extension to real-world estimates is reliant on precision estimates. Considering the range of deciduous RMSE values from 50.39 m^3/ha to 61.72 m^3/ha (volume) and 37.41 Mg/ha to 48.61 Mg/ha (biomass), extension to operational applications cannot to be excluded. It could be argued that RMSE values are of greater importance to operational implementation than R^2 values, providing both an estimate and its associated precision to the forest manager.

Although coniferous results were worse than those for a plot-level lidar study of Popescu *et al*. (2004) in the same area, adjusted R^2 values for deciduous types were significantly higher. The highest values for this study were 0.59 (2-class; 5.632 ha/segment) and 0.62 (3-class; 5.632 ha/segment) vs. an unadjusted R^2 of 0.36 found by Popescu *et al*. (2004). This latter result is of

importance, indicating the potential of a segment-based approach to deciduous volume- and biomass modeling. Segment-based approaches might be better suited to deciduous modeling due to the more diverse structure of deciduous growth, as opposed to pure coniferous stands. Small-radius plot-level deciduous volume and biomass modeling are potentially problematic due to possible stand variability and the large size (crown width) of old-growth deciduous trees.

Derivation and use of segments as measurement units could potentially encapsulate deciduous units better than a fixed plot-based approach. On the other hand, the lower coniferous adjusted R^2 values found in this study were attributed to the diversity in coniferous segments. A 2-class (deciduous-coniferous) modeling approach lent itself to inclusion of segments with only marginally more coniferous than deciduous basal area. This in turn resulted in reduced adjusted R^2 values due to increased within-segment variability. It should be noted that an *n*-class modeling approach, where $n > 2$, also has potential disadvantages. These include smaller segment numbers per class, along with reduced feasibility for operational applications due to more complex model fitting requirements. Although a 3-class approach has disadvantages when compared to a 2-class scheme, there was no definitive difference between 2- and 3-class model metrics. Deciduous adjusted R^2 values ranged between 0.51 and 0.59 (2-class) and 0.52 and 0.62 (3-class), while coniferous values ranged between 0.46 and 0.66 (2-class) and 0.47 and 0.67 (3-class). Adjusted R^2 values for the mixed class in the 3-class scheme ranged between 0.43 and 0.74. These results indicated that a simpler, 2-class approach was well-suited to the study area, but the option of a 3-class scheme is open to the user, depending on operational implications.

Model fit statistics did not exhibit vast differences among different segment sizes, but generally did slightly deteriorate with increasing average segment size. Except for the 5.632 ha/segmentation result, with high adjusted R^2 values for coniferous ($R^2 \approx 0.66$), deciduous ($R^2 \approx 0.59$), and combined models ($R^2 \approx 0.54$), the general trend was for fit statistics to become worse with increasing segment size. This was true for segment sizes between 0.091 ha/segment and 3.942 ha/segment with a general decreasing trend in adjusted R^2 values and an increasing trend in RMSE values. A clear reason for the increase in adjusted R^2 values in the case on 5.632 ha/segment was not evident. This increase perhaps could be attributed to better representation of BAF plot data at the larger segment size, while within-segment variability remained adequately

small to obtain acceptable model fit statistics. This lack of distinct differences among average segment sizes also is an artifact of the hierarchical segmentation algorithm. Minimization of within-segment variance occurred at the smaller average segment sizes, while these smaller segments form the building blocks for larger segments. Within-segment variance therefore has been minimized at smaller segment size levels, with the hierarchical structure not further contributing significantly in reducing this within-segment variance. Figure 3.7 shows the adjusted R^2 values for all segmentation levels and 2-class volume and biomass modeling.

Model metrics for the operational Appomattox stands were distinctly lower than those found in the case of segmentation applications for the 2-class scheme. Deciduous adjusted R^2 and RMSE values for volume modeling were 0.44 and 63.56 m^3/ha, while values for coniferous stands were 0.48 and 55.61 m^3/ha, respectively. Overall modeling results were 0.42 and 62.36 m^3/ha, which indicated that segmentation has distinct advantages over current defined stands in the study area. These trends are evident in Figure 3.7, with the modeling improvement due to segmentation being attributed to definition of modeling units using the same data source and units having lower within-segment variance. Although adjusted R^2 values for 3-class coniferous volume ($R^2 \approx 0.73$) and biomass ($R^2 \approx 0.70$) were high relative to other results, this was attributed to the current stand definition being based on homogenous, even-aged coniferous stands. This came at the cost of low adjusted R^2 values for deciduous stands for volume ($R^2 \approx 0.46$) and biomass ($R^2 \approx 0.46$).

Adjusted R^2 was found to be a metric well suited to model evaluation, especially considering a relatively large number of height distribution variables (≤ 7) needed to explain the variability in the dependent variables. Adjusted R^2 values of up to 0.59 (2-class deciduous volume), 0.67 (3-class coniferous volume), and 0.59 (combined volume) were deemed acceptable given the variability associated with public forests in the Virginia Piedmont. Stands often are mixed to a large degree, resulting in large within-stand variation. Adjusted R^2 values for biomass were as high as 0.62 (deciduous), 0.63 (coniferous), and 0.66 (combined). Although no definitive trend was prevalent, adjusted R^2 values were higher while RMSE values were lower in most model cases for smaller average segment sizes (0.035 – 0.318 ha/segment) as opposed to larger

segments (0.141 - 5.632 ha/segment). Smaller segments have smaller within-segment variability with associated higher between-segment variance.

Figure 3.7 Adjusted R^2 values for (a) 2-class volume and (b) 2-class biomass modeling

Evaluation of residual values for all models confirmed model validity, with no alarming trends in residual values. Although most plotted residuals were evenly distributed around zero, with only a

limited number of outliers, a slight increase in residuals with increasing predicted values were detected for deciduous models. This minor variance heteroscedasticity was attributed to two main factors. Biomass models have been shown to have a "fan-shaped" residual trend with increasing predicted values. This is especially true for predicted biomass values greater than 100 Mg/ha (Parresol, 1999). The biomass equations used to model tree biomass were based only on tree diameter, and not on measured height values (Schroeder *et al.,* 1997). This is a common practice (Schroeder *et al.,* 1997; Parresol, 1999), but could have contributed to poor residual distributions due to independent variable differences between biomass equations and lidar-based biomass models. Lidar data is an inherently height-based data source, while most biomass models do not include height as an independent variable. Volume and biomass modeling using logarithmic transformations of the dependent and independent variables were attempted, but the heteroscedasticity effect was not substantially reduced. Figures 3.8 – 3.13 graphically represent the 2-class models for deciduous, coniferous, and combined volume and biomass models for the 0.035 ha/segment segmentation, the smallest selected segmentation result with associated low within-segment variability. The variability in estimates is evident from these graphs, but a linear trend in field-measured vs. predicted values exists in each case. Residual values associated with each model also are shown. Field-measured vs. predicted value plots for all segmentation results are show in **Appendix H**.

Table 3.16 lists the volume and biomass estimates for the total study area according to basal area plots, and lidar and BAF plot modeling methods. Lidar model and BAF plot estimates were relatively similar, with differences of only 2.84% and 4.09% between model volume estimates for 27,050 (0.035 ha/segment) and 6,687 (0.141 ha/segment) segments, respectively, and the BAF plot estimate. Differences in model and BAF plot biomass values, namely 4.46% for 27,050 (0.035 ha/segment) segments and 3.91% for 6,687 (0.141 ha/segment) segments, were similar when compared with volume differences. Such small estimate differences were encouraging, even when the low precision in estimates was considered. The precision of volume and aboveground biomass estimates, as a percentage of the estimate, was lower in the case of modeled (40% - 41% and 43% - 46%, respectively) versus field-measured (59% and 69%, respectively) values. This increased precision through the addition of lidar as a modeling data source further highlighted the potential of the lidar modeling approach. The small loss in

precision between a smaller (0.035 ha/segment) and larger (0.141 ha/segment) segment size application also was noticeable.

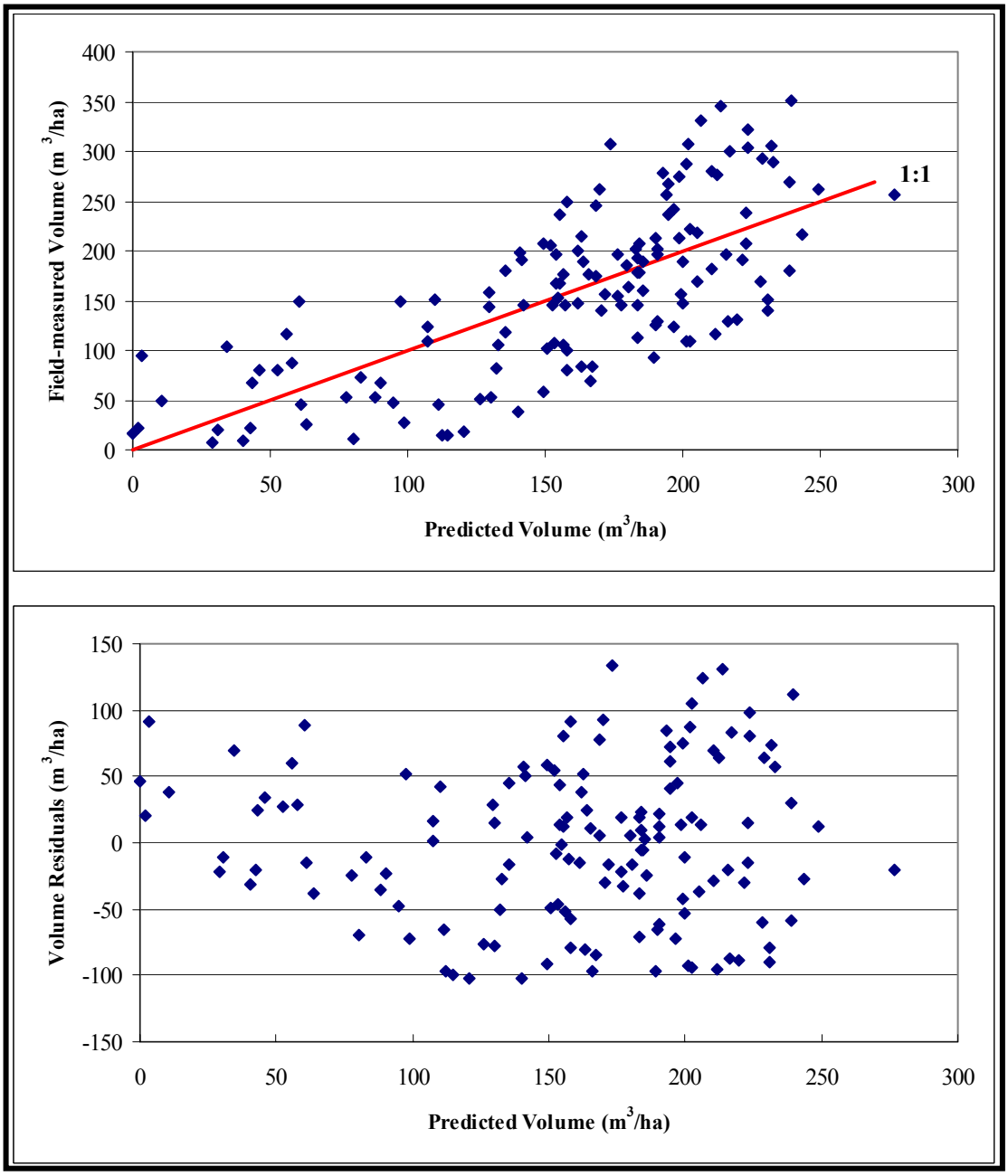


Figure 3.8 2-class volume model (27,050 segments; 0.035 ha/segment): Field-measured vs. predicted volume/ha values and residuals for deciduous plots (adjusted $R^2 = 0.51$)

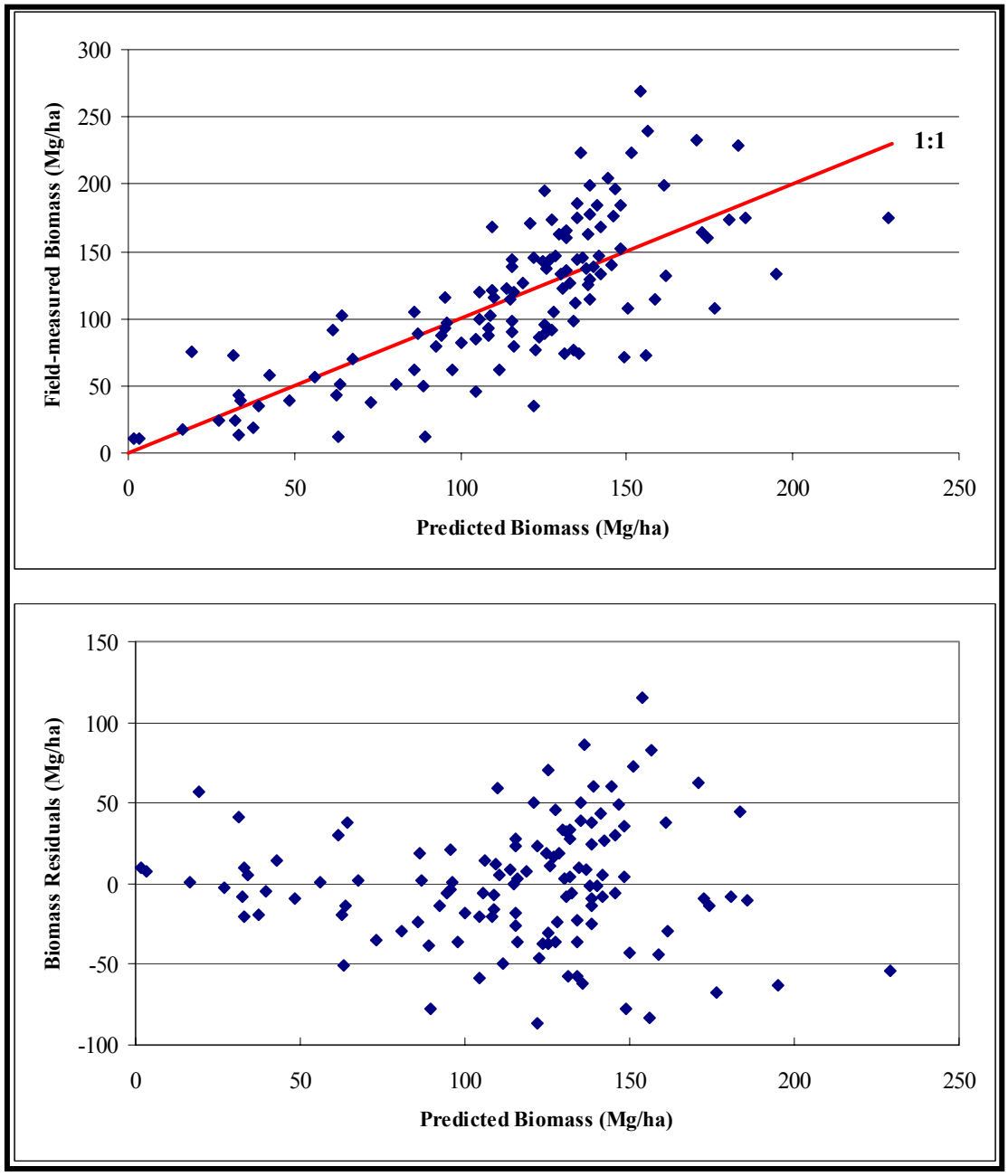


Figure 3.9 2-class biomass model (27,050 segments; 0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.54$)

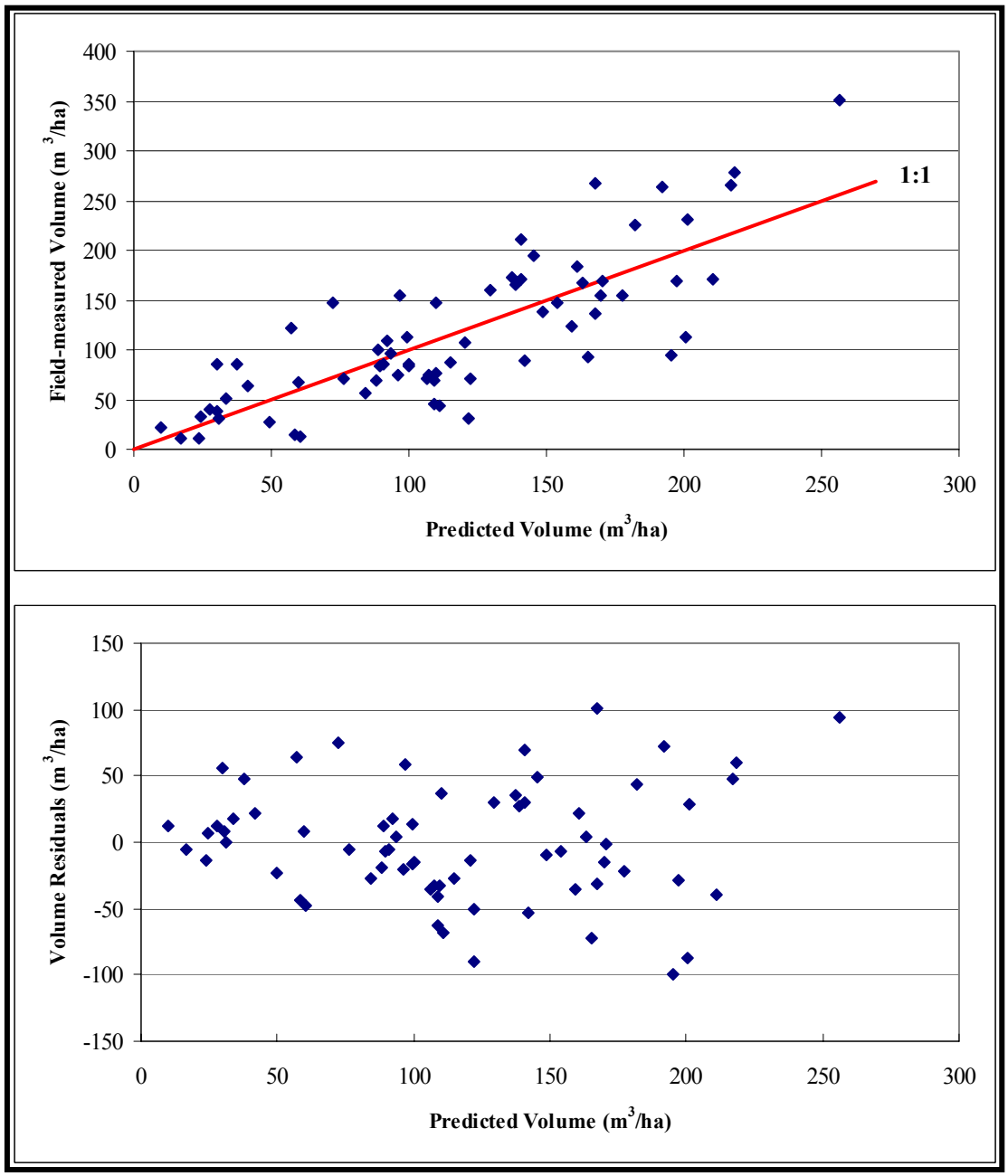


Figure 3.10 2-class volume model (27,050 segments; 0.035 ha/segment): Field-measured vs. predicted volume/ha values and residuals for coniferous plots (adjusted $R^2 = 0.62$)

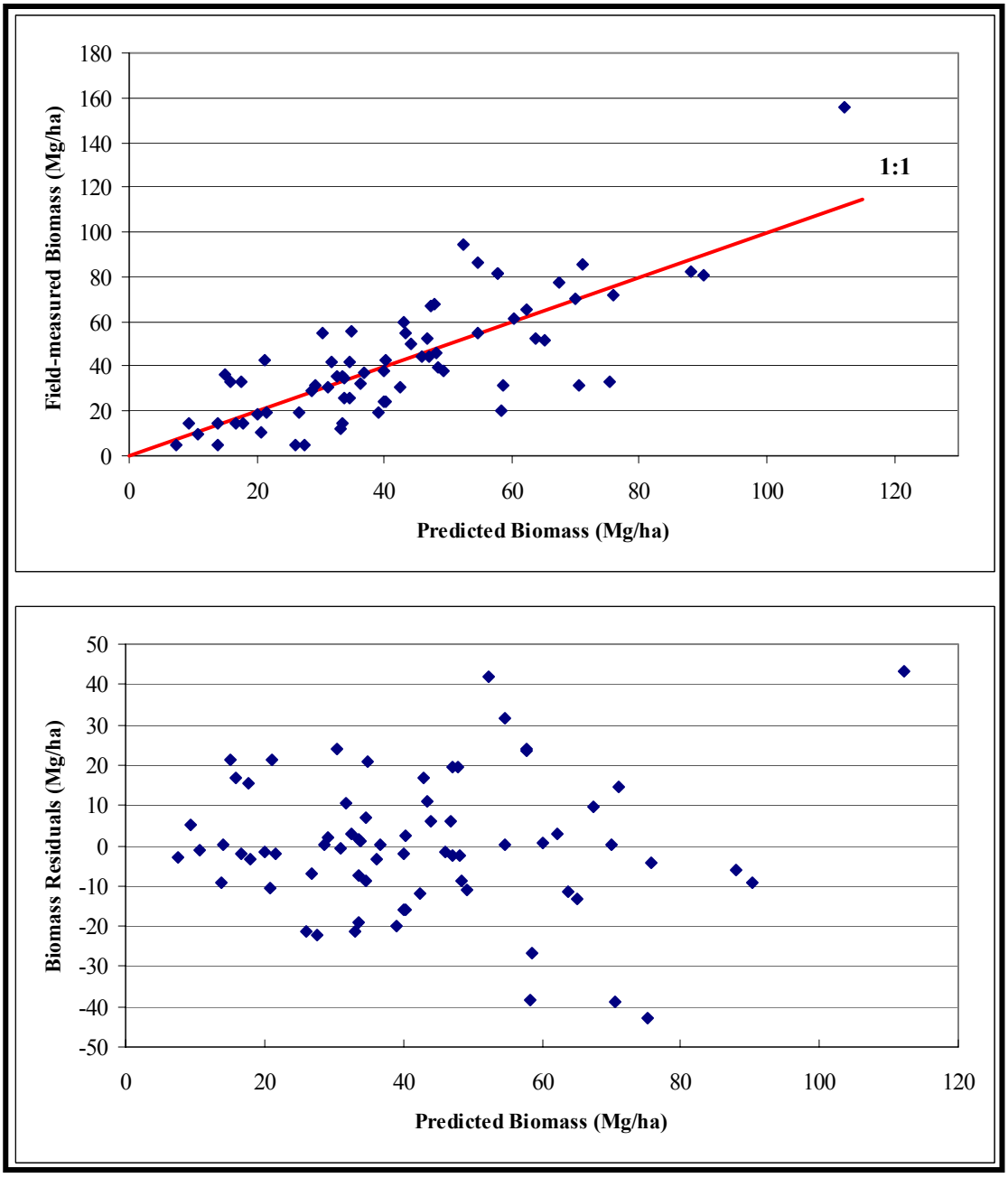


Figure 3.11 2-class biomass model (27,050 segments; 0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.57$)

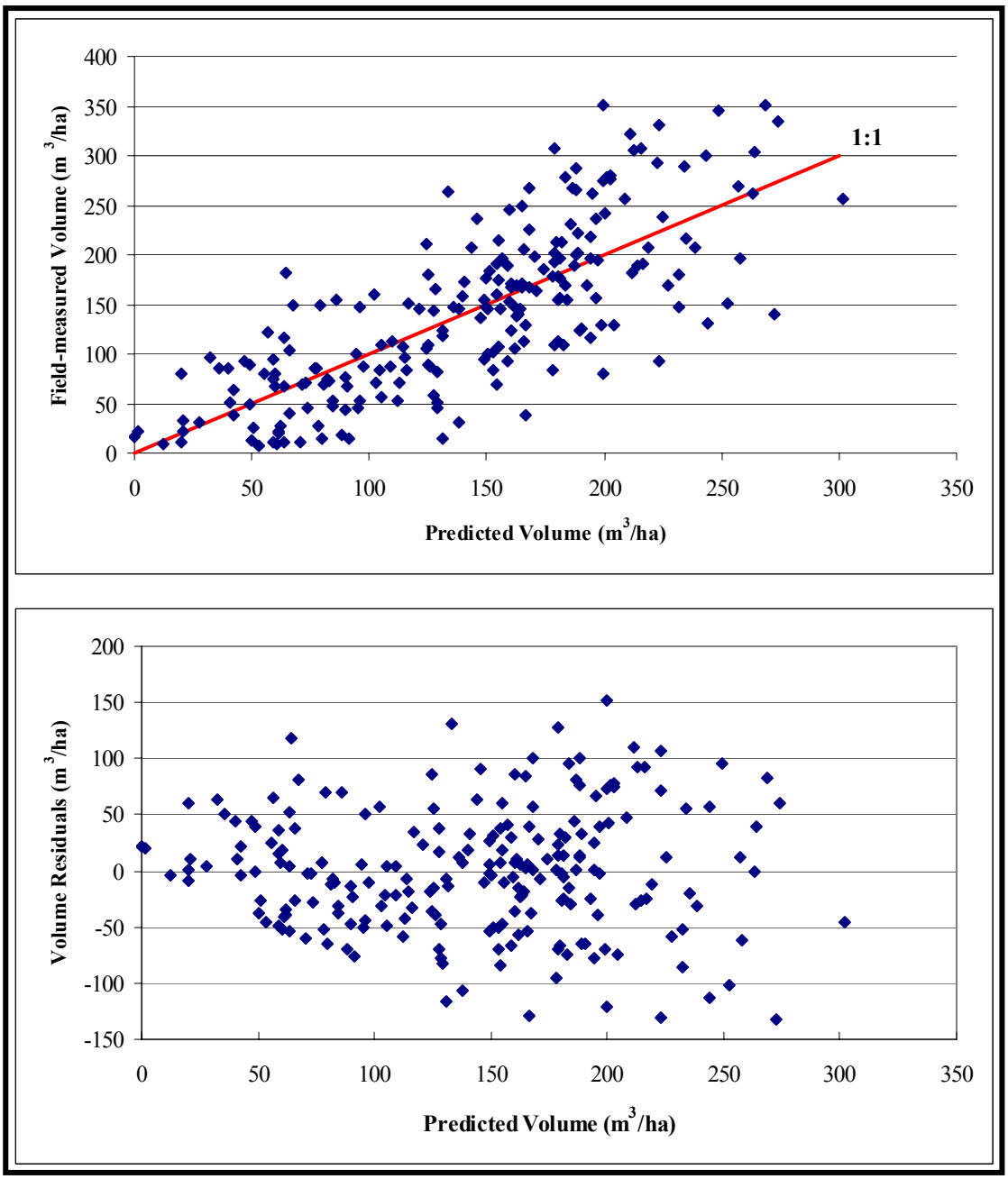


Figure 3.12 2-class volume model (27,050 segments; 0.035 ha/segment): Field-measured vs. predicted volume/ha values and residuals for all plots (adjusted $R^2 = 0.56$)

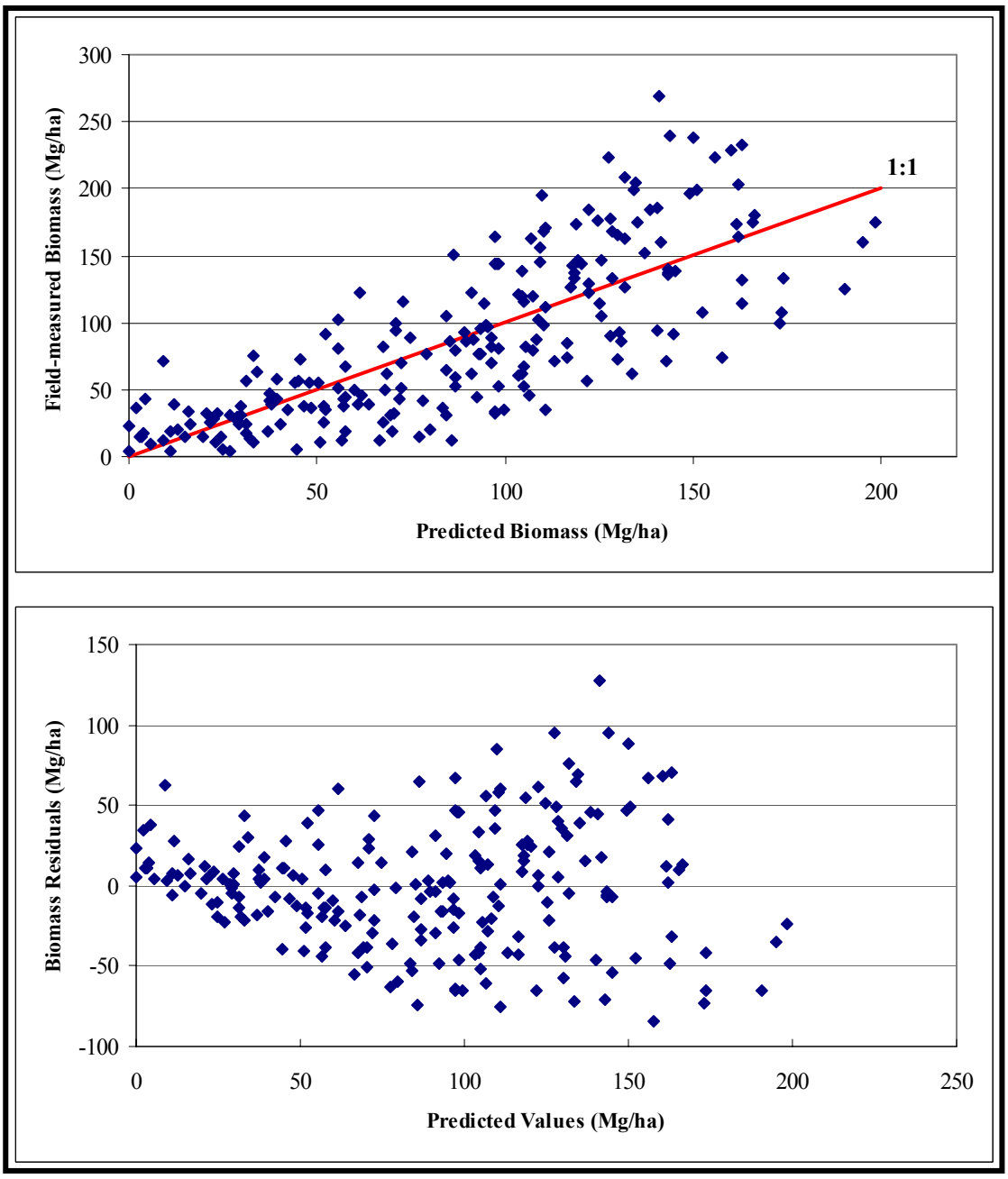


Figure 3.13 2-class biomass model (27,050 segments; 0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.60$)

The larger segment size only decreased precision by 0.5% and 1.6% in the case of volume and biomass, respectively. This could indicate that the modeling approach was similar across the two selected segment sizes. The choice of operational or applied segment size within this range is therefore potentially open to the user, and can be based on operational (management) and scaling factors. It can therefore be concluded that volume and biomass modeling, based on a single

remote sensing data source, coupled with BAF field plots, closely matched traditional, field-based estimates. A lidar-based approach, from forest segmentation and per-segment volume and biomass modeling, through to object-oriented segment classification, could encapsulate a comprehensive, unbiased inventory tool.

Table 3.16 Lidar model and BAF plot volume and biomass estimates for the entire study area

Method (945.80 ha total)	**Model**	**27,050 segments (0.035 ha/segment) 35.10 ha adjustment***	**6,687 segments (0.141 ha/segment) 13.07 ha adjustment***
Lidar + BAF plot model estimates	Volume (m^3/ha)	130,556.23 ± 52,200.69	128,881.96 ± 52,442.62
	Biomass (Mg/ha)	79,142.21 ± 36,271.91	86,075.24 ± 36,880.10
BAF plot estimates	Volume (m^3/ha)	134,372.81 ± 79,030.58	
	Biomass (Mg/ha)	82,832.46 ± 57,132.78	

* Segment area missing independent variable values and hence not modeled. Adjusted by using average per-segment volume/ha and biomass/ha values

3.4 Conclusions

Grid-cell volume and biomass modeling based on lidar distributions have been implemented successfully by Means *et al*. (2000) and Næsset (2002). These studies were limited to coniferous species, but R^2 values upwards of 0.90 bode well for future lidar distributions studies. This study explored an extension of the grid-cell approach to unique forest segments and a deciduous-coniferous forest mix. Hierarchical, multiresolution segmentation results were used as homogenous units for the extraction of lidar distributions, while basal area plots were used as field data for model fitting and validation. No distinct differences were found for volume and biomass modeling attempts across increasing segment sizes (0.035 – 5.632 ha/segment, although adjusted R^2 values generally decreased and RMSE values generally increased with increasing segment size.

This lack of modeling differences across varying segment sizes was attributed to the hierarchical nature of the segmentation algorithm, which resulted in small homogenous segments that served as building blocks for larger segments. Within-segment homogeneity already was minimized at smaller average segment sizes, resulting in no definitive difference in modeling results as segment size increased through recombination of smaller segments. However, segment-based

modeling efforts were distinctly better than those found for existing, operational forest stands in the study area. This was attributed to the larger within-stand height variation in the case of existing stands when compared to the variation found within homogenous segments.

Modeling results were very promising, even though coniferous and combined adjusted R^2 values for volume and biomass were lower than those found in other published studies. Lower coniferous R^2 values were attributed in part to a smaller range of volume and biomass observed values, as well to the inherent variability found in Virginia Piedmont forests. Adjusted R^2 values for deciduous segments were higher than those found for a comparable, plot-level study in the same area. This result indicated that a segment-level approach to deciduous volume and biomass modeling is a potential improvement over plot-based approaches. Given that volume- and biomass modeling were performed by using only height-related values, high R^2 values in the context of this study were unlikely. This was due to the nature of the modeled field data, which were based on diameter-at-breast-height (biomass) as well as height (volume).

RMSE values compared favorably with those found in other distributional modeling studies. Low RMSE values indicated that models could find applicability in an operational context, even when low R^2 values were considered. A comparison between estimates from modeling based on lidar and BAF plot data, and stand-alone BAF plot data for the 945 ha study area was promising, with differences smaller than 5%. This indicated that stand-alone plot and lidar model estimates were in fact relatively similar, again boding well for possible future application of models in an operational context.

Forward and Mallow's Cp selection proved successful in the reduction of independent variables from as many as 75 initial height distributional variables to the fewer than 10 used for final modeling. Further variable reduction through correlation analysis proved critical to the process of reducing variables. Final model selection from all candidate models was based on Mallow's Cp, adjusted R^2, RMSE values, and model simplicity. All criteria proved useful and even necessary in order to select a single best option from many Mallow's Cp recombined variable models. Final variables spanned the whole spectrum of possibilities, from general mean and range height values, to more abstract coefficient of variation and standard deviation-type

variables. Percentiles, both regular and canopy cover percentiles, also were well represented. The inclusion of reflectance variables was interesting since few studies (Means *et al*., 1999; Brandtberg *et al*., 2003) have included reflectance values as part of forest biophysical modeling. The wide range of variables indicated that sophisticated lidar scanners, that can record multiple returns and reflectance associated with each lidar hit, might well be necessary for effective modeling of variation in more complex forests.

Per-segment volume and biomass modeling has the potential of constituting part of a complete lidar-based inventory. Segmentation, volume and biomass modeling, and object-oriented classification could form a cohesive approach to forest inventory using remote sensing data, specifically lidar technology. Segmentation of lidar-derived data has the benefit of establishing homogenous objects for subsequent volume and biomass modeling, resulting in scalable units that could be conglomerated along with all associated per-segment estimates. A variable forest stand could thus be modeled at a more homogenous sub-stand level. Although this was not investigated, it could be that stand-level estimates will be more precise due to such a scalable, integrated approach. It seems likely that limited fieldwork will be required for any given region. Fieldwork might include limited segmentation verification, establishment of volume-lidar distribution regression equations, and collection of forest type information. Established distributional volume and lidar equations could be applied for future stands and derived segments, with periodic verification using either fixed or variable plots. Models likely would have to be calibrated or even re-developed for different regions, as it seems that results are geographically dependent (Means *et al*., 2000; Makela and Pekkarinen, 2001; Næsset, 2002; Pekkarinen, 2002).

Issues that potentially are critical to operational implementation include determination of the number of plots required for proper model fitting and the ideal segmentation size for model development and application. Although average segment size did not affect modeling results for this study area and approach, lower within-segment variances at smaller segment sizes theoretically define a range of segment sizes that could be better suited to volume and biomass modeling. Differences among segment sizes were not evident due the hierarchical nature of the segmentation algorithm, but it is likely that segment size selection for modeling is dependent on

stand-makeup, with more diverse stands requiring smaller segments with less within-segment variability, and *vice versa*.

Forest managers strive to obtain estimates of volume-by-type with high economical and statistical efficiency. Volume-by-segment estimations presented here potentially can be extended to object-oriented classification and subsequent volume-by-type assignment. Such an approach could constitute a stand-alone forest inventory based on remote sensing inputs, but the associated precision and cost are two issues that will need further consideration. Extension to net primary productivity modeling efforts at larger scales is likely, with remote sensing data well suited to scaling of measurements and results.

3.5 Acknowledgements

This research was made possible by funding from NASA (grant # NG65-10548; NG613-03019), the McIntire-Stennis Research Program (grant # VA-136589), the Forestry Department and Graduate Student Association at Virginia Polytechnic Institute and State University, and the Potomac chapter of the American Society for Photogrammetry and Remote Sensing. Field data collection was supported by the Virginia Department of Forestry, specifically Dr. John Scrivani, Todd Edgerton, Ralph Toddy, and Wayne Bowman (VDOF). Drs. Richard Oderwald (Virginia Polytechnic Institute and State University) and Sorin Popescu (Texas A&M University) provided invaluable assistance with statistical and lidar analyses. Amy Zhang from the Statistics department at Virginia Polytechnic Institute and State University served as statistical consult.

3.6 Literature Cited

Avery, T.E., and H.E. Burkhart, 1994. Forest Measurements. 4th Edition. McGraw-Hill, Boston, USA. 408 pp.

Baatz, M., and A. Schäpe. 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: STROBL, J. *et al.* (Hrsg.): Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg, September 2000, Karlsruhe, Herbert Wichmann Verlag: 12–23.

Brandtberg, T., T.A. Warner, R.E. Landenberger, and J.B. McGraw, 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment* 85 (2003): 290-303.

Burrough, P.A., and R.A. McDonnel, 1998. Principles of geographical information systems. Oxford University Press. Oxford, England. 333 pp.

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.

CORS, 2000. National Geodetic Survey's (NGS) Network of Continuously Operating Reference Stations (CORS). Web site: http://www.ngs.noaa.gov/CORS/. Accessed: January 19, 2004.

Douglas, T.E., D.L. Evans, K.L. Belli, and S.D. Roberts, 2003. Classification of pine and hardwood by the distribution and intensity of lidar returns. ISPRS "Three dimensional mapping workshop from InSAR and LIDAR", June 17-19, 2003, Portland, Oregon, USA. 5pp.

Drake, J.B., R.O. Dubayah, D.B. Clark, R.G. Knox, J.B. Blair, M.A. Hofton, R.L. Chazdon, J.F. Weishampel, and S.D. Prince, 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment* 79 (2002): 305-319.

Draper, N.R., and H. Smith, 1981. Applied regression analysis. 2nd Edition. John Wiley and Sons, Inc. New York, USA. 709 pp. (p. 294 – 303).

eCognition User's Manual, 2003. Definiens Imaging (www.definiens.com).

Engdahl, M.E., J. Pulliainen, and M. Hallikainen, 2003. Combined land-cover classification and stem volume estimation using multitemporal ERS tandem INSAR data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Hodgson, M.E., J.R. Jensen, L. Schmidt, S. Schill, and B. Davis, 2003. An evaluation of LIDAR- and IFSAR-derived digital elevation models in leaf-on conditions with USGS Level 1 and Level 2 DEMs. *Remote Sensing of Environment* 84 (2003): 295–308.

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

Kayitakire, F., C. Farcy, and P. Defourny, 2002. Ikonos-2 imagery potential for forest stands mapping. ForestSAT Symposium, Heriot Watt University, Edinburgh, August 5-9, 2002. 11 pp.

Kellndorfer, J.M., and F.T. Ulaby, 2003. Forest biomass inversion from SAR using object oriented image analysis techniques. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Kressler, F., Y. Kim, and K. Steinnocher, 2003. Object-oriented land cover classification of panchromatic KOMPSAT-1 and SPOT-5 data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 7 pp.

Lefsky, M.A., W.B. Cohen, S.A. Acker, G.G. Parker, T.A. Spies, and D. Harding, 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir and western hemlock forests. *Remote Sensing of Environment* 70: 339-361.

Lefsky M.A., W.B. Cohen, G.G. Parker, and D.J. Harding, 2002a. Lidar remote sensing for ecosystem studies. *Bioscience* 52 (1): 19-30.

Lefsky M.A., W.B. Cohen, D.J. Harding, G.G. Parker, S.A. Acker, and S.T. Gower, 2002b. Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology & Biogeography* 11 (2002): 393-399.

Li W., G.B. Benie, D.C. He, S. Wang, D. Ziou, and Q.H.J. Gwyn, 1999. Watershed-based hierarchical SAR image segmentation. *International Journal of Remote Sensing* 20 (17): 3377-3390.

Maclean, G.A., and W.B. Krabil, 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.

Makela H., and A. Pekkarinen, 2001. Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing of Environment* 77:66-75.

Means, J.E., S.A. Acker, D.J. Harding, J.B. Blair, M.A. Lefsky, W.B. Cohen, M.E. Harmon, and 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.

Means J.E., S.A. Acker, B.J. Fitt, M. Renslow, L. Emerson, and C.J. Hendrix, 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering & Remote Sensing* 66 (11): 1367-1371.

Montgomery, D.C, E.A. Peck, and G.G. Vining, 2001. Introduction to linear regression analysis. 3rd Edition. John Wiley and Sons, Inc. New York, USA. 641 pp. (P 299-301).

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80 (2002):88-99.

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

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

Nugroho, M., D.H. Hoekman, and R. De Kok, 2002a. Analysis of the forests spatial structure using SAR and Ikonos data. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 10 pp.

Nugroho, M., D.H. Hoekman, and R. de Kok, 2002b. Analysis of forest spatial structure using spatial decision rule. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 8 pp.

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

Pekkarinen A., 2002. Image segment-based spectral features in the estimation of timber volume. *Remote Sensing of Environment* 82: 349-359.

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

Popescu, S.C., R.H. Wynne, and J.A. Scrivani, 2004. Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, U.S.A. *Forest Science* (In press)

Riaño, D., E. Meierc, B. Allgöwerc, E. Chuviecoa, and S.L. Ustin, 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment* 86 (2003): 177–186.

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.

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.

Seielstad, C.A. and L.P. Queen, 2003. Using airborne laser altimetry to determine fuel models for estimating fire behavior. *Journal of Forestry* 101(4): 10-15.

Shandley, J., J. Franklin, and T. White, 1996. Testing the Woodcock-Harward image segmentation algorithm in an area of southern California chaparral and woodland vegetation. *International Journal of Remote Sensing* 17 (5): 983-1004.

Shankar, B.U., C.A. Murthy, and S.K. Pal, 1998. A new gray level based Hough transform for region extraction: An application to IRS images. *Pattern Recognition Letters* 19 (1998): 197-204.

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

Weishampel, J.F., K.J. Ranson, and D.J. Harding, 1996. Remote sensing of forest canopies. *Selbyana* 17: 6-14.

CHAPTER 4

OBJECT-ORIENTED CLASSIFICATION OF FOREST SEGMENTS USING SMALL-FOOTPRINT LIDAR DISTRIBUTIONAL AND CANOPY HEIGHT MODEL DATA

Abstract. This study evaluated the potential of a lidar-based (small-footprint, multiple return), object-oriented approach to deciduous and coniferous forest classification. The study area is located in Appomattox Buckingham State Forest in the Piedmont physiographic province of Virginia, U.S.A, at 78°41' W, 37°25' N. Vegetation is composed of various coniferous, deciduous, and mixed forest stands. The eCognition segmentation algorithm was used to segment a lidar-derived canopy height model (CHM). Between- and within-segment variances of the CHM were used to select segments for subsequent classification efforts, ranging in size from 0.035 ha/segment to 5.632 ha/segment. Basal area factor plots were used to assign segments to 2-class (deciduous-coniferous) and 3-class (deciduous-coniferous-mixed) forest definitions. Per-segment lidar point-height and CHM distributional parameters were used as input to a discriminant classification. The 2-class classification scheme afforded better accuracies than the 3-class approach for discriminant analysis. Lidar point-height-based classification yielded better overall accuracies (89.2%) than the CHM-based classification (79%) for the 2-class scheme. The lack of significant differences between accuracies for varying segment sizes, and between segment- and stand-based classifications, were attributed to the hierarchical nature of the segmentation algorithm and the existing definition of operational forest stands on a per-species or type basis, respectively. Variables that were useful for discrimination between deciduous and coniferous groups included the standard deviation of second return vegetation heights, various reflectance (1,064 nm) metrics, as well as height percentiles. These selected variables hinted at the importance that mid-canopy structural information and type-specific, near-infrared intensity values have to such a classification approach. Lidar-based, object-oriented classification of deciduous and coniferous types has significant potential, especially when combined with forest biophysical parameter modeling based on lidar data as a single source input.

4.1 Introduction

Accurate forest type discrimination is crucial to forest inventory, pest and environmental stress management, carbon sequestration monitoring, wildlife habitat definition, and management of human impacts on a forest environment. Forest classification is especially important for vast tracts of land, where accessibility may be limited, while remote sensing products are available. Such a classification enables managers to derive forest type maps from remotely sensed imagery, or ecologists to attribute carbon stores to specific forest groups. Traditional approaches have been based on multispectral remote sensing inputs (Nelson *et al.*, 1984; Shen *et al.*, 1985; Franklin, 1994; White *et al.*, 1995), while type definition using lidar height densities (hits/m^3) and reflectance also has come to the fore (Douglas *et al*., 2003). Such approaches mainly have been pixel- or stand-based, while segment-based classification has gained popularity as an alternative method (Willhauck, 2000; Heyman *et al*., 2003).

This latter approach results in real-world object classification, with procedures applied to homogenous units. Results also are representative of real-world objects and very often are devoid of the salt-and-pepper appearance so often found with pixel-based classifiers. While accuracies of pixel-based and object-oriented approaches are similar in many cases, the visual, realistic classification rendering of the object-oriented approach is of importance. An object refers to a spatial entity that is homogenous in terms of a selected property, as opposed to the traditional, continuous fields approach found in spatial analysis (Burrough and McDonnel, 1998). Willhauck (2000) compared standard maximum likelihood classifiers vs. object-oriented classification in the Argentine Nothofagus forests (water, non-forest, Nothofagus Pumilio, and Nothofagus Antarctica classes) using SPOT data. The object-oriented classification was based on multiresolution, hierarchical segmentation (eCognition algorithm). The object-oriented approach performed better than the maximum likelihood approaches in terms of classification accuracy at 93% and 96%, respectively. The author concluded that the object-oriented approach also resulted in a better visual result, while the maximum likelihood classification had a distinct salt-and-pepper appearance. Heyman *et al*. (2003) used a per-segment approach to improve aspen (*Populus tremuloides*) mapping in Oregon, USA. Segmentation of aerial color-infrared photographs (USGS National High Altitude Photography Program) was based on a histogram

thresholding method, using hue and saturation values. Segments were classified using unsupervised ISODATA clustering. The authors achieved 88% accuracy for mapping aspen segments into three broad categories (no aspens; 0 - 50% aspens; 50 - 100% aspens).

Citing forest mapping and classification as main goals, Hill (1999) applied a segmentation-based approach to delineate swamp-forest, lower-, middle, and upper floodplain forests in southeast Peru. Spatial low-pass filtering, followed by an edge detection and region growing segmentation algorithm, were applied to Landsat TM data. An accuracy of 91% was obtained when lower-level segments were aggregated into six meaningful forest classes. These classes included lower-, middle, and upper floodplain forest, *terra firme* (clay) forest, permanently flooded swamp forest, and seasonally flooded swamp forest. The author underlined the usefulness of aggregation and disaggregation of segments at various scales to ecological managers. Therefore high accuracies, real-world object extraction, and hierarchical aggregation, important to scaling attempts, can all be listed as advantages of object-oriented classification.

Derivation of unique objects or segments is required as a precursor to any object-oriented classification. Segmentation approaches include segmentation based on graph theory (Cheevasuvit, 1990), knowledge-based segmentation (Ton *et al.*, 1991), unsupervised segmentation using nonlinear regression (Acton, 1996), the Woodcock-Harward centroid-linkage algorithm (Shandley *et al.*, 1996), Markov random field model-based segmentation (Smits and Dellepiane, 1997; Sarkar *et al.*, 2002), a Hough transformed-based approach (Shankar *et al.*, 1998), watershed-based hierarchical segmentation (Li *et al.*, 1999), multiresolution, hierarchical segmentation (Baatz and Schäpe, 2000), and iterative edge-region co-operation (Kermad and Chehdi, 2002). Many studies applied segmentation as a preprocessing step to forest classification, as well as other object-based analyses, e.g., per-segment volume estimation. Object-oriented approaches to forest classification generally performed well when compared to more traditional, pixel-based approaches.

Jaakkola (1989) used an edge-preserving smoothing filter, followed by gradient filters for object delineation and a histogram segmentation method to generate segment signatures for forest compartment delineation. Maximum likelihood classification of segments performed slightly

worse than a contextual maximum likelihood classifier (77% vs. 88%), but executed up to 15 times faster. Schwarz *et al.* (2001) compared a pixel-based, parallelepiped supervised classification to eCognition's object-oriented approach for storm loss detection (forest vs. damaged forest) in Swiss alpine forests. The two approaches resulted in similar accuracies of 97% for IKONOS imagery with high spatial resolutions of 4 m (multispectral) and 1 m (panchromatic). Similar accuracies also were found when using SPOT imagery with a 10 m resolution (96%). This result confirmed that object-oriented approaches could yield accuracies comparable to pixel-based classification methods when high spatial resolution data are used. Abkar *et al.* (2000) used a likelihood-based segmentation approach, whereby posterior probabilities were calculated on a per-object basis, rather than per-pixel. This method was used to estimate deforestation extent and location in Thailand by comparing it to a per-pixel maximum likelihood classification of the same area. The segmentation approach improved accuracy by 10%, which was deemed significant. Kayitakire *et al.* (2002) used the same algorithm to map mixed oak (*Quercus* spp.), spruce (*Picea* spp.), beech (*Fagus* spp.), and pine (*Pinus* spp.) forest stands in Belgium. Overall accuracies of 88% (per-pixel clustering) and 83.3% (per-parcel derived map) were found. It should be noted that, for all object-oriented classifications, wrong classification of a parcel results in all the parcel pixels being misclassified, as opposed to single pixel misclassification.

Dorren *et al*. (2003) compared Landsat-based object-oriented forest classification (eCognition) to a pixel-based maximum-likelihood approach in Austria. Classification was performed for broadleaf (*Fagus silvatica*, *Acer pseudoplatanus*, *Tilia cordata*, *Fraxinus excelsior*), mixed (*Picea abies*, *F. silvatica*, *Abies alba*) and spruce (*P. abies*) forests. The pixel-based maximum-likelihood approach performed marginally better than the object-oriented classification approach (73% vs. 70%). However, the authors found that forest stand type maps based on object-oriented classification showed better agreement with observed in-field trends, especially in the case of variable stands on steep slopes.

Although previous studies have found acceptable per-object classification accuracies, most incorporated traditional multispectral data, e.g., Landsat TM and SPOT imagery. Multiple data sources, such as optical and lidar data, are often required if the goal extends past classification

(Willhauck, 2000; Dorren *et al*., 2003) to biophysical parameter modeling (Lefsky *et al.,* 2002; Næsset, 2002; Popescu *et al*., 2004). Douglas *et al.* (2003), however, used lidar canopy density and reflectance, derived from small footprint lidar data, to classify mature pine (15 plots), immature pine (14 plots), and mature hardwood (15 plots) stands in Mississippi. Analyses were limited to the forest canopy by only using lidar returns in the upper 50% of the total tree height. The authors implemented a discriminant classification with the number of lidar hits per cubic meter and the variance of the intensity data within the canopy of each plot as variables. Accuracies of 100% (mature pine), 85.7% (immature pine), and 93.3% (mature hardwood) were achieved, with an overall accuracy of 86.4%. Overall accuracy was 65.9% when only using the number of hits per cubic meter as independent variable. Such an approach bodes well for the application of lidar data, a structural data source due to its height information, to forest classification. Logical extensions of such a lidar-based classification include (i) the use of complete lidar distributional data, not limited to canopy returns, (ii) implementation of a range of distributional parameters, e.g., range, mode, skewness, kurtosis, and percentiles, and (iii) the evaluation of lidar-based, object-oriented forest classification, an attractive goal given the scalability of results through recombination of objects. Given the applicability of lidar data to the measurement of forest biophysical parameters, extension of this data source to classification has far reaching consequences. Lidar data conceivably could form the basis of a complete per-object, forest inventory.

Object-level estimates of forest biophysical characteristics need to be assigned to at least forest types, a task suited to object-oriented classification. Previous results have proven the usefulness of object-oriented approaches to classification, even highlighting its correspondence in accuracies to traditional, pixel-based classification. One could argue that object-oriented classification likely is a better proposition than pixel-based classification in most forest inventories, given the nature of the input data as defined objects and the necessity for homogenous results, as opposed to salt-and-pepper output. However, it should be noted that a species-level classification and volume assignment will likely require either detailed object delineation, or pixel-based, hyperspectral classification, or both. The main objective of this study was to assess the potential utility of lidar-based, object-oriented classification for mapping 2-class (deciduous-coniferous) and 3-class (deciduous-coniferous-mixed) segments in the Virginia

Piedmont. A secondary objective involved the comparison of discriminant classification using lidar point-height distributions on the one hand, and height distributions derived from a 1 m canopy height model (CHM) on the other. Addressing these two goals could help in determining the extent to which lidar data can be used as single remote sensing data source for forest biophysical parameter modeling (Lefsky *et al.,* 2002; Næsset, 2002; Popescu *et al*., 2004), followed by forest type classification in an operational forestry context.

4.2 Material and Methods

4.2.1 Study Area

The 946 ha (2,338 acres) study area is located in Appomattox Buckingham State Forest (Appomattox County) in the Piedmont physiographic province of Virginia, southeastern U.S.A at 78°41' W, 37°25' N (Figure 4.1). The mean elevation of the study area is 185 m (606 ft.), with minimum and maximum elevations of 133 m (436 ft.) and 225 m (738 ft.), respectively. Local topography can best be described as gentle rolling slopes and flat terrain. Vegetation is composed of various coniferous (*Pinus taeda, P. virginiana, P. echinata*, and *P. strobus*), deciduous (*Quercus coccinea, Q. alba,* and *Liriodendron tulipifera*), and mixed forest stands.

4.2.2 Available Data

Field data consisted of 256 mapped basal area plots (BAF; basal area factor 10) on a 16 columns by 16 rows, 201.17 m (10 chains) grid. Field data were collected during the summer, fall, and winter months (May – December) of 2003. A Magellan Sportrak Pro GPS unit (WAAS enabled) was used to navigate to within 6 feet of each designated plot center, after which a Corvallis Microtechnologies, Inc. (CMT) March II GPS unit was used to accurately map the established plot center (120 second static point collection). All GPS plot center locations were differentially corrected using data from the National Geodetic Survey's Continually Operating Reference Stations (CORS, 2000) and Corvallis Microtechnologies, Inc. PC-GPS software (Version 3.7; Corvallis Microtechnologies, Inc.).

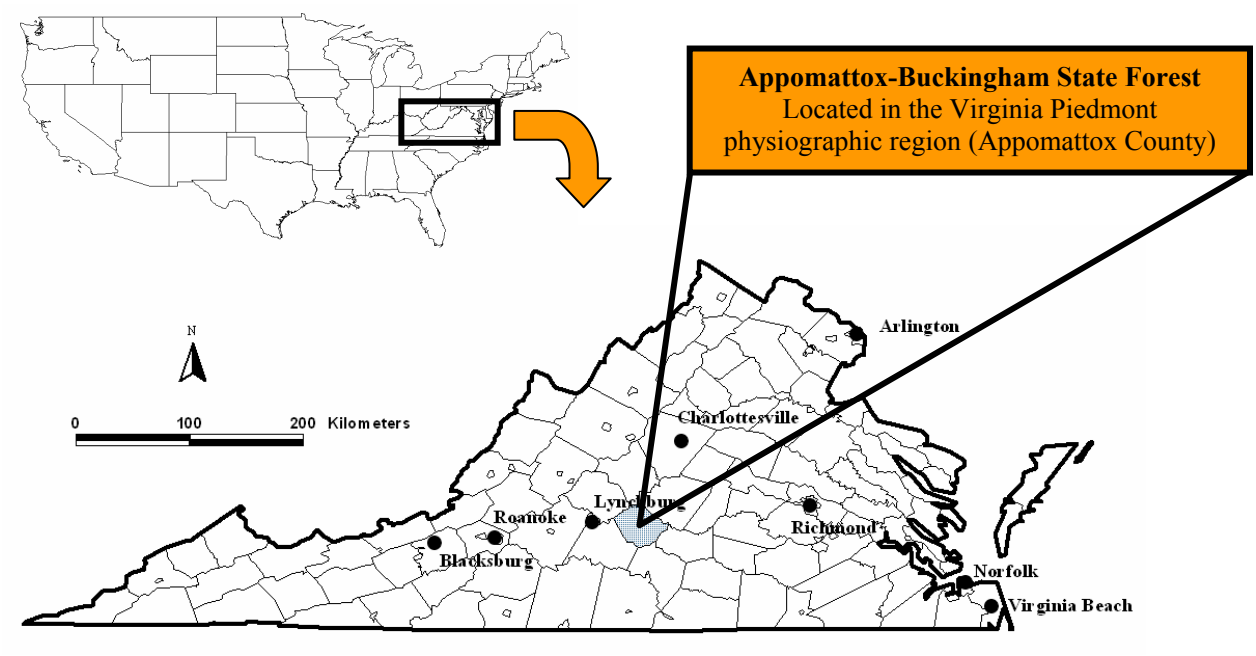


Figure 4.1 Study Area: Appomattox Buckingham State Forest

The following reference stations from the CORS-network were used based on data availability:

- Richmond, VA (37° 32' 16.42936" N; 77° 25' 46.77568" W)
- Fan Mountain, VA (37° 52' 43.46536" N; 78° 41' 37.24955" W)
- Blacksburg, VA (37° 12' 21.63726" N; 80° 24' 52.27622" W)

For each sampling point, the following data were collected (**Appendix A**, an actual data sheet):

- Plot basal area ("in-tree" count); diameter at breast height (dbh) > 5 inches (12.7 cm) (10-factor prism)
- Dbh and height for all plot trees tallied (diameter tape and Vertex hypsometer)
- Azimuth and distance from plot center to each tallied tree (SUUNTO compass and Vertex hypsometer's range finding function)
- Species codes of tallied trees (**Appendix B**)
- Differentially corrected GPS point at plot center (CMT's March II GPS unit)

Each basal area plot was mapped based on plot center coordinates, azimuths, and distances to tallied trees. A total of 37 non-forest BAF plots had to be discarded due to their location on private land or having volume and biomass values of zero. Zero-value plots could not be used in subsequent model development since it was impossible to assign a forest type in such cases, with no trees being tallied. This left a total of 219 BAF plots (Figure 4.2) that ultimately were used in the statistical analysis (**Appendix C**). Descriptive statistics for all basal area plots are given in Table 4.1.

Table 4.1 General descriptive information for deciduous, coniferous, and mixed plots, related to volume, biomass, and basal area properties

Class	Type	Parameter	Minimum	Maximum	Average	σ
2-class	Deciduous plots (140)	Volume/ha (m^3/ha)	6.94	350.65	157.64	84.14
		Biomass/ha (kg/ha)	11,105.69	269,006.23	113,599.00	58,602.63
		Basal area/ha (m^2/ha)	2.30	34.44	16.32	7.84
	Coniferous plots (79)	Volume/ha (m^3/ha)	8.32	350.93	114.49	75.44
		Biomass/ha (kg/ha)	4668.06	155,558.33	41,468.40	26,641.42
		Basal area/ha (m^2/ha)	2.30	36.73	14.24	7.91
3-class	Deciduous plots (112)	Volume/ha (m^3/ha)	6.94	350.65	156.16	89.32
		Biomass/ha (kg/ha)	11,105.69	269,006.23	117,312.63	62,534.74
		Basal area/ha (m^2/ha)	2.30	34.44	15.97	8.21
	Coniferous plots (56)	Volume/ha (m^3/ha)	8.32	278.99	100.45	66.42
		Biomass/ha (kg/ha)	4,668.06	81,645.10	33,655.42	19,952.03
		Basal area/ha (m^2/ha)	2.30	36.73	13.61	8.11
	Mixed plots (51)	Volume/ha (m^3/ha)	31.68	350.93	156.85	72.60
		Biomass/ha (kg/ha)	20,062.53	175,747.21	81,493.05	38,927.59
		Basal area/ha (m^2/ha)	4.59	36.73	16.84	6.68

Plots were assigned to 2- and 3-class forest type schemes based on basal area percentages. “Deciduous” or “Coniferous” types were defined as plots that had 50% or greater basal area contribution from either deciduous or coniferous species, respectively. A “Mixed” class was added to the 3-class type designation for plots that had less than 90% basal area contribution for either deciduous and coniferous species. A 90% cut-off was based on sample numbers for the 2- and 3-class schemes. The 2-class analysis consisted of 140 deciduous and 79 coniferous plots, while the 3-class analysis consisted of 112 deciduous, 56 coniferous, and 51 mixed plots. This

allowed for volume and biomass model development based on adequate plot samples (> 30) for both the 2- and 3-class analysis. There were only 25 mixed plots when a 75% cut-off was used, making this class redundant and too small for viable statistical analysis.

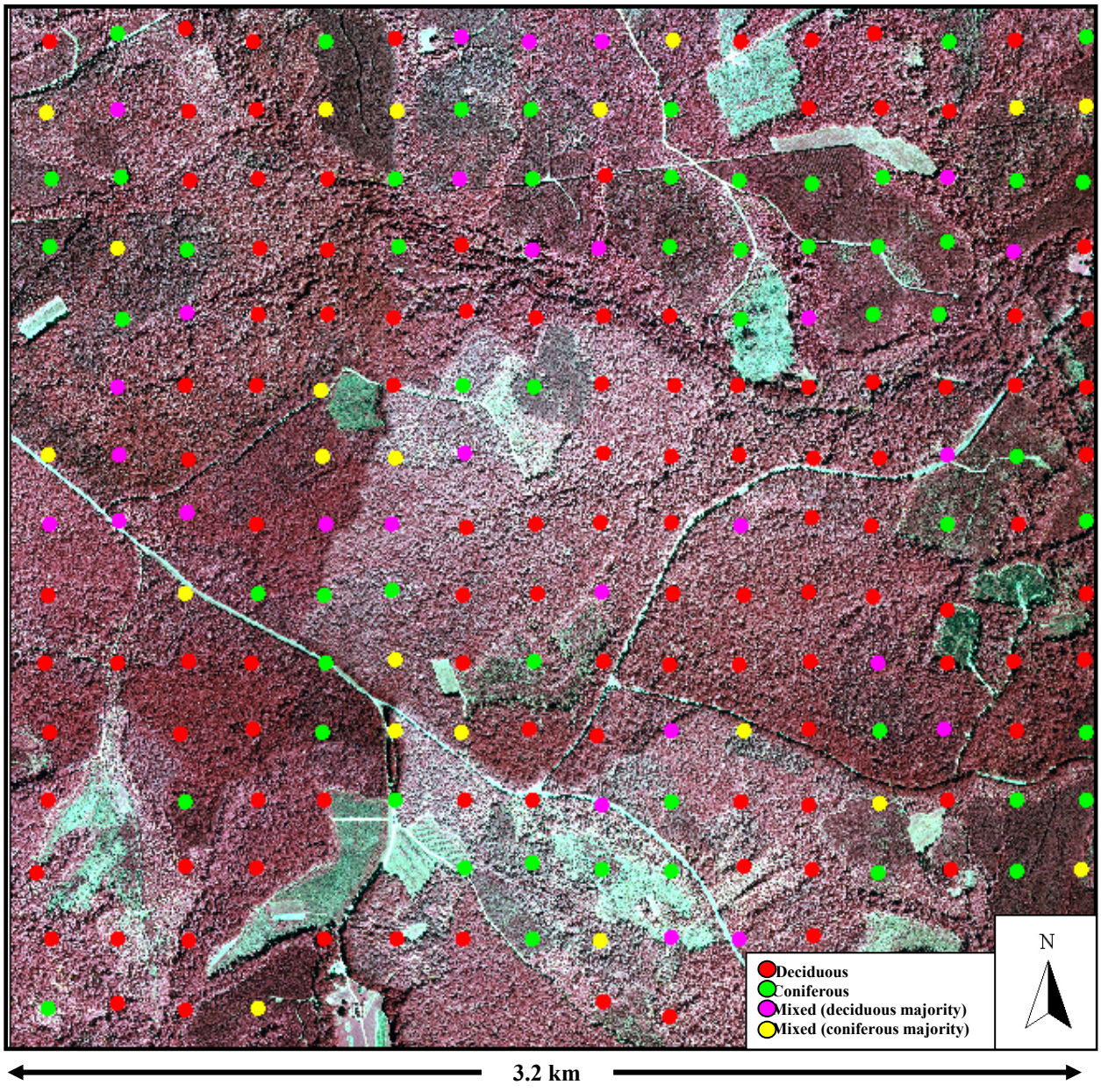


Figure 4.2 Mapped BAF plot locations on a 1999, leaf-on, color-infrared aerial photograph of the study area (bottom-middle plots missing plots due to locations on private land)

Lidar data were acquired by Spectrum Mapping, LLC using the DATIS II (small-footprint, high-density, multiple return) system. The lidar data were acquired on September 9, 2002, centered at 78°40'30" W, 37°25'9" N, and covered an area of approximately 958 ha (2,367 acres). Specifications of the lidar data set are given in Table 4.2.

Table 4.2 DATIS II lidar data set characteristics

Characteristics	Specification
Laser altitude	2,000 m above ground level
Laser scan field-of-view	75° maximum
Swath width and centerline spacing	800 m (2,625 ft.) and 400 m (1,312 ft.)
Scan rate	25 Hz
Laser pulse rate	35 kHz
Scan angle	± 13.5°
Returns	≤ 5
Resolvable distance between returns	0.75 m
Footprint	0.46 m (1.51 ft.)
Spacing across / along track	1 m (3.3 ft.) / 2 m (6.6 ft.)
Accuracy (X,Y,Z)	X,Y: 0.5 m; Z: 0.15 m
Wavelength	1,064 nm

4.2.3 Lidar Data Pre-Processing

A canopy height model (CHM) was needed for segment derivation as a precursor to object-oriented classification. First and last (vegetation-removed) returns from the lidar data set were extracted and corrected for possible errors (suspect low and high, or "bird" hits). Peripheral outlier height values with a low frequency and a distinct difference (> 6 m) from the next smallest or largest value were removed as outliers. This resulted in the removal of one return smaller than –75 m and six returns larger than 31 m. The first returns were median-filtered by 1 m grid cells in order to remove per-cell values that were redundant to subsequent interpolation procedures. First and last returns were interpolated to a 1 m spatial resolution grid using regular Kriging, since Popescu *et al.* (2002) found this to be the most accurate interpolation technique using similar data over the same study area. This approach effectively addressed instances where a 1 m grid cell lacked an original input value. The resultant 1 m resolution was detailed enough to detect road and stand breaks in the segmentation process. It had the additional benefits of requiring less computing power, as opposed to a 0.5 m grid, and likely produced a smoother canopy digital terrain model (DTM) and ground digital elevation model (DEM). Interpolation was performed using Surfer 7.0 software (Golden Software, Inc.). It was assumed that first return

data is representative of top-most canopy heights, while last, vegetation–removed returns were attributed to ground hits. The differenced first- and last return surface (CHM) was used as input to the eCognition segmentation and object-oriented classification algorithm. This allowed for extraction of forest segments based on height homogeneity and distinct stand breaks, e.g., roads and slope breaks.

The distributional classification approach, based on height distributional parameters, required that lidar data be processed on a per-return basis in order to retain information related to the return hierarchy. Peripheral outlier height values again were removed for all return data sets, based on the same approach as in the case of the CHM. Ground hits were removed using Terrascan V. 003.002 (Terrasolid, Inc.) and MicroStation V. 08.00.04.01 (Bentley Systems, Inc.) software. This algorithm identifies ground hits based on iterative slope analysis of lidar returns. *Grid cell size* and *maximum slope of the area* are required input parameters. *Grid cell size* is the smallest cell size for which a ground return can be extracted. A cell size of 10 m was used in order to extract a maximum number of ground returns for the first (31,294,660), second (11,101,215) and third (2,121,989) returns. Grid cell sizes of 39 m and 119 m were used for the fourth (175,093) and fifth (5,379) returns, respectively. Larger grid cell sizes were required for the last two categories due to the small number of returns in each case. Each of these two cell sizes resulted from cases where the number of ground hits reached a maximum for the fourth and fifth returns, based on the assumption that most of the hits from these categories would be ground hits due their ranking in the return hierarchy. A slope percentage parameter of 35% was used as a maximum for the area, obtained from a USGS DEM. Ground returns constitute an important component of overall lidar distributional patterns and were retained as data sets on a per-return basis.

Non-ground hits, designated as vegetation hits, were normalized for terrain by calculating the actual return height above a lidar-derived 1 m digital elevation model (DEM) of the study area. The actual height of each vegetation hit was calculated as the difference between the vegetation hit and the bilinear interpolated height of the four corner cells of the DEM cell directly beneath each hit. This was done using Surfer V. 8.1 software (Golden Software, Inc.). This process

normalized all vegetation hits for varying terrain elevations, thereby enabling discriminant classification to incorporate actual lidar point-heights (Means *et al*., 2000).

4.2.4 Segmentation of the Study Area

Segmentation was performed using a multiresolution, hierarchical algorithm (eCognition) applied to the lidar-derived CHM of the study area. Lidar data were considered a structural component, ideally suited to defining unique structural segments. The eCognition algorithm required Color:Shape and Smoothness:Compactness ratios as input parameters. The Color:Shape ratio was set at 0.8:0.2, based on the recommendation of the developers (Baatz and Schäpe, 2000; eCognition, 2003) and evaluation of alternative parameter inputs. Smoothness of shape was considered more important than shape in a forestry context, since smooth, boundary-following segments are preferable to compact, blocky segments. The Smoothness:Compactness weight combination therefore was set at 0.8:0.2.

Although eCognition was chosen as the preferred segmentation approach, one could argue that the segmentation method is subordinate in importance to the utility that resultant objects have to analyses. Even though a multitude of segmentation approaches exist in literature, e.g., the Woodcock-Harward (centroid linkage) algorithm (Shandley *et al.*, 1996), a Hough transform-based approach (Shankar *et al.*, 1998), and watershed-based hierarchical segmentation (Li *et al.*, 1999), it is ultimately of great importance that segmentation results are robust. Other important factors are ease of operational use, widespread availability, and adequate software support. eCognition also was the preferred approach to segmentation because of its hierarchical nature, correspondence to input data, and because results from this algorithm have been proven in the natural resources context (Kayitakire *et al.*, 2002; Nugroho *et al.*, 2002a; Nugroho *et al.* 2002b; Engdahl *et al.*, 2003; Kellndorfer and Ulaby, 2003; Kressler *et al.*, 2003).

The decision of which segmentation results to use for model development was based on between- and within-segment variability of the CHM. Classification was performed on segmentation results where within-segment variability was smaller than between-segment variability, so as to minimize within-segment variability. The smallest selected segment size

corresponded to the area represented by average tallied tree distance from field-collected BAF plot centers, plus one and two standard deviations. This ensured that segments were representative of plot-level field data, based on corresponding areas. Ten average segment sizes, ranging from 0.035 ha/segment to 3.942 ha/segment, were chosen for subsequent forest type classification. These selections corresponded to segment sizes where within-segment variance was smaller than between-segment variance of the CHM heights. This was done in order to evaluate classification accuracies across a range of average segment sizes. The current Appomattox stand map (167 segments; 5.666 ha/stand) also was selected, as well as the segmentation result that corresponded to the number of operational stands (168 segments; 5.632 ha/segment). Operational stands were used in order to compare segmentation-based classification to stand-based classification. Figure 4.3 shows the 1 m CHM used for segmentation of the study area.

Vegetation and ground lidar data sets were extracted on a per-segment basis for all segmentation results using ARCGIS V. 8.3 software (ESRI). Resultant data sets were exported to SAS V. 8.02 software (Level 02M0; SAS, Inc.) for subsequent discriminant classification. BAF plots were assigned to the segment in which they were located. This was done through post-stratification after the determination of the best segmentation results (small within segment variance, large between segment variance). Segments without BAF plots were excluded from the classification process.

4.2.5 Derivation of Per-segment Lidar Point-Height and CHM Distributions

Lidar point-height and CHM distributions were extracted from segments that contained a BAF plot to be used in subsequent discriminant classification. Forest type classification was based on the assumption that distinct forest cover and structural types have different, unique distributions or lidar hit densities (Douglas *et al*., 2003). These distributions could be characterized through construction of height distributions for vegetation returns per-segment using the small-footprint DATIS II lidar data (Table 4.2). Lidar distributions should be representative of unique stand structural characteristics such as canopy closure and stand height distribution. Intermediate return distributions also could be useful, since these returns represent mid-canopy forest

structure. A multi-tiered forest structure will theoretically have many intermediate returns, while an even-aged, weed-controlled, and thinned coniferous stand might exhibit a majority of first and last returns, with few intermediate hits.

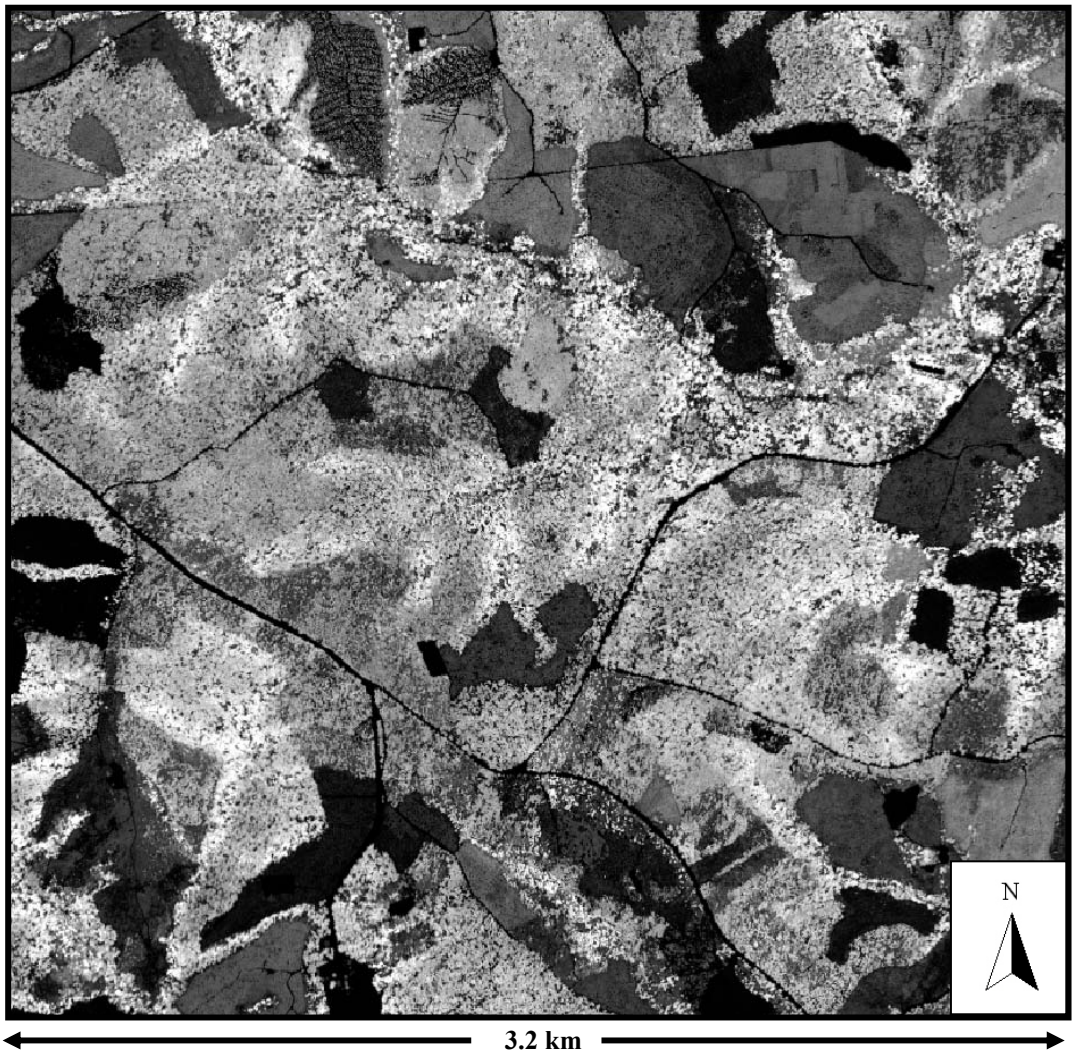


Figure 4.3 Segmentation results for 6,687 segments (0.141 ha/segment) overlaid on the canopy height model of the study area (946 ha)

Only first and second return variables from vegetation return data sets were used because there were many segments with missing values for the third through fifth return data sets. Distributional parameters included the mean, coefficient of variation, kurtosis, maximum, minimum, mode, range, standard error of the mean, skewness, standard deviation, number of observations, height percentile points at 10% intervals of height values, and canopy cover

percentiles. Canopy cover percentiles were based on the proportion of first returns smaller than a given percentage of maximum height. The ratio of the number of vegetation or ground hits and the total number of lidar hits per segment also was calculated. This was done for second, and third through fifth group vegetation hits, as well as first, second, and third through fifth group ground hits. The vegetation ratio for each segment was calculated as the ratio of the number of vegetation hits per segment and the total hits for that segment. Lidar intensity (Figure 4.4) distributional parameters values for the first and second returns included the intensity mean, median, coefficient of variation, maximum, minimum, range, standard error of the mean, and standard deviation.

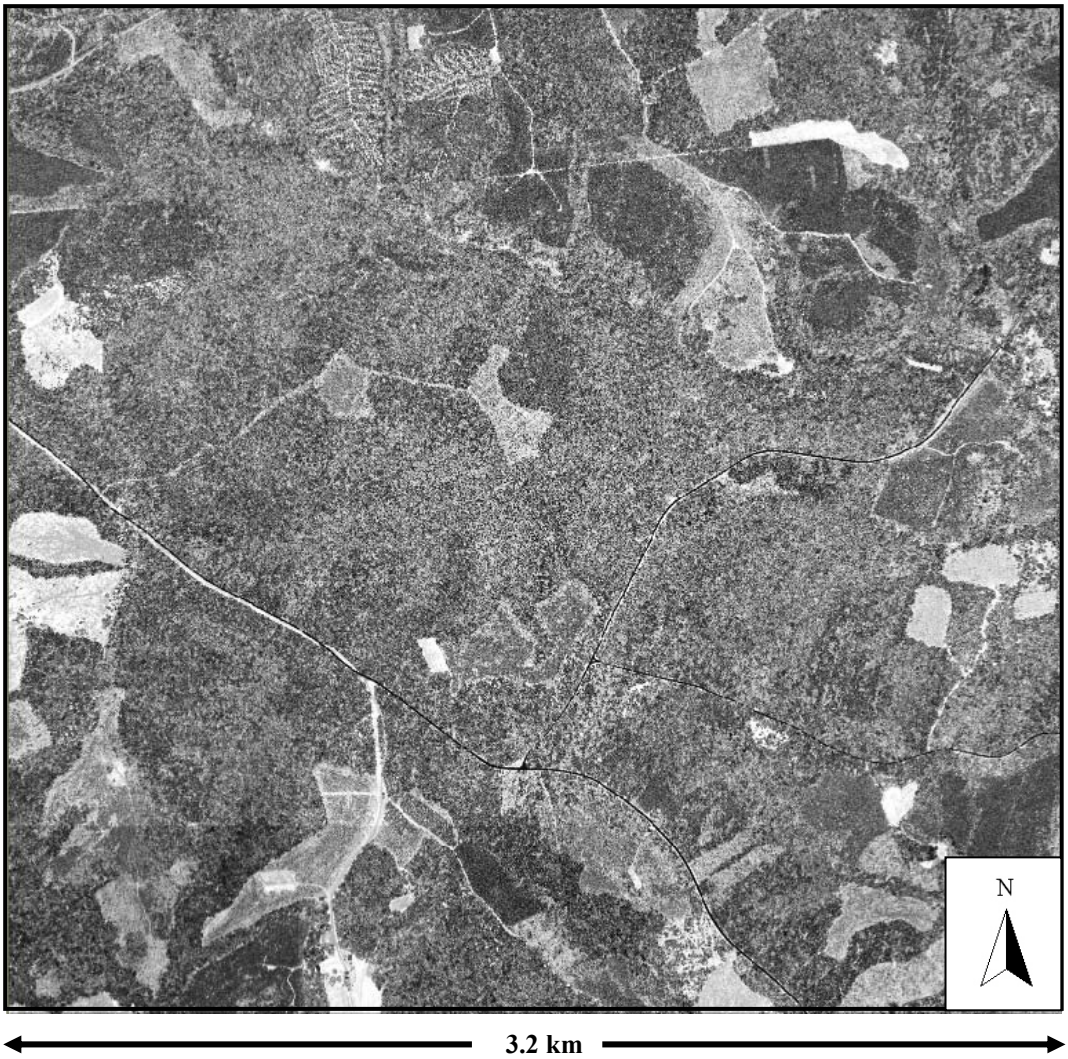


Figure 4.4 Lidar 1st return intensity image. Brighter tones are indicative of higher intensities

CHM distributions were derived in a similar fashion, but were constrained to only per-segment CHM heights. Distributional parameters were limited to first return types and canopy cover percentiles, due to the singular nature of the CHM, as opposed to multiple returns found in the case of lidar height points. Lidar point-height and CHM distributional parameters are shown in Table 4.3. These types of distribution metrics have been shown to be useful descriptors of stand characteristics for 10x10 m grid cells in Douglas-fir (*Pseudotsuga menziesii*), western Oregon stands (Means *et al.*, 2000) and 200 m^2 sample plots in Norway spruce, and Scots pine stands in southeast Norway (Næsset, 2002).

Table 4.3 Distributional variables used as input to discriminant classification. Only underlined variables were used in CHM-based distributional classification

1st Return Variables	2nd Return Variables	Other Variables
Height	**Height**	**Ratios and Canopy percentiles**
MeanVeg1 CVVeg1 KurtosisVeg1 MaxVeg1 MinVeg1 ModeVeg1 RangeVeg1 StdMeanVeg1 SkewnessVeg1 StdVeg1 MedianVeg1 P_Veg1_10 P_Veg1_20 P_Veg1_25 P_Veg1_30 P_Veg1_40 P_Veg1_50 P_Veg1_60 P_Veg1_70 P_Veg1_75 P_Veg1_80 P_Veg1_90	MeanVeg2 CVVeg2 KurtosisVeg2 MaxVeg2 MinVeg2 ModeVeg2 RangeVeg2 StdMeanVeg2 SkewnessVeg2 StdVeg2 MedianVeg2 P_Veg2_10 P_Veg2_20 P_Veg2_25 P_Veg2_30 P_Veg2_40 P_Veg2_50 P_Veg2_60 P_Veg2_70 P_Veg2_75 P_Veg2_80 P_Veg2_90	Veg2ratio Veg3_5ratio Grnd1ratio Grnd2ratio Grnd3_5ratio Vegratio Canopy10P Canopy20P Canopy30P Canopy40P Canopy50P Canopy60P Canopy70P Canopy80P Canopy90P
Reflectance (intensity)	**Reflectance (intensity)**	
MeanRef1 CVRef1 MaxRef1 MinRef1 RangeRef1 StdMeanRef1 StdRef1 MedianRef1	MeanRef2 CVRef2 MaxRef2 MinRef2 RangeRef2 StdMeanRef2 StdRef2 MedianRef2	

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

4.2.6 Classification Approach

A discriminant classification approach, similar to the one used by van Aardt and Wynne (2001), was applied to the lidar distributional parameters. Discriminant approaches, as opposed to non-parametric classifiers, have been shown to be better suited to the classification of high-resolution images where training data have a high degree of overlap in the feature space (Cortijo and De la

Blanca, 1999). Two data type approaches to discriminant object-oriented classification were used in an attempt to evaluate the usefulness of more complex lidar point-height-based and a simpler CHM-based technique. The first method consisted of discriminant classification using multiple returns and their associated distributions, while the second was based on distributional parameters derived from the 1 m CHM. The first approach constituted a very detailed, "high-resolution" effort that had more preprocessing and hardware requirements than the simpler CHM-based distributional attempt. Both classification approaches were applied to 2-class (deciduous-coniferous) and 3-class (deciduous- coniferous-mixed) classification schemes.

General classification statistics, including overall accuracy, user's and producer's accuracies, and Kappa-statistics (Congalton and Green, 1999), were calculated for all segment sizes. Cross-validation was used as accuracy assessment for both the point-height-based and CHM-based distributional approach. Comparison between different average segment sizes within each classification approach was based on a normalized *z*-test statistic derived from the proportion of correctly classified samples (Foody, 2004). Samples were treated as independent, since the number of samples for each class was not constant across average segment sizes. The standardized normal test statistic for cases with independent test samples are given by:

$$z = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad [1]$$

where

x_1, x_2 = correctly allocated number in two independent samples of size n_1 and n_2, respectively

p = $(x_1 + x_2)/(n_1 + n_2)$ (Foody, 2004)

A significance value of $\alpha = 0.05$ were used in all cases, with the null hypothesis (H_0) being that there was no difference between classification accuracies for different segment sizes. H_0 was

rejected in favor of the alternative hypothesis (H_a), that there were in fact differences in classification accuracies, when *z*-values were larger than the value associated with $\alpha = 0.05$, namely $z = 1.96$. Significance tests were performed for both 2- and 3-class classification schemes within each classification approach. The significance test was iteratively repeated for the highest and lowest accuracies and for the lowest accuracies in ascending order if the two extreme accuracies were significantly different from each other. This was done until no significant differences between the highest accuracy and the other accuracies were found, in order to establish accuracy significance. It should be noted that accuracies could be found significantly different by chance alone, given the number of possible comparisons among twelve segmentation treatments.

Significance tests within each approach were extended to test significance between the 2- and 3-class schemes within each approach and to significance testing for the 2-class and 3-class schemes between the two methods. These tests were based on a standard T-test with paired samples (average segment size) for differences between classification accuracy means (Ott, 1993). Accuracies were regarded as significantly different if the calculated T-value was higher than the T-value for $\alpha = 0.05$.

Lidar point-height and CHM grid height distributional parameters were extracted from each segment with a known forest type, based on the field BAF plot data. Stepwise discriminant variable reduction was followed by discriminant classification based on the reduced variable sets. The set of 75 possible point-height classification variables and 31 CHM variables (Table 4.3) were reduced to 10 or fewer variables through stepwise discriminant techniques using an α-level between 0.1 and 0.35. This procedure reduced the variables to those that maximized between group variability, while minimizing within group variability for a given α-level. A simple correlation analysis was performed to evaluate correlations among discriminant variables, with high inter-correlations defined as Pearson's coefficients of 0.8 or higher. A value of 0.8 was chosen based on data characteristics, with the knowledge that all lidar-derived variables are height-related, and hence some correlation was to be expected. This allowed for additional variable reduction to only those variables that defined the discriminant feature space independent of each other. The variables with the highest significance to type discrimination were retained

based on partial R^2 values, while redundant variables were removed. Reduced variable sets were entered into discriminant analyses for 2- and 3-class classification, for each of the segmentation results and for each distributional data type. A linear discriminant function was generated for each forest type per classification. Discriminant formulae serve as classifier functions in that height distributional variables of an observation can be entered in all functions, with the observation ultimately assigned to the type-function with the highest score. All discriminant analyses were performed using SAS V. 8.02 software (Level 02M0; SAS, Inc.).

Accuracy reports for all discriminant analyses were generated using a cross-validation routine within the SAS discriminant procedure. Single observations were iteratively removed from the analysis, followed by discriminant function development, and classification of the separate variable into one of the pre-defined classes. Accuracy reports were based on combination of all removed observation classifications during successive runs. This type of approach lent itself to traditional accuracy assessment techniques, with reference and user (map) classes (Congalton and Green, 1999).

4.3 Results and Discussion

4.3.1 Discriminant Classification using Lidar Point-Height Distributional Variables

Stepwise discriminant analysis was suited to reduction of classification variables from 75 original lidar point-height distributional variables (Table 4.3) to fewer than 10 variables in each classification attempt. Identified variables were those that best separated classes, while maintaining a low within class variation. The final stepwise-selected variables that were entered into the discriminant analysis and their partial R^2 values are shown in Table 4.5.

Correlation analysis resulted in removal of as few as zero to as many as four variables in each classification attempt. Most of the removed variables were related in terms of distributional characteristics, e.g., closely spaced percentile values, while correlations among other variables, e.g., skewness of first return vegetation heights and the 40^{th} canopy cover percentile (3-class; 27,050 segments; 0.035 ha/segment), were less intuitive. This underlined the importance of

Table 4.5 Final variables entered into point-height-based discriminant classification (*2-class = Deciduous-Coniferous; 3-class = Deciduous-Coniferous-Mixed).* Partial R^2 values are given after each variable as an indicator of relative importance to the classification

Classification		Final variables entered into discriminant analysis
27,050 segments (0.035 ha/segment)	2-class	MaxVeg1(0.40) MedianRef1 (0.09) Grnd2ratio (0.09) MinRef2 (0.03) StdMeanRef2 (0.02)
	3-class	MedianRef1 (0.12) KurtosisVeg2 (0.10) ModeVeg2 (0.07) Canopy40P (0.05) MeanVeg2 (0.05) Veg2ratio (0.03)
10,352 segments (0.091 ha/segment)	2-class	StdVeg2 (0.32) MedianRef1 (0.17) Grnd2ratio (0.05) MinRef2 (0.02) StdMeanRef2 (0.02) Vegratio (0.02) Canopy50P (0.01)
	3-class	StdVeg2 (0.34) MedianRef1 (0.19) StdRef2 (0.05) Veg2ratio (0.04) P_Veg2_10 (0.04) Canopy40P (0.03) MinRef2 (0.03)
6,687 segments (0.141 ha/segment)	2-class	StdVeg2 (0.33) MedianRef1 (0.17) Veg2ratio (0.05) P_Veg2_30 (0.03) MinVeg2 (0.02) RangeRef2 (0.02) SkewnessVeg1 (0.01)
	3-class	StdVeg2 (0.35) Veg2ratio (0.05) StdRef2 (0.05) MeanRef1 (0.04) P_Veg2_10 (0.03) P_Veg2_40 (0.03) Canopy70P (0.03) Canopy40P (0.03) MinVeg1 (0.02) RangeRef1 (0.02)
2,972 segments (0.318 ha/segment)	2-class	StdVeg2 (0.33) MedianRef1 (0.15) Grnd1ratio (0.04) MaxRef1 (0.04) CVVeg1 (0.02)
	3-class	StdVeg2 (0.36) MedianRef1 (0.17) MaxRef1 (0.05) Grnd1ratio (0.04) StdMeanRef2 (0.04) P_Veg2_40 (0.04) P_Veg2_10 (0.03) Canopy80P (0.02)
1,473 segments (0.642 ha/segment)	2-class	StdVeg2 (0.33) MedianRef1(0.14) ZeroNVeg2ratio (0.03) P_Veg2_30 (0.03) RangeRef1 (0.02) StdRef1 (0.02) Canopy80P (0.01)
	3-class	StdVeg2 (0.35) MedianRef1 (0.15) RangeRef1 (0.04) ZeroNgrnd1ratio (0.04) Canopy80P (0.03) SkewnessVeg1 (0.02)
981 segments (0.964 ha/segment)	2-class	StdVeg2 (0.35) MedianRef1 (0.13) ZeroNgrnd1ratio (0.04) RangeRef1 (0.05) StdMeanVeg2 (0.03) CVVeg2 (0.02) P_Veg2_90 (0.02) KurtosisVeg2 (0.01)
	3-class	StdVeg2 (0.36) MedianRef1 (0.15) MinRef1 (0.07) ZeroNgrnd1ratio (0.05) StdMeanRef2 (0.04) Canopy70P (0.03) ZeroNgrnd3_5ratio (0.02)
749 segments (1.263 ha/segment)	2-class	StdVeg2 (0.34) MedianRef1 (0.13) ZeroNgrnd1ratio (0.05) RangeRef1 (0.05) RangeVeg2 (0.02) Canopy20P (0.02) MinVeg1 (0.01) MaxRef2 (0.01)
	3-class	StdVeg2 (0.38) MedianRef1 (0.14) MinRef1 (0.10) ZeroNgrnd1ratio (0.05) P_Veg2_40 (0.05) ZeroNVeg3_5ratio (0.03)
502 segments (1.885 ha/segment)	2-class	StdVeg2 (0.3693) MeanRef1 (0.1658) ZeroNgrnd1ratio (0.0242) StdRef2 (0.0325) RangeVeg2 (0.0141) StdMeanRef2 (0.0190) CVVeg1 (0.0232) MinVeg2 (0.0113) RangeRef1 (0.0112)
	3-class	StdVeg2 (0.42) MeanRef1 (0.17) MinRef1 (0.12) P_Veg2_40 (0.04) ZeroNgrnd1ratio (0.08) StdRef2 (0.04) Canopy40P (0.04)
374 segments (2.530 ha/segment)	2-class	StdVeg2 (0.35) MeanRef1 (0.11) Canopy60P (0.06) SkewnessVeg1 (0.01) MinVeg2 (0.01) ModeVeg2 (0.01)
	3-class	StdVeg2 (0.37) MeanRef1 (0.15) P_Veg1_10 (0.03) ModeVeg1 (0.03) StdRef2 (0.03) ZeroNgrnd2ratio (0.03) ModeVeg2 (0.02) MinVeg2 (0.02) Canopy80P (0.02)
240 segments (3.942 ha/segment)	2-class	StdVeg2 (0.33) MeanRef1 (0.13) RangeVeg2 (0.02) ModeVeg1 (0.02) MinVeg1 (0.02) Canopy10P (0.02)
	3-class	StdVeg2 (0.37) MeanRef1 (0.19) P_Veg2_10 (0.04) P_Veg2_40 (0.05) MinVeg2 (0.04) StdMeanRef2 (0.03) MedianRef1 (0.04) Canopy30P (0.03)
168 segments (5.632 ha/segment)	2-class	StdVeg2 (0.34) MedianRef1 (0.15) StdMeanRef1 (0.03) MaxRef1 (0.05) MaxVeg2 (0.04) ModeVeg2 (0.03) RangeRef2 (0.03) Canopy10P (0.01) Canopy30P (0.02)
	3-class	StdVeg2 (0.38) MeanRef1 (0.17) StdRef2 (0.05) StdMeanRef1 (0.06) P_Veg1_10 (0.03) P_Veg2_40 (0.05) Canopy60P (0.03)
167 Appomattox forest stands (5.666 ha/segment)	2-class	MedianRef1 (0.28) P_Veg2_80 (0.18) P_Veg2_10 (0.05) Canopy60P (0.04) Canopy10P (0.03)
	3-class	MedianRef1 (0.35) MeanRef1 (0.08) P_Veg2_10 (0.09) P_Veg2_80 (0.07) P_Veg1_80 (0.05) MaxVeg2 (0.05) MaxRef1 (0.07) CVRef2 (0.03) ZeroNgrnd2ratio (0.03) Canopy30P (0.04)

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

complete evaluation of all correlations in order to define the feature space with uncorrelated variables, which were most important to class separation, based on partial R^2 values.

All types of distributional data types were well represented by the final selected variables. It was interesting to note the representation of reflectance values (near-infrared: 1,064 nm) as part of selected variables, with especially the median reflectance of first return vegetation heights present in all data sets. This highlighted the importance that per-object, lidar-associated reflectance (intensity) data have to lidar-based classification approaches. Near-infrared wavelengths have been shown to be highly discriminant among vegetation types, especially between deciduous and coniferous species (Martin *et al.*, 1998; Fung *et al.*, 1999; van Aardt and Wynne, 2001). Other well-represented, important variables (based on partial R^2 values) included the standard deviation of second return vegetation heights and the number of first return ground hits as a ratio of total first returns. These variables were indicators of vegetative cover. Second return metrics, which are associated with intermediate returns and by extension lower canopy cover indicators, played a role in defining different forest structures. Large standard deviations for second return vegetation heights can be interpreted to be indicative of a mid-canopy structure with a more variable height range, while the opposite is true of smaller standard deviations.

Certain forest types are more likely to have a variable canopy structure (larger variation), measured from the top of the canopy, than other types. Prime examples of stands with high and low height variation are all-aged deciduous stands with a variable canopy structure, as opposed to even-aged coniferous stands with a more uniform canopy height range. The number of first return ground hits, as a ratio of total first returns, also helped to define forest structure as an indicator of canopy cover. Stands with closed canopies result in a lower first return ground hit ratio, while deciduous stands are likely to contain gaps, e.g. successional or wind damage gaps, with an associated increase in first return ground hit ratio. Both these metrics, the standard deviation of second return vegetation heights and the ratio of first return ground hits and total number of first returns, had important discrimination characteristics needed to separate deciduous from coniferous objects.

Classification accuracies for the point-height-based discriminant classification approach are shown in Table 4.6. Overall accuracies ranged from 81.4% to 89.2% for the 2-class, deciduous-coniferous classification. Producer's and user's accuracies for the deciduous and coniferous classes indicated that deciduous class assignment was more reliable than for coniferous objects, both from a map producer's and a map user's perspective. Kappa statistics for the 2-class classification ranged from 60.2% to 76.7%. Although overall and Kappa statistics peaked at the 502 segments (1.885 ha/segment) application, significant differences were only found between 1.885 ha/segment (89.2%) and 0.035 ha/segment (82.2%) and 0.964 ha/segment (81.4%) (α = 0.05). Overall accuracies for the 3-class, deciduous-coniferous-mixed classification ranged from 61.6% to 70.8%, while the Kappa statistics varied between 40.4% and 53.5%. Peak values again were found at the 502 segments (1.885 ha/segment) segment-level. No significant differences were found in the case of the 3-class classification scheme. Significance testing indicated that there were minor to non-existing differences within the 2-class and 3-class schemes, but the T-test revealed a significance difference between the mean classification accuracies for the 2- and 3-class schemes at α = 0.05. Producer's accuracies generally were highest for coniferous, followed by deciduous and mixed objects, in that order. User's accuracies typically decreased from deciduous to coniferous to mixed object classification.

Although maximum classification accuracies were found at the 1.885 ha/segment size, these maxima were not statistical improvements over accuracy results for other segment sizes. This indicated that average segment size did not influence classification for the study area. This result was attributed to the hierarchical nature of the segmentation algorithm which merges smaller, homogenous segments to form larger segments at higher hierarchical levels. Within-segment variance was therefore already minimized before combination of smaller segments, leading to similar accuracies for all segment sizes.

Table 4.6 Accuracies associated with 2- and 3-class lidar point-height-based discriminant classification for all segmentation applications

Table 4.6	Classification		Producer's Accuracy	User's Accuracy	Overall Accuracy	Kappa-statistic
2-class	27,050 segments (0.035 ha/segment)	Deciduous	85.7%	86.3%	82.2%	61.5%
		Coniferous	75.9%	75%		
	10,352 segments (0.091 ha/segment)	Deciduous	87.1%	88.4%	84.5%	66.5%
		Coniferous	79.4%	77.8%		
	6,687 segments (0.141 ha/segment)	Deciduous	88.6%	88.6%	85.4%	68.3%
		Coniferous	79.7%	79.7%		
	2,972 segments (0.318 ha/segment)	Deciduous	88.6%	87.3%	84.5%	66.2%
		Coniferous	77.2%	79.2%		
	1,473 segments (0.642 ha/segment)	Deciduous	82.1%	89.8%	82.6%	63.6%
		Coniferous	83.5%	72.5%		
	981 segments (0.964 ha/segment)	Deciduous	83.1%	87.6%	81.4%	60.2%
		Coniferous	78.4%	71.6%		
	749 segments (1.263 ha/segment)	Deciduous	86.6%	87.9%	83.5%	63.9%
		Coniferous	77.8%	75.7%		
	502 segments (1.885 ha/segment)	Deciduous	89.8%	93.4%	89.2%	76.7%
		Coniferous	88.2%	82.12%		
	374 segments (2.530 ha/segment)	Deciduous	81.4%	90.6%	82.6%	63.7%
		Coniferous	84.8%	71.8%		
	240 segments (3.942 ha/segment)	Deciduous	81.5%	88.2%	81.9%	62.5%
		Coniferous	82.5%	73.4%		
	168 segments (5.632 ha/segment)	Deciduous	90.5%	89.3%	87.0%	71.4%
		Coniferous	80.5%	82.5%		
	167 Appomattox forest stands (5.666 ha/segment)	Deciduous	86.9%	86.9%	83.2%	63.4%
		Coniferous	76.5%	76.5%		
3-class	27,050 segments (0.035 ha/segment)	Deciduous	68.4%	78.8%	67.5%	49.2%
		Coniferous	79.6%	76.5%		
		Mixed	53.2%	43.1%		
	10,352 segments (0.091 ha/segment)	Deciduous	70.5%	77.5%	66.7%	47.2%
		Coniferous	71.4%	74.1%		
		Mixed	52.9%	42.9%		
	6,687 segments (0.141 ha/segment)	Deciduous	71.4%	82.5%	70.3%	53.5%
		Coniferous	75%	79.2%		
		Mixed	62.7%	46.4%		
	2,972 segments (0.318	Deciduous	72.3%	81%	68.5%	50.3%

Table 4.6 Classification			Producer's Accuracy	User's Accuracy	Overall Accuracy	Kappa-statistic
	ha/segment)	Coniferous	71.4%	72.7%		
		Mixed	56.9%	45.3%		
	1,473 segments (0.642 ha/segment)	Deciduous	62.5%	76.9%	61.6%	40.6%
		Coniferous	69.6%	69.6%		
		Mixed	51%	36.1%		
	981 segments (0.964 ha/segment)	Deciduous	69.4%	79.8%	65.7%	46.1%
		Coniferous	69.2%	69.2%		
		Mixed	54%	42.2%		
	749 segments (1.263 ha/segment)	Deciduous	68.9%	77.7%	66.0%	46.3%
		Coniferous	76%	71.7%		
		Mixed	50%	42.4%		
	502 segments (1.885 ha/segment)	Deciduous	76.2%	81.9%	70.8%	53.3%
		Coniferous	72.9%	79.5%		
		Mixed	56.5%	45.6%		
	374 segments (2.530 ha/segment)	Deciduous	69.7%	79.5%	64.7%	45.3%
		Coniferous	67.4%	67.4%		
		Mixed	53.1%	43.3%		
	240 segments (3.942 ha/segment)	Deciduous	73.4%	71.2%	66.4%	48.0%
		Coniferous	71.4%	73.5%		
		Mixed	54%	55.1%		
	168 segments (5.632 ha/segment)	Deciduous	70%	67.3%	61.7%	40.4%
		Coniferous	69.6%	59.3%		
		Mixed	47.6%	55.6%		
	167 Appomattox forest stands (5.666 ha/segment)	Deciduous	72.1%	79.5%	68.4%	51.3%
		Coniferous	68.2%	65.2%		
		Mixed	63.3%	57.6%		

The significance of the highest accuracies for the 2-class vs. 3-class scheme indicated that a deciduous-coniferous forest delineation was better suited to the Virginia Piedmont, as opposed to the inclusion of a mixed category as well. This was attributed to the forest types found in the study area, with most of the stands represented by one distinct taxonomic group and few completely mixed stands. Only 25 (11.4%) of the BAF plots were inherently part of a mixed class when a 75% basal area purity cut-off was used, further corroborating this conclusion.

There also was no significant difference between segment-based classification and classification based on existing forest stands in the study area, in both the 2- and 3-class schemes. Although unexpected, this result was ascribed to the definition of forest stands in an operational context. Forest stands are more often defined by their species make-up (coniferous-deciduous-mixed) than by their height structure (even-aged vs. all-aged). This effectively made the Appomattox stand map a thematic species map. However, the classification of segments was based on forest structure and not forest type definition. Segment-based classification accuracies therefore were deemed very encouraging.

4.3.2 Discriminant Classification using CHM Height Distributional Variables

Stepwise discriminant analysis again was effective in reducing classification variables from 31 original CHM distributional variables (Table 4.3) to fewer than 10 variables for both 2- and 3-class classification schemes. Final stepwise-selected variables are shown in Table 4.7. As was the case for the point-height-based approach, all distributional data types were well represented, e.g., maximum height, canopy cover percentiles, and the standard deviation of height. Maximum height, the 90^{th} height percentile, and the 20^{th} and 30^{th} canopy cover percentiles were particularly well represented. These variables indicated that although the maximum height served as a discriminant factor between types, structural information, represented by the percentile variables, also was important to forest type definition.

Classification accuracies for the CHM-based discriminant classification approach (Table 4.8) ranged between 64.4% and 79% (2-class) and 54.3% and 64.3% (3-class). Only 1.885 ha/segment (79% and 0.141 ha/segment (64.4%) were significantly different from each other, while there were no significant differences in the 3-class classification scheme ($\alpha = 0.05$). The 2- and 3-class mean accuracies were significantly different from each other at $\alpha = 0.05$ (paired T-test). This corroborated the results from the point-height-based approach that a 2-class (deciduous-coniferous) scheme is better suited to the study area than a 3-class (deciduous-coniferous –mixed) scheme.

There again was no significant difference between segment-based classification and classification based on exiting forest stands in the study area, further corroborating results from the point-height-based approach. Kappa statistics (2-class: 28.7% – 53.6%; 3-class: 32.1% – 44.8%) were distinctly lower than those found for the point-height-based approach, indicating that classification using lidar point-heights resulted in a generally better forest type assignment than the CHM-based classification.

Table 4.7 Final variables entered into CHM-based discriminant classification (*2-class = Deciduous-Coniferous; 3-class = Deciduous-Coniferous-Mixed).* Partial R^2 values are given after each variable as an indicator of relative importance to the classification

Classification		Final variables entered into discriminant analysis
27,050 segments (0.035 ha/segment)	2-class	Canopy30P (0.04) P_Veg1_90 (0.014) Canopy60P (0.024) RangeVeg1 (0.02)
	3-class	MaxVeg1 (0.26) Canopy60P (0.02)
10,352 segments (0.091 ha/segment)	2-class	Canopy30P (0.06) StdMeanVeg1 (0.01) P_Veg1_25 (0.02) CVVeg1 (0.01)
	3-class	Canopy30P (0.08) StdMeanVeg1 (0.02) CVVeg1 (0.02) P_Veg1_60 (0.04)
6,687 segments (0.141 ha/segment)	2-class	Canopy20P (0.05) StdMeanVeg1 (0.02) Canopy60P (0.02) MinVeg1 (0.01)
	3-class	MaxVeg1 (0.29) Canopy30P (0.08) Canopy90P (0.04)
2,972 segments (0.318 ha/segment)	2-class	StdMeanVeg1 (0.04) Canopy40P (0.03) MinVeg1 (0.03)
	3-class	MaxVeg1 (0.27) StdMeanVeg1 (0.07) Canopy40P (0.05) MinVeg1 (0.02) Canopy70P (0.01)
1,473 segments (0.642 ha/segment)	2-class	MaxVeg1 (0.25) Canopy30P (0.05) MinVeg1 (0.01) CVVeg1 (0.03) Canopy80P (0.01) P_Veg1_30 (0.01)
	3-class	Canopy80P (0.03) Canopy30P (0.04) MaxVeg1 (0.02)
981 segments (0.964 ha/segment)	2-class	P_Veg1_90 (0.27) Canopy30P (0.08)
	3-class	P_Veg1_90 (0.29) Canopy30P (0.09) Canopy70P (0.08) StdMeanVeg1 (0.03)
749 segments (1.263 ha/segment)	2-class	P_Veg1_90 (0.27) Canopy20P (0.08) KurtosisVeg1 (0.01) Canopy80P (0.02)
	3-class	Canopy30P (0.10) Canopy70P (0.06) RangeVeg1 (0.01) StdMeanVeg1 (0.01)
502 segments (1.885 ha/segment)	2-class	P_Veg1_90 (0.28) Canopy20P (0.12) KurtosisVeg1 (0.02) MinVeg1 (0.01) StdMeanVeg1 (0.02)
	3-class	RangeVeg1 (0.34) Canopy20P (0.08) P_Veg1_60 (0.02)
374 segments (2.530 ha/segment)	2-class	P_Veg1_90 (0.27) Canopy20P (0.04) Canopy40P (0.03) SkewnessVeg1 (0.02)
	3-class	P_Veg1_90 (0.29) Canopy30P (0.08) Canopy60P (0.05) Canopy80P (0.03)
240 segments (3.942 ha/segment)	2-class	Canopy20P (0.10) StdMeanVeg1 (0.02) MaxVeg1 (0.01) P_Veg1_20 (0.02) StdVeg1 (0.01)
	3-class	P_Veg1_90 (0.25) Canopy20P (0.15) Canopy70P (0.03)
168 segments (5.632 ha/segment)	2-class	Canopy20P (0.12) MaxVeg1 (0.01) KurtosisVeg1 (0.02) StdMeanVeg1 (0.02) StdVeg1 (0.02)
	3-class	RangeVeg1 (0.28) Canopy20P (0.11) Canopy60P (0.06) Canopy90P (0.06)
167 Appomattox forest stands (5.666 ha/segment)	2-class	StdVeg1 (0.21) CVVeg1 (0.03) P_Veg1_20 (0.05) RangeVeg1 (0.08)
	3-class	StdVeg1 (0.30) CVVeg1 (0.08) P_Veg1_25 (0.05) KurtosisVeg1 (0.04) RangeVeg1 (0.04)

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table 4.8 Accuracies associated with 2- and 3-class CHM-based discriminant classification for all segmentation applications

Table 4.8	Classification		Producer's Accuracy	User's Accuracy	Overall Accuracy	Kappa-statistic
2-class	27,050 segments (0.035 ha/segment)	Deciduous	71.4%	82%	71.7%	41.5%
		Coniferous	72.2%	58.8%		
	10,352 segments (0.091 ha/segment)	Deciduous	79.3%	84.7%	77.6%	52.7%
		Coniferous	74.7%	67%		
	6,687 segments (0.141 ha/segment)	Deciduous	61.4%	78.2%	64.4%	28.7%
		Coniferous	69.6%	50.5%		
	2,972 segments (0.318 ha/segment)	Deciduous	70.7%	82.5%	71.7%	41.8%
		Coniferous	73.4%	58.6%		
	1,473 segments (0.642 ha/segment)	Deciduous	75.7%	81.5%	73.5%	44.1%
		Coniferous	69.6%	61.8%		
	981 segments (0.964 ha/segment)	Deciduous	79.4%	82.4%	75.7%	47.6%
		Coniferous	68.9%	64.6%		
	749 segments (1.263 ha/segment)	Deciduous	79.1%	83.5%	76.2%	48.8%
		Coniferous	70.8%	64.6%		
	502 segments (1.885 ha/segment)	Deciduous	84.3%	83.6%	79%	53.6%
		Coniferous	69.1%	70.1%		
	374 segments (2.530 ha/segment)	Deciduous	83.1%	81.7%	77.2%	50.1%
		Coniferous	66.7%	68.8%		
	240 segments (3.942 ha/segment)	Deciduous	76.1%	82.4%	75.2%	48.6%
		Coniferous	73.7%	65.6%		
	168 segments (5.632 ha/segment)	Deciduous	70.3%	80%	69.6%	36.8%
		Coniferous	68.3%	56%		
	167 Appomattox forest stands (5.666 ha/segment)	Deciduous	75.4%	80.7%	72.6%	42%
		Coniferous	67.6%	60.5%		
3-class	27,050 segments (0.035 ha/segment)	Deciduous	57.1%	69.6%	59.4%	36.8%
		Coniferous	71.4%	58.8%		
		Mixed	51%	44.1%		
	10,352 segments (0.091 ha/segment)	Deciduous	64.3%	75%	61.6%	39.9%
		Coniferous	69.6%	61.9%		
		Mixed	47.1%	40%		
	6,687 segments (0.141 ha/segment)	Deciduous	46.4%	72.2%	54.8%	32.4%
		Coniferous	66.1%	60.7%		
		Mixed	60.8%	36%		
	2,972 segments (0.318	Deciduous	53.6%	71.4%	58.9%	37.2%

Table 4.8 Classification			Producer's Accuracy	User's Accuracy	Overall Accuracy	Kappa-statistic
	ha/segment)	Coniferous	71.4%	65.6%		
		Mixed	56.9%	39.2%		
	1,473 segments (0.642 ha/segment)	Deciduous	44.6%	72.5%	54.3%	32.1%
		Coniferous	69.6%	59.1%		
		Mixed	58.8%	35.7%		
	981 segments (0.964 ha/segment)	Deciduous	65.7%	77.2%	63.8%	43.3%
		Coniferous	65.4%	60.7%		
		Mixed	58%	46.8%		
	749 segments (1.263 ha/segment)	Deciduous	63.2%	77%	63.6%	43.3%
		Coniferous	70%	62.5%		
		Mixed	58%	46%		
	502 segments (1.885 ha/segment)	Deciduous	67.3%	72.3%	62.6%	40.1%
		Coniferous	70.8%	64.2%		
		Mixed	43.5%	41.7%		
	374 segments (2.530 ha/segment)	Deciduous	66.3%	78.7%	62.5%	42.3%
		Coniferous	69.6%	56.1%		
		Mixed	49%	46.2%		
	240 segments (3.942 ha/segment)	Deciduous	68.8%	67.7%	60.4%	39.3%
		Coniferous	65.7%	54.8%		
		Mixed	46%	54.8%		
	168 segments (5.632 ha/segment)	Deciduous	66%	67.3%	64.3%	44.8%
		Coniferous	73.9%	60.7%		
		Mixed	57.1%	63.2%		
	167 Appomattox forest stands (5.666 ha/segment)	Deciduous	76.7%	73.3%	58.9%	35.7%
		Coniferous	50%	50%		
		Mixed	40%	42.9%		

The 2- and 3-class classification approaches based on distributions from lidar point-heights and the 1 m CHM were significantly different from each other. This result showed that the point-height-based approach resulted in better 2- and 3-class accuracies than the CHM-based approach. However, the first method is both computationally and financially expensive, requiring advanced computer hardware and lidar sensors. The latter approach can be based solely on a forest canopy height model. Results for significance tests are shown in Table 4.9.

Table 4.9 Significance results for point-height-based and CHM-based discriminant classifications across all average segment sizes (significance at $\alpha = 0.05$ is indicated by *)

Significance Test		Point-Height-Based Discriminant Classification	CHM-Based Discriminant Classification
z- test: Between average segment sizes (Best 2-class: 1.885 ha/segment; 89.2%) (Best 3-class: 1.885 ha/segment; 70.8%)	2-class	2.20* (0.964 ha/segment: 81.4%) 2.03* (0.035 ha/segment: 82.2%)	3.27* (0.141 ha/segment: 64.4%)
	3-class	None	None
Between 2-class and 3-class means (critical two-tail T-test: 2.20)		19.64*	12.96*
Between 2-class means (critical two-tail T-test: 2.20)		7.47*	
Between 3-class means (critical two-tail T-test: 2.20)		4.47*	

A caveat of any distributional approach, namely unclassified segments due to missing classification variables, generally can be circumvented by limiting independent variables to those types that are well represented for all segment sizes. Fifth lidar height return values, for instance, were not used as part of the independent variable set, thereby avoiding problems with unrepresented variables for selected segments. However, only as few as 5 out of the total 6,687 (0.141 ha/segment) segments will remain unclassified due to missing second return variables. This was attributed to segments, devoid of vegetation, where only first lidar returns were recorded. Unclassified segments were not used as part of the cross-validation accuracy assessment, but operationally should still be addressed through reclassification. Possible solutions to situations where segments have missing classification variables include post-classification photo-interpretation of unclassified segments, or type determination based on field visits.

Classification results indicated that although a 3-class designation resulted in increased forest type resolution, a 2-class is more accurate than a 3-class division of the feature space. This was attributed to more confusion in the case of a 3-class division, with the addition of a third, mixed

category only serving to detract from a simpler 2-class approach. It also could be argued that definition of a mixed class is not warranted for the study area. This is based on the number of mixed BAF plots, with only 51 mixed plots out of 219 samples when a 90% basal area purity was required for pure class assignment. This number dropped to 25 out of 219 objects when the basal area requirement for a pure class definition was reduced to 75%. Since a 75% pure class definition is closer than a 90% definition to what one would encounter in practice, the small number of mixed plots at 75% purity makes this class unsuitable for further consideration and statistical analysis. However, when only considering the 3-class definition, coniferous producer's accuracies were higher than deciduous accuracies. This was due to a "purer" definition of the coniferous class in the 3-class approach. A 90% basal area coniferous majority, as opposed to 51% for the 2-class approach, resulted in a definitively purer class. This came at the cost of reduced overall accuracy due to the addition of a mixed class. Class definitions ultimately are the choice of the user, who might be inclined to define Virginia Piedmont forests as a 2-class forest biome, thereby increasing classification accuracies, but sacrificing a more detailed class definition.

Discriminant functions for the point-height-based 2- and 3-class classifications and all segment sizes are listed in **Appendix I**. Appendix J lists discriminant functions for the CHM-based distributional classification approach.

4.4 Conclusions

Object-oriented classification results, from both lidar point-height-based and CHM-based distributional discriminant classifications, were promising when one considers the type of input data, the variability in natural ecosystems based on basal area contributions for each class, and the classification approach used. Accuracies as high as 89.2% for the point-height-based discriminant classification and 79% for the CHM-based approach bode well for lidar-based object-oriented classification. Higher accuracies for the point-height-based approach were attributed to the increased variable type range, which included more structural (2^{nd} returns) and spectral (reflectance) variables than the CHM-based approach.

Although the more complex, the point-height-based approach resulted in an increase of 10.2% in overall accuracy, the CHM-based approach performed surprisingly well. Accuracies for the latter approach were acceptable when considering that classification was based on a single layer, canopy height model data source. However, discriminant classification using lidar point-heights was shown to be significantly better than the CHM-based approach, and was considered the better choice, even when its increased complexity was taken into account. High accuracies for between group classifications, with only structural data as input, showed potential when one considers that the primary uses of lidar data are associated with forest biophysical characterization (Means et al., 2000; Lefsky et al., 2002; Naesset, 2004; Popescu *et al*., 2004). Extension of lidar data to forest type definition therefore presents an opportunity to base a complete forest inventory on one remote sensing data source.

There were no significant differences within classification approaches, indicating that average segment size did not matter. The lack of decreasing classification accuracies at larger segment sizes was attributed to the hierarchical nature of the segmentation algorithm. The algorithm was based on the minimization of within-segment variance at all segment levels, but since smaller segments constituted the building blocks for larger objects, variance minimization already was adequately addressed at lower levels. Stand-based classification also was not significantly different from segment-based approaches. This was due to the definition of operational forest stands on a per-species or type basis, resulting in an already existing forest type map.

Stepwise discriminant analysis, as part of discriminant classification, was confirmed (van Aardt and Wynne, 2001) as an effective method for the reduction of classification variables from more than seventy to fewer than ten. This procedure reduced the feature space to those variables that maximized between-group separation while minimizing the variability within each group. Selected variables included a broad range of distributional data types, while reflectance values (near-infrared) were present. The inclusion of the median reflectance for lidar first return vegetative hits highlighted the importance of per-object, lidar-associated reflectance data to lidar-based classification approaches. Inclusion of reflectance variables corroborated wavelengths used for more traditional hyperspectral classification approaches (Martin *et al.*, 1998; Fung *et al.*, 1999; van Aardt and Wynne, 2001). The standard deviation of second return vegetation hits and

the first return ground hits as a ratio of total first returns were also well represented variables. Both these metrics were important indicators of vegetative cover.

Two-class forest type definition (deciduous-coniferous) resulted in better accuracies than a 3-class (deciduous-coniferous-mixed) approach for the study area. This was attributed to high basal area percentages, in terms of deciduous-coniferous mixtures, that had to be used for mixed class definition. The mixed class eventually was closely associated with the deciduous group, thereby reducing deciduous producer's accuracies in the 3-class scheme. Lower basal area percentages for mixed-class definition resulted in an almost negligible number of mixed sample objects. Higher producer's accuracies for the coniferous class in the 3-class approach indicated that the deciduous-mixed classes were main contributors to lower overall accuracies with increased between-group confusion.

Although traditional approaches achieved adequate classification accuracies when multispectral data were used, a lidar-based object-oriented approach has multiple benefits. Lidar data have been used extensively for the modeling of forest biophysical parameters, and coupled with classification, could enable a forest inventory approach that is based on one data source. Due to their hierarchical nature, object-oriented approaches also have the benefit of recombination of object-level results to fit almost any required scale. Lastly, the simplicity of the approach is attractive, since only per-object height values are potentially required, while co-registration errors between optical and lidar data also are minimized. Extension of accurate classifications to lidar-derived forest volume or biomass-by-type is appealing, and could prove useful in forest inventories. The biggest hurdles to large-scale application are data cost and processing time, which likely will drop as more data providers emerge and technology improves. Irrespective of these caveats, analytical approaches continuously are evolving to a stage where multi-source remote sensing data might no longer be required for complete forest inventories.

4.5 Acknowledgements

This research was made possible by funding from NASA (grant # NG65-10548; NG613-03019), the McIntire-Stennis Research Program (grant # VA-136589), the Forestry Department and

Graduate Student Association at Virginia Polytechnic Institute and State University, and the Potomac chapter of the American Society for Photogrammetry and Remote Sensing. Field data collection was supported by the Virginia Department of Forestry, specifically Dr. John Scrivani, Todd Edgerton, Ralph Toddy, and Wayne Bowman (VDOF). Drs. Richard Oderwald (Virginia Polytechnic Institute and State University) and Sorin Popescu (Texas A&M University) provided invaluable assistance with statistical and lidar analyses. Amy Zhang from the Statistics department at Virginia Polytechnic Institute and State University served as statistical consult.

4.6 Literature Cited

Abkar A., M.A. Sharifi, and N.J. Mulder, 2000. Likelihood-based image segmentation and classification: A framework for the integration of expert knowledge in image classification procedures. *Journal of Algebraic Geometry* 2 (2): 16 pp.

Acton, S.T., 1996. On supervised segmentation of remotely sensed imagery using nonlinear regression. *International Journal of Remote Sensing* 17 (7): 1407-1415.

Baatz, M., and A. Schäpe. 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: STROBL, J. *et al.* (Hrsg.): Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg, September 2000, Karlsruhe, Herbert Wichmann Verlag: 12–23.

Burrough, P.A., and R.A. McDonnel, 1998. Principles of geographical information systems. Oxford University Press. Oxford, England. 333 pp.

Cheevasuvit F., 1990. Image segmentation based on graph theory. Proceedings of the 23rd International Symposium on Remote Sensing of Environment, Bangkok, Thailand, April 18-25, 1990. pp. 1329-1334.

Congalton, R.G. and Green, K. (1999), Assessing the accuracy of remotely sensed data: Principles and practices. Lewis Publishers, New York, 137 p.

Cortijo, F.J., and N.P. De la Blanca, 1999. The performance of regularized discriminant analysis versus non-parametric classifiers applied to high-dimensional image classification. *International Journal of Remote Sensing* 20 (17): 3345-3365.

Dorren, L.K.A, Maier, B., and A.C. Seijmonsbergen, 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology and Management* 183 (2003) 31-46.

Douglas, T.E., D.L. Evans, K.L. Belli, and S.D. Roberts, 2003. Classification of pine and hardwood by the distribution and intensity of lidar returns. ISPRS "Three dimensional mapping workshop from InSAR and LIDAR", June 17-19, 2003, Portland, Oregon, USA. 5pp.

eCognition User's Manual, 2003. Definiens Imaging (www.definiens.com).

Engdahl, M.E., J. Pulliainen, and M. Hallikainen, 2003. Combined land-cover classification and stem volume estimation using multitemporal ERS tandem INSAR data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Foody, G.M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing* 70 (5): 627-633.

Franklin, S.E., 1994. Discrimination of subalpine forest species and canopy density using digital CASI, SPOT PLA and Landsat TM data. *Photogrammetric Engineering & Remote Sensing* 60 (10): 1233-1241.

Fung, T., F.Y. Ma, and W.L. Sui, 1999. Hyperspectral data analysis for subtropical tree species identification. Proceedings: 1999 ASPRS Annual Conference, Portland Oregon.

Heyman, O., G.G. Gaston, A.J. Kimerling, and J.T. Campbell, 2003. A per-segment approach to improving Aspen mapping from high-resolution remote sensing imagery. *Journal of Forestry* (June, 2003): 29-33.

Hill, R.A., 1999. Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of Remote Sensing* 20 (5): 1039-1044.

Jaakkola S.P., 1989. Applicability of SPOT for forest management. *Advanced Space Research* 9 (1): 135-141.

Kayitakire, F., C. Farcy, and P. Defourny, 2002. Ikonos-2 imagery potential for forest stands mapping. ForestSAT Symposium, Heriot Watt University, Edinburgh, August 5-9, 2002. 11 pp.

Kellndorfer, J.M., and F.T. Ulaby, 2003. Forest biomass inversion from SAR using object oriented image analysis techniques. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Kermad C.D., and K. Chehdi, 2002. Automatic image segmentation system through iterative edge-region co-operation. *Image and Vision Computing* 20: 541-555.

Kressler, F., Y. Kim, and K. Steinnocher, 2003. Object-oriented land cover classification of panchromatic KOMPSAT-1 and SPOT-5 data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 7 pp.

Lefsky M.A., W.B. Cohen, D.J. Harding, G.G. Parker, S.A. Acker, and S.T. Gower, 2002. Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology & Biogeography* 11 (2002): 393-399.

Li W., G.B. Benie, D.C. He, S. Wang, D. Ziou, and Q.H.J. Gwyn, 1999. Watershed-based hierarchical SAR image segmentation. *International Journal of Remote Sensing* 20 (17): 3377-3390.

Lyman Ott, R, 1993. An introduction to statistical methods and data analysis. Duxbury Press, Belmont, California, USA. 1051 pp.

Martin, M.E., S.D. Newman, J.D. Aber, and R.G. Congalton, 1998. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment* 65: 249 – 254.

Means J.E., S.A. Acker, B.J. Fitt, M. Renslow, L. Emerson, and C.J. Hendrix, 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering & Remote Sensing* 66 (11): 1367-1371.

Naesset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80 (2002):88-99.

Nelson, R.F., R.S. Latty, and G. Mott, 1984. Classifying northern forests using Thematic Mapper simulator data. *Photogrammetric Engineering & Remote Sensing* 50 (5): 607-617.

Nugroho, M., D.H. Hoekman, and R. De Kok, 2002a. Analysis of the forests spatial structure using SAR and Ikonos data. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 10 pp.

Nugroho, M., D.H. Hoekman, and R. de Kok, 2002b. Analysis of forest spatial structure using spatial decision rule. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 8 pp.

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

Popescu, S.C., R.H. Wynne, and J.A. Scrivani, 2004. Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, U.S.A. *Forest Science* (In press)

Sarkar A., M.K. Biswas, B. Kartikeyan, V. Kumar, K.L. Majumder, and D.K. Pal, 2002. A MRF model-based segmentation approach to classification for multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 40 (5): 1102-1113.

Schwarz, M., C.H. Steinmeier, and L. Waser, 2001. Detection of storm losses in alpine forest areas by different methodic approaches using high-resolution satellite data. Proceedings of the 21st EARSeL Symposium, Paris. 7pp.

Shandley, J., J. Franklin, and T. White, 1996. Testing the Woodcock-Harward image segmentation algorithm in an area of southern California chaparral and woodland vegetation. *International Journal of Remote Sensing* 17 (5): 983-1004.

Shankar B.U., C.A. Murthy, and S.K. Pal, 1998. A new gray level based Hough transform for region extraction: An application to IRS images. *Pattern Recognition Letters* 19 (1998): 197-204.

Shen, S.S., G.D. Badwar, and J.G. Carnes, 1985. Separability of boreal forest species in the Lake Jennette area, Minnesota. *Photogrammetric Engineering and Remote Sensing* 51 (11): 1775-1783.

Smits P.C., and S.G. Dellepaine, 1997. An irregular region label model for multi-channel image segmentation. *Patter Recognition Letters* 18 (1997): 1133-1142.

Ton J., J. Sticklen, and A.K. Jain, 1991. Knowledge-based segmentation of Landsat images. *IEEE Transactions on Geoscience and Remote Sensing* 29 (2): 222-232.

van Aardt, J.A.N., and R.H. Wynne, 2001. Spectral separability among six southern tree species. *Photogrammetric Engineering & Remote Sensing* 67 (12): 1367-1375.

White, J.D., C.K. Glenn, and J.E. Pinder III, 1995. Forest mapping at Lassen Volcanic National Park, California, using Landsat TM data and a geographical information system. *Photogrammetric Engineering and Remote Sensing* 61 (3): 299-305.

Willhauck, G., 2000. Comparison of object oriented classification techniques and standard image analysis for the use of change detection between SPOT multispectral satellite images and aerial photos. ISPRS XXXIII conference, Amsterdam, Netherlands, 2000. 8 pp.

CHAPTER 5

MULTIRESOLUTION, HIERARCHICAL SEGMENTATION OF SMALL-FOOTPRINT LIDAR DATA AS A FORESTRY INVENTORY PRECURSOR

Abstract. Forest inventory analyses have traditionally been stand-based, with intensive field surveys. Remote sensing segment-based approaches have shown potential to reduce error, while being scalable to operational sizes. However, criteria for selection of segment sizes that are suitable to further forest inventory analyses are lacking in the literature. This study evaluated a lidar-based segmentation approach as a precursor to subsequent object-oriented forest measurement and type characterization. The study area is located in Appomattox Buckingham State Forest in the Piedmont physiographic province of Virginia, U.S.A, at 78°41' W, 37°25' N. Vegetation is composed of various coniferous, deciduous, and mixed forest stands. The eCognition segmentation algorithm was used to segment a canopy height model (CHM) derived from small-footprint lidar data. Mapped basal area factor (BAF) plots were used to determine field volume selected segments. Selection of segmentation outcomes for further analyses was based on between- and within-segment variances of the CHM, and BAF plot size. Segments ranging in size from 0.035 ha/segment to 5.632 ha/segment were selected as potential candidates for extension to further analyses. The F-statistic for each segmentation result was used to determine the limit to which smaller segments could be scaled. Lidar distribution-based object-oriented volume estimation and forest type classification (deciduous-coniferous) were used to validate segment selections. Adjusted R^2 and RMSE values, and classification accuracies did not corroborate the observed increasing within-segment variance trend found with increasing segment size. This was attributed to poor segment representation by BAF plot measurements, a small range in observed volume values (6.94 – 350.93 m^3/ha), and the hierarchical nature of the segmentation algorithm. However, adjusted R^2 (0.52 – 0.59) and RMSE values (52.16 – 57.25 m^3/ha) were better than for volume models using existing forest stands (R^2 = 0.42; RMSE = 62.36 m^3/ha). A conceptual decision protocol was developed by which potential users can determine which segmentation outcomes are suitable for application in further analyses, but further investigation for larger observed dependent variable ranges is warranted. Expert knowledge and field validation can never be excluded from the decision process.

5.1 Introduction

The traditional way to gauge forest resources involves large-scale, labor-intensive forest inventories, often incorporating intricate sampling schemes and extrapolation efforts (Avery and Burkhart, 1994). Alternative approaches to forest inventory have become available with the advent of new remote sensing technologies, specifically light detection and ranging (lidar) (Lefsky *et al*., 2002a). Lidar has enabled users to extract forest structural information, e.g., mean and maximum heights, as well as height distributions (Means *et al.,* 2000; Lefsky *et al.,* 2002b; Næsset, 2002; Popescu *et al*., 2004). Although aerial photography has long been used to estimate forest volume through stand-volume tables, these methods are time consuming, subjective, and more applicable to smaller areas (Avery and Burkhart, 1994). Lidar remote sensing provides a relatively novel approach to large-scale forest inventories. Such remote sensing approaches lend themselves to methods that are repeatable and objective, resulting in yield estimates that approach acceptable accuracy and precision. However, precision and accuracy are two metrics that also can be closely linked to the variability of processing units (Avery and Burkhart, 1994; Shivers and Borders, 1996). A potentially useful approach would entail the definition of such processing objects in terms of the very characteristics that need to be inventoried, namely height, basal area, and volume. One logical approach would be object definition through the same structurally-related data source that will later be incorporated in inventory procedures, i.e., lidar data. This in turn leads to the next hurdle, namely definition of structurally homogeneous objects as analysis units using the chosen data source.

Forest stand maps have long been used to define the units of analysis and management in forestry practice. Existing stands have been derived through photo interpretation, site-growth assessment, and field experience. Although forest stands are often based on site-quality assessments and species composition, one could argue that forest structural homogeneity is crucial to unbiased and precise estimates of forest biophysical parameters (Avery and Burkhart, 1994; Shivers and Borders, 1996). This raises the question of whether or not stands in their current form are necessarily the best unit for the estimation of forest parameters. Since within-stand structural homogeneity is an integral prerequisite to the use of forest objects as unique

units for measurements (Avery and Burkhart, 1994; Shivers and Borders, 1996), derivation of homogenous objects is critical to subsequent analyses.

Image segmentation is a technique well suited to the derivation of such objects in a forestry context (Nugroho *et al.,* 2002a; Engdahl *et al.,* 2003; Heyman *et al*., 2003). Segmentation is the partitioning of an image into regions that share common properties. Segmentation needs to be both exhaustive (whole image is partitioned into regions) and exclusive (a cell can only belong to one region), with cells or points partitioned in a region sharing at least one common property. Such regions therefore have two key properties, (i) they are distinct in spectral, structural, or binary response and (ii) they are spatially distinct. Regions in images are important because they correspond to unique units within a scene (Wilson and Spann, 1988; Jain *et al.*, 1995; Russ, 1995; Darwish *et al.*, 2003).

Segmentation approaches have been used extensively for forest inventory purposes, mainly related to classification and per-object biophysical parameter estimation. The idea of distinct forest units came to the fore in a study by McCormick and Folving (1998), with a segmentation algorithm central in the proposed design to estimate biodiversity per forest stand using fractal and dominance methods. The authors proposed the use of any medium-resolution satellite imagery (Landsat, SPOT, etc.) to obtain viable segmentation results using an edge-preserving smoothing algorithm, based on the linkage of each edge pixel in an image to its neighboring pixel with the minimum edge value. Heyman *et al*. (2003) used a per-segment approach to improve aspen (*Populus tremuloides*) mapping in Oregon, USA. They applied a histogram thresholding method based on hue and saturation values of high-resolution color-infrared photographs. The authors achieved an 88% accuracy for mapping aspen segments into three broad categories (no aspens; 0 - 50% aspens; 50 - 100% aspens). Kayitakire *et al.* (2002) used the same algorithm to map forest stands in Belgium (mixed oak, spruce, beech, and pine stands). Overall accuracies of 88% (per-pixel clustering) and 83.3% (per-parcel derived map) were found, but it should be noted that wrong classification of a parcel results in all the parcel pixels being misclassified, as opposed to singe pixel misclassification. While segmentation-based classification studies are plentiful, application of segments have been extended to forest biophysical parameter estimation as well.

Häme and Tomppo (1987), Jaakkola (1989), Woodcock *et al.* (1990), and Ryherd and Woodcock (1990) used segmentation approaches to delineate management objects for inventory and assessment purposes. Woodcock *et al.* (1994) applied the Woodcock-Harward algorithm, followed by unsupervised classification, to estimate conifer forest stand attributes in the Stanislaus National Forest, California. The error for the estimate of total timber volume was 4.6%, with the conclusion that such mapping projects were suited to large-scale mapping. Although volume estimate errors for Landsat-based segmentation were relatively high (≈ 59%), Kilpeläinen and Tokola (1999) found that that segmentation can ultimately be used to stratify forests, thereby reducing variation and increasing sampling efficiency. Landsat TM data were used to estimate stand mean volume on a per-segment basis. Various pine (*Pinus sylvestris, Picea abies*), birch (*Betula pendula, B. pubescens*), and other broadleaved species, namely alder (*Alnus incana*) and aspen (*Populus tremula*), were studied. The authors stated that results for total mean volume of growing stock in southern Finland were dependent on the data and validation method used. Hyyppä and Inkinen (1999) were able to derive accurate tree locations and crown areas from segmented lidar data, using a modified watershed segmentation procedure. This approach ultimately led to accurate height and volume estimations, using the crown area and stems-per-hectare derivations. Engdahl *et al.* (2003) incorporated the eCognition, multiresolution-hierarchical segmentation algorithm in a design aimed at estimating accurate stem volumes of Scots pine and Norwegian spruce in southern Finland. INSAR-based (ERS-1/2 SAR data) stem volume estimates were comparable to Finnish National Forest Inventory (NFI) estimates, with root mean square error (RMSE) values of 101 m^3/ha and 115 m^3/ha, and correlation values (r) of 0.79 and 0.69, respectively. The authors concluded that it is possible to produce an accurate, segmented land cover classification and stem volume estimates for the forest segments.

Of particular interest to this study are two Finnish studies that attempted "segment-aided" timber volume estimation. The first by Makela and Pekkarinen (2001) attempted a Landsat TM plot-level volume-by-species estimation. The authors used a measurement space-guided clustering, defined as an ISODATA classification followed by connected component labeling (CCL), which in turn was based on edge detection and linking. A directed trees approach, based on gradient analysis and edge detection, was applied as an alternative segmentation method. Spectral

features used for volume estimation were extracted in two ways, namely (i) from a fixed window around the field sample plot, and (ii) from the pixels in the fixed window that belonged to the same segment as the sample plot. The ISODATA CCL approach yielded the most accurate volume estimation for Scots pine and Norway spruce (*Pinus sylvestris* and *Picea abies*), as well as for the total volume. The directed trees algorithm had the best results for broad-leaf species (*Betula pendula*, *B. pubescens,* and *Populus tremula*). Improvements from the fixed window approach to segment-based approaches were 1% to 11.3% (relative RMSE), though RMSE values remained high. The authors did conclude, however, that the segmentation of forest areas into stratified units was very suitable for the estimation of forest variables. Errors could be minimized by extracting estimates from more homogenous segments (Makela and Pekkarinen, 2001).

The second study by Pekkarinen (2002) was done as a follow-up to investigate the stand-level errors. Departing from the premise that the magnitude of errors was related to the limited spatial resolution of sensors such as Landsat TM, the author investigated the use of very high resolution images for image-based multi-source forest inventory (MSFI). The author concluded that the segmentation algorithm succeeded in delineating distinct forest areas for feature extraction. The segment-based approach performed better than reference data, extracted from square-shaped windows. Segmentation resulted in a decrease of up to 10% in the case of broad-leaved species RMSEs, and up to 8% in the case of spruce species. The author suggested that segment-level data could decrease associated segment-level errors (Pekkarinen, 2002).

Deciding on the segment size to use for further analyses has been a difficult question to address to date. Most authors used classification accuracy assessment (Antunes *et al*., 2003; Darwish *et al*., 2003; Heyman *et al*., 2003; Neubert, 2001) and visual correspondence to input data, along with number of segments, as decision criteria (Schiewe *et al*., 2001; Schiewe, 2002). Most approaches have either been tree- (Hyyppä and Inkinen, 1999; Nugroho *et al.,* 2002a), plot- (Makela and Pekkarinen, 2001), or stand-based (Woodcock *et al.,* 1994; Kayitakire *et al.,* 2002), indicating a possible need for a decision rule for selection of the segmentation result best suited to subsequent applications. Since the improvement of forest classification or biophysical

parameter estimation often is used as justification for segment-based analyses, a clear need exists for definition of the segmentation result best suited to such analyses.

Current segment-based results are encouraging, proving that many authors have successfully implemented delineation of uniform forest objects, followed by parameter estimation per defined unit (Woodcock *et al.*, 1994; Makela and Pekkarinen, 2001; Pekkarinen, 2002; Engdahl *et al.*, 2003; Kellndorfer and Ulaby, 2003). More reliable parameter estimates might be obtained from homogenous, data-derived stand units, instead of using traditionally defined forest stands as units for parameter extraction (Makela and Pekkarinen, 2001; Pekkarinen, 2002). Although existing forest stands are based on site (soil, topography, and micro-climate), age class, and species composition, it cannot be assumed that such stands are in fact structurally homogenous. The use of ancillary structural (lidar) data in defining units of measurement therefore has potential if accurate, scaleable estimates are required for large tracts of land using remotely sensed data.

The primary objective of this study was to assess the utility of small-footprint lidar data for the derivation of structurally homogenous segments as a precursor to per-segment forest inventory. This assessment was based on the selection of segmentation results that were best suited to subsequent object-oriented volume modeling and forest type classification, using lidar-distributional data. An object refers to a spatial entity that is homogenous in terms of a selected property, as opposed to the traditional, continuous fields approach found in spatial analysis (Burrough and McDonnel, 1998). It was assumed that per-segment analysis will be based on traditional sampling approaches, e.g., variable or fixed plots within segments, and not on complete segment measurement. Small-footprint lidar data have been shown to be effective in the prediction of plot-, grid-cell-, and stand-level volume and aboveground biomass (Hyyppä and Inkinen, 1999; Means *et al.*, 2000; Næsset, 2002; Holmgren *et al.*, 2003; Popescu *et al.*, 2004). Lidar-based forest type classification has also proven viable (Douglas *et al.,* 2003), illustrating the relevance of lidar as input data source to forest inventory. Two secondary objectives were addressed, namely (i) determination of the applicability that within- and between-segment variance have to selection of segmentation results for extended analyses, and (ii) evaluation of a conceptual decision model to select the appropriate segment size for further object-oriented analyses. Such segment-based forest inventory analyses require segments that are homogenous in

terms of a structural characteristic related to the estimated metric (Pekkarinen, 2002). It is clear from past studies that optimal segmentation must lead to stratification into structurally homogenous stand units (Woodcock *et al.*, 1994; Makela and Pekkarinen, 2001; Hyyppa and Inkinen, 2002; Engdahl *et al.*, 2003). Segments also need to be of manageable size for further analysis, particularly (i) objects need to be larger than field plots, and (ii) segmentation results have to be scalable to match operational stand conditions. Extension of segmentation results to lidar-based analyses is an attractive proposition, as this would result in the dependence of both unit definition and analyses on a single remote sensing data source.

5.2 Methods

5.2.1 Study Area

The 946 ha (2,338 acres) study area is located in Appomattox Buckingham State Forest (Appomattox County) in the Piedmont physiographic province of Virginia, southeastern U.S.A at 78°41' W, 37°25' N (Figure 5.1). The mean elevation of the study area is 185 m (606 ft.), with minimum and maximum elevations of 133 m (436 ft.) and 225 m (738 ft.), respectively. Local topography can best be described as gentle rolling slopes and flat terrain. Vegetation is composed of various coniferous (*Pinus taeda, P. virginiana, P. echinata*, and *P. strobus*), deciduous (*Quercus coccinea, Q. alba,* and *Liriodendron tulipifera*), and mixed forest stands.

5.2.2 Available Data

The lidar data set (Table 5.1) was acquired by Spectrum Mapping, LLC using the DATIS II, small-footprint, high-density, multiple return system. Data were acquired on September 9, 2002, centered at 78°40'30" W, 37°25'9" N, and covered an area of approximately 958 ha (2,367 acres).

Basal area factor (BAF) plots were used for validation of segmentation results. BAF plot data consisted of 256 mapped BAF plots (basal area factor 10) on a 16 columns by 16 rows, 201.17 m (10 chains) grid Field data were collected during the summer, fall, and winter months (May –

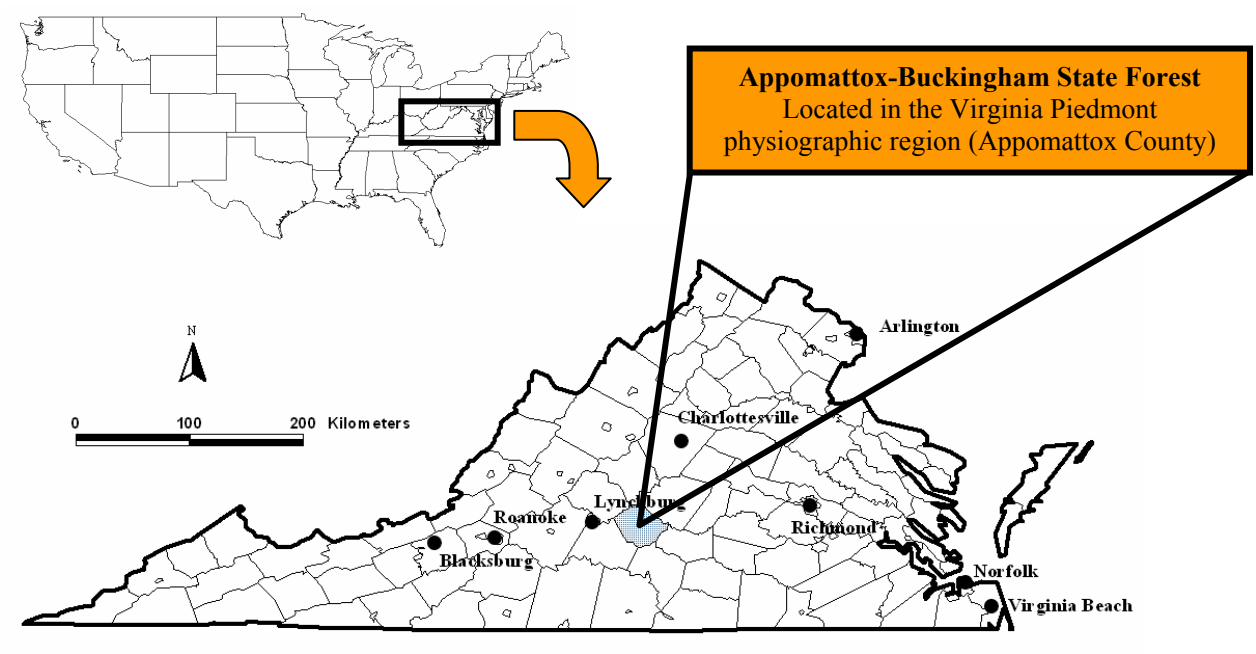


Figure 5.1 Study Area: Appomattox Buckingham State Forest

Table 5.1 DATIS II lidar data set characteristics

Characteristics	Specification
Laser altitude	2,000 m (6,562 ft.) above ground level
Laser scan field-of-view	75° maximum
Swath width and centerline spacing	800 m (2,625 ft.) and 400 m (1,312 ft.)
Scan rate	25 Hz
Laser pulse rate	35 kHz
Scan angle	± 13.5°
Returns	≤ 5
Resolvable distance between returns	0.75 m
Footprint	0.46 m (1.51 ft.)
Spacing across / along track	1 m (3.3 ft.) / 2 m (6.6 ft.)
Accuracy (X,Y,Z)	X,Y: 0.5 m; Z: 0.15 m (X,Y: < 1.6 ft.; Z: < 0.49 ft.)
Post-processed GPS accuracy	< 0.05 m
Wavelength	1,064 nm

December) of 2003. All GPS plot center locations were differentially corrected using data from the National Geodetic Survey's Continually Operating Reference Stations (CORS, 2000) and Corvallis Microtechnologies, Inc. PC-GPS software (Version 3.7; Corvallis Microtechnologies, Inc.). For each sampling point, the following data were collected (**Appendix A**, an actual data sheet):

- Plot basal area ("in-tree" count); diameter at breast height (dbh) > 5 inches (12.7 cm) (10-factor prism)
- Dbh and height for all plot trees tallied (diameter tape and Vertex hypsometer)
- Azimuth and distance from plot center to each tallied tree (SUUNTO compass and Vertex hypsometer's range finding function)
- Species codes of tallied trees (**Appendix B**)
- Differentially corrected GPS point at plot center (CMT's March II GPS unit)

A total of 37 BAF plots were discarded due to their location on private land or having volume and biomass values of zero. This left a total of 219 BAF plots that were used in statistical analyses (**Appendix C**).

Each basal area plot was geographically mapped based on plot center coordinates, azimuths, and distances to tallied trees. Plots were assigned to a 2-class forest type scheme based on basal area percentages. "Deciduous" or "Coniferous" types were defined as plots that had 50% or greater basal area contribution from either deciduous or coniferous species, respectively. This 2-class analysis consisted of 140 deciduous and 79 coniferous plots.

BAF plot measurements were expanded to a per-hectare volume for each segment using standard BAF expansion equations:

$$\text{Volume/hectare} = \frac{\sum (VBAR * 10)}{\sum Samples} * Metric\ Conversion\ Factor \quad \ldots[1]$$

where

Volume/hectare = Volume ($m^3.ha^{-1}$) per hectare for each segment

VBAR = Volume-Basal-Area Ratio

$$= \frac{Volume}{\pi\left(\frac{D}{2}\right)^2 * \frac{1}{144}} \quad \text{(dbh in inches)}$$

$$\text{Metric Conversion Factor} = \frac{1}{0.4046856} \rightarrow \text{(acre to hectare)} \quad \text{(Avery and Burkhart, 1994)}$$

Descriptive statistics for deciduous and coniferous BAF plots are given in Table 5.2.

Table 5.2 General descriptive information for deciduous and coniferous BAF plots

Type	Parameter	Minimum	Maximum	Average	σ
Deciduous plots (140)	Volume/ha (m^3/ha)	6.94	350.65	157.64	84.14
	Biomass/ha (Mg/ha)	11.11	269.01	113.60	58.60
	Basal area/ha (m^2/ha)	2.30	34.44	16.32	7.84
Coniferous plots (79)	Volume/ha (m^3/ha)	8.32	350.93	114.49	75.44
	Biomass/ha (Mg/ha)	4.67	155.56	41.47	26.64
	Basal area/ha (m^2/ha)	2.30	36.73	14.24	7.91

5.2.3 Lidar Data Pre-Processing

A canopy height model (CHM) was needed for segment derivation as a precursor to per-segment volume modeling and classification. First and last (vegetation-removed) returns from the lidar data set were extracted and corrected for possible errors (suspect low and high, or "bird" hits). Peripheral outlier height values with a low frequency and a distinct difference (> 6 m) from the next smallest or largest value were removed as outliers. This resulted in the removal of one return smaller than –75 m and six returns larger than 31 m. The first returns were median-filtered by 1 m grid cells in order to remove per-cell values that were redundant to subsequent interpolation procedures. First and last returns were interpolated to a 1 m spatial resolution grid using regular Kriging, since Popescu *et al.* (2002) found this to be the most accurate interpolation technique using similar data over the same study area. This approach effectively addressed instances where a 1 m grid cell lacked an original input value. The resultant 1 m resolution was detailed enough to detect road and stand breaks in the segmentation process. It had the additional benefits of requiring less computing power, as opposed to a 0.5 m grid, and likely produced a smoother canopy digital terrain model (DTM) and ground digital elevation model (DEM). Interpolation was performed using Surfer 7.0 software (Golden Software, Inc.). The differenced first- and last return surface (CHM) was used as input to the eCognition

segmentation algorithm. This allowed for extraction of forest segments based on height homogeneity and distinct stand breaks, e.g., roads and slope breaks. The CHM is shown in Figure 5.2.

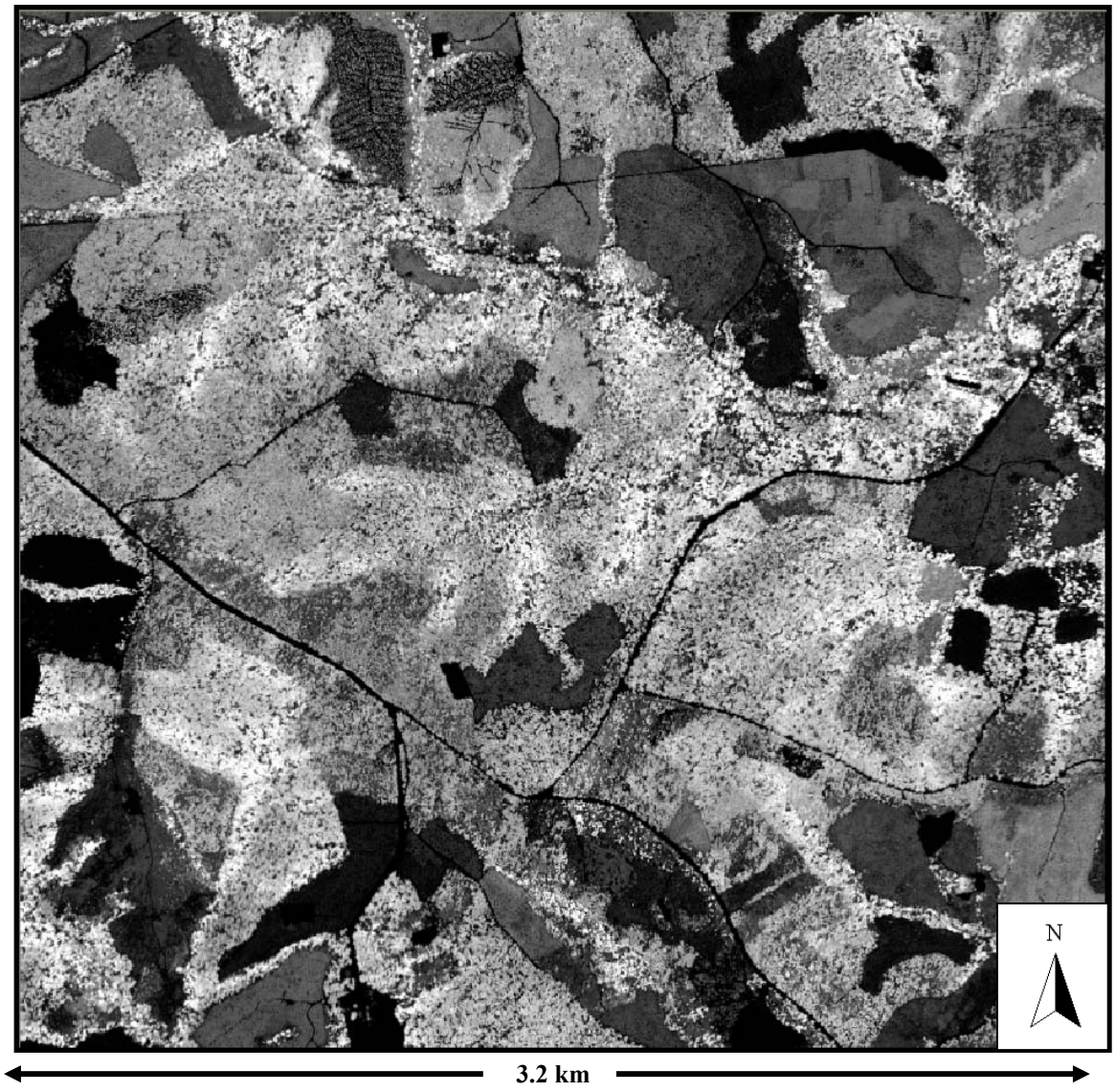


Figure 5.2 A 1 m canopy height model of the study area. Brighter tones correspond to taller trees, and *vice versa*

Lidar data distributional characteristics were required for per-segment volume modeling and classification, as evaluation of segmentation results. Such a distributional approach to analyses required that lidar data be processed on a per-return basis in order to retain information related to the return hierarchy. Peripheral outlier height values again were removed for all return data sets, based on the same approach as in the case of the CHM. Ground hits were removed using

Terrascan V. 003.002 (Terrasolid, Inc.) and MicroStation V. 08.00.04.01 (Bentley Systems, Inc.) software. This algorithm identifies ground hits based on iterative slope analysis of lidar returns. *Grid cell size* and *maximum slope of the area* are required input parameters. *Grid cell size* is the smallest cell size for which a ground return can be extracted. A cell size of 10 m was used in order to extract a maximum number of ground returns for the first (31,294,660), second (11,101,215) and third (2,121,989) returns. Grid cell sizes of 39 m and 119 m were used for the fourth (175,093) and fifth (5,379) returns, respectively. Larger grid cell sizes were required for the last two categories due to the small number of returns in each case. Each of these two cell sizes resulted from cases where the number of ground hits reached a maximum for the fourth and fifth returns, based on the assumption that most of the hits from these categories would be ground hits due their ranking in the return hierarchy. A slope percentage parameter of 35% was used as a maximum for the area, obtained from a USGS DEM. Ground returns constitute an important component of overall lidar distributional patterns and were retained as data sets on a per-return basis.

Non-ground hits, designated as vegetation hits, were normalized for terrain by calculating the actual return height above a lidar-derived 1 m digital elevation model (DEM) of the study area. The actual height of each vegetation hit was calculated as the difference between the vegetation hit and the bilinear interpolated height of the four corner cells of the DEM cell directly beneath each hit. This was done using Surfer V. 8.1 software (Golden Software, Inc.). This process normalized all vegetation hits for varying terrain elevations, thereby enabling analyses to incorporate actual lidar point heights (Means *et al*., 2000).

5.2.4 Segmentation of the Study Area

5.2.4.1 Multiresolution, Hierarchical Segmentation (eCognition algorithm)

eCognition (Definiens Imaging, GmbH) was chosen as the preferred segmentation approach for this study. The eCognition algorithm focuses on objects or regions instead of single pixels as basic processing units, and is therefore defined as an object-oriented approach. Segmentation is initiated with one-pixel objects, which are merged into bigger objects as the algorithm

progresses. The underlying concept is the minimization of the weighted heterogeneity of image objects – in each step adjacent components that define the smallest growth in heterogeneity are merged. This process is simultaneously applied across the whole image to obtain objects of comparable size and quality. Although generally considered a "black-box" approach, publications that describe the quantitative basis of the algorithm are abundant (Baatz and Schäpe, 2000; Neubert, 2001; Schiewe, 2002; Antunes *et al.*, 2003; Darwish *et al.*, 2003; Wong *et al.*, 2003). Various studies have implemented the eCognition approach in forestry contexts, with applications ranging from classification to biophysical parameter estimation (Kayitakire *et al.*, 2002; Nugroho *et al.*, 2002a; Nugroho *et al.* 2002b; Engdahl *et al.*, 2003; Kellndorfer and Ulaby, 2003; Kressler *et al.*, 2003).

Although eCognition was chosen as the preferred segmentation approach, one could argue that the segmentation method is subordinate in importance to the utility that resultant objects have to analyses. There are a multitude of alternative approaches to segmentation. These include knowledge-based segmentation (Ton *et al.*, 1991), unsupervised segmentation using nonlinear regression (Acton, 1996), the Woodcock-Harward (centroid linkage) algorithm (Shandley *et al.*, 1996), a Hough transform-based approach (Shankar *et al.*, 1998), watershed-based hierarchical segmentation (Li *et al.*, 1999), and iterative edge-region co-operation (Kermad and Chehdi, 2002). It is ultimately of great importance that segmentation results are robust. Other important factors are ease of operational use, widespread availability, and adequate software support.

5.2.4.2 Multiresolution Segmentation Methods

Segmentation of lidar data was performed prior to other analyses since units of observation needed to be defined first. Practical, algorithm-related issues such as the scale parameter, related to segment size, and color and shape weights, related to segment quality, were the first priority. Mixing color and shape weighting will have a distinct user-defined outcome, since stands of trees usually are structurally definitive, as well as spatially recognizable. Even-aged, managed coniferous stands are prime examples of this. They have a distinctly different structural definition than deciduous stands or open areas. Well-managed coniferous stands often have a

single-story crown canopy, as opposed to multiple sub-canopy layers that are often found in deciduous stands.

Parameter inputs to the segmentation algorithm were varied slightly to obtain results that were visually similar to the CHM, while ensuring that segment borders were neither too smoothed nor fractally shaped. It should be noted that general, recommended principles (eCognition, 2003) were followed, so as not to shift the focus from the selection of optimal segmentation results for further analyses, to segmentation procedures. The most important parameter inputs and their variations were:

- *Scale parameter:* The scale parameter was changed in increments of 1 from 10 (large number of small segments) to 212 (larger and fewer segments), and in increments of 5 from 215 to 240 (less than 100 segments). This ensured that segments could be evaluated across a wide range of average segment sizes. Output was compared to an existing Appomattox Buckingham State Forest stand map (Figure 5.3; 167 stands) to ensure that segments corresponded to manageable units. Although the existing stand map was useful for defining operational segment sizes, the optimal unit of analysis could be too small to manage operationally. The hierarchical nature of the algorithm therefore was important since subsequent analysis could be applied at a smaller segment level, while the results could be scaled for operational use. Minimum and

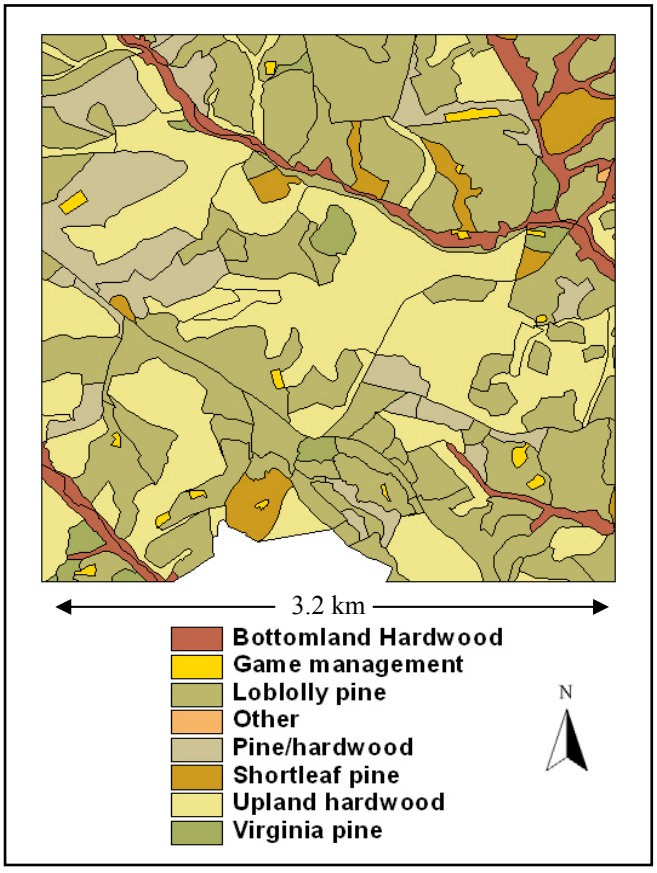


Figure 5.3 Existing Appomattox Buckingham State Forest stand map for study area

maximum scale parameter limits were evident from visual inspection, with as many as 55,908 segments (0.027 ha/segment; scale parameter = 10) and as few as 82 segments (11.538 ha/segment; scale parameter = 240). Such a broad range was useful for the calculation and visualization of within- and between-segment variance trends.

- *The color and shape weights*: As per recommendation of the algorithm developers (Baatz and Schäpe, 2000; eCognition, 2003), color homogeneity was weighted more heavily than shape homogeneity. It should be noted that *a priori* knowledge is required to configure settings optimally, while this effort followed established guidelines to define objects, along with quantitative and qualitative methods to evaluate results. Color-shape input was evaluated for three settings:
    - Color:Shape = 0.9:0.1
    - Color:Shape = 0.8:0.2
    - Color:Shape = 0.7:0.3

  This range made it possible to evaluate segmentation results over a range of inputs where color was weighted distinctly higher than shape. Smoothness of shape was considered more important than compactness of shape in a forestry context, since smooth, boundary-following segments are preferable to compact, blocky segments. The Smoothness:Compactness weight combination therefore was kept constant at 0.8:0.2.

5.2.4.3 Evaluation of Segmentation Results

A conceptual protocol was developed by which potential users can evaluate segmentation results for subsequent object-oriented analyses. The protocol implemented (i) visual correspondence to the CHM input data, (ii) within- and between-segment variance for initial segment selection, (iii) a plot size indicator for final segment selection, (iv) the between-within segment variance ratio (F-statistic) as an indicator of the upper limit to which segments could be scaled, and (v) validation of segment selection based on per-segment volume modeling and forest type classification:

(i) Segmentation results were overlaid on the CHM data used as input to the segmentation algorithm. Results were visually evaluated at both ends of the spectrum, at the small segment scale (10,000+ segments) and at the large segment scale (<200 segments). Although it is not quantifiable, such a visual comparison is often best at discerning valid segmentation results (Schiewe *et al*., 2001; Schiewe, 2002).

(ii) Segmentation results also were assessed based on between vs. within-segment variability. For each segmentation output during each run, the within- and between-segment variability was calculated for the CHM (Microsoft Visual C++, V. 6.0; Microsoft Corporation). Such a variance estimate could be closely related to variation in forest structure, specifically per-segment forest height. It also provided a standard variance comparison among segmentation procedures. The formulas for calculating between and within-segment variability were derived (C++ code: **Appendix K**) from the general variance formula for clusters of unequal sizes:

$$\frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i}\left(y_{ij}-\bar{\bar{y}}\right)^2}{\sum_{i=1}^{N}M_i-1} \qquad \ldots[2]$$

(Sukhatme and Sukhatme, 1970)

$$= \frac{\sum_{i}^{N}\sum_{j}^{M_j}\left[y_{ij}-\bar{y}_i+\bar{y}_i-\bar{\bar{y}}\right]^2}{\sum_{i}^{N}M_i-1}$$

$$= \frac{\sum_{i}^{N}\sum_{j}^{M_j}\left[\left(y_{ij}-\bar{y}_i\right)^2+\left(\bar{y}_i-\bar{\bar{y}}\right)^2+2(0)\right]}{\sum_{i}^{N}M_i-1}$$

$$= \frac{\sum_{i}^{N}\sum_{j}^{M_j}\left[\left(y_{ij}-\bar{y}_i\right)^2+\left(\bar{y}_i-\bar{\bar{y}}\right)^2\right]}{\sum_{i}^{N}M_i-1}$$

$$= \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_j}\left[\left(y_{ij}-\bar{y}_i\right)^2+\left(\bar{y}_i-\bar{\bar{y}}\right)^2\right]}{\sum_{i=1}^{N}M_i-1}$$

$$= \frac{\sum_{i=1}^{N}\left[\sum_{j=1}^{M_j}\left(y_{ij}-\bar{y}_i\right)^2\frac{\left(M_i-1\right)}{\left(M_i-1\right)}+M_i\left(\bar{y}_i-\bar{\bar{y}}\right)^2\right]}{\left(\sum_{i=1}^{N}M_i-1\right)}$$

$$= \frac{\sum_{i=1}^{N}\left[\left(M_i-1\right)s_i^2+M_i\left(\bar{y}_i-\bar{\bar{y}}\right)^2\right]}{\left(\sum_{i=1}^{N}M_i-1\right)}$$

$$= \frac{\sum_{i=1}^{N}\left[\left(M_i-1\right)s_i^2\right]}{\left(\sum_{i=1}^{N}M_i-1\right)}+\frac{\sum_{i=1}^{N}M_i\left(\bar{y}_i-\bar{\bar{y}}\right)^2}{\sum_{i=1}^{N}M_i-1}$$

$$= \boxed{\frac{\sum_{i=1}^{N}\left[\left(M_i-1\right)s_i^2\right]}{\left(\sum_{i=1}^{N}M_i-1\right)}} + \boxed{\frac{\left(N-1\right)}{\left(\sum_{i=1}^{N}M_i-1\right)}*\frac{\sum_{i=1}^{N}M_i\left(\bar{y}_i-\bar{\bar{y}}\right)^2}{N-1}} \qquad \ldots[3]$$

$\downarrow$ **Within-segment variation**

$\downarrow$ **Between-segment variation**

….(Oderwald, 2003)

where

N = # of segments

M_i = # of elements per segment

y_{ij} = observation *j* in segment *i*

$\bar{y}_i$ = mean of segment *i*

$\bar{\bar{y}}$ = overall arithmetic mean

s_i = variance for segment *I*

Theoretically, the preferable number of segments correspond to the variance-range in overall segment variance where within-segment variance < between-segment variance. This approach ensures that statistically different, homogenous segments are selected. Homogeneity is conducive to estimates with low associated RMSEs (Avery and Burkhart, 1994; Shivers and Borders, 1996), while statistically distinct groups tend to be more separable based on the measured parameter (van Aardt and Wynne, 2001). Such an approach also ensures that limited overall variance is still attributable to between-segment variability. As such, an acceptable segmentation result has small within-segment variance and large between-segment variance. Such a result therefore ensures structurally unique segments, but comes at the cost of small, unmanageable units. Although segment size also is an important consideration, the hierarchical nature of the segmentation results lent itself to scaling of segments from unmanageably small to operational sizes (Figure 5.3). The intersection of within and between-segment variance graphs for all three segmentation runs (Color:Shape weight variations) occurred at larger and hence fewer segments. Much of the overall model variance could therefore still be absorbed by between-segment variance, which made the segmentation result at this intersection not well suited to subsequent analyses based specifically on the CHM. The between-within variance intersection was therefore considered to be better suited to determining the level to which smaller segments could be scaled.

(iii) Given that the between-within ratio constituted an F-value, the F-statistic for each segmentation was calculated and used as a decision criterion for scaling level. Degrees of freedom were associated with number of segments and total number of elements for each segmentation run. The segmentation result where the between-within ratio (F-value) became significant ($\alpha = 0.05$), served as an indication of the larger level to which smaller segments could be scaled. Such an upper scaling limit indicated where between-segment variance had become significantly larger than within-segment variance, resulting in homogeneous segments in terms of overall model variance. Subsequent per-segment analyses were potentially more robust in

terms of precision, given this significant homogeneity. A large coniferous stand could by extension incorporate many smaller, unique segment components. After applying an analysis to smaller segments, results can be scaled to constitute the larger area, which is still considered homogenous in terms of overall variance.

(iv) The preferable segmentation result used for subsequent per-segment analyses therefore would be one where (i) structural within-segment variance was smaller than between-segment variance and (ii) the within-between variance plots reached asymptotes (all variance absorbed). The variance plots, however, were not truly asymptotic for the range of segmentation results. Although not quantitatively asymptotic, plots visually leveled off at smaller segment scales. Segment-based analysis incorporating in situ data requires average segment size to not be smaller than average field plot size. Average segment size should therefore encapsulate the size of an average field plot, resulting in true segment-plot representation. It was assumed that any plot that fell within a segment in its totality, was a representation of that segment. This assumption was based on BAF plot sampling within homogeneous units (Avery and Burkhart, 1994; Shivers and Borders, 1996).

A simple approach, based on the distance of the majority of trees from BAF plot centers, was used to select the initial segmentation result for per-segment analyses. Average plot size ensured that choice of segmentation result would correspond to general field plot sizes, with low within-segment variability. The chosen plot size metric was based on average tree distance from BAF plot center plus one and two standard deviations. This ensured that, for normally distributed tree distances from plot centers, at least 95% of trees per plot would be encapsulated by the plot size as defined by the average plus two standard deviations. The segmentation result with an average segment size just smaller than the lower plot size definition (average distance plus one standard deviation) and the result with an average segment size just larger than the larger plot definition were selected. This was done to encapsulate the plot-size selection range, from one to two standard deviations plus average tree distance from plot center. Given the non-normal distribution of distances from BAF plot centers (Figure 5.4.a), the distribution for $\log_{10}$-transformed distance data (Figure 5.4.b) was used to determine the initial segmentation size.

(v) Final validation of selected segmentation outcomes was based on object-oriented volume modeling and forest type classification using selected segment sizes. Ten average segment sizes, ranging from 0.035 ha/segment to 3.942 ha/segment, were chosen for subsequent volume model development and classification. These selections corresponded to segment sizes where within-segment variance was smaller than between-segment variance of the CHM heights. This was done in order to evaluate model performance across a range of average segment sizes. The current Appomattox stand map (167 segments; 5.666 ha/stand) also was selected, as well as the segmentation result that corresponded to the number of operational stands (168 segments; 5.632 ha/segment). Operational stands were used in order to compare segmentation-based modeling and classification to stand-based analyses. The lower end for average segment size was based on the plot size metric, while the upper limit was defined by the significance of the F-statistic.

Height distributional parameters were extracted for each average segment size, for those segments that contained actual BAF plot measurements. Only first and second return variables from vegetation return data sets were used, because many segments had missing values for the third through fifth return data sets. Distributional parameters included the mean, coefficient of variation, kurtosis, maximum, minimum, mode, range, standard error of the mean, skewness, standard deviation, number of observations, height percentile points at 10% intervals of height values, and canopy cover percentiles. Canopy cover percentiles were based on the proportion of first returns smaller than a given percentage of maximum height. The ratio of the number of vegetation or ground hits and the total number of lidar hits per segment also was calculated. This was done for second, and third through fifth group vegetation hits, as well as first, second, and third through fifth group ground hits. The vegetation ratio for each segment was calculated as the ratio of the number of vegetation hits per segment and the total hits for that segment (Chapter 3).

The derived lidar distributional parameters were used for object-oriented volume modeling and discriminant classification for average segment sizes of 0.035 ha/segment, 0.091 ha/segment, 0.141 ha/segment, 0.318 ha/segment, 0.642 ha/segment, 0.964 ha/segment, 1.263 ha/segment, 1.885 ha/segment, 2.53 ha/segment, 3.942 ha/segment, 5.632 ha/segment, and the Appomattox Forest stands (5.666 ha/segment).

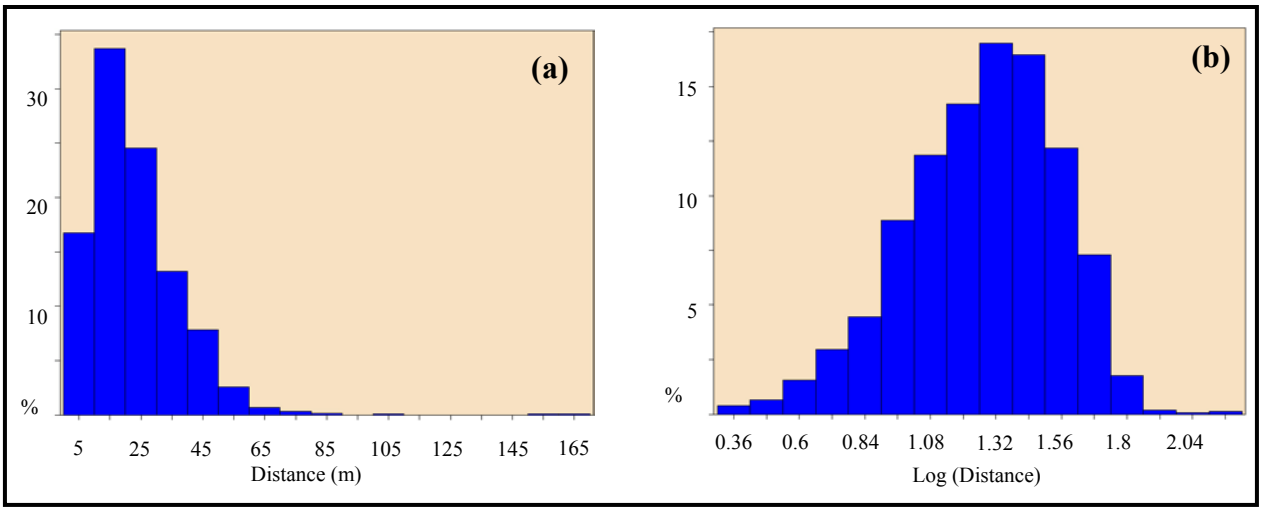


Figure 5.4 (a) Distribution of the distance of individual trees from BAF plot centers and (b) distribution of the $\log_{10}$ of the distance of individual trees from BAF plot centers

Stepwise discriminant analysis was used to reduce the set of independent variables to fewer than 10 (van Aardt and Wynne, 2001) for each analysis-per-segment size. Further variable reduction and validation were achieved through correlation analyses, by which variables with correlations of higher than 0.8 were removed. Only the variables with the highest correlation to the dependent variable (volume modeling) and largest partial R^2 contribution to the classification were retained. This was followed by the selection of a volume model based on adjusted R^2 and Mallow's Cp values (Chapter 3). Given the goal of evaluating model performance across a range of segment sizes, variables were selected from existing defined independent variable sets. Discriminant analysis was used to perform object-oriented classification. Both modeling and classification were performed using SAS V. 8.02 software (Level 02M0; SAS, Inc.).

Per-segment BAF plot volume estimates were averaged in cases where segments of larger average size contained more than one BAF plot. Volume models across average segment sizes were compared based on adjusted R^2 and RMSE values, while classification outcomes were evaluated based on classification accuracies and significance testing. Accuracies were obtained though cross-validation within the discriminant classification routine. Significance tests were performed for the 2-class scheme using a *z*-score test ($\alpha = 0.05$) based on the proportion of correctly assigned samples, with the assumption of samples being independent (Foody, 2004).

This latter assumption was rooted in the fact that different average segment sizes resulted in varying sample numbers used for classification. Volume model comparisons were extended to operational Appomattox Buckingham State Forest stands to evaluate the usefulness of segmentation versus the current definition management units. The same BAF plot averaging and lidar distributional approaches were taken in this latter case, with stands essentially treated as segments.

5.3 Results and Discussion

A proposed conceptual decision model, by which segmentation results can be evaluated and selected for further analyses, is shown if Figure 5.5. The five criteria used for segmentation evaluation proved to be useful in determining whether or not segmentation results were viable, which segmentation outcomes were potentially useful as selections for further applications, and specifying a level to which smaller segments could be scaled. However, confirmation of results through applied analyses did not conform to the expectation that within- and between-segment variance would be integral to segmentation evaluation. Visual correspondence to the CHM input data will be discussed first, followed by the quantitative statistical applications of CHM between and within-segment variance, applicability of the F-statistic for each average segment size, and the plot size metric, as an indicator of the segmentation results suited to further analyses. Segmentation evaluation based on per-segment volume modeling forms the final part of this section.

5.3.1 Visual Segmentation Results

The eCognition segmentation algorithm performed extremely well from a visual perspective. Segmentation boundaries followed the CHM's boundary definitions and detected stand and road breaks. Figures 5.6 – 5.8 show the segmentation results for the study area for Color:Shape combinations of 0.7:0.3 – 0.9:01. The first segmentation result that was significant at $\alpha = 0.05$ (F-value) and the first segmentation result where the number of segments was greater than 2,000 (0.473 ha/segment), are shown to serve as comparison in each case.

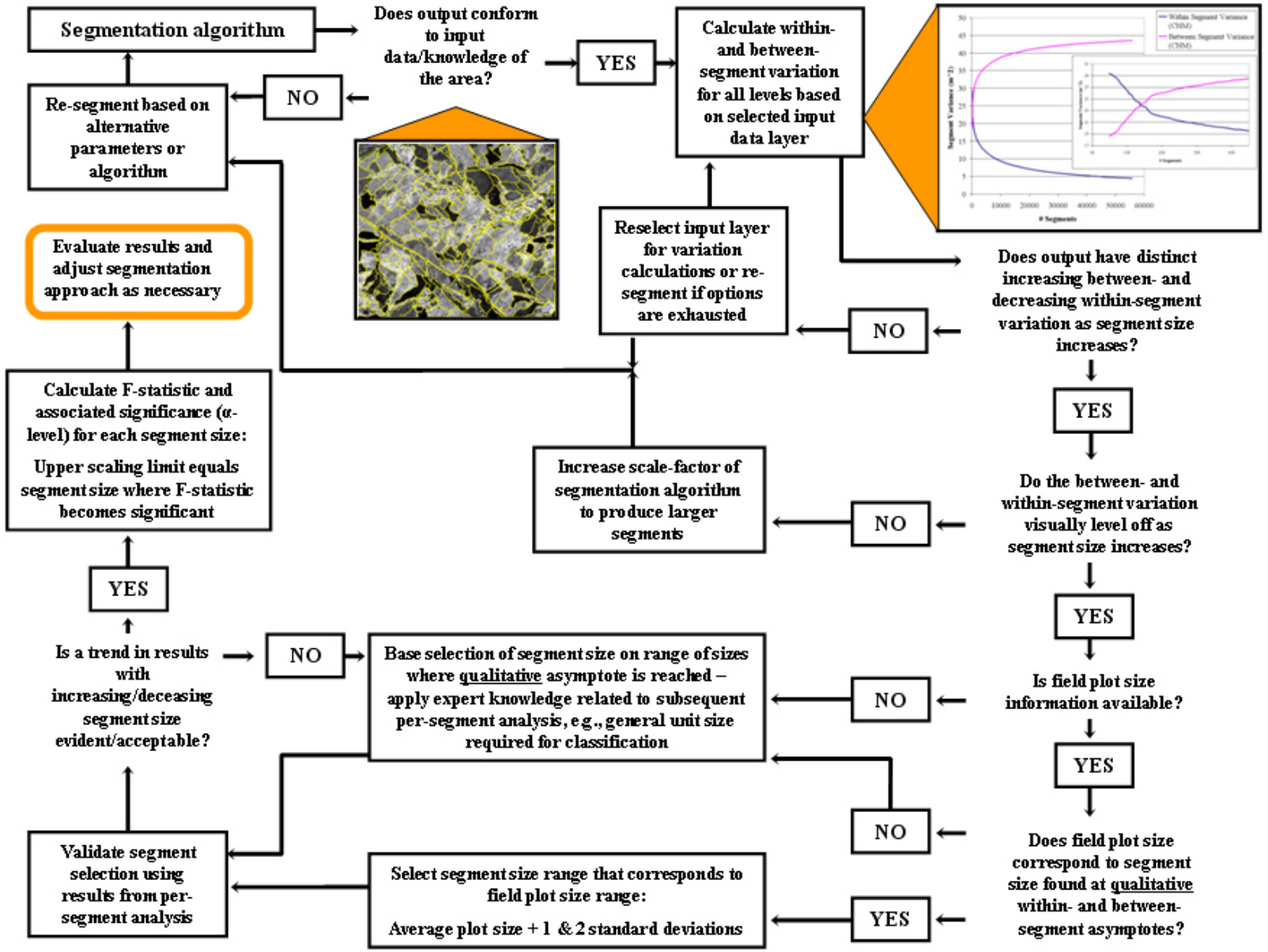


Figure 5.5 Conceptual protocol for selection of segmentation results for subsequent analyses, e.g. per-segment volume modeling and classification

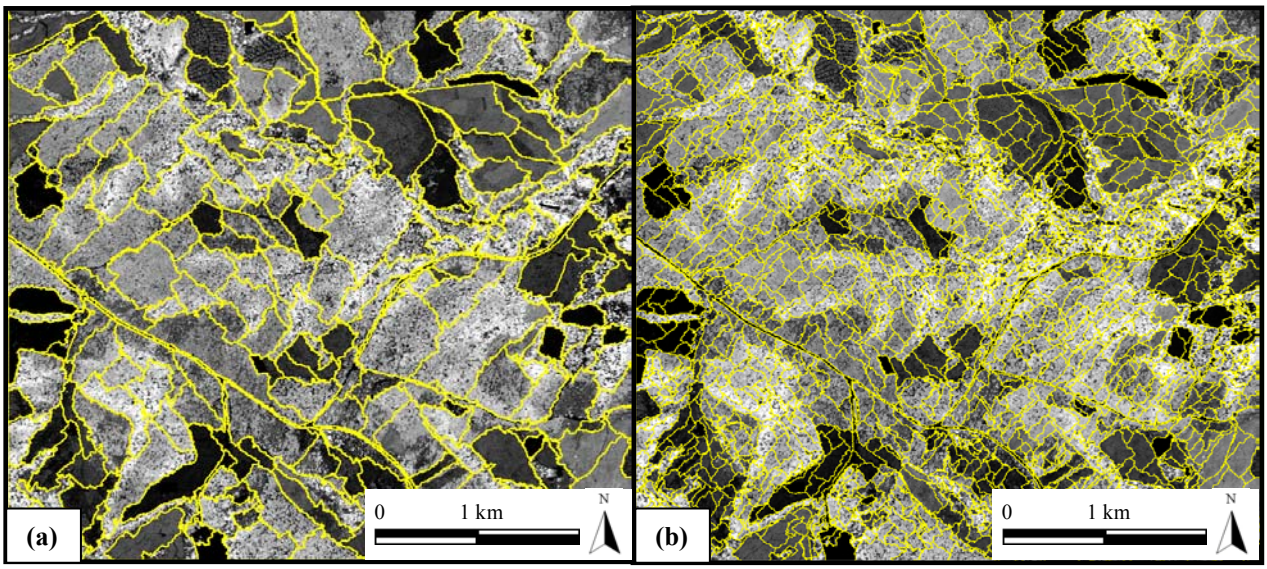


Figure 5.6 Segmentation results (Color:Shape = 0.7:0.3) for (a) 232 (4.078 ha/segment) and (b) and 2,037 (0.464 ha/segment) segments with the lidar CHM as backdrop

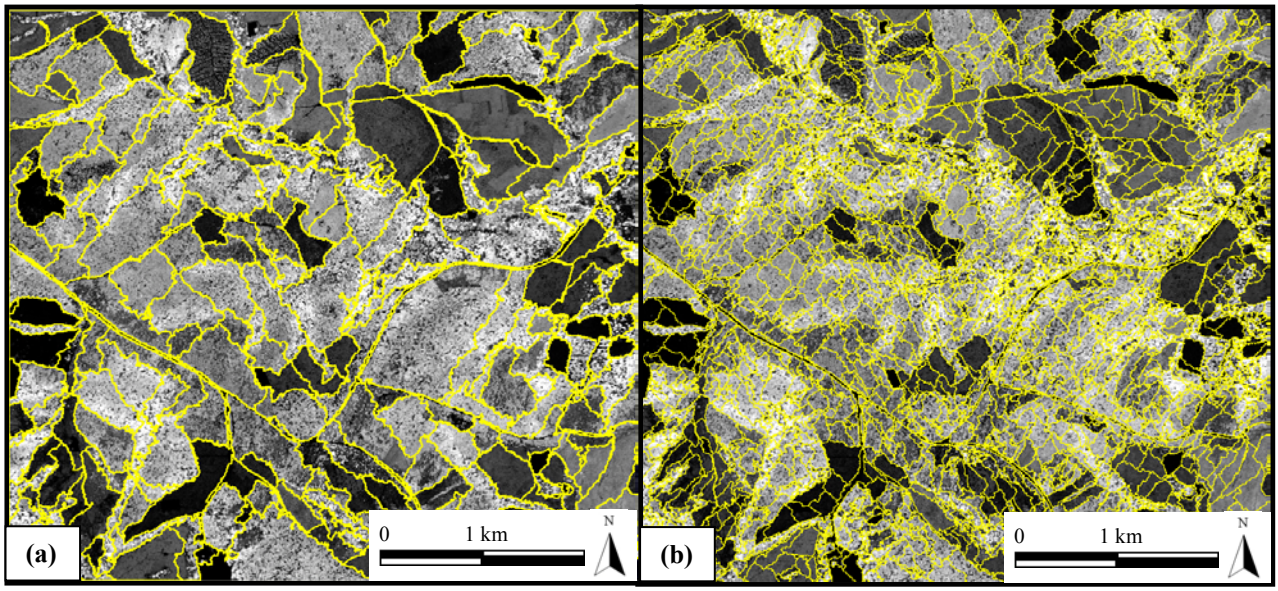


Figure 5.7 Segmentation results (Color:Shape = 0.8:0.2) for (a) 240 (3.942 ha/segment) and (b) and 2,002 (0.473 ha/segment) segments with the lidar CHM as backdrop

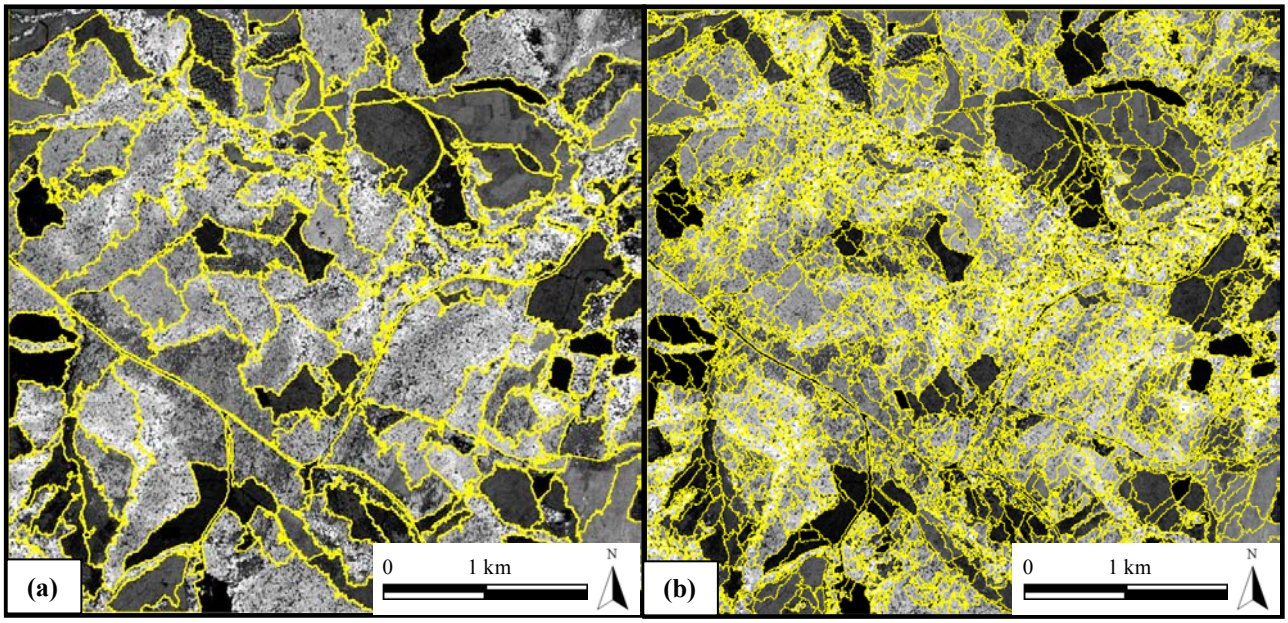


Figure 5.8 Segmentation results (Color:Shape = 0.9:0.1) for (a) 167 (5.666 ha/segment) and (b) and 2,005 (0.472 ha/segment) segments with the lidar CHM as backdrop

The larger segmentation result is a probable upper scaling limit, while the 2,000 (0.473 ha/segment) segment result represented an outcome where each BAF plot represents only one segment. It is clear from the larger and smaller segmentation results in Figures 5.6 - 5.8 that the Color:Shape weights had a distinct influence on algorithm performance. On the one extreme, where Color:Shape = 0.7:0.3, segments were smoother, or less fractally shaped. As the Color:Shape ratio increased to 0.9:0.1, it became evident that the segments were less smooth, or more fractally shaped (Figure 5.8). The effect of an increase in shape weight resulted in deviations from the color information (structural CHM), with the benefit of more appealing segment boundaries. However, segmentation correspondence to the structural height input of the CHM was considered critical, while visual appearance was secondary. The segmentation result where Color:Shape = 0.8:0.2 was chosen as preferred candidate, given the trade-off in correspondence to input data and visual appeal.

5.3.2 Evaluation of Between- and Within CHM Variance

Between- and within CHM segment variance [Equation 3] have potential as a method to gauge segmentation results appropriate for further per-segment analyses. Figures 5.9 - 5.11 show that in each case (Color:Shape weights) there was a definite intersection where between-segment variability started, and continued to exceed within-segment variability as segment size decreased. This intersection varied with algorithm parameter inputs, but in all cases approximately corresponded to the operational stand map (167 stands). Intersections occurred at relatively few and large segment outcomes, after which the curves followed either a growing (between-variance) or decreasing (within-variance) exponential-type curve. Except for the location of this intersection of between- and within variance curves, there appeared to be no distinct difference in curve shapes for the three Color:Shape segmentation inputs. It was initially thought that segment results suited to further per-segment analyses would be where the both within- and between variance curves started to level off. This was based on the assumption that most of the overall variance has been absorbed by between-segment variance, while within-segment variance was approaching a minimum. In theory this would ensure structurally homogenous, but statistically distinct segments. Such a selection therefore is amenable to both per-segment volume modeling and classification, two likely segment-based extensions, due to low within-segment variability and high between-segment variability. These assumptions were not corroborated by the ensuing per-segment analyses. Although general BAF plot size was chosen as an indicator of segment sizes useful for further analyses, this metric only was suited to define the smallest possible average segment size.

5.3.3 The Between-Within-segment Ratio as an F-statistic

Since the between-within variance ratio is by definition an F-value, it was used in all three segmentation runs to determine when average segment size became significant. The commonly used α-value of 0.05 was selected to indicate significance (Foody, 2004). Table 5.3 lists the segmentation scenarios for each run where the F-value became significant in terms of the F-statistic. These values indicate the average segment size where between-segment variance became significantly larger than within-segment variance. Table 5.3 shows that significant

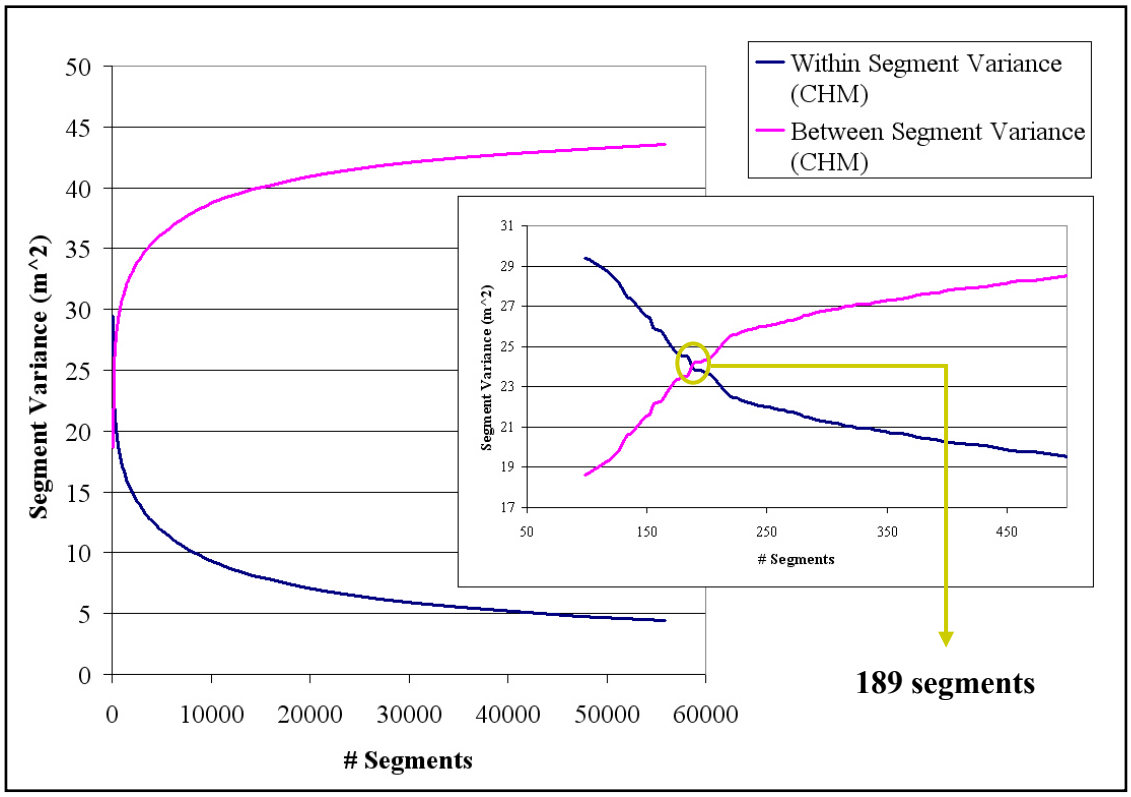


Figure 5.9 Within and between CHM segment variances for eCognition (Color:Shape = 0.7:0.3)

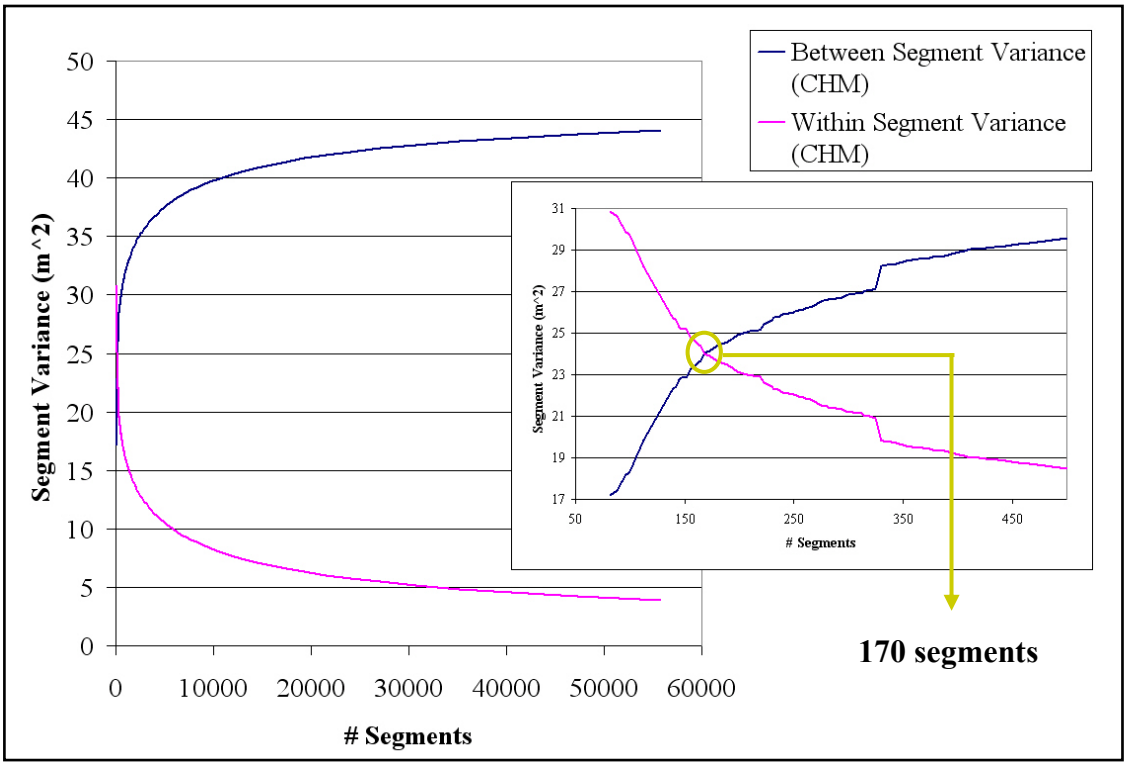


Figure 5.10 Within and between CHM segment variances for eCognition (Color:Shape = 0.8:0.2)

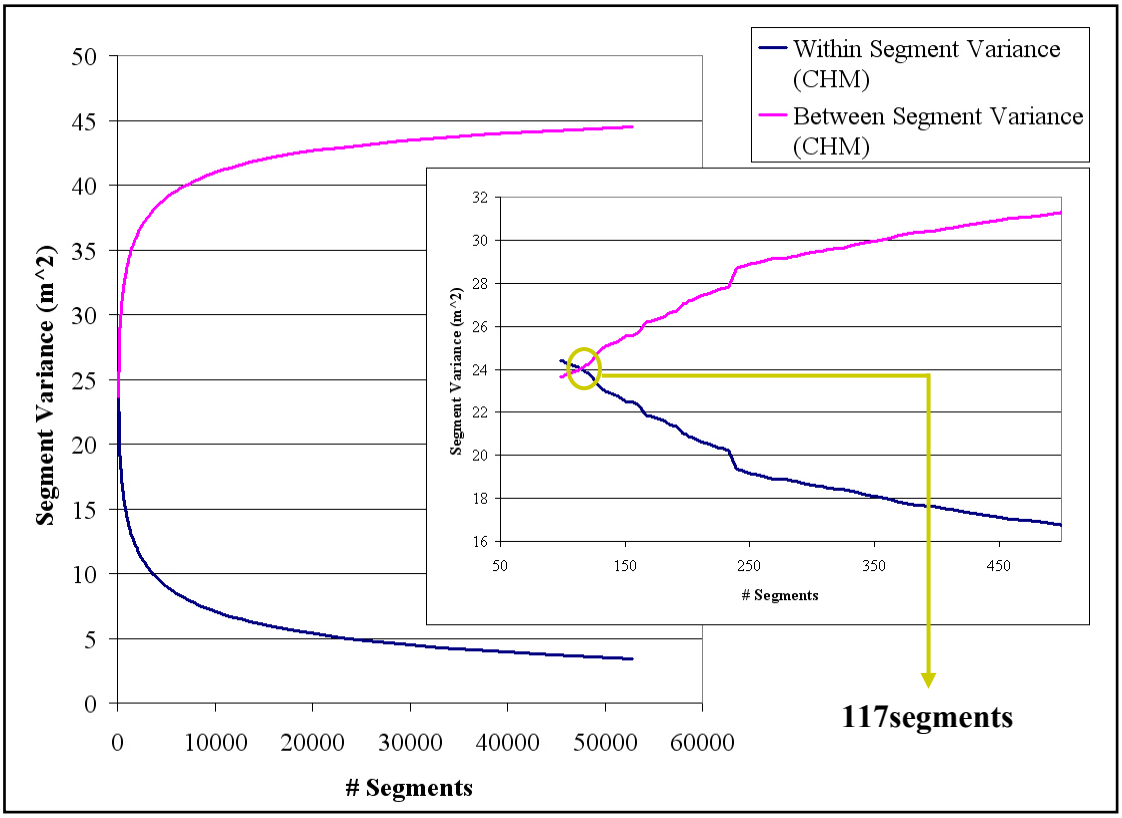


Figure 5.11 Within and between CHM segment variances for eCognition (Color:Shape = 0.9:0.1)

segmentation results did not occur at the exact between-within-segment variance intersections. In all three segmentation runs the between-within variance ratio only became significant at larger segment numbers and smaller sizes. These significant segment numbers were all operationally viable, being either slightly fewer (Color:Shape = 0.9:0.1) or more (Color:Shape = 0.7:0.3 and 0.8:0.2) than the 167 existing forest stands.

Table 5.3 Number of segments, F-value (α = 0.05; degrees of freedom associated with number of segments and total number of pixels), and between-within variance ratio for the first significant segmentation outcome

Segmentation Run 1 Color:Shape = 0.7:0.3			**Segmentation Run 2 Color:Shape = 0.8:02**			**Segmentation Run 3 Color:Shape = 0.9:0.1**		
Segments	F-value (α=0.05)	Variance Ratio	Segments	F-value (α=0.05)	Variance Ratio	Segments	F-value (α=0.05)	Variance Ratio
232 (4.078 ha/segment)	1.158	1.16	240 (3.942 ha/segment)	1.155	1.169	167 (5.666 ha/segment)	1.187	1.2

The F-statistic proved to be theoretically useful for determining statistical significance of segmentation results, or stated differently, the upper scaling limit for operational use. However, both modeling and classification results at segment sizes greater than the largest significant size were viable. This indicated that this upper scaling limit should be carefully evaluated by the user. It is based on a theoretically sound decision, but results could be influenced by the chosen segmentation algorithm and characteristics of the geographical area. Although the significance level of $\alpha = 0.05$ is frequently used in statistical tests (Foody, 2004), this value is somewhat arbitrary and can be adjusted based on expert knowledge. The variance plots were in turn useful for visual evaluation of segment variance behavior. Between- and within CHM variance therefore were critical to the decision of scaling limit (statistical significance) and visual inspection of where variance curves leveled off.

5.3.4 Determination of Segment Size for Extension to Further Analyses

Due to the non-asymptotic nature of the between-within CHM variance plots (Figures 5.9 – 5.11), an alternative method was required to estimate the minimum segment size for subsequent analyses. Since variance plots qualitatively leveled off in both variance terms from approximately 25,000 segments upwards, segmentation results of equal or greater numbers potentially would be viable options for further extensions. This assumption was based on the degree to which overall variance had been distributed between the two secondary variance components at these segment sizes. Between-segment variance reached a maximum, while within-segment variance was at an approximate minimum, albeit only qualitatively.

Implementation of a decision rule based on the general BAF plot size proved useful in defining the lower limit of average segment size for subsequent per-segment analyses. The distances of trees from BAF plot centers were not normally distributed (Figure 5.4.a), but log-transformed distances were approximately normal based on visual inspection (Figure 5.4.b). Due to the large number of observations and associated degrees of freedom (> 1,500), a test for normality could not be performed and had to be based on visual inspection. Table 5.4 lists the segment sizes for all three segmentation runs, represented by the radii defined by the average distance from plot center plus σ and 2σ.

A plot defined by the average distance plus σ was almost 4 times smaller than a plot based on average distance plus 2σ. This resulted in a plot size for 1σ of 0.037 ha, or 25,667 segments in the study area, compared to a plot size for 2σ of 0.135 ha, or 7,022 segments. The number of segments represented by these two plot sizes are shown in Figure 5.12, plotted on the variance curves for the Color:Shape = 0.8:0.2 segmentation run. From Figure 5.12 it is clear that the plot size indicated by average tree distance plus σ, resulted in a decision rule that will likely be more viable for subsequent analyses as opposed to adding 2σ to the average distance.

In the first case both variance curves have started to level off, while much of the overall variance could still be absorbed by the between-segment variance component in the case of 2σ. A decision rule, based on the average tree distance from plot center plus σ, therefore was a logical choice as opposed to a larger segment size with more variability.

Table 5.4 Decision rule to determine segmentation results to be used for subsequent volume and biomass model fitting, as well as associated segment numbers and sizes for segmentation runs

Decision Rule Plot Size (log-normal)	Segmentation Run 1 Color:Shape = 0.7:0.3		Segmentation Run 2 Color:Shape = 0.8:02		Segmentation Run 3 Color:Shape = 0.9:0.1	
$\bar{x}+\sigma$ =10.83 m (0.037ha; 25,667 segments)	27,049 segments	0.035 ha/segment	27,050 segments	0.035 ha/segment	28,988 segments	0.033 ha/segment
$\bar{x}+2\sigma$ =20.71 m (0.135 ha; 7,022 segments)	6,870 segments	0.138 ha/segment	6,687 segments	0.141 ha/segment	6,870 segments	0.138 ha/segment

5.3.5 Validation of Segmentation Choices

Object-oriented volume modeling and discriminant classification (2-class: deciduous-coniferous) were used as validation of segmentation selection criteria across a range of average segment sizes from 0.035 ha/segment – 5.632 ha/segment, as well as for the existing forest stands. Table 5.5 shows adjusted R^2 and RMSE values, the overall classification accuracies, and the F-statistic (between-within variance ratio) for eleven segmentation results across this segment size range, as well as for the existing forest stands in the study area. Figures 5.13 – 5.16 graphically show the trends, if any, for each of these metrics as segment size increases.

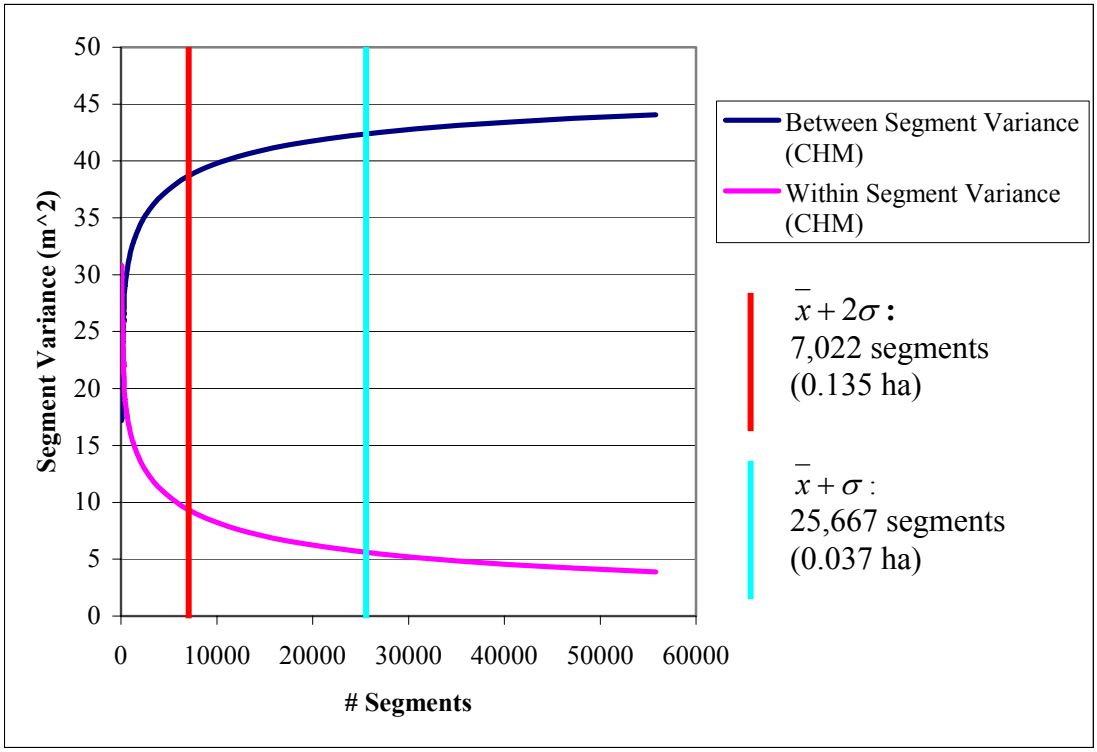


Figure 5.12 Between-within CHM variance plots for Color:Shape = 0.8:0.2 with BAF plot size decision rules. Average tree distance + 1σ resulted in a segment size that possibly was most viable for further analysis

It is clear from Figures 5.14 – 5.16 that the trend of decreasing F-statistic values with increasing segment size, shown in Figure 5.13, was not reflected in other metrics. This decreasing trend in F-statistic was statistically defined by distinct between-segment variance decreases at the cost of increasing within-segment variance. It was expected that the quality of model and classification metrics also would decrease as within-segment variance increased with average segment size.

Increasing within-unit variance generally leads to less precise estimates (Avery and Burkhart, 1994; Shivers and Borders, 1996). Adjusted R^2 values (Figure 5.14) showed a limited decrease as segment size increased, and although only marginal differences, RMSE values (Figure 5.15) increased as average segment size increased. There were very few significant differences between the highest accuracy (1.885 ha/segment: 89.2%) and lower accuracies (0.964 ha/segment: 81.4%; 0.035 ha/segment: 82.2%).

Table 5.5 Adjusted R^2 and RMSE values, and classification accuracies for all selected segmentation outcomes

Segmentation result	Volume Adjusted R^2	Volume RMSE (m^3/ha)	Overall Classification Accuracy (%)	F-statistic (Between-within variance ratio)
27,050 segments (0.035 ha/segment)	0.59	52.78	82.2	7.7717
10,352 segments (0.091 ha/segment)	0.59	53.75	84.5	6.0146
6,687 segments (0.141 ha/segment)	0.56	55.18	85.4	4.9126
2,972 segments (0.318 ha/segment)	0.56	55.74	84.5	4.0494
1,473 segments (0.642 ha/segment)	0.54	56.73	82.6	2.9034
981 segments (0.964 ha/segment)	0.54	57.25	81.4	2.5186
749 segments (1.263 ha/segment)	0.55	56.85	83.5	2.2555
502 segments (1.885 ha/segment)	0.57	54.79	89.2	1.9848
374 segments (2.530 ha/segment)	0.54	55.26	82.6	1.808
240 segments (3.942 ha/segment)	0.52	56.23	81.9	1.5994
168 segments (5.632 ha/segment)	0.54	52.16	87.0	1.472
167 Appomattox forest stands (5.666 ha/segment)	0.42	62.36	83.2	None

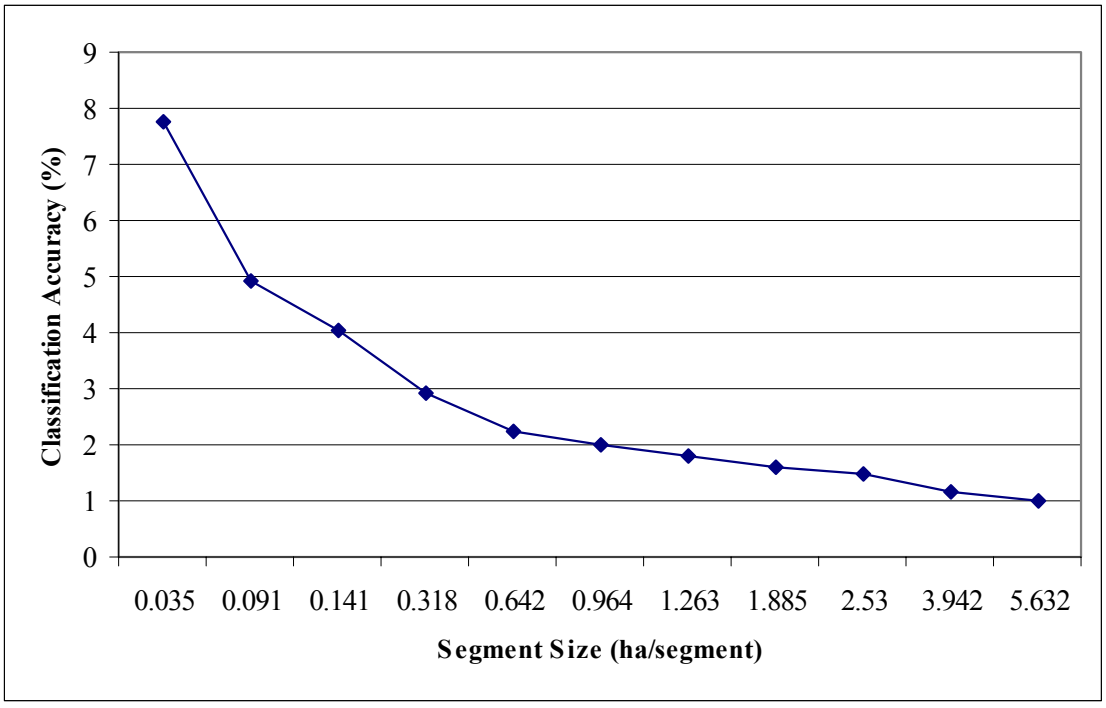


Figure 5.13 F-statistic (between-within variance ratio) with increasing average segment size

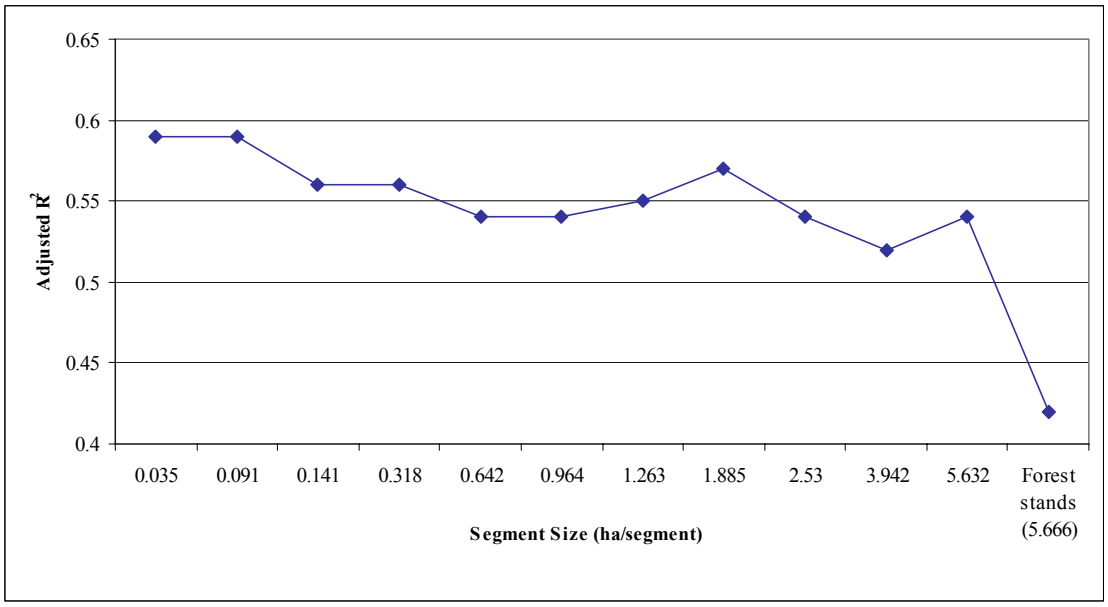


Figure 5.14 Overall volume modeling adjusted R^2 with increasing average segment size

Figure 5.15 Overall volume modeling RMSE with increasing average segment size

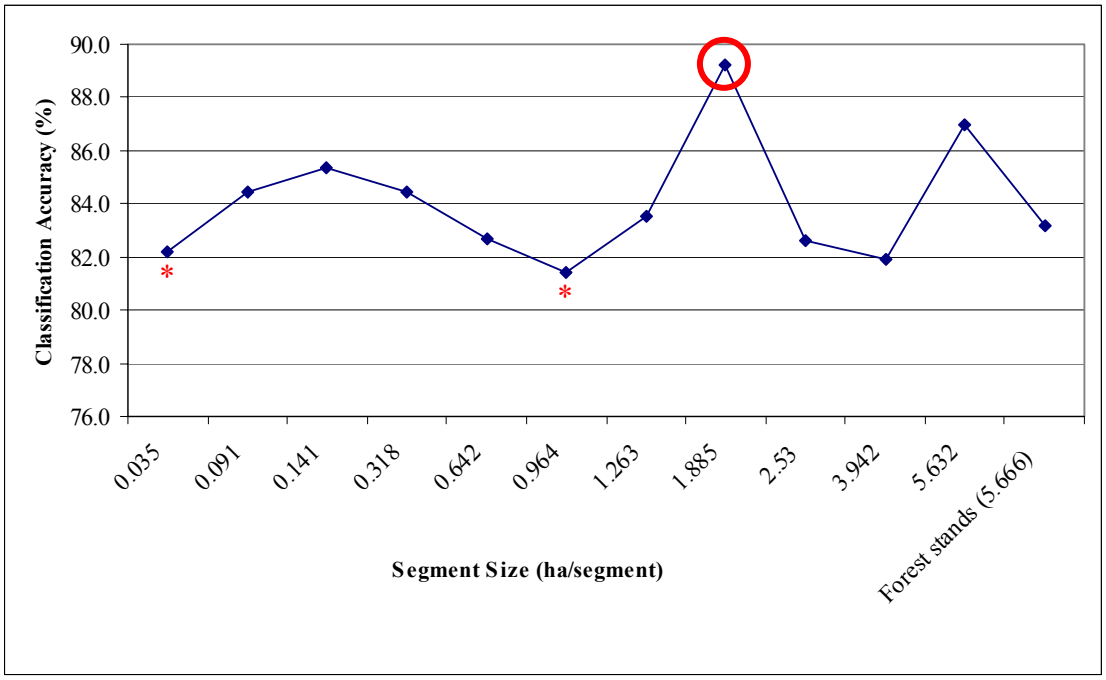


Figure 5.16 Overall deciduous-coniferous classification accuracies with increasing average segment size (significant differences between the highest classification accuracy (◯) are shown by *

Differences between segmentation results and adjusted R^2 and RMSE values for the existing forest stands were more pronounced (Table 5.5). There was a distinct decrease in adjusted R^2 and increase in RMSE values when the volume model was applied to the 167 operational forest stands. This indicated that segmentation of the study area using lidar-based data resulted in a better object-definition, in terms of the modeling results. Structural segmentation therefore had an advantage over traditional stand delineation methods (site, forest type, etc.) in terms of the analyses. However, both forest segmentation and subsequent analyses were based on lidar height data, while forest stands were extracted by other methods. This could have contributed to the poorer analysis metrics in the latter case.

It is inevitable that final validation of selected segmentation results can only come from subsequent per-segment analyses. Although defined criteria have been presented with which to select candidate segment sizes, application of these selected segmentation outcomes to a real-world application have not proven effective. This was attributed to three main factors. Firstly, BAF plots were assumed to represent entire segments or stands, while a full-segment inventory was not operationally feasible. However, such a full inventory could corroborate the expected

decreasing fit metrics as within-segment variance decreases, since a full inventory will result in complete segment description. Secondly, the results from this study require further investigation, given the small range of measured volume values (6.94 – 350.93 m^3/ha) and relatively small range in CHM segment variances.

Previous lidar distributional volume modeling studies (Means et al., 2000; Næsset, 2002) investigated volume ranges of between 18 m^3/ha and 2,051 m^3/ha, and 41 m^3/ha and 639.8 m^3/ha, respectively. Evaluation of segmentation-based modeling for such observed ranges therefore is warranted to conclusively determine the utility of a segment-based approach. Overall CHM variance was on the order of 48.05 m. Within-segment variance ranged from 5.47 m^2 to 22.06 m^2, while between-segment variance ranged from 25.99 m^2 to 42.52 m^2 within the evaluated average segment size range. This variability range is not exceedingly large, which might have contributed to larger segments, with associated increased within-segment variance, not being operationally different from smaller average segment sizes. Although these differences were significant based on F-tests ($\alpha = 0.05$), it was not reflected in the modeling exercise. Lastly, per-segment analyses were based on segments derived through a hierarchical segmentation procedure. The observed lack of poorer model metrics and classification accuracies at larger segment sizes was attributed to the hierarchical nature of the segmentation algorithm. The algorithm was based on the minimization of within-segment variance at all segment levels, but since smaller segments constituted the building blocks for larger objects, variance minimization already was adequately addressed at lower levels.
Therefore, all segmentation evaluation criteria should be seen as indicators of possible solutions, while expert knowledge and experience should never be discounted. The conceptual protocol presented in Figure 5.5 has a sound statistical basis, but its applicability to other case studies warrants further investigation.

5.4 Conclusions

Multiresolution, hierarchical segmentation was regarded as a precursor to subsequent per-segment analysis. Accurate segmentation of a forested area is needed before per-segment analyses, e.g., volume model development and object-oriented classification, can be performed.

Homogenous segments (small within-segment variability) that are distinctly different from each other (large between-segment variance) are therefore required (Avery and Burkhart, 1994; Shivers and Borders, 1996). This necessitates not only segmentation as pre-processing activity, but also the evaluation and validation of segmentation results.

The use of the lidar-derived canopy height model (CHM) enabled segment delineation based on structural properties, important if subsequent applications include of volume and biomass model development, both of which have a structural basis. This can be contrasted with spectral segmentation, where specific wavelength information is the determinant of the segmentation outcome. Although the latter approach has merit when species specific information is required, a structural segmentation would likely follow not only structural boundaries, but also forest type definitions. This is due to forest stands (deciduous vs. coniferous) being unique in their structural properties (Douglas *et al.,* 2003), as well as the fact that most forest stands are delineated by stand breaks or forest roads. The multiresolution, hierarchical segmentation method used in this study (eCognition algorithm) performed well, both from a qualitative and quantitative perspective. It also was applicable to possible per-segment analyses due to the scalable nature of results, based on the hierarchical approach of the algorithm. This resulted in topologically sound segmentation results at all scales, with smaller segments constituting the building blocks for larger segments. In terms of visual correspondence the eCognition algorithm conformed to structural boundaries evident from the CHM image. Stand and road breaks were accurately detected, but managerial objects, e.g., fire breaks and power lines, could hinder homogenous object definition in cases where such a feature intersected a stand. Visual correspondence was consistent at all segmentation scales, from small to large segments. Although not definitive, visual correspondence to input data is a critical factor, and always should be considered (Schiewe *et al*., 2001; Schiewe, 2002).

Segmentation results also were valid from a statistical perspective. Between- and within-segment variance plots of the CHM (height) illustrated that within-segment variability increased as between-segment variability decreased with increasing average segment size. Smaller segments were thus potentially better suited to statistical manipulation like volume model fitting, resulting in models with lower RMSEs. This was due to the lower structural within-segment variation

when smaller, homogeneous segments were compared to larger, more heterogeneous segments. The between-within variance intersect also provided a guideline as to the level of scalability of segmentation results. Coupled with an F-statistic for each segmentation outcome, this intersect was critical to determination of the eventual, operational segment size. These two metrics were well suited to defining eventual operational stand sizes, and along with structurally-based segmentation, could help in redefining more homogenous forest stand better suited to management and inventory.

Deciding on the segment size to use for further application also has been a difficult question to address to date. Most authors used classification accuracy assessment (Antunes *et al*., 2003; Darwish *et al*., 2003; Heyman *et al*., 2003; Neubert, 2001) visual correspondence to input data along with number of segments as decision criteria for thematic output (Schiewe *et al*., 2001; Schiewe, 2002). Segment selection for volume- and biomass modeling mainly has been based on tree- (Hyyppä and Inkinen, 1999) or stand approximations (Woodcock *et al.,* 1994), or was circumvented by selecting plots within a segment (Makela and Pekkarinen, 2001; Pekkarinen 2002). The use of the average tree distance from BAF plot center plus one or two standard deviations resulted in a useful method for defining the minimum segment size and number at which statistical models could be developed. This radius equaled a plot size that approximated the segment size and number where between- and within-segment variance curves leveled off. This outcome resulted in a segment size where little of the overall variance could still be absorbed by the between-segment variance, i.e. resulting in homogenous segments that were different from each other in terms of height. A conceptual protocol was presented by which a user can evaluate segmentation results based on the presented criteria, namely within- and between-segment ratio and field plot size, while the significance of the F-statistic can be used as an indicator of upper scaling limit.

However, segment size selection based on the provided criteria was not corroborated by subsequent object-oriented volume modeling and forest type classification. There were no distinct reductions in model fit metrics and classification accuracies as average segment size increased. This was contrary to expectation, but was attributed to three main factors. While BAF plots were assumed to represent entire segments or stands, a full per-segment inventory could

corroborate the expected decreasing fit metrics as within-segment variance decreases. Such a complete inventory will result in complete segment description. Validation results for this study also were based on a relatively small measured volume range, while other lidar distributional studies used a larger range of observed values. Extension of the developed methods for segment selection to larger observed ranges therefore is warranted to conclusively determine the utility of the proposed selection criteria. Lastly, the hierarchical nature of the segmentation algorithm resulted in larger segments consisting of smaller, homogenous segments with low within-segment height variance. Minimization of within-segment variance therefore was already addressed at smaller segment levels, with larger segments consisting of these homogenous building blocks. Analyses were ultimately applied to segment levels with minimized within-segment variability across the entire range of average segment sizes.

However, segment-based volume modeling was shown to improve model fit metrics and RMSE values as opposed to modeling based on existing operational stands in the study area. This indicated that segmentation of the forested study area was indeed warranted. Whereas current stand delineation was based on site, forest types, and growth-and-yield measurements, segmentation resulted in a true structural definition of the study area, using lidar data. Resultant structurally homogenous segments were more conducive to an object-oriented volume modeling approach than defined stands, proving that the current definition of measurement and management units should be reconsidered.

In conclusion, the use of between- and within-segment variation, an F-statistic calculated on a per-segment outcome, and average BAF plot size, have potential as criteria for deciding which segment level should be useful in subsequent analyses. It should be noted, however, that visual segment correspondence to input data is critical. A human interpreter usually has a keen eye for delineating unique image areas. Due to the lack of objectivity, repeatability, and speed, segment delineation based on human interpretation is not an option. However, human interpretation of results is useful to verify segmentation output. This study focused on the selection of the best individual segmentation outcome, within a set of segmentation results. The conceptual decision model that was presented has potential, but as with any such approach, the user should practice caution due to its “fuzzy” nature. In many cases only a range of average segment sizes will be

defined, while selection and verification could require field visits. It is recommended that future work should focus on validation of selection criteria for increased dependent variable ranges, as well as development of other possible decision rules.

5.5 Acknowledgements

This research was made possible by funding from NASA (grant # NG65-10548; NG613-03019), the McIntire-Stennis Research Program (grant # VA-136589), the Forestry Department and Graduate Student Association at Virginia Polytechnic Institute and State University, and the Potomac chapter of the American Society for Photogrammetry and Remote Sensing. Field data collection was supported by the Virginia Department of Forestry, specifically Dr. John Scrivani, Todd Edgerton, Ralph Toddy, and Wayne Bowman (VDOF). Drs. Richard Oderwald (Virginia Polytechnic Institute and State University) and Sorin Popescu (Texas A&M University) provided invaluable assistance with statistical and lidar analyses. Amy Zhang from the Statistics department at Virginia Polytechnic Institute and State University served as statistical consult.

5.6 Literature Cited

Acton, S.T., 1996. On supervised segmentation of remotely sensed imagery using nonlinear regression. *International Journal of Remote Sensing* 17 (7): 1407-1415.

Antunes, A.F., C. Lingnau, and J.C. Da Silva, 2003. Object oriented analysis and semantic network for high resolution image classification. In: Anais XI SBSR, Belo Horizonte, Brazil, April 5-10, 2003, INPE. pp. 273-279.

Avery, T.E., and H.E. Burkhart, 1994. Forest Measurements. 4th Edition. McGraw-Hill, Boston, USA. 408 p.

Baatz, M., and A. Schäpe. 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: STROBL, J. *et al.* (Hrsg.): Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg, September 2000, Karlsruhe, Herbert Wichmann Verlag: 12–23.

Burrough, P.A., and R.A. McDonnel, 1998. Principles of Geographical Information Systems. Oxford University Press. Oxford, England. 333 pp.

Darwish, A., K. Leukert, and W. Reinhardt, 2003. Image segmentation for the purpose of object-based classification. Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse 3 pp.

Douglas, T.E., D.L. Evans, K.L. Belli, and S.D. Roberts, 2003. Classification of pine and hardwood by the distribution and intensity of lidar returns. ISPRS "Three dimensional mapping workshop from InSAR and LIDAR", June 17-19, 2003, Portland, Oregon, USA. 5pp.

eCognition User's Manual, 2003. Definiens Imaging (www.definiens.com).

Engdahl, M.E., J. Pulliainen, and M. Hallikainen, 2003. Combined land-cover classification and stem volume estimation using multitemporal ERS tandem INSAR data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Foody, G.M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing* 70 (5): 627-633.

Häme T., and E. Tomppo, 1987. Stand based forest inventory from SPOT image – first experiments. SPOT first in-flight results.

Heyman, O., G.G. Gaston, A.J. Kimerling, and J.T. Campbell, 2003. A per-segment approach to improving Aspen mapping from high-resolution remote sensing imagery. *Journal of Forestry* (June, 2003): 29-33.

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

Hyyppä J., and M. Inkinen, 1999. Detecting and estimating attributes for single trees using laser scanner. *Photogrammetric Journal of Finland* 16: 27-42.

Jaakkola S.P., 1989. Applicability of SPOT for forest management. *Advanced Space Research* 9 (1): 135-141.

Jain, R., K. Rangachar, and B.G. Schunck, 1995. Machine vision. MIT Press and McGraw-Hill, Inc., New York. 549 p.

Kayitakire, F., C. Farcy, and P. Defourny, 2002. Ikonos-2 imagery potential for forest stands mapping. ForestSAT Symposium, Heriot Watt University, Edinburgh, August 5-9, 2002. 11 pp.

Kellndorfer, J.M., and F.T. Ulaby, 2003. Forest biomass inversion from SAR using object oriented image analysis techniques. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 3 pp.

Kermad C.D., and K. Chehdi, 2002. Automatic image segmentation system through iterative edge-region co-operation. *Image and Vision Computing* 20: 541-555.

Kilpeläinen, P., and T. Tokola, 1999. Gain to be achieved from stand delineation in Landsat TM image-based estimates of stand volume. *Forest Ecology and Management* 124 (1999): 105-111.

Kressler, F., Y. Kim, and K. Steinnocher, 2003. Object-oriented land cover classification of panchromatic KOMPSAT-1 and SPOT-5 data. In: Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse. 7 pp.

Lefsky M.A., W.B. Cohen, G.G. Parker, and D.J. Harding, 2002a. Lidar remote sensing for ecosystem studies. *Bioscience* 52 (1): 19-30.

Lefsky M.A., W.B. Cohen, D.J. Harding, G.G. Parker, S.A. Acker, and S.T. Gower, 2002b. Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology & Biogeography* 11 (2002): 393-399.

Li W., G.B. Benie, D.C. He, S. Wang, D. Ziou, and Q.H.J. Gwyn, 1999. Watershed-based hierarchical SAR image segmentation. *International Journal of Remote Sensing* 20 (17): 3377-3390.

Makela H., and A. Pekkarinen, 2001. Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing of Environment* 77:66-75.

McCormick N., and S. Folving, 1998. Monitoring European forest biodiversity at regional scales using satellite remote sensing. Assessment of biodiversity for improved forest planning, 283-289, P. Bachman *et al.* (eds.). Kluwer Academic Publishers, Netherlands.

Means J.E., S.A. Acker, B.J. Fitt, M. Renslow, L. Emerson, and C.J. Hendrix, 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering & Remote Sensing* 66 (11): 1367-1371.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80 (2002):88-99.

Neubert, M., 2001. Segment-based analysis of high resolution satellite and laser scanning data. In: Hilty, L. M. & Gilgen, P. W. (Eds.): Sustainability in the Information Society. Proceedings of the 15th International Symposium Informatics for Environmental Protection, Zurich, October 10-12, 2001, Marburg. pp. 379-386.

Nugroho, M., D.H. Hoekman, and R. De Kok, 2002a. Analysis of the forests spatial structure using SAR and Ikonos data. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 10 pp.

Nugroho, M., D.H. Hoekman, and R. de Kok, 2002b. Analysis of forest spatial structure using spatial decision rule. Presented at ForestSAT Symposium Heriot Watt University, Edinburgh, August 5-9, 2002. 8 pp.

Oderwald, R.G., 2003. Formula derivation for within and between-segment variation. *Personal communication*.

Pekkarinen A., 2002. Image segment-based spectral features in the estimation of timber volume. *Remote Sensing of Environment* 82: 349-359.

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

Popescu, S.C., R.H. Wynne, and J.A. Scrivani, 2004. Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, U.S.A. *Forest Science* (In press)

Russ J.C., 1995. The image processing handbook. CRC Press, Ann Arbor, pp. 674.

Ryherd S., and C. Woodcock, 1990. The use of texture in image segmentation for the definition of forest stand boundaries. Proceedings of the 23rd International Symposium on Remote Sensing of Environment, Bangkok, Thailand, April 18-25, 1990. pp. 1209-1213.

Schiewe, J., L. Tufte, and M. Ehlers, 2001. Potential and problems of multi-scale segmentation methods in remote sensing. *GeoBIT/GIS* 6: 34-39.

Schiewe, J., 2002. Segmentation of high-resolution remotely sensed data - concepts, applications and problems. Proceedings of the Joint International Symposium on Geospatial Theory, Processing, and Applications. Ottawa, Canada, July 9-12, 2002. 6 pp.

Shandley, J., J. Franklin, and T. White, 1996. Testing the Woodcock-Harward image segmentation algorithm in an area of southern California chaparral and woodland vegetation. *International Journal of Remote Sensing* 17 (5): 983-1004.

Shankar, B.U., C.A. Murthy, and S.K. Pal, 1998. A new gray level based Hough transform for region extraction: An application to IRS images. *Pattern Recognition Letters* 19 (1998): 197-204.

Shivers, B.D., and B.E. Borders, 1996. Sampling techniques for forest resource inventory. John Wiley and Sons, Inc. New York, USA. 356 p.

Ton J., J. Sticklen, and A.K. Jain, 1991. Knowledge-based segmentation of Landsat images. *IEEE Transactions on Geoscience and Remote Sensing* 29 (2): 222-232.

van Aardt, J.A.N., and R.H. Wynne, 2001. Spectral separability among six southern tree species. *Photogrammetric Engineering & Remote Sensing* 67 (12): 1367-1375.

Wilson R., and M. Spann, 1988. Image segmentation and uncertainty. Research Studies Press Ltd., Letchworth, Hertfordshire, England, 180 pp.

Wong T.H., S.B. Mansor, M.R. Mispan, N. Ahmad, and W.N.A. Sulaiman, 2003. Feature extraction based on object oriented analysis. Proceedings of the 2003 ATC Conference, May 20-21, 2003, Malaysia. 10 pp.

Woodcock, C.E., V. Jakabhazy, S. Macomber, S. Ryherd, A.H. Strahler, and Y. Wu, 1990. Timber inventory using Landsat Thematic Mapper imagery and canopy reflectance models. Proceedings of the 23rd International Symposium on Remote Sensing of Environment, Bangkok, Thailand, April 18-25, 1990. pp. 937-943.

Woodcock, C.E., J.B. Collins, S. Gopal, V.D. Jakabhazy, X. Li, S. Macomber, S. Ryherd, V.J. Harward, J. Levitan, Y. Wu, and R. Warbington, 1994. Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment* 50: 240-254.

CHAPTER 6

CONCLUSIONS

6.1 Introduction

An object-oriented approach to volume- and biomass-by-type has been shown to have potential for operational forest inventory. Segmentation of the study area, as well as subsequent per-segment modeling and classification, was based on lidar data as the input data source. Segmentation of a canopy height model (CHM), followed by modeling and classification analyses, resulted in model fit statistics and classification accuracies comparable to those found for other approaches in literature. However, this research was unique in that (i) basic processing units were defined as unique, homogenous segments in terms of the CHM (ii) segments were derived from the data source used for all forest inventory analyses, and (iii) results for subsequent analyses were scalable to operational sizes due to the hierarchical nature of the segmentation algorithm. Specific conclusions related to object-oriented volume- and biomass modeling, deciduous-coniferous classification, and selection of segmentation outcomes for further analyses are discussed next.

6.2 Conclusions: Object-oriented Volume- and Biomass Modeling

Grid-cell volume and biomass modeling based on lidar distributions have been implemented successfully by Means *et al*. (2000) and Næsset (2002). These studies were limited to coniferous species, but R^2 values upwards of 0.90 bode well for future lidar distributions studies. This study explored an extension of the grid-cell approach to unique forest segments and a deciduous-coniferous forest mix. Hierarchical, multiresolution segmentation results were used as homogenous units for the extraction of lidar distributions, while basal area plots were used as field data for model fitting and validation. No distinct differences were found for volume and biomass modeling attempts across increasing segment sizes (0.035 – 5.632 ha/segment, although adjusted R^2 values generally decreased and RMSE values generally increased with increasing segment size.

This lack of modeling differences across varying segment sizes was attributed to the hierarchical nature of the segmentation algorithm, which resulted in small homogenous segments that served as building blocks for larger segments. Within-segment homogeneity already was minimized at smaller average segment sizes, resulting in no definitive difference in modeling results as segment size increased through recombination of smaller segments. However, segment-based modeling efforts were distinctly better than those found for existing, operational forest stands in the study area. This was attributed to the larger within-stand height variation in the case of existing stands when compared to the variation found within homogenous segments.

Modeling results were very promising, even though coniferous and combined adjusted R^2 values for volume and biomass were lower than those found in other published studies. Lower coniferous R^2 values were attributed in part to a smaller range of volume and biomass observed values, as well to the inherent variability found in Virginia Piedmont forests. Adjusted R^2 values for deciduous segments were higher than those found for a comparable, plot-level study in the same area. This result indicated that a segment-level approach to deciduous volume and biomass modeling is a potential improvement over plot-based approaches. Given that volume- and biomass modeling were performed by using only height-related values, high R^2 values in the context of this study were unlikely. This was due to the nature of the modeled field data, which were based on diameter-at-breast-height (biomass) as well as height (volume).

RMSE values compared favorably with those found in other distributional modeling studies. Low RMSE values indicated that models could find applicability in an operational context, even when low R^2 values were considered. A comparison between estimates from modeling based on lidar and BAF plot data, and stand-alone BAF plot data for the 945 ha study area was promising, with differences smaller than 5%. This indicated that stand-alone plot and lidar model estimates were in fact relatively similar, again boding well for possible future application of models in an operational context.

Forward and Mallow's Cp selection proved successful in the reduction of independent variables from as many as 75 initial height distributional variables to the fewer than 10 used for final modeling. Further variable reduction through correlation analysis proved critical to the process

of reducing variables. Final model selection from all candidate models was based on Mallow's Cp, adjusted R^2, RMSE values, and model simplicity. All criteria proved useful and even necessary in order to select a single best option from many Mallow's Cp recombined variable models. Final variables spanned the whole spectrum of possibilities, from general mean and range height values, to more abstract coefficient of variation and standard deviation-type variables. Percentiles, both regular and canopy cover percentiles, also were well represented. The inclusion of reflectance variables was interesting since few studies (Means *et al*., 1999; Brandtberg *et al*., 2003) have included reflectance values as part of forest biophysical modeling. The wide range of variables indicated that sophisticated lidar scanners, that can record multiple returns and reflectance associated with each lidar hit, might well be necessary for effective modeling of variation in more complex forests.

Per-segment volume and biomass modeling has the potential of constituting part of a complete lidar-based inventory. Segmentation, volume and biomass modeling, and object-oriented classification could form a cohesive approach to forest inventory using remote sensing data, specifically lidar technology. Segmentation of lidar-derived data has the benefit of establishing homogenous objects for subsequent volume and biomass modeling, resulting in scalable units that could be conglomerated along with all associated per-segment estimates. A variable forest stand could thus be modeled at a more homogenous sub-stand level. Although this was not investigated, it could be that stand-level estimates will be more precise due to such a scalable, integrated approach. It seems likely that limited fieldwork will be required for any given region. Fieldwork might include limited segmentation verification, establishment of volume-lidar distribution regression equations, and collection of forest type information. Established distributional volume and lidar equations could be applied for future stands and derived segments, with periodic verification using either fixed or variable plots. Models likely would have to be calibrated or even re-developed for different regions, as it seems that results are geographically dependent (Means *et al*., 2000; Makela and Pekkarinen, 2001; Næsset, 2002; Pekkarinen, 2002).

Issues that potentially are critical to operational implementation include determination of the number of plots required for proper model fitting and the ideal segmentation size for model

development and application. Although average segment size did not affect modeling results for this study area and approach, lower within-segment variances at smaller segment sizes theoretically define a range of segment sizes that could be better suited to volume and biomass modeling. Differences among segment sizes were not evident due the hierarchical nature of the segmentation algorithm, but it is likely that segment size selection for modeling is dependent on stand-makeup, with more diverse stands requiring smaller segments with less within-segment variability, and *vice versa*.

Forest managers strive to obtain estimates of volume-by-type with high economical and statistical efficiency. Volume-by-segment estimations presented here potentially can be extended to object-oriented classification and subsequent volume-by-type assignment. Such an approach could constitute a stand-alone forest inventory based on remote sensing inputs, but the associated precision and cost are two issues that will need further consideration. Extension to net primary productivity modeling efforts at larger scales is likely, with remote sensing data well suited to scaling of measurements and results.

6.3 Conclusions: Object-oriented Deciduous-Coniferous Classification

Object-oriented classification results, from both point-height-based and CHM-based distributional discriminant classifications, were promising when one considers the type of input data, the variability in natural ecosystems based on basal area contributions for each class, and the classification approach used. Accuracies as high as 89.2% for the point-height-based discriminant classification and 79% for the CHM-based approach bode well for lidar-based object-oriented classification. Higher accuracies for the point-height-based approach were attributed to the increased variable type range, which included more structural (2^{nd} returns) and spectral (reflectance) variables than the CHM-based approach.

Although the more complex, the point-height-based approach resulted in an increase of 10.2% in overall accuracy, the CHM-based approach performed surprisingly well. Accuracies for the latter approach were acceptable when considering that classification was based on a single layer, canopy height model data source. However, the point-height-based discriminant classification

was shown to be significantly better than the CHM-based approach, and was considered the better choice, even when its increased complexity was taken into account. High accuracies for between group classifications, with only structural data as input, showed potential when one considers that the primary uses of lidar data are associated with forest biophysical characterization (Means et al., 2000; Lefsky et al., 2002; Naesset, 2004; Popescu *et al*., 2004). Extension of lidar data to forest type definition therefore presents an opportunity to base a complete forest inventory on one remote sensing data source.

There were no significant differences within classification approaches, indicating that average segment size did not matter. The lack of decreasing classification accuracies at larger segment sizes was attributed to the hierarchical nature of the segmentation algorithm. The algorithm was based on the minimization of within-segment variance at all segment levels, but since smaller segments constituted the building blocks for larger objects, variance minimization already was adequately addressed at lower levels. Stand-based classification also was not significantly different from segment-based approaches. This was due to the definition of operational forest stands on a per-species or type basis, resulting in an already existing forest type map.

Stepwise discriminant analysis, as part of discriminant classification, was confirmed (van Aardt and Wynne, 2001) as an effective method for the reduction of classification variables from more than seventy to fewer than ten. This procedure reduced the feature space to those variables that maximized between-group separation while minimizing the variability within each group. Selected variables included a broad range of distributional data types, while reflectance values (near-infrared) were present. The inclusion of the median reflectance for lidar first return vegetative hits highlighted the importance of per-object, lidar-associated reflectance data to lidar-based classification approaches. Inclusion of reflectance variables corroborated wavelengths used for more traditional hyperspectral classification approaches (Martin *et al.*, 1998; Fung *et al.*, 1999; van Aardt and Wynne, 2001). The standard deviation of second return vegetation hits and the first return ground hits as a ratio of total first returns were also well represented variables. Both these metrics were important indicators of vegetative cover.

Two-class forest type definition (deciduous-coniferous) resulted in better accuracies than a 3-class (deciduous-coniferous-mixed) approach for the study area. This was attributed to high basal

area percentages, in terms of deciduous-coniferous mixtures, that had to be used for mixed class definition. The mixed class eventually was closely associated with the deciduous group, thereby reducing deciduous producer's accuracies in the 3-class scheme. Lower basal area percentages for mixed-class definition resulted in an almost negligible number of mixed sample objects. Higher producer's accuracies for the coniferous class in the 3-class approach indicated that the deciduous-mixed classes were main contributors to lower overall accuracies with increased between-group confusion.

Although traditional approaches achieved adequate classification accuracies when multispectral data were used, a lidar-based object-oriented approach has multiple benefits. Lidar data have been used extensively for the modeling of forest biophysical parameters, and coupled with classification, could enable a forest inventory approach that is based on one data source. Due to their hierarchical nature, object-oriented approaches also have the benefit of recombination of object-level results to fit almost any required scale. Lastly, the simplicity of the approach is attractive, since only per-object height values are potentially required, while co-registration errors between optical and lidar data also are minimized. Extension of accurate classifications to lidar-derived forest volume or biomass-by-type is appealing, and could prove useful in forest inventories. The biggest hurdles to large-scale application are data cost and processing time, which likely will drop as more data providers emerge and technology improves. Irrespective of these caveats, analytical approaches continuously are evolving to a stage where multi-source remote sensing data might no longer be required for complete forest inventories.

6.4 Conclusions: Precursory Selection of Segmentation Results for Forest Inventory

Multiresolution, hierarchical segmentation was regarded as a precursor to subsequent per-segment analysis. Accurate segmentation of a forested area is needed before per-segment analyses, e.g., volume model development and object-oriented classification, can be performed. Homogenous segments (small within-segment variability) that are distinctly different from each other (large between-segment variance) are therefore required (Avery and Burkhart, 1994; Shivers and Borders, 1996). This necessitates not only segmentation as pre-processing activity, but also the evaluation and validation of segmentation results.

The use of the lidar-derived canopy height model (CHM) enabled segment delineation based on structural properties, important if subsequent applications include of volume and biomass model development, both of which have a structural basis. This can be contrasted with spectral segmentation, where specific wavelength information is the determinant of the segmentation outcome. Although the latter approach has merit when species specific information is required, a structural segmentation would likely follow not only structural boundaries, but also forest type definitions. This is due to forest stands (deciduous vs. coniferous) being unique in their structural properties (Douglas *et al.,* 2003), as well as the fact that most forest stands are delineated by stand breaks or forest roads. The multiresolution, hierarchical segmentation method used in this study (eCognition algorithm) performed well, both from a qualitative and quantitative perspective. It also was applicable to possible per-segment analyses due to the scalable nature of results, based on the hierarchical approach of the algorithm. This resulted in topologically sound segmentation results at all scales, with smaller segments constituting the building blocks for larger segments. In terms of visual correspondence the eCognition algorithm conformed to structural boundaries evident from the CHM image. Stand and road breaks were accurately detected, but managerial objects, e.g., fire breaks and power lines, could hinder homogenous object definition in cases where such a feature intersected a stand. Visual correspondence was consistent at all segmentation scales, from small to large segments. Although not definitive, visual correspondence to input data is a critical factor, and always should be considered (Schiewe *et al*., 2001; Schiewe, 2002).

Segmentation results also were valid from a statistical perspective. Between- and within-segment variance plots of the CHM (height) illustrated that within-segment variability increased as between-segment variability decreased with increasing average segment size. Smaller segments were thus potentially better suited to statistical manipulation like volume model fitting, resulting in models with lower RMSEs. This was due to the lower structural within-segment variation when smaller, homogeneous segments were compared to larger, more heterogeneous segments. The between-within variance intersect also provided a guideline as to the level of scalability of segmentation results. Coupled with an F-statistic for each segmentation outcome, this intersect was critical to determination of the eventual, operational segment size. These two metrics were

well suited to defining eventual operational stand sizes, and along with structurally-based segmentation, could help in redefining more homogenous forest stand better suited to management and inventory.

Deciding on the segment size to use for further application also has been a difficult question to address to date. Most authors used classification accuracy assessment (Antunes *et al*., 2003; Darwish *et al*., 2003; Heyman *et al*., 2003; Neubert, 2001) visual correspondence to input data along with number of segments as decision criteria for thematic output (Schiewe *et al*., 2001; Schiewe, 2002). Segment selection for volume- and biomass modeling mainly has been based on tree- (Hyyppä and Inkinen, 1999) or stand approximations (Woodcock *et al.,* 1994), or was circumvented by selecting plots within a segment (Makela and Pekkarinen, 2001; Pekkarinen 2002). The use of the average tree distance from BAF plot center plus one or two standard deviations resulted in a useful method for defining the minimum segment size and number at which statistical models could be developed. This radius equaled a plot size that approximated the segment size and number where between- and within-segment variance curves leveled off. This outcome resulted in a segment size where little of the overall variance could still be absorbed by the between-segment variance, i.e. resulting in homogenous segments that were different from each other in terms of height. A conceptual protocol was presented by which a user can evaluate segmentation results based on the presented criteria, namely within- and between-segment ratio and field plot size, while the significance of the F-statistic can be used as an indicator of upper scaling limit.

However, segment size selection based on the provided criteria was not corroborated by subsequent object-oriented volume modeling and forest type classification. There were no distinct reductions in model fit metrics and classification accuracies as average segment size increased. This was contrary to expectation, but was attributed to three main factors. While BAF plots were assumed to represent entire segments or stands, a full per-segment inventory could corroborate the expected decreasing fit metrics as within-segment variance decreases. Such a complete inventory will result in complete segment description. Validation results for this study also were based on a relatively small measured volume range, while other lidar distributional studies used a larger range of observed values. Extension of the developed methods for segment

selection to larger observed ranges therefore is warranted to conclusively determine the utility of the proposed selection criteria. Lastly, the hierarchical nature of the segmentation algorithm resulted in larger segments consisting of smaller, homogenous segments with low within-segment height variance. Minimization of within-segment variance therefore was already addressed at smaller segment levels, with larger segments consisting of these homogenous building blocks. Analyses were ultimately applied to segment levels with minimized within-segment variability across the entire range of average segment sizes.

However, segment-based volume modeling was shown to improve model fit metrics and RMSE values as opposed to modeling based on existing operational stands in the study area. This indicated that segmentation of the forested study area was indeed warranted. Whereas current stand delineation was based on site, forest types, and growth-and-yield measurements, segmentation resulted in a true structural definition of the study area, using lidar data. Resultant structurally homogenous segments were more conducive to an object-oriented volume modeling approach than defined stands, proving that the current definition of measurement and management units should be reconsidered.

In conclusion, the use of between- and within-segment variation, an F-statistic calculated on a per-segment outcome, and average BAF plot size, have potential as criteria for deciding which segment level should be useful in subsequent analyses. It should be noted, however, that visual segment correspondence to input data is critical. A human interpreter usually has a keen eye for delineating unique image areas. Due to the lack of objectivity, repeatability, and speed, segment delineation based on human interpretation is not an option. However, human interpretation of results is useful to verify segmentation output. This study focused on the selection of the best individual segmentation outcome, within a set of segmentation results. The conceptual decision model that was presented has potential, but as with any such approach, the user should practice caution due to its “fuzzy” nature. In many cases only a range of average segment sizes will be defined, while selection and verification could require field visits. It is recommended that future work should focus on validation of selection criteria for increased dependent variable ranges, as well as development of other possible decision rules.

6.5 Final Considerations

Object-oriented volume and aboveground biomass modeling and forest type classification, based on small-footprint lidar-distributional data, have been shown to have significant potential for operational extension. Lidar-based segmentation was used to derive segmentation outcomes for these analyses, but further research is required for the selection of average segment size most appropriate for per-segment analyses. Although a conceptual decision model was presented which facilitated segment selection, validation of segment choices did not corroborate the method used. This was attributed to the hierarchical nature of the segmentation algorithm, which by definition led to larger segments being defined by smaller segments. Within-segment variance was therefore already minimized at the smallest segment size, before combination into larger segments took place. It is recommended that future research focus on investigating whether or not the statistically-based selection criteria are effective for in the case of a non-hierarchical segmentation approach and across a wider range of the dependent variable. Segment-based modeling was found to be superior to modeling based on current stand delineations for the study area, but the application of segmentation for modeling and classification analyses needs to be investigated for alternative geographical areas, with varying conditions.

6.6 Literature Cited

Antunes, A.F., C. Lingnau, and J.C. Da Silva, 2003. Object oriented analysis and semantic network for high resolution image classification. In: Anais XI SBSR, Belo Horizonte, Brazil, April 5-10, 2003, INPE. pp. 273-279.

Avery, T.E., and H.E. Burkhart, 1994. Forest Measurements. 4th Edition. McGraw-Hill, Boston, USA. 408 p.

Darwish, A., K. Leukert, and W. Reinhardt, 2003. Image segmentation for the purpose of object-based classification. Proceedings of IGARSS 2003 IEEE, July 2003, Toulouse 3 pp.

Douglas, T.E., D.L. Evans, K.L. Belli, and S.D. Roberts, 2003. Classification of pine and hardwood by the distribution and intensity of lidar returns. ISPRS "Three dimensional mapping workshop from InSAR and LIDAR", June 17-19, 2003, Portland, Oregon, USA. 5pp.

Fung, T., F.Y. Ma, and W.L. Sui, 1999. Hyperspectral data analysis for subtropical tree species identification. Proceedings: 1999 ASPRS Annual Conference, Portland Oregon.

Heyman, O., G.G. Gaston, A.J. Kimerling, and J.T. Campbell, 2003. A per-segment approach to improving Aspen mapping from high-resolution remote sensing imagery. *Journal of Forestry* (June, 2003): 29-33.

Hyyppä J., and M. Inkinen, 1999. Detecting and estimating attributes for single trees using laser scanner. *Photogrammetric Journal of Finland* 16: 27-42.

Makela H., and A. Pekkarinen, 2001. Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing of Environment* 77:66-75.

Martin, M.E., S.D. Newman, J.D. Aber, and R.G. Congalton, 1998. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment* 65: 249 – 254.

Means, J.E., S.A. Acker, D.J. Harding, J.B. Blair, M.A. Lefsky, W.B. Cohen, M.E. Harmon, and 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.

Means J.E., S.A. Acker, B.J. Fitt, M. Renslow, L. Emerson, and C.J. Hendrix, 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering & Remote Sensing* 66 (11): 1367-1371.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80 (2002):88-99.

Neubert, M., 2001. Segment-based analysis of high resolution satellite and laser scanning data. In: Hilty, L. M. & Gilgen, P. W. (Eds.): Sustainability in the Information Society. Proceedings of the 15th International Symposium Informatics for Environmental Protection, Zurich, October 10-12, 2001, Marburg. pp. 379-386.

Pekkarinen A., 2002. Image segment-based spectral features in the estimation of timber volume. *Remote Sensing of Environment* 82: 349-359.

Schiewe, J., L. Tufte, and M. Ehlers, 2001. Potential and problems of multi-scale segmentation methods in remote sensing. *GeoBIT/GIS* 6: 34-39.

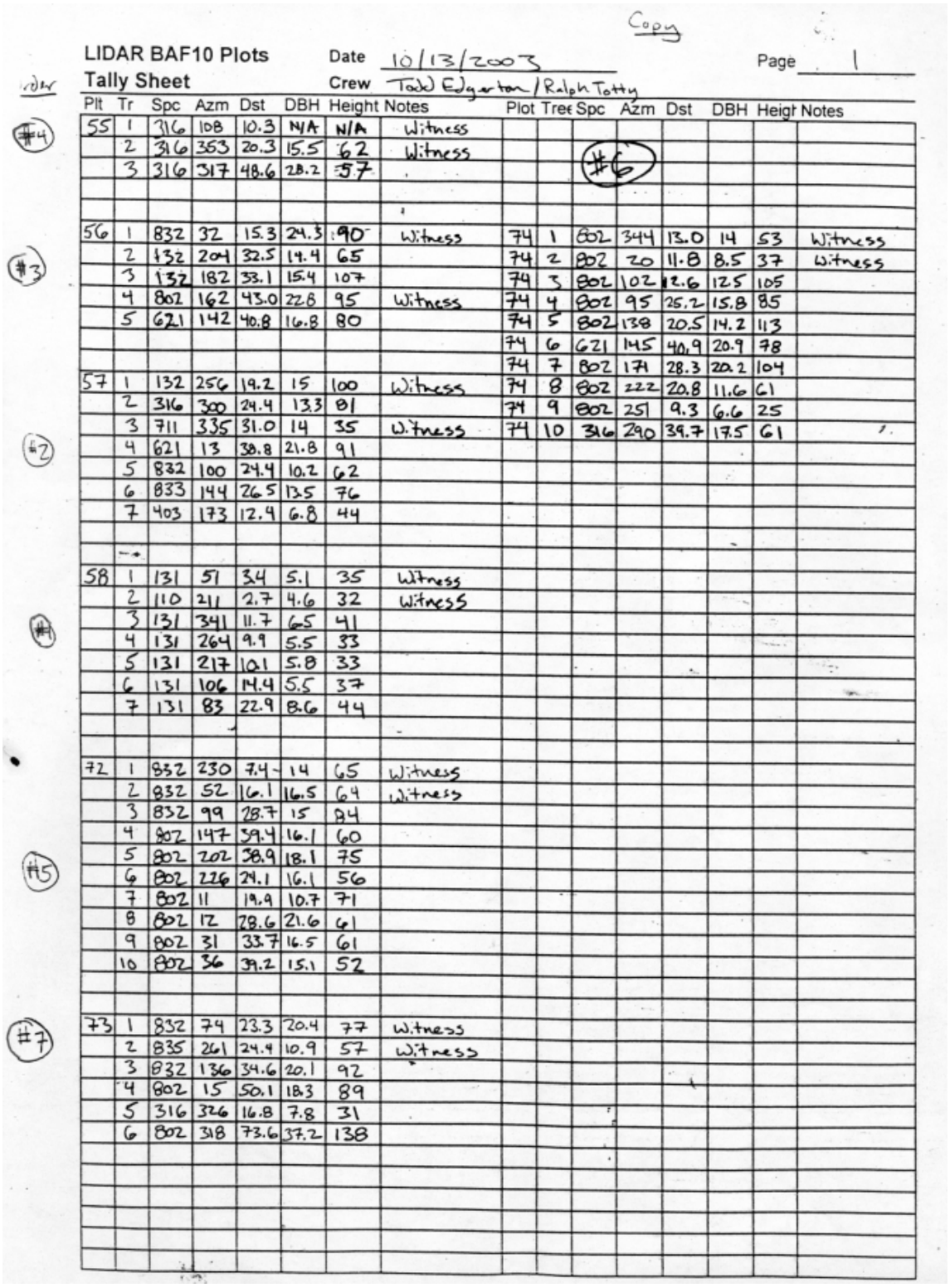
Schiewe, J., 2002. Segmentation of high-resolution remotely sensed data - concepts, applications and problems. Proceedings of the Joint International Symposium on Geospatial Theory, Processing, and Applications. Ottawa, Canada, July 9-12, 2002. 6 pp.

Shivers, B.D., and B.E. Borders, 1996. Sampling techniques for forest resource inventory. John Wiley and Sons, Inc. New York, USA. 356 p.

van Aardt, J.A.N., and R.H. Wynne, 2001. Spectral separability among six southern tree species. *Photogrammetric Engineering & Remote Sensing* 67 (12): 1367-1375.

Woodcock, C.E., J.B. Collins, S. Gopal, V.D. Jakabhazy, X. Li, S. Macomber, S. Ryherd, V.J. Harward, J. Levitan, Y. Wu, and R. Warbington, 1994. Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment* 50: 240-254.

Appendix A
Example Data Sheet



LIDAR BAF10 Plots — Date 10/13/2003 — Page 1
Tally Sheet — Crew Todd Edgerton / Ralph Totty
Copy

Plt	Tr	Spc	Azm	Dst	DBH	Height	Notes
55	1	316	108	10.3	N/A	N/A	Witness
	2	316	353	20.3	15.5	62	Witness
	3	316	317	48.6	28.2	57	
56	1	832	32	15.3	24.3	90	Witness
	2	132	204	32.5	14.4	65	
	3	132	182	33.1	15.4	107	
	4	802	162	43.0	22.8	95	Witness
	5	621	142	40.8	16.8	80	
57	1	132	256	19.2	15	100	Witness
	2	316	300	24.4	13.3	81	
	3	711	335	31.0	14	35	Witness
	4	621	13	38.8	21.8	91	
	5	832	100	24.4	10.2	62	
	6	833	144	26.5	13.5	76	
	7	403	173	12.4	6.8	44	
58	1	131	51	3.4	5.1	35	Witness
	2	110	211	2.7	4.6	32	Witness
	3	131	341	11.7	6.5	41	
	4	131	264	9.9	5.5	33	
	5	131	217	10.1	5.8	33	
	6	131	106	14.4	5.5	37	
	7	131	83	22.9	8.6	44	
72	1	832	230	7.4	14	65	Witness
	2	832	52	16.1	16.5	64	Witness
	3	832	99	28.7	15	84	
	4	802	147	39.4	16.1	60	
	5	802	202	38.9	18.1	75	
	6	802	226	24.1	16.1	56	
	7	802	11	19.9	10.7	71	
	8	802	12	28.6	21.6	61	
	9	802	31	33.7	16.5	61	
	10	802	36	39.2	15.1	52	
73	1	832	74	23.3	20.4	77	Witness
	2	835	261	24.4	10.9	57	Witness
	3	832	136	34.6	20.1	92	
	4	802	15	50.1	18.3	89	
	5	316	326	16.8	7.8	31	
	6	802	318	73.6	37.2	138	

Plot	Tree	Spc	Azm	Dst	DBH	Height	Notes
74	1	802	344	13.0	14	53	Witness
74	2	802	20	11.8	8.5	37	Witness
74	3	802	102	12.6	12.5	105	
74	4	802	95	25.2	15.8	85	
74	5	802	138	20.5	14.2	113	
74	6	621	145	40.9	20.9	78	
74	7	802	171	28.3	20.2	104	
74	8	802	222	20.8	11.6	61	
74	9	802	251	9.3	6.6	25	
74	10	316	290	39.7	17.5	61	

Appendix B
Common Species Codes for Field Data Collection

Common name	Botanical name	Species code
Eastern red cedar	*(Juniperus virginiana)*	067
Shortleaf pine	*(Pinus echinata)*	110
Eastern white pine	*(Pinus strobus)*	129
Loblolly pine	*(Pinus taeda)*	131
Virginia pine	*(Pinus virginiana)*	132
Red maple	*(Acer rubrum)*	316
Hickory	*(Carya spp.)*	403
Dogwood	*(Cornus florida)*	491
Yellow poplar	*(Liriodendron tulipifera)*	621
Black gum	*(Nyssa sylvatica)*	693
Sourwood	*(Oxydendrum arboretum)*	711
Black cherry	*(Prunus serotina)*	762
White oak	*(Quercus alba)*	802
Scarlet oak	*(Quercus coccinea)*	806
Southern red oak	*(Quercus falcata)*	812
Chestnut oak	*(Quercus prinus)*	832
Northern red oak	*(Quercus rubra)*	833
Post oak	*(Quercus stellata)*	835
Black oak	*(Quercus velutina)*	837
Black locust	*(Robinia pseudoacacia)*	901

Appendix C
Basal Area Plot Values for Volume-Per-Hectare, Biomass-Per-Hectare, Basal Area-Per-Hectare, and Type

Table C.1 BAF plot locations and all associated values used for volume and biomass modeling (*Type 1 = Deciduous; 2 = Coniferous; 3 = Mixed*). Coordinates are in the UTM projection, zone 17N (Datum: NAD83; Spheroid: GRS1980)

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
1	704253.63	4145346.19	205.90	143.90	22.96	-	0.65	0.65	-	100.00	1	1
2	704451.17	4145368.68	146.96	54.49	25.25	0.25	-	0.25	100.00	-	2	2
3	704645.77	4145381.29	191.23	137.50	18.37	-	1.43	1.43	-	100.00	1	1
4	704840.42	4145342.51	118.61	87.89	11.48	-	0.88	0.88	-	100.00	1	1
5	705051.89	4145345.23	67.51	19.47	9.18	0.08	-	0.08	100.00	-	2	2
6	705252.10	4145353.40	52.13	50.76	6.89	-	0.50	0.50	-	100.00	1	1
7	705442.26	4145356.69	123.68	51.25	16.07	0.17	1.16	1.33	13.00	87.00	1	3
8	705640.64	4145345.95	200.25	88.32	20.66	0.24	0.43	0.67	35.83	64.17	1	3
9	705850.53	4145343.89	140.87	82.54	16.07	0.21	0.26	0.48	44.39	55.61	1	3
10	706056.98	4145350.35	167.83	85.78	18.37	0.23	0.17	0.40	56.52	43.48	2	3
11	706252.65	4145344.88	198.61	163.71	22.96	-	0.99	0.99	-	100.00	1	1
12	706456.89	4145349.83	178.45	143.20	20.66	-	0.90	0.90	-	100.00	1	1
13	706640.34	4145366.50	307.81	204.14	27.55	-	1.45	1.45	-	100.00	1	1
14	706853.88	4145343.92	225.26	61.05	22.96	0.56	-	0.56	100.00	-	2	2
15	707046.81	4145350.56	105.38	70.20	11.48	-	0.22	0.22	-	100.00	1	1
16	707251.34	4145357.15	69.48	19.08	6.89	0.19	-	0.19	100.00	-	2	2
17	704241.20	4145136.34	85.11	42.31	9.18	0.21	0.05	0.25	81.13	18.87	2	3
18	704449.94	4145149.00	157.97	89.11	18.37	0.17	0.25	0.42	40.21	59.79	1	3
19	704653.87	4145145.25	168.14	122.23	20.66	-	0.37	0.37	-	100.00	1	1
20	704851.24	4145149.34	81.75	78.96	11.48	-	0.65	0.65	-	100.00	1	1
21	705052.52	4145149.28	97.17	35.70	11.48	0.14	0.04	0.18	79.46	20.54	2	3
22	705259.24	4145143.58	56.98	25.61	6.89	0.07	0.05	0.12	58.41	41.59	2	3
23	705443.26	4145149.77	169.07	51.98	18.37	0.60	-	0.60	100.00	-	2	2
24	705643.79	4145149.48	109.41	37.94	16.07	0.26	-	0.26	100.00	-	2	2

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
25	705846.64	4145147.21	350.93	155.56	36.73	0.68	0.51	1.19	56.87	43.13	2	3
26	706049.91	4145149.45	121.82	42.61	13.77	0.36	0.02	0.38	95.94	4.06	2	2
28	706447.40	4145151.15	180.60	151.19	20.66	-	0.91	0.91	-	100.00	1	1
29	706660.92	4145154.02	245.84	195.09	27.55	-	1.15	1.15	-	100.00	1	1
30	706854.47	4145144.39	45.38	43.06	6.89	-	0.16	0.16	-	100.00	1	1
31	707051.90	4145153.06	70.25	41.63	11.48	0.09	0.06	0.15	61.71	38.29	2	3
32	707251.27	4145155.50	31.68	20.06	4.59	0.05	0.04	0.09	55.25	44.75	2	3
33	704257.25	4144947.10	264.46	81.21	36.73	0.40	-	0.40	100.00	-	2	2
34	704457.36	4144952.31	39.97	14.57	6.89	0.06	-	0.06	100.00	-	2	2
35	704660.46	4144942.36	129.71	114.34	13.77	-	1.05	1.05	-	100.00	1	1
36	704857.52	4144947.06	116.13	93.70	11.48	-	0.93	0.93	-	100.00	1	1
37	705057.62	4144946.14	57.87	61.81	9.18	-	0.27	0.27	-	100.00	1	1
38	705250.51	4144947.55	100.84	34.85	13.77	0.28	-	0.28	100.00	-	2	2
39	705438.33	4144944.90	197.16	95.43	20.66	0.24	0.39	0.63	38.86	61.14	1	3
40	705649.48	4144945.78	165.40	54.57	22.96	0.34	-	0.34	100.00	-	2	2
41	705860.86	4144954.92	68.06	63.73	9.18	-	0.42	0.42	-	100.00	1	1
42	706051.41	4144950.53	84.81	31.14	13.77	0.16	-	0.16	100.00	-	2	2
43	706248.58	4144940.53	42.76	14.65	6.89	0.06	-	0.06	100.00	-	2	2
44	706458.36	4144933.63	71.39	24.52	11.48	0.11	-	0.11	100.00	-	2	2
45	706661.17	4144953.53	21.29	9.77	4.59	0.04	-	0.04	100.00	-	2	2
46	706850.01	4144952.36	276.98	138.92	20.66	0.45	1.70	2.14	20.84	79.16	1	3
47	707051.84	4144940.29	93.31	31.87	13.77	0.20	-	0.20	100.00	-	2	2
48	707244.15	4144938.89	278.99	77.25	32.14	0.49	-	0.49	100.00	-	2	2
49	704251.00	4144749.41	159.50	55.84	25.25	0.28	-	0.28	100.00	-	2	2
50	704450.75	4144747.92	230.57	86.37	22.96	0.85	0.22	1.06	79.74	20.26	2	3
51	704649.09	4144740.95	112.77	32.73	11.48	0.37	-	0.37	100.00	-	2	2
52	704859.86	4144746.07	148.90	122.05	18.37	-	0.58	0.58	-	100.00	1	1
53	705056.74	4144739.45	212.01	152.12	20.66	-	1.00	1.00	-	100.00	1	1
54	705260.29	4144752.70	95.73	30.00	13.77	0.14	-	0.14	100.00	-	2	2
55	705441.89	4144755.45	26.48	39.19	4.59	-	0.52	0.52	-	100.00	1	1

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
56	705649.27	4144743.58	131.62	74.07	11.48	0.23	0.71	0.93	24.20	75.80	1	3
57	705839.54	4144745.57	144.49	105.37	16.07	0.11	0.60	0.71	16.01	83.99	1	3
58	706048.34	4144752.38	75.37	28.72	13.77	0.12	-	0.12	100.00	-	2	2
59	706251.10	4144742.64	8.32	4.70	2.30	0.02	-	0.02	100.00	-	2	2
60	706449.86	4144748.66	26.92	14.79	6.89	0.07	-	0.07	100.00	-	2	2
61	706650.81	4144752.05	71.45	26.08	11.48	0.14	-	0.14	100.00	-	2	2
62	706850.06	4144765.13	107.67	37.08	16.07	0.21	-	0.21	100.00	-	2	2
63	707041.88	4144738.16	151.63	91.39	18.37	0.08	0.30	0.38	21.77	78.23	1	3
64	707249.20	4144748.80	149.08	99.82	13.77	-	0.55	0.55	-	100.00	1	1
66	704460.97	4144542.76	123.56	44.41	13.77	0.36	0.02	0.38	93.91	6.09	2	2
67	704651.89	4144558.28	145.04	92.42	18.37	0.08	0.22	0.30	27.65	72.35	1	3
68	704856.00	4144556.89	195.58	144.91	20.66	-	0.74	0.74	-	100.00	1	1
69	705056.63	4144552.99	185.19	144.62	18.37	-	1.11	1.11	-	100.00	1	1
70	705249.63	4144547.51	124.59	89.69	11.48	-	0.69	0.69	-	100.00	1	1
71	705459.64	4144562.06	288.14	208.16	27.55	-	1.58	1.58	-	100.00	1	1
72	705660.75	4144544.06	191.78	179.73	22.96	-	1.33	1.33	-	100.00	1	1
73	705855.78	4144548.94	140.53	108.24	13.77	-	1.38	1.38	-	100.00	1	1
74	706044.59	4144551.87	207.66	168.64	22.96	-	1.12	1.12	-	100.00	1	1
75	706251.32	4144541.85	14.38	10.25	4.59	0.05	-	0.05	100.00	-	2	2
76	706448.55	4144546.89	151.20	114.54	16.07	0.22	0.92	1.13	18.97	81.03	1	3
77	706634.84	4144556.31	136.56	54.97	25.25	0.26	-	0.26	100.00	-	2	2
78	706827.94	4144552.65	89.34	39.65	16.07	0.36	-	0.36	100.00	-	2	2
79	707043.72	4144550.56	22.15	18.28	2.30	-	0.14	0.14	-	100.00	1	1
80	707256.90	4144539.54	108.92	116.04	16.07	-	1.04	1.04	-	100.00	1	1
82	704451.06	4144344.34	202.26	122.06	22.96	0.31	0.50	0.81	38.14	61.86	1	3
83	704643.77	4144347.00	237.79	163.92	20.66	-	1.73	1.73	-	100.00	1	1
84	704851.80	4144346.84	155.08	111.68	16.07	0.06	0.80	0.86	7.24	92.76	1	1
85	705038.76	4144333.70	183.14	65.07	18.37	0.42	0.07	0.48	85.85	14.15	2	3
86	705248.33	4144347.62	52.80	43.10	6.89	-	0.17	0.17	-	100.00	1	1
87	705448.21	4144347.40	82.86	30.71	9.18	0.24	0.01	0.25	94.15	5.85	2	2

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
88	705653.64	4144346.30	170.54	70.25	27.55	0.33	0.02	0.35	92.94	7.06	2	2
89	705850.21	4144352.62	188.56	165.41	22.96	0.02	1.32	1.34	1.14	98.86	1	1
90	706060.79	4144347.91	181.72	184.59	22.96	-	1.73	1.73	-	100.00	1	1
91	706243.11	4144349.91	79.47	71.73	9.18	-	0.85	0.85	-	100.00	1	1
92	706447.81	4144346.07	22.64	17.97	2.30	-	0.12	0.12	-	100.00	1	1
93	706635.82	4144357.76	37.68	46.12	6.89	-	0.28	0.28	-	100.00	1	1
94	706846.40	4144343.44	256.65	174.57	22.96	-	1.43	1.43	-	100.00	1	1
95	707045.06	4144348.93	19.00	14.68	2.30	-	0.05	0.05	-	100.00	1	1
96	707252.32	4144346.16	94.50	71.67	9.18	-	0.61	0.61	-	100.00	1	1
97	704246.75	4144149.23	171.72	82.03	16.07	0.50	0.28	0.78	64.51	35.49	2	3
98	704454.47	4144148.08	153.02	87.44	16.07	0.33	0.49	0.82	39.98	60.02	1	3
99	704656.00	4144136.22	102.12	76.68	11.48	-	0.33	0.33	-	100.00	1	1
101	705044.32	4144143.46	154.88	71.61	13.77	0.30	0.29	0.59	51.22	48.78	2	3
102	705251.08	4144139.74	146.86	52.58	16.07	0.38	0.05	0.43	87.39	12.61	2	3
103	705452.89	4144152.18	176.58	119.45	20.66	0.11	0.61	0.73	15.63	84.37	1	3
105	705853.05	4144150.86	83.12	62.17	9.18	-	0.30	0.30	-	100.00	1	1
106	706049.30	4144142.03	202.65	173.65	22.96	-	1.28	1.28	-	100.00	1	1
107	706246.09	4144148.78	273.82	198.70	25.25	-	2.48	2.48	-	100.00	1	1
108	706462.60	4144138.47	322.01	199.46	25.25	-	1.90	1.90	-	100.00	1	1
109	706653.19	4144139.49	168.43	107.61	13.77	-	0.80	0.80	-	100.00	1	1
110	706849.96	4144147.53	69.52	35.45	6.89	0.25	0.32	0.57	44.19	55.81	1	3
111	707052.75	4144141.47	31.85	14.62	6.89	0.06	-	0.06	100.00	-	2	2
112	707253.39	4144150.60	262.07	160.42	18.37	-	2.43	2.43	-	100.00	1	1
113	704250.65	4143947.71	92.74	61.72	9.18	0.06	0.46	0.52	11.76	88.24	1	3
114	704452.58	4143956.27	50.36	34.76	6.89	0.07	0.09	0.16	42.28	57.72	1	3
115	704649.68	4143981.27	188.36	143.77	25.25	0.06	0.44	0.50	11.42	88.58	1	3
116	704851.06	4143949.03	99.84	77.11	11.48	-	0.32	0.32	-	100.00	1	1
117	705053.27	4143945.14	167.28	98.06	16.07	0.17	0.58	0.75	23.13	76.87	1	3
118	705243.81	4143949.20	112.77	74.36	11.48	0.17	0.42	0.59	28.63	71.37	1	3
119	705457.05	4143939.81	168.77	136.19	18.37	-	1.19	1.19	-	100.00	1	1

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
120	705657.92	4143946.13	175.84	121.65	16.07	-	0.86	0.86	-	100.00	1	1
121	705844.56	4143953.98	279.89	185.47	27.55	0.06	1.01	1.06	5.35	94.65	1	1
122	706053.00	4143952.16	178.83	129.05	16.07	-	1.30	1.30	-	100.00	1	1
123	706249.63	4143944.16	235.97	114.43	20.66	0.42	0.48	0.90	46.52	53.48	1	3
124	706458.94	4143966.68	242.23	162.60	20.66	-	1.22	1.22	-	100.00	1	1
125	706631.52	4143942.83	300.45	238.68	32.14	-	1.74	1.74	-	100.00	1	1
126	706850.91	4143949.65	38.44	18.47	9.18	0.07	-	0.07	100.00	-	2	2
127	707054.09	4143947.78	19.96	17.54	2.30	-	0.11	0.11	-	100.00	1	1
128	707250.85	4143956.36	31.35	14.19	6.89	0.05	-	0.05	100.00	-	2	2
129	704250.27	4143742.92	146.41	126.91	16.07	-	0.96	0.96	-	100.00	1	1
131	704646.45	4143746.58	155.42	50.16	16.07	0.33	0.04	0.37	88.80	11.20	2	3
132	704856.49	4143746.48	91.84	27.14	9.18	0.36	-	0.36	100.00	-	2	2
133	705048.00	4143740.72	128.86	33.35	11.48	0.40	-	0.40	100.00	-	2	2
134	705243.37	4143757.71	75.12	23.90	9.18	0.21	-	0.21	100.00	-	2	2
135	705446.62	4143742.07	155.86	115.28	16.07	-	0.61	0.61	-	100.00	1	1
136	705664.00	4143744.51	218.68	159.44	20.66	-	1.24	1.24	-	100.00	1	1
137	705852.03	4143753.52	249.32	143.86	27.55	0.20	0.71	0.90	22.05	77.95	1	3
138	706057.14	4143748.53	108.34	86.44	11.48	-	0.63	0.63	-	100.00	1	1
139	706259.17	4143740.17	212.54	177.11	25.25	-	1.03	1.03	-	100.00	1	1
140	706450.40	4143751.39	256.02	168.47	25.25	0.08	1.06	1.14	6.92	93.08	1	1
141	706636.25	4143738.88	179.99	131.96	16.07	-	1.33	1.33	-	100.00	1	1
142	706850.06	4143698.70	48.63	39.14	4.59	-	0.40	0.40	-	100.00	1	1
144	707252.59	4143747.08	146.29	92.29	11.48	-	0.76	0.76	-	100.00	1	1
145	704239.49	4143546.39	195.93	139.69	18.37	-	1.32	1.32	-	100.00	1	1
146	704451.13	4143543.87	14.88	11.76	2.30	-	0.02	0.02	-	100.00	1	1
147	704654.54	4143548.91	156.59	133.73	18.37	-	0.80	0.80	-	100.00	1	1
148	704838.34	4143544.40	269.45	173.15	20.66	-	1.68	1.68	-	100.00	1	1
149	705050.58	4143546.27	80.86	22.31	6.89	0.39	-	0.39	100.00	-	2	2
150	705254.14	4143556.53	210.50	94.34	22.96	0.43	0.22	0.65	65.89	34.11	2	3
151	705448.63	4143546.72	116.06	102.22	13.77	-	0.86	0.86	-	100.00	1	1

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
152	705652.70	4143549.32	50.61	19.68	9.18	0.09	-	0.09	100.00	-	2	2
153	705852.85	4143551.31	67.29	62.07	9.18	-	0.35	0.35	-	100.00	1	1
154	706048.57	4143540.64	144.64	97.68	13.77	0.08	0.72	0.80	9.74	90.26	1	1
155	706244.91	4143542.35	124.33	104.35	13.77	-	0.67	0.67	-	100.00	1	1
156	706451.98	4143552.17	214.34	171.20	22.96	-	1.14	1.14	-	100.00	1	1
157	706649.15	4143547.67	145.09	84.40	16.07	0.28	0.37	0.64	42.83	57.17	1	3
158	706847.53	4143544.33	195.50	146.67	20.66	-	0.83	0.83	-	100.00	1	1
159	707042.86	4143551.38	108.43	72.16	9.18	-	0.56	0.56	-	100.00	1	1
160	707245.05	4143556.62	304.96	203.07	25.25	-	1.93	1.93	-	100.00	1	1
161	704253.02	4143343.01	145.45	120.20	16.07	-	0.97	0.97	-	100.00	1	1
163	704631.26	4143341.77	345.03	223.51	29.84	-	1.64	1.64	-	100.00	1	1
164	704839.35	4143353.77	207.97	132.59	18.37	-	0.96	0.96	-	100.00	1	1
165	705040.56	4143347.01	168.61	52.82	16.07	0.50	0.03	0.53	95.01	4.99	2	2
166	705253.16	4143348.66	264.84	80.90	25.25	0.69	0.09	0.77	88.77	11.23	2	3
167	705444.87	4143344.61	77.00	38.17	9.18	0.18	0.05	0.23	79.02	20.98	2	3
168	705638.89	4143356.64	160.90	133.86	16.07	-	1.29	1.29	-	100.00	1	1
169	705837.37	4143335.61	15.31	12.20	2.30	-	0.02	0.02	-	100.00	1	1
170	706053.03	4143348.55	277.67	175.75	27.55	0.14	0.89	1.03	13.23	86.77	1	3
171	706259.98	4143350.63	173.02	67.57	20.66	0.42	0.25	0.66	62.87	37.13	2	3
172	706447.09	4143352.00	173.92	138.56	18.37	-	1.07	1.07	-	100.00	1	1
173	706654.43	4143350.21	46.24	11.98	4.59	0.10	-	0.10	100.00	-	2	2
174	706841.90	4143352.45	188.49	146.99	22.96	0.18	0.93	1.11	16.45	83.55	1	3
175	707050.40	4143350.92	24.83	24.85	4.59	-	0.05	0.05	-	100.00	1	1
176	707253.68	4143345.71	195.21	66.75	20.66	0.40	0.04	0.44	90.17	9.83	2	2
177	704250.28	4143150.64	350.65	233.32	29.84	-	2.03	2.03	-	100.00	1	1
179	704646.22	4143145.92	182.44	46.93	16.07	0.58	-	0.58	100.00	-	2	2
180	704857.41	4143148.50	163.99	126.53	18.37	-	0.71	0.71	-	100.00	1	1
181	705047.36	4143148.79	79.28	57.38	6.89	-	0.52	0.52	-	100.00	1	1
182	705254.16	4143148.33	113.35	42.64	11.48	0.51	0.03	0.54	94.03	5.97	2	2
183	705452.44	4143149.58	45.36	38.05	6.89	-	0.08	0.08	-	100.00	1	1

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
184	705651.06	4143147.24	6.94	11.30	2.30	-	0.02	0.02	-	100.00	1	1
185	705851.27	4143134.77	82.87	56.46	9.18	0.17	0.36	0.53	32.30	67.70	1	3
186	706049.33	4143145.84	95.41	31.51	11.48	0.34	-	0.34	100.00	-	2	2
187	706250.12	4143144.06	146.59	97.15	16.07	-	0.30	0.30	-	100.00	1	1
188	706447.61	4143134.05	192.66	136.92	20.66	-	0.72	0.72	-	100.00	1	1
189	706653.95	4143137.81	87.55	35.21	9.18	0.24	0.08	0.32	75.86	24.14	2	3
190	706851.85	4143147.36	72.34	56.59	9.18	-	0.17	0.17	-	100.00	1	1
191	707050.96	4143147.61	12.38	5.13	2.30	0.03	-	0.03	100.00	-	2	2
192	707252.59	4143147.51	154.76	45.74	18.37	0.34	-	0.34	100.00	-	2	2
193	704213.58	4142940.03	289.97	175.18	22.96	-	1.30	1.30	-	100.00	1	1
195	704643.81	4142957.04	47.25	39.32	6.89	-	0.09	0.09	-	100.00	1	1
196	704851.61	4142951.54	206.54	144.98	18.37	-	1.11	1.11	-	100.00	1	1
199	705452.85	4142953.78	86.25	36.40	18.37	0.12	-	0.12	100.00	-	2	2
200	705647.35	4142966.16	87.15	25.80	11.48	0.14	-	0.14	100.00	-	2	2
201	705851.84	4142949.70	11.60	4.70	2.30	0.02	-	0.02	100.00	-	2	2
202	706049.29	4142944.25	10.16	4.67	2.30	0.02	-	0.02	100.00	-	2	2
203	706262.87	4142955.15	53.15	50.44	9.18	-	0.14	0.14	-	100.00	1	1
204	706458.99	4142949.91	222.25	174.67	25.25	-	1.05	1.05	-	100.00	1	1
205	706650.61	4142937.57	84.99	33.16	16.07	0.13	-	0.13	100.00	-	2	2
206	706860.10	4142950.04	79.71	23.69	4.59	0.02	1.31	1.33	1.28	98.72	1	1
207	707052.47	4142943.32	138.91	44.31	18.37	0.30	-	0.30	100.00	-	2	2
208	707237.12	4142947.87	63.43	32.77	11.48	0.09	0.03	0.12	72.40	27.60	2	3
209	704257.12	4142745.83	293.45	228.59	29.84	-	1.71	1.71	-	100.00	1	1
210	704447.70	4142746.59	15.47	12.63	2.30	-	0.03	0.03	-	100.00	1	1
211	704653.51	4142741.16	303.63	196.20	29.84	-	0.91	0.91	-	100.00	1	1
213	705048.29	4142748.63	267.41	222.68	32.14	-	1.12	1.12	-	100.00	1	1
214	705256.66	4142746.89	330.60	239.42	32.14	0.05	1.95	2.00	2.74	97.26	1	1
215	705449.80	4142748.72	235.83	184.17	22.96	-	1.61	1.61	-	100.00	1	1
216	705647.89	4142746.94	83.11	30.57	13.77	0.15	-	0.15	100.00	-	2	2
217	705846.16	4142741.19	335.01	99.74	25.25	1.20	0.16	1.36	88.31	11.69	2	3

Plot #	X coordinate	Y coordinate	Volume/ha (m^3/ha)	Biomass/ha (Mg/ha)	Basal area (m^2/ha)	Actual coniferous basal area (m^2/ha)	Actual deciduous basal area (m^2/ha)	Total basal area (m^2/ha)	Coniferous %	Deciduous %	Majority type	90% Majority type
218	706050.90	4142750.54	104.92	79.55	13.77	0.07	0.28	0.36	20.54	79.46	1	3
219	706246.46	4142746.59	261.72	163.00	29.84	0.11	0.48	0.59	18.74	81.26	1	3
220	706463.35	4142758.78	128.74	91.17	11.48	-	0.71	0.71	-	100.00	1	1
223	707050.14	4142740.84	89.26	35.72	11.48	0.17	0.03	0.19	85.76	14.24	2	3
224	707240.98	4142755.00	267.90	81.65	34.44	0.52	-	0.52	100.00	-	2	2
225	704250.36	4142544.84	10.23	5.09	2.30	0.02	-	0.02	100.00	-	2	2
226	704447.51	4142561.07	307.18	269.01	34.44	-	2.30	2.30	-	100.00	1	1
227	704645.89	4142541.67	107.09	101.66	13.77	-	0.65	0.65	-	100.00	1	1
228	704854.35	4142546.57	68.76	32.61	11.48	0.11	0.02	0.13	83.14	16.86	2	3
233	705854.84	4142563.59	10.45	12.88	2.30	-	0.03	0.03	-	100.00	1	1
234	706048.38	4142523.19	86.37	72.94	9.18	-	0.56	0.56	-	100.00	1	1
236	706444.63	4142542.82	147.52	59.81	13.77	0.48	0.21	0.69	69.87	30.13	2	3
237	706672.34	4142549.20	9.15	11.11	2.30	-	0.02	0.02	-	100.00	1	1
239	707048.46	4142560.55	104.53	75.94	11.48	-	0.45	0.45	-	100.00	1	1
240	707258.04	4142545.88	215.72	125.28	16.07	-	1.12	1.12	-	100.00	1	1

Appendix D
Individual Tree Volume and Biomass Equations for Loblolly, Shortleaf, and Virginia Pine, and Hardwood Species

Volume equations

Loblolly pine single tree volume: $V_t = b_0 D^{b_1} H^{b_2}$ …[1]

where

V_t = Total outside bark volume (units^3 H)

D = diameter at breast height (same units as H)

H = total tree height (same units as D)

b_0, b_1, b_2 = equation parameters, with $b_2 + b_1 = 3$, and $b_0 = 0.83937$ and $b_1 = 2.18530$,

(Sharma and Oderwald, 2001)

Other southern pine and hardwood single tree volume:

* $V_t = b_0 \left(D^2 H\right)^{b_1}$ *dbh* < 12.7 cm (5"; pines), *dbh* < 28 cm (11"; hardwoods) …[2]

* $V_t = b_0 \left(D^2\right)^{b_1} H^{b_2}$ *dbh* > 12.7 cm (5"; pines), *dbh* > 28 cm (11"; hardwoods) …[3]

where

V_t = Total outside bark volume (ft^3)

D = diameter at breast height (inches)

H = total tree height (feet)

b_0, b_1, b_2 = equation parameters (Table C.1)

(Saucier and Clark, 1985; Clark *et al.,* 1986)

* Metric conversion factor = 0.0283168 (ft^3 to m^3)

Table C.1 Equation parameters for single tree volume equations [14; 15] (Saucier and Clark, 1985; Clark *et al.,* 1986)

Species	Equation parameters				
	Dbh < 12.7 cm (pines) Dbh < 28 cm (hardwoods)		**Dbh >12.7 cm (pines) Dbh > 28 cm (hardwoods)**		
	b_0	b_1	b_0	b_1	b_2
Southern pines (except loblolly)	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

Biomass equations

Single tree above-ground biomass:

Deciduous: $\text{Biomass} = 0.5 + \frac{25000 dbh^{2.5}}{dbh^{2.5} + 246872}$; R^2=0.99 …[4]

Coniferous: $\text{Biomass} = 0.5 + \frac{15000 dbh^{2.5}}{dbh^{2.5} + 364946}$ **;** R^2=0.98 …[5]

where

dbh = diameter at breast height (cm)
Biomass = single tree biomass (kg)

Schroeder *et al.* (1997)

Appendix E
SAS Program Code for Forward Variable Selection, Correlation Analysis, and Mallow's Cp Selection

```
************************************
* PROC REG WITH FORWARD SELECTION;
************************************;

title1 'Forward selection to reduce independent variables';
proc reg data = < Distributional data set >;
model Volume =

MeanVeg1 CVVeg1 KurtosisVeg1 MaxVeg1 MinVeg1 ModeVeg1 RangeVeg1 StdMeanVeg1 SkewnessVeg1 StdVeg1
MedianVeg1 P_Veg1_10 P_Veg1_20 P_Veg1_25 P_Veg1_30 P_Veg1_40 P_Veg1_50 P_Veg1_60 P_Veg1_70 P_Veg1_75
        P_Veg1_80 P_Veg1_90

MeanRef1 CVRef1 MaxRef1 MinRef1 RangeRef1 StdMeanRef1 StdRef1 MedianRef1

MeanVeg2 CVVeg2 KurtosisVeg2 MaxVeg2 MinVeg2 ModeVeg2 RangeVeg2 StdMeanVeg2 SkewnessVeg2 StdVeg2
MedianVeg2 P_Veg2_10 P_Veg2_20 P_Veg2_25 P_Veg2_30 P_Veg2_40 P_Veg2_50 P_Veg2_60 P_Veg2_70 P_Veg2_75
        P_Veg2_80 P_Veg2_90

MeanRef2 CVRef2 MaxRef2 MinRef2 RangeRef2 StdMeanRef2 StdRef2 MedianRef2

ZeroNVeg2ratio ZeroNVeg3_5ratio ZeroNgrnd1ratio ZeroNgrnd2ratio ZeroNgrnd3_5ratio Vegratio Canopy10P
Canopy20P Canopy30P Canopy40P Canopy50P Canopy60P Canopy70P Canopy80P Canopy90P /

selection = forward slentry = < α-level >;
run;
```

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

```
*************
* PROC CORR;
*************;

title1 'Forward selected variables entered into correlation analysis';
proc corr data = < Distributional data set >;
var < List Dependent and Independent Variables >;
run;

*******************************
* PROC REG WITH CP SELECTION;
*******************************;

title1 'Non-correlated, forward-selected variables entered into Mallow's Cp selection';
proc reg data = < Distributional data set >;

model Volume =  < List uncorrelated independent variables > /

selection = cp adjrsq rmse vif collinoint influence;
output out = < Output data set > r=res p=pred rstudent= Rstudent;
run;
```

Appendix F
Final Variables Entered into Mallow's Cp Regression Selection and Variables Removed Based on High Pearson's Correlation Values

Table F1.1 2-class: Significant variables for 27,050 segments (0.035 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_60 MeanRef1 CVVeg2 MinVeg2 Canopy80P	0.15	P_Veg2_60
	C	MeanVeg1 KurtosisVeg1 SkewnessVeg1 MaxRef1 MedianRef1 ModeVeg2 Grnd3_5ratio Canopy70P	0.135	P_Veg2_90
	A	P_Veg1_60 MedianRef1 ModeVeg2 P_Veg2_10 CVRef2 RangeRef2 StdRef2 Vegratio Canopy30P	0.125	None
Biomass	D	StdMeanVeg1 P_Veg1_70 StdRef1 CVVeg2 ModeVeg2 P_Veg2_40	0.15	P_Veg2_20 Canopy30P
	C	MeanVeg1 KurtosisVeg1 SkewnessVeg1 ModeVeg2 P_Veg2_10 P_Veg2_30 RangeRef2 Canopy70P	0.15	P_Veg2_25
	A	P_Veg1_70 MinRef2 Canopy30P	0.1	P_Veg1_50

Table F1.2 3-class: Significant variables for 27,050 segments (0.035 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_70 MeanRef1 MaxRef1 MinVeg2 ModeVeg2 P_Veg2_20 MeanRef2 Canopy80P	0.15	MedianRef2 Canopy30P
	C	P_Veg1_30 StdMeanVeg2 Canopy10P Canopy90P	0.1	None
	M	CVVeg1 MinRef1 MedianRef1 MinVeg2 P_Veg2_10 P_Veg2_25 P_Veg2_60 RangeRef2	0.225	P_Veg2_20
Biomass	D	P_Veg1_70 CVRef1 CVVeg2 MinVeg2 ModeVeg2 P_Veg2_10 P_Veg2_30	0.175	P_Veg2_25 Canopy30P
	C	P_Veg1_30 StdMeanVeg2 Grnd1ratio	0.075	None
	M	MeanVeg1 MinRef1 StdMeanRef1 MinVeg2 SkewnessVeg2 P_Veg2_10 StdRef2	0.25	ModeVeg1 StdMeanVeg2 Canopy40P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F2.1 2-class: Significant variables 10,352 segments (0.091 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	StdMeanVeg1 P_Veg1_60 MaxRef1 CVVeg2 MinVeg2 P_Veg2_25 Canopy80P	0.215	P_Veg2_60
	C	MeanVeg1 MinVeg1 MedianRef1 StdRef2 Vegratio Canopy60P	0.2	StdVeg2 Canopy40P
	A	CVVeg1 MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 MinRef2 Canopy30P	0.085	P_Veg1_70
Biomass	D	P_Veg1_60 MaxRef1 CVVeg2 StdRef2	0.2	None
	C	CVVeg1 P_Veg1_20 MedianRef1 P_Veg2_40 StdRef2 Vegratio Canopy40P Canopy70P	0.2	None
	A	StdMeanVeg1 MaxRef1 MinVeg2 StdVeg2 MinRef2 Canopy30P	0.2	P_Veg1_50

Table F2.2 3-class: Significant variables for 10,352 segments (0.091 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_70 MaxRef1 MedianRef1 CVVeg2 MinVeg2 StdMeanVeg2 MinRef2	0.25	Canopy30P RangeRef2
	C	P_Veg1_20 P_Veg2_20 P_Veg2_80 StdRef2 Grnd1ratio Canopy60P	0.125	P_Veg1_60 P_Veg2_75 Canopy40P Canopy80P
	M	MinVeg1 P_Veg1_10 MeanRef1 MinRef1 StdMeanRef1 SkewnessVeg2 P_Veg2_10 MaxRef2 Canopy40P	0.225	None
Biomass	D	P_Veg1_70 MaxRef1 CVVeg2 MinVeg2 StdMeanVeg2	0.2	P_Veg1_10 P_Veg1_75
	C	MaxVeg1 StdMeanVeg1 P_Veg1_20 StdMeanVeg2 P_Veg2_80 StdRef2 Grnd1ratio	0.175	None
	M	MinVeg1 StdMeanVeg1 P_Veg1_20 MaxRef1 ModeVeg2 P_Veg2_20 P_Veg2_40 StdMeanRef2 Canopy20P	0.19	P_Veg1_40

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F3.1 2-class: Significant variables for 6,687 segments (0.141 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_60 MaxRef1 CVVeg2 ModeVeg2 P_Veg2_10 P_Veg2_40 Grnd1ratio	0.2	P_Veg1_10
	C	MeanVeg1 StdVeg1 MedianRef1 P_Veg2_20 RangeRef2 StdRef2 Vegratio Canopy40P	0.135	None
	A	CVVeg1 MinVeg1 MedianVeg1 MaxRef1 RangeRef1 MedianRef1 ModeVeg2 P_Veg2_10Canopy30P	0.2	None
Biomass	D	P_Veg1_60 MeanRef1 MedianRef2	0.2	P_Veg1_20 P_Veg1_30 P_Veg1_40 MedianVeg1 MeanRef2 Canopy30P
	C	StdMeanVeg1 P_Veg1_40 MaxRef1 ModeVeg2 RangeRef2 Grnd1ratio Canopy30P Canopy90P	0.2	None
	A	CVVeg2 ModeVeg2 P_Veg2_80 Grnd2ratio	0.1	P_Veg1_10 P_Veg1_20 P_Veg1_30 P_Veg1_40 P_Veg1_50

Table F3.2 3-class: Significant variables for 6,687 segments (0.141 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_70 MedianRef1 ModeVeg2 StdMeanVeg2 MinRef2	0.2	SkewnessVeg2
	C	StdVeg1 P_Veg1_20 MedianRef1 ModeVeg2 Grnd1ratio Canopy90P	0.2	MeanVeg1
	M	P_Veg1_10 MaxRef1 MinRef1 MedianRef1 P_Veg2_40 MaxRef2 Canopy40P	0.2	None
Biomass	D	StdMeanVeg1 P_Veg1_80 ModeVeg2 P_Veg2_10	0.25	SkewnessVeg2
	C	P_Veg1_20 ModeVeg2 P_Veg2_80 StdRef2 Vegratio Canopy90P	0.095	P_Veg1_40
	M	MaxVeg1 StdMeanVeg1 P_Veg1_20 MaxRef1 KurtosisVeg2 MaxRef2 StdMeanRef2 Canopy70P	0.15	StdMeanVeg2 P_Veg1_40

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F4.1 2-class: Significant variables for 2,972 segments (0.318 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MeanVeg1 RangeRef1 StdMeanRef1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy20P Canopy90P	0.175	StdVeg2 P_Veg2_75
	C	P_Veg1_40 MinRef2 StdRef2 Vegratio	0.3	None
	A	StdMeanVeg1 P_Veg1_40 ModeVeg2 P_Veg2_10 MinRef2 StdRef2 Vegratio Canopy90P	0.2	Canopy30P
Biomass	D	StdMeanVeg1 P_Veg1_50 ModeVeg2 P_Veg2_10 Canopy80P	0.15	Canopy20P Canopy90P
	C	P_Veg1_40 MaxRef1 ModeVeg2 MinRef2 StdRef2 Vegratio	0.325	MedianVeg1
	A	MeanVeg1 CVVeg1 MedianRef1 ModeVeg2 P_Veg2_30 Canopy80P	0.1	Canopy90P

Table F4.2 3-class: Significant variables for 2,972 segments (0.318 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MinVeg1 MedianVeg1 MeanRef1 RangeRef1 ModeVeg2 P_Veg2_25 MaxRef2	0.315	MeanVeg1 Canopy20P P_Veg2_10
	C	MeanVeg1 MaxRef1 StdVeg2 StdRef2 Vegratio Canopy80P	0.235	P_Veg1_25 P_Veg1_40 P_Veg1_50 P_Veg1_90
	M	P_Veg1_10 MinRef1 StdMeanRef1 MaxRef2 MinRef2	0.3	None
Biomass	D	P_Veg1_60 ModeVeg2 P_Veg2_10 MaxRef2 Canopy80P	0.2	MeanRef2 Canopy90P
	C	P_Veg1_30 MinRef1 StdMeanVeg2 StdRef2 Vegratio Canopy80P	0.25	ModeVeg1
	M	P_Veg1_20 MedianRef1 P_Veg2_60 MinRef2 Grnd1ratio Canopy70P Canopy90P	0.2	SkewnessVeg1 P_Veg1_90 P_Veg2_75 Canopy50P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F5.1 2-class: Significant variables for 1,473 segments (0.642 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_40 MaxRef1 StdMeanRef1 RangeVeg2 P_Veg2_10 MaxRef2 Canopy90P	0.15	Canopy60P
	C	P_Veg1_40 MedianRef1 MeanRef2 MinRef2 ZeroNgrnd1ratio	0.175	MedianVeg1 P_Veg1_10 P_Veg1_20 CVRef2
	A	StdMeanVeg1 P_Veg1_40 P_Veg2_10 MinRef2 StdRef2 ZeroNgrnd1ratio Canopy90P	0.135	RangeRef2
Biomass	D	P_Veg1_50 RangeVeg2 P_Veg2_10 ZeroNgrnd1ratio	0.2	MaxVeg1 Canopy60P
	C	P_Veg1_40 MinRef2 Vegratio	0.2	P_Veg1_10
	A	CVVeg1 MedianVeg1 MedianRef1 MinVeg2 P_Veg2_30	0.175	MeanVeg1

Table F5.2 3-class: Significant variables for 1,473 segments (0.642 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	CVVeg1 MedianVeg1 RangeRef1 RangeVeg2 MaxRef2	0.15	SkewnessVeg1 CVVeg2 MaxVeg1
	C	P_Veg1_30 StdMeanRef2 StdRef2 MedianRef2 ZeroNgrnd1ratio Canopy70P	0..2	None
	M	MaxRef1 MinRef1 StdMeanRef1 MedianRef1 MaxRef2 MinRef2 Canopy60P	0.225	P_Veg1_10 Canopy40P
Biomass	D	P_Veg1_60 CVVeg2 Canopy80P	0.2	Canopy90P P_Veg1_90
	C	P_Veg1_25 StdMeanVeg2 CVRef2 StdRef2 ZeroNgrnd1ratio Canopy70P	0.15	None
	M	KurtosisVeg1 ModeVeg1 P_Veg2_40 ZeroNVeg3_5ratio Canopy60P Canopy90P	0.3	SkewnessVeg1 Canopy70P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F6.1 2-class: Significant variables for 981 segments (0.964 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MinVeg1 StdMeanVeg1 P_Veg1_10 P_Veg1_40 MaxRef1 MedianRef1 MinRef2 ZeroNgrnd1ratio	0.2	Canopy30P
	C	MeanVeg1 P_Veg1_10 MedianRef1 MaxVeg2 P_Veg2_30 MaxRef2 ZeroNgrnd1ratio Canopy30P	0.15	P_Veg2_50 P_Veg2_80
	A	MinVeg1 StdMeanVeg1 P_Veg1_40 MinRef1 MedianRef1 MinRef2 ZeroNgrnd1ratio	0.115	P_Veg1_10 Canopy30P Canopy50P
Biomass	D	StdMeanVeg1 P_Veg1_40 MaxRef1 MinVeg2 MinRef2 ZeroNgrnd1ratio	0.3	P_Veg1_20 Canopy30P
	C	MeanVeg1 P_Veg1_10 MaxVeg2 MinVeg2 ZeroNgrnd1ratio Canopy30P	0.2	None
	A	CVVeg1 MinVeg1 StdMeanVeg1 P_Veg1_40 CVRef1 MedianRef1 MinVeg2 MinRef2 Canopy70P	0.15	None

Table F6.2 3-class: Significant variables for 981 segments (0.964 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MinVeg1 P_Veg1_40 CVVeg2 ModeVeg2 ZeroNgrnd1ratio	0.3	MeanVeg1 Canopy30P Canopy50P
	C	MeanVeg1 StdVeg2 StdRef2 Vegratio Canopy80P	0.12	None
	M	MeanRef1 StdMeanRef1 P_Veg2_60 Canopy60P	0.25	MeanVeg2 StdVeg2 MedianVeg2 Canopy40P
Biomass	D	MinVeg1 StdMeanVeg1 MedianVeg1 CVRef1 CVVeg2 ZeroNgrnd1ratio Vegratio	0.3	P_Veg1_20
	C	P_Veg1_30 P_Veg2_80 StdRef2 Vegratio	0.12	None
	M	RangeVeg1 P_Veg2_60 MinRef2 Canopy60P Canopy70P	0.25	MeanVeg2 Canopy70P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F7.1 2-class: Significant variables for 749 segments (1.263 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MinVeg1 P_Veg1_10 P_Veg1_40 MeanRef1 MaxRef1 MinRef1 MinRef2 StdMeanRef2	0.295	Canopy20P
	C	MeanVeg1 P_Veg1_10 MedianRef1 ZeroNgrnd1ratio	0.05	MedianRef2
	A	P_Veg1_40 MedianRef1 MinVeg2 StdMeanRef2 ZeroNgrnd1ratio	0.125	None
Biomass	D	MinVeg1 P_Veg1_40 MaxRef1 MinRef1 StdRef1 CVRef2	0.3	P_Veg1_50
	C	P_Veg1_10 MedianRef1 MinVeg2 SkewnessVeg2 Vegratio Canopy50P	0.15	MeanRef2 Canopy40P
	A	CVVeg1 P_Veg1_40 MeanRef1 MaxRef1 MinVeg2 StdMeanRef2	0.25	Canopy60P

Table F7.2 3-class: Significant variables for 749 segments (1.263 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	CVVeg1 MinVeg1 P_Veg1_10 P_Veg1_40 MaxRef1 MedianRef1 StdMeanRef2 Canopy80P	0.275	Canopy30P
	C	MeanVeg1 StdVeg1 MinVeg2 ZeroNgrnd1ratio	0.145	None
	M	MedianRef1 ModeVeg2 StdMeanVeg2 MedianVeg2 CVRef2 MinRef2 Canopy60P	0.175	StdMeanVeg1 P_Veg2_80
Biomass	D	CVVeg1 MinVeg1 P_Veg1_40 MaxRef1	0.3	RangeRef1
	C	P_Veg1_20 MinRef1 CVVeg2 MinVeg2 StdMeanVeg2 StdRef2 Vegratio	0.25	MedianRef2
	M	ModeVeg2 Canopy10P Canopy60P Canopy90P	0.35	Canopy40P Canopy50P Canopy70P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F8.1 2-class: Significant variables for 502 segments (1.885 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	StdMeanVeg1 P_Veg1_30 MinRef1 MedianRef1 P_Veg2_20 MaxRef2	0.25	P_Veg2_60 MeanRef2 Canopy20P
	C	P_Veg1_10 MeanRef1	0.15	MeanRef2 MedianRef2
	A	P_Veg1_40 MinRef1 MedianRef1 MeanRef2 ZeroNgrnd1ratio	0.125	P_Veg1_10 MeanVeg2 P_Veg2_60 MedianRef2
Biomass	D	RangeVeg1 P_Veg1_40 MinRef1 ModeVeg2 StdMeanVeg2 ZeroNgrnd1ratio	0.3	ModeVeg1
	C	P_Veg1_40 MedianRef1 SkewnessVeg2 MedianVeg2 ZeroNVeg2ratio Canopy30P Canopy50P Canopy70P	0.15	MeanVeg2
	A	CVVeg1 P_Veg1_40 MeanRef1 MinRef1 MinVeg2 ModeVeg2 MaxRef2 MedianRef2	0.25	None

Table F8.2 3-class: Significant variables for 502 segments (1.885 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_30 MinRef1 MedianRef1 MaxRef2 Canopy70P	0.35	MedianRef2 Canopy50P
	C	MeanVeg1 StdVeg2 MinRef2 StdRef2 Vegratio Canopy30P Canopy50P Canopy70P	0.175	Canopy60P
	M	MinRef1 MedianRef1 MinRef2 ZeroNgrnd1ratio Canopy50P Canopy90P	0.15	Canopy30P
Biomass	D	P_Veg1_40 MinRef1 ModeVeg2 StdMeanVeg2 ZeroNgrnd1ratio Canopy70P	0.225	ModeVeg1 Canopy40P
	C	P_Veg1_30 P_Veg2_30 P_Veg2_80 StdRef2 Vegratio	0.25	MeanVeg2
	M	KurtosisVeg1 StdMeanVeg1 MinRef2 ZeroNgrnd1ratio Canopy60P	0.35	None

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F9.1 2-class: Significant variables for 374 segments (2.530 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_50 StdRef1 Canopy90P	0.25	None
	C	P_Veg1_40 MeanRef1 MaxRef1 ModeVeg2 RangeVeg2 StdRef2 Canopy30P Canopy90P	0.15	None
	A	CVVeg1 MinVeg1 P_Veg1_40 MaxRef1 MedianRef1 MaxVeg2 Canopy90P	0.185	Canopy30P
Biomass	D	MedianVeg1 MinRef1 ModeVeg2 CVRef2 Canopy90P	0.3	ModeVeg1
	C	MinVeg1 P_Veg1_40 MeanRef1 RangeRef1 StdVeg2 MedianVeg2 Canopy80P	0.2	P_Veg1_90 MeanVeg2
	A	CVVeg1 SkewnessVeg1 P_Veg1_50 MeanRef1 P_Veg2_20 Canopy80P	0.225	ModeVeg1 Canopy40P

Table F9.2 3-class: Significant variables for 374 segments (2.530 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MinVeg1 P_Veg1_50 RangeRef1 CVVeg2 MaxRef2 MedianRef2 Canopy70P Canopy90P	0.275	ModeVeg1
	C	KurtosisVeg1 P_Veg1_25 MeanRef1 RangeRef1 P_Veg2_70 StdRef2 ZeroNgrnd1ratio Canopy30P	0.2	P_Veg1_20 MeanVeg2
	M	SkewnessVeg2 Canopy60P Canopy90P	0.085	P_Veg2_80 Canopy50P
Biomass	D	MinVeg1 MedianVeg1 MaxRef1 ModeVeg2 MaxRef2 MedianRef2 Canopy90P	0.25	ModeVeg1
	C	P_Veg1_25 MinVeg2 P_Veg2_70 ZeroNgrnd1ratio Canopy40P	0.2	None
	M	SkewnessVeg1 P_Veg1_10 MinRef1 MedianRef1 Canopy60P Canopy80P	0.12	None

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F10.1 2-class: Significant variables for 240 segments (3.942 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_60 StdMeanRef1 MedianRef1 RangeRef2 ZeroNVeg2ratio Canopy90P	0.15	P_Veg1_50 P_Veg2_70
	C	RangeVeg1 P_Veg1_40 ModeVeg2 RangeVeg2 P_Veg2_25 Canopy90P	0.35	P_Veg2_20
	A	P_Veg1_40 StdMeanRef1 MedianRef1 P_Veg2_20 ZeroNgrnd1ratio Canopy90P	0.25	P_Veg2_60 Canopy30P
Biomass	D	StdMeanVeg1 P_Veg1_60 MaxVeg2 P_Veg2_10 RangeRef2 ZeroNVeg2ratio Canopy90P	0.2	MedianVeg1 P_Veg1_80
	C	P_Veg1_30 RangeRef1 RangeVeg2 MaxRef2 Canopy90P	0.3	None
	A	StdMeanRef1 CVVeg2 P_Veg2_75 MinRef2 ZeroNgrnd3_5ratio Canopy90P	0.15	P_Veg1_30

Table F10.2 3-class: Significant variables for 240 segments (3.942 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_60 KurtosisVeg2 ZeroNgrnd2ratio	0.25	StdVeg2
	C	MinVeg1 P_Veg1_25 RangeVeg2 ZeroNgrnd2ratio Canopy30P	0.3	
	M	MinVeg1 P_Veg1_30 MinRef1 CVVeg2 KurtosisVeg2 Canopy10P Canopy60P Canopy90P	0.1	ModeVeg1 Canopy50P Canopy80P
Biomass	D	StdMeanVeg1 P_Veg1_60 MaxVeg2 ModeVeg2 MinRef2 ZeroNVeg2ratio Canopy90P	0.25	MedianVeg1 StdRef1
	C	P_Veg1_25 MedianRef1 StdRef2 MedianRef2 ZeroNgrnd3_5ratio	0.3	None
	M	MeanVeg1 MinVeg1 RangeRef1 Vegratio Canopy70P	0.13	None

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F11.1 2-class: Significant variables for 168 segments (5.632 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg2_10 P_Veg2_70 MinRef2 ZeroNgrnd3_5ratio Canopy20P Canopy80P	0.3	SkewnessVeg2 SkewnessVeg1
	C	ModeVeg1 P_Veg1_40 RangeVeg2 MedianVeg2 StdRef2 ZeroNgrnd3_5ratio	0.3	Canopy30P
	A	P_Veg1_40 StdMeanVeg2 StdRef2 ZeroNgrnd3_5ratio	0.25	ModeVeg1 P_Veg1_70
Biomass	D	MinVeg1 P_Veg2_10 P_Veg2_75 CVRef2 ZeroNVeg2ratio Canopy20P Canopy70P	0.3	KurtosisVeg2 Canopy80P
	C	P_Veg1_40 MaxVeg2 ModeVeg2 P_Veg2_40 Canopy80P	0.25	P_Veg2_20 P_Veg2_25
	A	CVRef1 CVVeg2 P_Veg2_10 P_Veg2_30 P_Veg2_75 ZeroNVeg3_5ratio Canopy80P	0.275	P_Veg1_30

Table F11.2 3-class: Significant variables for 168 segments (5.632 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MinRef1 MaxVeg2 ModeVeg2 P_Veg2_70 ZeroNgrnd3_5ratio	0.35	CVVeg2 MinRef2
	C	KurtosisVeg1 P_Veg1_40 CVVeg2 P_Veg2_40	0.2	P_Veg1_90 P_Veg2_70 P_Veg2_80
	M	ModeVeg1 MinRef1 MinVeg2 StdMeanRef2 ZeroNgrnd1ratio Canopy10P Canopy50P Canopy80P	0.135	None
Biomass	D	ModeVeg2 P_Veg2_75 RangeRef2 StdMeanRef2 ZeroNgrnd3_5ratio Canopy90P	0.35	KurtosisVeg2 MinRef2
	C	ModeVeg1 RangeVeg1 P_Veg1_30 RangeVeg2	0.2	None
	M	KurtosisVeg1 MinVeg1 ModeVeg1 P_Veg1_10 CVRef1 RangeRef1 ZeroNVeg3_5ratio Canopy60P Canopy80P	0.115	Canopy50P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table F12.1 2-class: Significant variables for 167 Appomattox forest stands (5.666 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	MaxVeg1 MedianVeg1 MaxRef1 MedianRef1 RangeVeg2 StdMeanVeg2 P_Veg2_10 ZeroNVeg2ratio Canopy10P Canopy80P	0.275	None
	C	P_Veg1_30 MaxRef1 StdRef1 MedianRef1 KurtosisVeg2 MinRef2 StdRef2 ZeroNVeg2ratio	0.3	P_Veg1_20
	A	CVVeg1 MedianVeg1 CVRef1 RangeRef1 MedianRef1 P_Veg2_90 StdMeanRef2 ZeroNgrnd1ratio Canopy90P	0.16	Canopy10P StdMeanVeg2
Biomass	D	MedianVeg1 P_Veg2_10 ZeroNVeg2ratio Canopy10P	0.25	P_Veg2_20
	C	P_Veg1_30 MedianRef1 Vegratio	0.3	None
	A	RangeRef1 StdRef1 MedianRef1 StdMeanVeg2 P_Veg2_75 MedianRef2 ZeroNVeg3_5ratio ZeroNgrnd1ratio Canopy80P Canopy90P	0.125	None

Table F12.2 3-class: Significant variables for 167 Appomattox forest stands (5.666 ha/segment) and all forest types as chosen by forward selection (*deciduous = D; coniferous = C; all segments/types = A*)

Model		Variables	Stepwise α-value	Removed
Volume	D	P_Veg1_50 MaxRef1 MedianRef1 RangeVeg2 P_Veg2_10 ZeroNgrnd2ratio Canopy10P	0.3	Canopy30P
	C	P_Veg1_25 MaxRef1 MaxVeg2 Canopy80P	0.135	None
	M	MedianVeg1 MinRef1 CVRef2 MinRef2 StdMeanRef2 Canopy90P	0.3	Canopy30P Canopy70P
Biomass	D	MedianVeg1 P_Veg2_10 ZeroNgrnd2ratio Canopy10P	0.35	CVVeg1 Canopy20P
	C	P_Veg1_20 MaxRef1 StdVeg2 P_Veg2_30 StdMeanRef2 StdRef2	0.2	P_Veg2_10 P_Veg2_25
	M	MeanVeg1 StdMeanRef1 MinVeg2 P_Veg2_20 Canopy90P	0.325	SkewnessVeg1 Canopy30P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Appendix G
Candidate Volume and Biomass Models for Deciduous, Coniferous, Mixed, and All Combined Types and Segmentation Treatments

Table G.1.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 27,050 segments (0.035 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.1.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_60 CVVeg2 Canopy80P	0.53	0.51	4.25	58.62
		P_Veg1_60 CVVeg2 MinVeg2 Canopy80P	0.53	0.52	4.60	58.48
		P_Veg1_60 MeanRef1 CVVeg2	0.52	0.51	4.65	58.71
		P_Veg1_60 CVVeg2	0.52	0.51	4.91	58.98
		* P_Veg1_60 MinVeg2	0.51	0.51	5.23	59.04
		P_Veg1_60 MeanRef1 CVVeg2 Canopy80P	0.53	0.51	5.31	58.64
		P_Veg1_60 MinVeg2 Canopy80P	0.52	0.51	5.35	58.86
		P_Veg1_60 MeanRef1 MinVeg2	0.52	0.51	5.36	58.86
		P_Veg1_60 CVVeg2 MinVeg2	0.52	0.51	5.47	58.88
		P_Veg1_60 MeanRef1	0.51	0.51	5.49	59.10
		P_Veg1_60 MeanRef1 CVVeg2 MinVeg2	0.53	0.51	5.63	58.71
		P_Veg1_60	0.51	0.50	5.99	59.41
		P_Veg1_60 MeanRef1 CVVeg2 MinVeg2 Canopy80P	0.53	0.52	6.00	58.57
		P_Veg1_60 MeanRef1 MinVeg2 Canopy80P	0.52	0.51	6.50	58.89
		P_Veg1_60 Canopy80P	0.51	0.50	6.65	59.34
		P_Veg1_60 MeanRef1 Canopy80P	0.52	0.50	7.07	59.23
	C	* MeanVeg1 KurtosisVeg1 SkewnessVeg1 MedianRef1 ModeVeg2 Canopy70P	0.65	0.62	7.95	45.43
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 MedianRef1 ModeVeg2 Grnd3_5ratio Canopy70P	0.66	0.63	7.99	45.09
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 MedianRef1 Grnd3_5ratio Canopy70P	0.65	0.62	8.63	45.68
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 MaxRef1 MedianRef1 ModeVeg2 Grnd3_5ratio Canopy70P	0.67	0.63	9.00	45.10
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 MaxRef1 MedianRef1 Grnd3_5ratio Canopy70P	0.66	0.62	9.27	45.56
		MeanVeg1 SkewnessVeg1 MedianRef1 Grnd3_5ratio Canopy70P	0.64	0.61	9.36	46.26
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 MaxRef1 MedianRef1 ModeVeg2 Canopy70P	0.66	0.62	9.44	45.62
		MeanVeg1 SkewnessVeg1 MedianRef1 ModeVeg2 Grnd3_5ratio Canopy70P	0.65	0.61	9.64	46.03
		MeanVeg1 SkewnessVeg1 MedianRef1 ModeVeg2 Canopy70P	0.65	0.61	9.72	46.39
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 MedianRef1 Canopy70P	0.63	0.60	9.88	46.44

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.1.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	A	P_Veg1_60 MedianRef1 ModeVeg2 P_Veg2_10 RangeRef2 StdRef2 Canopy30P	0.60	0.59	9.31	52.64
		P_Veg1_60 MedianRef1 ModeVeg2 P_Veg2_10 RangeRef2 StdRef2 Vegratio Canopy30P	0.61	0.59	9.52	52.53
		P_Veg1_60 MedianRef1 ModeVeg2 P_Veg2_10 CVRef2 RangeRef2 StdRef2 Vegratio Canopy30P	0.61	0.59	10.00	52.45
		* P_Veg1_60 MedianRef1 P_Veg2_10 RangeRef2 StdRef2 Canopy30P	0.60	0.58	10.00	52.87
		P_Veg1_60 MedianRef1 P_Veg2_10 RangeRef2 StdRef2 Vegratio Canopy30P	0.60	0.59	10.33	52.78
		P_Veg1_60 MedianRef1 ModeVeg2 P_Veg2_10 CVRef2 RangeRef2 StdRef2 Canopy30P	0.60	0.59	10.76	52.70
		P_Veg1_60 MedianRef1 P_Veg2_10 CVRef2 RangeRef2 StdRef2 Vegratio Canopy30P	0.60	0.58	11.12	52.75
		P_Veg1_60 MedianRef1 RangeRef2 StdRef2 Vegratio Canopy30P	0.59	0.58	11.19	53.04
		P_Veg1_60 MedianRef1 P_Veg2_10 CVRef2 RangeRef2 StdRef2 Canopy30P	0.60	0.58	11.60	52.96
		P_Veg1_60 MedianRef1 CVRef2 RangeRef2 StdRef2 Vegratio Canopy30P	0.60	0.58	11.72	52.98
		P_Veg1_60 MedianRef1 P_Veg2_10 StdRef2 Canopy30P	0.59	0.58	12.27	53.32
		P_Veg1_60 MedianRef1 RangeRef2 StdRef2 Canopy30P	0.59	0.59	12.30	53.32
Biomass	D	* StdMeanVeg1 P_Veg1_70 StdRef1 CVVeg2 ModeVeg2	0.56	0.54	5.30	38.33
		StdMeanVeg1 P_Veg1_70 StdRef1 CVVeg2 ModeVeg2 P_Veg2_40	0.56	0.54	7.00	38.44
		StdMeanVeg1 P_Veg1_70 StdRef1 CVVeg2 P_Veg2_40	0.55	0.53	7.80	38.74
		StdMeanVeg1 P_Veg1_70 StdRef1 CVVeg2	0.54	0.52	8.23	38.96
		StdMeanVeg1 P_Veg1_70 StdRef1 ModeVeg2 P_Veg2_40	0.54	0.52	10.84	39.23
		StdMeanVeg1 P_Veg1_70 CVVeg2 ModeVeg2	0.53	0.51	10.87	39.38
		StdMeanVeg1 P_Veg1_70 StdRef1	0.52	0.51	11.62	39.65
		StdMeanVeg1 P_Veg1_70 StdRef1 ModeVeg2	0.53	0.51	11.70	39.51
	C	* MeanVeg1 SkewnessVeg1 ModeVeg2 P_Veg2_30 RangeRef2 Canopy70P	0.61	0.57	8.59	17.56
		MeanVeg1 SkewnessVeg1 ModeVeg2 P_Veg2_10 P_Veg2_30 RangeRef2 Canopy70P	0.62	0.58	8.67	17.43
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 ModeVeg2 P_Veg2_10 P_Veg2_30 RangeRef2 Canopy70P	0.63	0.58	9.00	17.34
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 ModeVeg2 RangeRef2 Canopy70P	0.60	0.57	9.22	17.64
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 ModeVeg2 P_Veg2_30 RangeRef2 Canopy70P	0.62	0.57	9.24	17.51
		MeanVeg1 SkewnessVeg1 ModeVeg2 P_Veg2_30 Canopy70P	0.59	0.56	9.30	17.78
		MeanVeg1 KurtosisVeg1 SkewnessVeg1 ModeVeg2 Canopy70P	0.59	0.56	9.36	17.79
		MeanVeg1 SkewnessVeg1 P_Veg2_10 P_Veg2_30 RangeRef2 Canopy70P	0.60	0.56	9.45	17.68
	A	* P_Veg1_70 Canopy30P	0.60	0.60	2.13	38.36
		P_Veg1_70 MinRef2 Canopy30P	0.60	0.60	4.00	38.43
		P_Veg1_70	0.58	0.58	12.08	39.32
		P_Veg1_70 MinRef2	0.58	0.57	14.01	39.40

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.1.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 27,050 segments (0.035 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.1.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_70 MinVeg2 ModeVeg2 P_Veg2_20	0.62	0.60	3.46	54.99
		* P_Veg1_70 ModeVeg2 P_Veg2_20	0.61	0.59	3.78	55.39
		MinVeg2 ModeVeg2 P_Veg2_20 Canopy80P	0.62	0.60	4.48	54.99
		P_Veg1_70 MeanRef1 MinVeg2 ModeVeg2 P_Veg2_20	0.62	0.60	4.52	55.01
		P_Veg1_70 MaxRef1 MinVeg2 ModeVeg2 P_Veg2_20	0.62	0.60	4.88	55.12
		P_Veg1_70 MinVeg2 ModeVeg2 P_Veg2_20 MeanRef2	0.62	0.60	4.96	55.14
		P_Veg1_70 MeanRef1 ModeVeg2 P_Veg2_20	0.61	0.59	5.08	55.48
		P_Veg1_70 ModeVeg2 P_Veg2_20 Canopy80P	0.61	0.59	5.32	55.55
		P_Veg1_70 MaxRef1 ModeVeg2 P_Veg2_20	0.61	0.59	5.49	55.60
		P_Veg1_70 ModeVeg2 P_Veg2_20 MeanRef2	0.61	0.59	5.66	55.65
	C	P_Veg1_30 Canopy10P	0.49	0.47	4.38	4843
		* P_Veg1_30 StdMeanVeg2 Canopy10P	0.50	0.47	4.91	48.23
		P_Veg1_30 StdMeanVeg2 Canopy10P Canopy90P	0.52	0.48	5.00	47.81
		P_Veg1_30 Canopy10P Canopy90P	0.50	0.47	5.61	48.54
		P_Veg1_30	0.44	0.43	7.94	50.37
		P_Veg1_30 Canopy90P	0.45	0.43	8.09	50.05
		P_Veg1_30 StdMeanVeg2 Canopy90P	0.46	0.42	9.84	50.42
		P_Veg1_30 StdMeanVeg2	0.44	0.41	9.93	50.84
	M	MinRef1 MedianRef1 MinVeg2 P_Veg2_25 P_Veg2_60 RangeRef2	0.62	0.57	5.77	47.55
		* MinRef1 MedianRef1 MinVeg2 P_Veg2_60 RangeRef2	0.60	0.56	6.07	48.27
		MinRef1 MedianRef1 P_Veg2_60 RangeRef2	0.58	0.54	7.13	49.34
		MinRef1 MedianRef1 MinVeg2 P_Veg2_10 P_Veg2_25 P_Veg2_60 RangeRef2	0.63	0.57	7.36	47.88
		CVVeg1 MinRef1 MedianRef1 MinVeg2 P_Veg2_25 P_Veg2_60 RangeRef2	0.62	0.56	7.56	47.99
		CVVeg1 MinRef1 MedianRef1 MinVeg2 P_Veg2_60 RangeRef2	0.60	0.55	7.93	48.74
		MinRef1 MedianRef1 MinVeg2 P_Veg2_10 P_Veg2_60 RangeRef2	0.60	0.55	8.02	48.79
		MinRef1 MedianRef1 P_Veg2_25 P_Veg2_60 RangeRef2	0.58	0.53	8.51	49.56
		MinRef1 MedianRef1 P_Veg2_10 P_Veg2_60 RangeRef2	0.58	0.53	8.75	49.69
		MinRef1 MedianRef1 P_Veg2_10 P_Veg2_25 P_Veg2_60 RangeRef2	0.60	0.54	8.76	49.19
		MedianRef1 MinVeg2 P_Veg2_60 RangeRef2	0.56	0.52	8.82	50.19
		MedianRef1 P_Veg2_60 RangeRef2	0.54	0.51	8.89	50.68

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.1.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	D	P_Veg1_70 CVRef1 CVVeg2 MinVeg2 ModeVeg2	0.59	0.57	4.07	39.57
		* P_Veg1_70 CVVeg2 MinVeg2 ModeVeg2	0.58	0.56	4.83	39.95
		P_Veg1_70 CVRef1 CVVeg2 ModeVeg2	0.58	0.56	5.15	40.02
		P_Veg1_70 CVVeg2	0.56	0.55	5.17	40.45
		P_Veg1_70 CVVeg2 ModeVeg2	0.57	0.56	5.33	40.27
		P_Veg1_70 CVRef1 CVVeg2	0.57	0.56	5.43	40.30
		P_Veg1_70 CVRef1 MinVeg2 ModeVeg2 P_Veg2_30	0.59	0.56	5.61	39.91
		P_Veg1_70 CVRef1 MinVeg2 ModeVeg2	0.58	0.56	5.83	40.17
		P_Veg1_70 CVRef1 CVVeg2 ModeVeg2 P_Veg2_10	0.59	0.56	5.95	39.98
		P_Veg1_70 CVRef1 CVVeg2 MinVeg2 ModeVeg2 P_Veg2_30	0.59	0.57	6.01	39.77
		P_Veg1_70 CVRef1 CVVeg2 MinVeg2 ModeVeg2 P_Veg2_10	0.59	0.57	6.03	39.78
		P_Veg1_70 CVVeg2 MinVeg2	0.57	0.55	6.20	40.46
	C	* P_Veg1_30 StdMeanVeg2 Grnd1ratio	0.53	0.50	4.00	14.13
		P_Veg1_30 Grnd1ratio	0.48	0.46	7.06	14.66
		P_Veg1_30	0.43	0.42	10.11	15.16
		P_Veg1_30 StdMeanVeg2	0.44	0.42	11.35	15.20
	M	* MeanVeg1 StdMeanRef1 MinVeg2	0.51	0.48	2.55	28.20
		MeanVeg1 MinVeg2	0.48	0.46	2.98	28.64
		MeanVeg1 StdMeanRef1 P_Veg2_10	0.50	0.47	3.08	28.36
		MeanVeg1 MinVeg2 SkewnessVeg2	0.50	0.47	3.36	28.45
		MeanVeg1 StdMeanRef1 MinVeg2 SkewnessVeg2	0.52	0.48	3.55	28.19
		MeanVeg1 StdMeanRef1 MinVeg2 P_Veg2_10	0.52	0.47	3.65	28.22
		MeanVeg1 MinRef1 StdMeanRef1 MinVeg2	0.51	0.47	4.33	28.43
		MeanVeg1 MinRef1 MinVeg2 SkewnessVeg2	0.51	0.47	4.42	28.46
		MeanVeg1 StdMeanRef1 MinVeg2 StdRef2	0.51	0.46	4.55	28.50
		MeanVeg1 MinVeg2 P_Veg2_10	0.48	0.45	4.58	28.82
		MeanVeg1 MinRef1 MinVeg2	0.48	0.45	4.64	28.84
		MeanVeg1 MinRef1 StdMeanRef1 P_Veg2_10	0.51	0.46	4.66	28.54

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.2.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 10,352 segments (0.091 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.2.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	StdMeanVeg1 P_Veg1_60 MaxRef1 CVVeg2 MinVeg2	0.57	0.55	6.65	56.17
		StdMeanVeg1 P_Veg1_60 MaxRef1 CVVeg2 MinVeg2 Canopy80P	0.58	0.56	6.75	55.98
		P_Veg1_60 MaxRef1 CVVeg2 MinVeg2 Canopy80P	0.57	0.55	6.77	56.20
		* P_Veg1_60 MaxRef1 CVVeg2 MinVeg2	0.56	0.55	6.96	56.44
		P_Veg1_60 MaxRef1 CVVeg2 MinVeg2 P_Veg2_25 Canopy80P	0.57	0.55	7.84	56.21
		StdMeanVeg1 P_Veg1_60 MaxRef1 CVVeg2 MinVeg2 P_Veg2_25	0.57	0.55	7.94	56.23
		P_Veg1_60 MaxRef1 CVVeg2	0.55	0.54	7.99	56.85
		StdMeanVeg1 P_Veg1_60 MaxRef1 CVVeg2 MinVeg2 P_Veg2_25 Canopy80P	0.58	0.56	8.00	56.03
		P_Veg1_60 MaxRef1 CVVeg2 Canopy80P	0.56	0.55	8.03	56.66
		P_Veg1_60 MaxRef1 CVVeg2 MinVeg2 P_Veg2_25	0.57	0.55	8.07	56.46
		P_Veg1_60 MaxRef1 CVVeg2 P_Veg2_25 Canopy80P	0.57	0.55	8.30	56.51
		P_Veg1_60 MaxRef1 CVVeg2 P_Veg2_25	0.56	0.55	8.35	56.72
		StdMeanVeg1 P_Veg1_60 CVVeg2 MinVeg2	0.56	0.54	9.00	56.86
		StdMeanVeg1 P_Veg1_60 MaxRef1 CVVeg2	0.56	0.54	9.48	56.96
	C	* MeanVeg1 MinVeg1 MedianRef1 StdRef2 Vegratio	0.67	0.64	5.25	45.07
		MeanVeg1 MinVeg1 MedianRef1 StdRef2 Vegratio Canopy60P	0.67	0.64	7.00	45.30
		MeanVeg1 MinVeg1 MedianRef1 Vegratio	0.64	0.62	7.96	46.20
		MeanVeg1 MinVeg1 StdRef2 Vegratio	0.64	0.62	8.45	46.35
		MeanVeg1 MinVeg1 StdRef2 Vegratio Canopy60P	0.65	0.63	8.82	46.17
		MeanVeg1 MinVeg1 MedianRef1 Vegratio Canopy60P	0.65	0.62	9.14	46.27
		MeanVeg1 MedianRef1 StdRef2 Vegratio	0.63	0.61	10.12	46.85
		MeanVeg1 StdRef2 Vegratio	0.62	0.61	10.33	47.18
		MeanVeg1 StdRef2 Vegratio Canopy60P	0.63	0.61	11.59	47.28
		MeanVeg1 MinVeg1 MedianRef1 StdRef2	0.63	0.61	12.01	47.40

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.2.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	A	CVVeg1 MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 MinRef2 Canopy30P	0.61	0.60	9.00	53.17
		CVVeg1 MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 Canopy30P	0.60	0.59	10.48	53.48
		CVVeg1 MinVeg1 P_Veg1_50 MedianRef1 MinVeg2 MinRef2 Canopy30P	0.60	0.59	11.35	53.59
		* CVVeg1 MinVeg1 P_Veg1_50 MedianRef1 MinVeg2 Canopy30P	0.60	0.59	11.60	53.75
		CVVeg1 MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinRef2 Canopy30P	0.60	0.59	11.99	53.67
		MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 MinRef2 Canopy30P	0.60	0.59	12.46	53.73
		CVVeg1 MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 MinRef2	0.60	0.59	12.72	53.76
		CVVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 MinRef2 Canopy30P	0.60	0.59	12.92	53.79
		CVVeg1 MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2	0.60	0.58	12.97	53.92
		CVVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 Canopy30P	0.59	0.58	13.10	53.93
		MinVeg1 P_Veg1_50 MaxRef1 MedianRef1 MinVeg2 Canopy30P	0.59	0.58	13.40	53.97
		CVVeg1 MinVeg1 P_Veg1_50 MedianRef1 MinVeg2	0.59	0.58	13.72	54.13
Biomass	D	P_Veg1_60 MaxRef1 CVVeg2 StdRef2	0.53	0.52	5.00	40.74
		* P_Veg1_60 MaxRef1 CVVeg2	0.52	0.51	5.45	40.95
		P_Veg1_60 CVVeg2	0.51	0.51	6.12	41.20
		41234 P_Veg1_60 CVVeg2 StdRef2	0.52	0.50	7.35	41.23
		41522 P_Veg1_60 MaxRef1 StdRef2	0.51	0.50	9.30	41.52
		P_Veg1_60	0.49	0.49	10.53	41.98
		P_Veg1_60 StdRef2	0.50	0.49	10.60	41.85
		P_Veg1_60 MaxRef1	0.50	0.49	11.21	41.94
	C	P_Veg1_20 MedianRef1 P_Veg2_40 Vegratio Canopy40P Canopy70P	0.62	0.59	8.86	16.99
		CVVeg1 P_Veg1_20 MedianRef1 P_Veg2_40 StdRef2 Vegratio Canopy40P Canopy70P	0.64	0.60	9.00	16.77
		* P_Veg1_20 MedianRef1 P_Veg2_40 Vegratio Canopy70P	0.61	0.59	9.31	17.15
		P_Veg1_20 MedianRef1 P_Veg2_40 StdRef2 Vegratio Canopy40P Canopy70P	0.63	0.60	9.40	16.94
		P_Veg1_20 MedianRef1 P_Veg2_40 Vegratio	0.60	0.58	9.45	17.27
		P_Veg1_20 MedianRef1 P_Veg2_40 StdRef2 Vegratio Canopy70P	0.62	0.59	9.49	17.06
		P_Veg1_20 MedianRef1 P_Veg2_40 StdRef2 Vegratio	0.61	0.58	9.53	17.17
		CVVeg1 P_Veg1_20 MedianRef1 P_Veg2_40 Vegratio Canopy40P Canopy70P	0.63	0.59	9.59	16.96
		P_Veg1_20 P_Veg2_40 StdRef2 Vegratio Canopy70P	0.61	0.58	9.70	17.19
		CVVeg1 P_Veg1_20 P_Veg2_40 StdRef2 Vegratio Canopy40P Canopy70P	0.63	0.59	9.77	16.98

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.2.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	A	* MaxRef1 StdVeg2 MinRef2	0.60	0.59	1.58	38.69
		MaxRef1 MinVeg2 StdVeg2 MinRef2	0.60	0.59	3.39	38.76
		MaxRef1 StdVeg2 MinRef2 Canopy30P	0.60	0.59	3.40	38.76
		StdMeanVeg1 MaxRef1 StdVeg2 MinRef2	0.60	0.59	3.47	38.77
		StdMeanVeg1 MaxRef1 MinVeg2 StdVeg2 MinRef2	0.60	0.59	5.12	38.83
		MaxRef1 MinVeg2 StdVeg2 MinRef2 Canopy30P	0.60	0.59	5.19	38.83
		StdMeanVeg1 MaxRef1 StdVeg2 MinRef2 Canopy30P	0.60	0.59	5.34	38.85
		StdMeanVeg1 MaxRef1 MinVeg2 StdVeg2 MinRef2 Canopy30P	0.60	0.59	7.00	38.91
		MaxRef1 StdVeg2	0.58	0.58	7.30	39.29
		MaxRef1 MinVeg2 StdVeg2	0.58	0.58	8.24	39.29
		StdVeg2 MinRef2	0.58	0.58	8.30	39.38
		MaxRef1 StdVeg2 Canopy30P	0.58	0.58	8.93	39.35
		StdMeanVeg1 MaxRef1 StdVeg2	0.58	0.58	9.18	39.37
		StdVeg2 MinRef2 Canopy30P	0.58	0.57	9.54	39.40

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.2.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 10,352 segments (0.091 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.2.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_70 MaxRef1 CVVeg2 MinVeg2	0.62	0.61	6.37	55.81
		P_Veg1_70 MaxRef1 CVVeg2 MinVeg2 StdMeanVeg2	0.63	0.61	6.81	55.66
		* P_Veg1_70 MaxRef1 CVVeg2	0.62	0.60	6.85	55.18
		P_Veg1_70 MedianRef1 CVVeg2 MinVeg2 StdMeanVeg2	0.63	0.61	7.00	55.71
		P_Veg1_70 CVVeg2 MinVeg2 StdMeanVeg2	0.62	0.61	7.22	56.02
		P_Veg1_70 MaxRef1 MedianRef1 CVVeg2 MinVeg2 StdMeanVeg2	0.63	0.61	7.34	55.54
		P_Veg1_70 MaxRef1 MedianRef1 CVVeg2 MinVeg2	0.63	0.61	7.38	55.81
		P_Veg1_70 MaxRef1 CVVeg2 MinVeg2 StdMeanVeg2 MinRef2	0.63	0.61	7.47	55.58
		P_Veg1_70 MedianRef1 MinVeg2 StdMeanVeg2 MinRef2	0.63	0.61	7.57	55.86
		P_Veg1_70 MaxRef1 MedianRef1 CVVeg2	0.62	0.61	7.57	56.11
	C	* P_Veg1_20 P_Veg2_20 P_Veg2_80 Grnd1ratio	0.58	0.54	5.51	44.82
		P_Veg1_20 P_Veg2_20 P_Veg2_80 StdRef2 Grnd1ratio	0.59	0.55	6.15	44.67
		P_Veg1_20 P_Veg2_20 P_Veg2_80 Grnd1ratio Canopy60P	0.59	0.55	6.19	44.69
		P_Veg1_20 StdRef2 Grnd1ratio	0.55	0.53	6.46	45.64
		P_Veg1_20 P_Veg2_20 P_Veg2_80 StdRef2 Grnd1ratio Canopy60P	0.60	0.55	7.00	44.60
		P_Veg1_20 P_Veg2_20 StdRef2 Grnd1ratio	0.56	0.53	7.11	45.51
		P_Veg1_20 StdRef2 Grnd1ratio Canopy60P	0.56	0.53	7.23	45.56
		P_Veg1_20 P_Veg2_20 StdRef2 Grnd1ratio Canopy60P	0.58	0.54	7.36	45.20
		P_Veg1_20 P_Veg2_20 Grnd1ratio Canopy60P	0.56	0.52	7.97	45.88
		P_Veg1_20 P_Veg2_80 StdRef2 Grnd1ratio	0.56	0.52	8.10	45.94
	M	* MinVeg1 P_Veg1_10 MeanRef1 StdMeanRef1 SkewnessVeg2 P_Veg2_10 MaxRef2 Canopy40P	0.69	0.63	9.56	44.29
		MinVeg1 P_Veg1_10 MeanRef1 MinRef1StdMeanRef1 SkewnessVeg2 P_Veg2_10 MaxRef2 Canopy40P	0.70	0.63	10.00	43.99
		MinVeg1 P_Veg1_10 MeanRef1 SkewnessVeg2 P_Veg2_10 MaxRef2 Canopy40P	0.66	0.61	10.73	45.36
		MinVeg1 P_Veg1_10 MeanRef1 StdMeanRef1 SkewnessVeg2 P_Veg2_10 Canopy40P	0.66	0.61	10.90	45.45
		MinVeg1 P_Veg1_10 MeanRef1 MinRef1 SkewnessVeg2 P_Veg2_10 MaxRef2 Canopy40P	0.67	0.60	12.20	45.63
		MinVeg1 P_Veg1_10 MeanRef1 MinRef1 StdMeanRef1 SkewnessVeg2 P_Veg2_10 Canopy40P	0.67	0.60	12.48	45.78
		MinVeg1 P_Veg1_10 MeanRef1 StdMeanRef1 SkewnessVeg2 MaxRef2 Canopy40P	0.65	0.60	12.99	46.47
		MinVeg1 P_Veg1_10 MeanRef1 StdMeanRef1 P_Veg2_10 MaxRef2 Canopy40P	0.64	0.58	13.67	46.80
		MinVeg1 P_Veg1_10 MeanRef1 StdMeanRef1 MaxRef2 Canopy40P	0.62	0.57	14.13	47.42

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.2.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	D	* P_Veg1_70 MaxRef1 CVVeg2	0.59	0.57	5.17	40.77
		P_Veg1_70 MaxRef1 CVVeg2 MinVeg2	0.59	0.58	5.26	40.60
		P_Veg1_70 MaxRef1 CVVeg2 MinVeg2 StdMeanVeg2	0.60	0.58	6.00	40.55
		P_Veg1_70 CVVeg2 MinVeg2 StdMeanVeg2	0.59	0.57	6.32	40.80
		P_Veg1_70 MaxRef1 CVVeg2 StdMeanVeg2	0.59	0.57	6.86	40.90
		P_Veg1_70 CVVeg2	0.57	0.56	6.93	41.28
		P_Veg1_70 CVVeg2 StdMeanVeg2	0.58	0.57	7.01	41.11
		P_Veg1_70 MinVeg2 StdMeanVeg2	0.58	0.57	7.44	41.19
		P_Veg1_70 CVVeg2 MinVeg2	0.58	0.56	8.06	41.31
		P_Veg1_70 MaxRef1 MinVeg2 StdMeanVeg2	0.58	0.57	8.38	41.19
	C	* MaxVeg1 StdMeanVeg1 P_Veg1_20 StdMeanVeg2 P_Veg2_80 StdRef2 Grnd1ratio	0.67	0.62	8.00	12.29
		StdMeanVeg1 P_Veg1_20 StdMeanVeg2 P_Veg2_80 StdRef2 Grnd1ratio	0.65	0.60	9.49	12.59
		MaxVeg1 StdMeanVeg1 P_Veg1_20 StdMeanVeg2 P_Veg2_80 Grnd1ratio	0.63	0.59	11.17	12.80
		MaxVeg1 StdMeanVeg1 P_Veg1_20 P_Veg2_80 StdRef2 Grnd1ratio	0.63	0.59	11.62	12.85
		StdMeanVeg1 P_Veg1_20 P_Veg2_80 StdRef2 Grnd1ratio	0.61	0.58	11.99	13.00
		StdMeanVeg1 P_Veg1_20 StdMeanVeg2 StdRef2 Grnd1ratio	0.61	0.57	12.10	13.01
		StdMeanVeg1 P_Veg1_20 StdMeanVeg2 P_Veg2_80 Grnd1ratio	0.61	0.57	12.48	13.06
		MaxVeg1 StdMeanVeg1 P_Veg1_20 StdMeanVeg2 StdRef2 Grnd1ratio	0.62	0.57	12.99	13.02
		P_Veg1_20 StdMeanVeg2 StdRef2 Grnd1ratio	0.59	0.55	13.96	13.32
		MaxVeg1 P_Veg1_20 P_Veg2_80 StdRef2 Grnd1ratio	0.60	0.56	14.21	13.26
	M	* StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 P_Veg2_40 StdMeanRef2 Canopy20P	0.61	0.55	6.74	26.18
		StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 P_Veg2_40 StdMeanRef2	0.59	0.53	7.08	26.59
		StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 StdMeanRef2 Canopy20P	0.58	0.53	7.85	26.82
		StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 StdMeanRef2	0.56	0.51	7.94	27.13
		MinVeg1 StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 P_Veg2_40 StdMeanRef2 Canopy20P	062	0.55	8.03	26.26
			0.60	0.53	8.31	26.66
		MinVeg1 StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 P_Veg2_40 StdMeanRef2	0.59	0.53	8.63	26.76
		MinVeg1 StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 StdMeanRef2 Canopy20P	0.57	0.52	8.68	27.07
		MinVeg1 StdMeanVeg1 P_Veg1_20 MaxRef1 P_Veg2_20 StdMeanRef2	0.61	0.54	8.70	26.47
		StdMeanVeg1 P_Veg1_20 MaxRef1 ModeVeg2 P_Veg2_20 P_Veg2_40 StdMeanRef2 Canopy20P	0.59	0.52	9.05	26.89
			0.57	0.51	9.32	27.26
		StdMeanVeg1 P_Veg1_20 MaxRef1 ModeVeg2 P_Veg2_20 P_Veg2_40 StdMeanRef2 P_Veg1_20 MaxRef1 P_Veg2_20 P_Veg2_40 StdMeanRef2 Canopy20P				

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.3.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 6,687 segments (0.141 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.3.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_60 CVVeg2 ModeVeg2 P_Veg2_10 P_Veg2_40	0.54	0.53	6.19	57.92
		P_Veg1_60 MaxRef1 CVVeg2 ModeVeg2 P_Veg2_10 P_Veg2_40	0.55	0.53	6.90	57.85
		* P_Veg1_60 CVVeg2 ModeVeg2 P_Veg2_10	0.53	0.52	7.13	58.33
		P_Veg1_60 CVVeg2 ModeVeg2	0.53	0.52	7.21	58.55
		P_Veg1_60 MaxRef1 CVVeg2 ModeVeg2 P_Veg2_10	0.54	0.52	7.30	58.16
		P_Veg1_60 CVVeg2 ModeVeg2 P_Veg2_10 P_Veg2_40 Grnd1ratio	0.55	0.53	7.45	57.97
		P_Veg1_60 ModeVeg2 P_Veg2_10	0.52	0.51	7.91	58.70
		P_Veg1_60 MaxRef1 CVVeg2 ModeVeg2 P_Veg2_10 P_Veg2_40 Grnd1ratio	0.55	0.53	8.00	57.88
		P_Veg1_60 CVVeg2	0.52	0.51	8.13	58.95
		P_Veg1_60 MaxRef1 ModeVeg2 P_Veg2_10	0.53	0.52	8.15	58.55
		P_Veg1_60 MaxRef1 CVVeg2 ModeVeg2	0.53	0.51	8.49	58.62
		P_Veg1_60 CVVeg2 ModeVeg2 Grnd1ratio	0.53	0.51	8.50	58.62
	C	* MeanVeg1 StdVeg1 MedianRef1 P_Veg2_20 RangeRef2 Vegratio Canopy40P	0.65	0.62	8.82	46.75
		MeanVeg1 StdVeg1 MedianRef1 P_Veg2_20 RangeRef2 StdRef2 Vegratio Canopy40P	0.66	0.62	9.00	46.48
		MeanVeg1 StdVeg1 MedianRef1 RangeRef2 StdRef2 Vegratio Canopy40P	0.64	0.61	10.12	47.17
		MeanVeg1 MedianRef1 RangeRef2 StdRef2 Vegratio Canopy40P	0.63	0.60	10.03	47.76
		MeanVeg1 StdVeg1 RangeRef2 StdRef2 Vegratio Canopy40P	0.63	0.60	11.05	47.77
		MeanVeg1 StdVeg1 MedianRef1 P_Veg2_20 RangeRef2 Vegratio	0.63	0.60	11.07	47.78
		MeanVeg1 StdVeg1 P_Veg2_20 RangeRef2 StdRef2 Vegratio Canopy40P	0.64	0.60	11.08	47.48
		MeanVeg1 StdVeg1 MedianRef1 StdRef2 Vegratio Canopy40P	0.63	0.60	11.44	47.89
		MeanVeg1 StdVeg1 P_Veg2_20 RangeRef2 Vegratio	0.62	0.59	11.47	48.19
		MeanVeg1 MedianRef1 StdRef2 Vegratio Canopy40P	0.62	0.59	11.52	48.21
		MeanVeg1 StdVeg1 MedianRef1 P_Veg2_20 Vegratio Canopy40P	0.63	0.60	11.52	47.92

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.3.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume		CVVeg1 MinVeg1 MedianVeg1 MaxRef1 RangeRef1 MedianRef1 ModeVeg2 P_Veg2_10	0.59	0.57	10.00	54.95
		Canopy30P	0.58	0.56	10.42	55.26
		CVVeg1 MedianVeg1 MaxRef1 MedianRef1 ModeVeg2 P_Veg2_10 Canopy30P	0.58	0.56	10.65	55.16
		CVVeg1 MinVeg1 MedianVeg1 MaxRef1 MedianRef1 ModeVeg2 P_Veg2_10 Canopy30P	0.57	0.56	10.81	55.44
		* CVVeg1 MedianVeg1 MaxRef1 MedianRef1 ModeVeg2 P_Veg2_10	0.58	0.56	10.81	55.18
		CVVeg1 MedianVeg1 MaxRef1 RangeRef1 MedianRef1 ModeVeg2 P_Veg2_10 Canopy30P	0.58	0.56	11.14	55.35
	A	CVVeg1 MinVeg1 MedianVeg1 MedianRef1 ModeVeg2 P_Veg2_10 Canopy30P	0.57	0.56	11.40	55.39
		CVVeg1 MinVeg1 MedianVeg1 MaxRef1 MedianRef1 ModeVeg2 P_Veg2_10	0.58	0.56	11.46	55.27
		CVVeg1 MinVeg1 MedianVeg1 MaxRef1 RangeRef1 MedianRef1 ModeVeg2 P_Veg2_10	0.57	0.56	11.62	55.42
		CVVeg1 MedianVeg1 MaxRef1 RangeRef1 MedianRef1 ModeVeg2 P_Veg2_10	0.57	0.56	11.95	55.58
		CVVeg1 MinVeg1 MedianVeg1 MedianRef1 ModeVeg2 P_Veg2_10	0.57	0.56	12.07	55.60
		CVVeg1 MedianVeg1 MaxRef1 RangeRef1 MedianRef1 Canopy30P	0.57	0.56	12.18	55.49
		CVVeg1 MedianVeg1 MaxRef1 RangeRef1 MedianRef1 ModeVeg2 Canopy30P				
			0.49	0.48	3.09	42.32
		* P_Veg1_60 MedianRef2	0.49	0.48	4.00	42.31
	D	P_Veg1_60 MeanRef1 MedianRef2	0.47	0.47	4.09	42.63
		P_Veg1_60	0.48	0.47	5.54	42.70
		P_Veg1_60 MeanRef1				
		P_Veg1_40 MaxRef1 ModeVeg2 RangeRef2 Grnd1ratio Canopy30P Canopy90P	0.62	0.58	7.94	17.31
		* P_Veg1_40 MaxRef1 ModeVeg2 RangeRef2 Grnd1ratio Canopy90P	0.60	0.57	8.02	17.44
		P_Veg1_40 MaxRef1 RangeRef2 Grnd1ratio Canopy30P Canopy90P	0.60	0.57	8.50	17.50
		StdMeanVeg1 P_Veg1_40 MaxRef1 ModeVeg2RangeRef2 Grnd1ratio Canopy30P	0.62	0.58	9.00	17.32
Biomass		Canopy90P				
	C	P_Veg1_40 MaxRef1 RangeRef2 Grnd1ratio Canopy90P	0.59	0.56	9.18	17.69
		StdMeanVeg1 P_Veg1_40 MaxRef1 ModeVeg2 RangeRef2 Grnd1ratio Canopy90P	0.61	0.57	9.20	17.46
		StdMeanVeg1 P_Veg1_40 MaxRef1 ModeVeg2 RangeRef2 Grnd1ratio	0.60	0.56	9.60	17.63
		StdMeanVeg1 P_Veg1_40 MaxRef1 RangeRef2 Grnd1ratio Canopy30P Canopy90P	0.60	0.57	9.97	17.56
		StdMeanVeg1 P_Veg1_40 MaxRef1 ModeVeg2 Grnd1ratio	0.58	0.55	10.13	17.80
		StdMeanVeg1 P_Veg1_40 ModeVeg2 Grnd1ratio	0.57	0.55	10.14	17.91
		* P_Veg2_80 Grnd2ratio	0.59	0.58	4.49	38.99
	A	CVVeg2 ModeVeg2 P_Veg2_80 Grnd2ratio	0.59	0.58	5.00	38.94
		CVVeg2 P_Veg2_80	0.58	0.57	9.00	39.48
		CVVeg2 ModeVeg2 P_Veg2_80	0.58	0.57	9.18	39.41

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.3.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 6,687 segments (0.141 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.3.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_70 MedianRef1 ModeVeg2 StdMeanVeg2 MinRef2	0.60	0.58	6.00	57.71
		* P_Veg1_70 ModeVeg2 StdMeanVeg2 MinRef2	0.59	0.58	6.22	58.04
		P_Veg1_70 MedianRef1 StdMeanVeg2 MinRef2	0.59	0.57	7.02	58.25
		P_Veg1_70 StdMeanVeg2 MinRef2	0.58	0.57	7.41	58.61
		P_Veg1_70 MedianRef1 StdMeanVeg2	0.58	0.57	7.65	58.68
		P_Veg1_70 StdMeanVeg2	0.57	0.56	7.98	59.02
		P_Veg1_70 MedianRef1 ModeVeg2 StdMeanVeg2	0.59	0.57	8.32	58.60
		P_Veg1_70 ModeVeg2 StdMeanVeg2	0.58	0.57	8.53	58.91
	C	P_Veg1_20 MedianRef1 ModeVeg2 Grnd1ratio Canopy90P	0.61	0.57	5.73	43.65
		* P_Veg1_20 MedianRef1 ModeVeg2 Grnd1ratio	0.59	0.56	5.75	44.09
		P_Veg1_20 ModeVeg2 Grnd1ratio Canopy90P	0.59	0.55	6.38	44.35
		P_Veg1_20 ModeVeg2 Grnd1ratio	0.60	0.54	6.86	44.95
		StdVeg1 P_Veg1_20 MedianRef1 ModeVeg2 Grnd1ratio Canopy90P	0.61	0.57	7.00	43.77
		StdVeg1 P_Veg1_20 MedianRef1 ModeVeg2 Grnd1ratio	0.59	0.55	7.75	44.52
		StdVeg1 P_Veg1_20 ModeVeg2 Grnd1ratio Canopy90P	0.59	0.55	7.85	44.57
		P_Veg1_20 Grnd1ratio	0.54	0.52	8.35	45.92
		P_Veg1_20 MedianRef1 Grnd1ratio	0.55	0.53	8.45	45.60
		StdVeg1 P_Veg1_20 ModeVeg2 Grnd1ratio	0.57	0.53	8.83	45.38
	M	* P_Veg1_10 MaxRef1 MinRef1 MedianRef1 P_Veg2_40 MaxRef2 Canopy40P	0.67	0.62	8.00	44.91
		P_Veg1_10 MinRef1 MedianRef1 P_Veg2_40 MaxRef2 Canopy40P	0.65	0.60	9.16	46.00
		P_Veg1_10 MaxRef1 MinRef1 MedianRef1 MaxRef2 Canopy40P	0.64	0.59	10.06	46.45
		P_Veg1_10 MedianRef1 P_Veg2_40 MaxRef2 Canopy40P	0.62	0.58	10.08	46.90
		P_Veg1_10 MaxRef1 MedianRef1 P_Veg2_40 MaxRef2 Canopy40P	0.64	0.59	10.48	46.65
		P_Veg1_10 MaxRef1 MinRef1 MedianRef1 P_Veg2_40 MaxRef2	0.63	0.58	10.80	46.81
		P_Veg1_10 MinRef1 MedianRef1 P_Veg2_40 MaxRef2	0.61	0.56	12.49	48.04
		P_Veg1_10 MaxRef1 MedianRef1 P_Veg2_40 MaxRef2	0.60	0.56	12.73	48.16
		P_Veg1_10 MedianRef1 P_Veg2_40 MaxRef2	0.59	0.55	12.80	48.57
Biomass	D	StdMeanVeg1 P_Veg1_80 ModeVeg2 P_Veg2_10	0.55	0.53	5.00	42864
		* StdMeanVeg1 P_Veg1_80	0.53	0.52	5.08	43272
		StdMeanVeg1 P_Veg1_80 P_Veg2_10	0.53	0.52	6.20	43298
		StdMeanVeg1 P_Veg1_80 ModeVeg2	0.53	0.52	6.49	43356
		P_Veg1_80 ModeVeg2 P_Veg2_10	0.53	0.52	6.83	43421
		P_Veg1_80	0.51	0.51	6.92	43812

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.3.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	C	P_Veg1_20 ModeVeg2 P_Veg2_80 StdRef2 Vegratio	0.63	0.59	6.46	12.71
		* P_Veg1_20 ModeVeg2 P_Veg2_80 Vegratio	0.62	0.59	6.51	12.84
		P_Veg1_20 ModeVeg2 P_Veg2_80 Vegratio Canopy90P	0.63	0.59	6.83	12.76
		P_Veg1_20 ModeVeg2 P_Veg2_80 StdRef2 Vegratio Canopy90P	0.64	0.60	7.00	12.65
		P_Veg1_20 ModeVeg2 StdRef2 Vegratio Canopy90P	0.63	0.59	7.02	12.78
		P_Veg1_20 ModeVeg2 Vegratio Canopy90P	0.61	0.58	7.75	12.99
		P_Veg1_20 ModeVeg2 StdRef2 Vegratio	0.60	0.57	8.28	13.06
		P_Veg1_20 P_Veg2_80 StdRef2 Vegratio	0.59	0.56	9.44	13.19
		P_Veg1_20 StdRef2 Vegratio	0.58	0.56	9.45	13.30
		P_Veg1_20 ModeVeg2 Vegratio	0.58	0.55	9.97	13.36
	M	* MaxVeg1 P_Veg1_20 MaxRef1 MaxRef2 StdMeanRef2	0.62	0.58	4.48	25.14
		StdMeanVeg1 P_Veg1_20 MaxRef1 MaxRef2 StdMeanRef2	0.62	0.58	5.06	25.31
		MaxVeg1 StdMeanVeg1 P_Veg1_20 MaxRef1 MaxRef2 StdMeanRef2	0.64	0.59	5.27	25.07
		MaxVeg1 MaxRef1 KurtosisVeg2 MaxRef2 StdMeanRef2	0.61	0.57	5.67	25.48
		MaxVeg1 P_Veg1_20 MaxRef1 KurtosisVeg2 MaxRef2 StdMeanRef2	0.63	0.58	5.77	25.21
		StdMeanVeg1 P_Veg1_20 MaxRef1 MaxRef2 StdMeanRef2 Canopy70P	0.63	0.58	5.92	25.26
		MaxVeg1 MaxRef1 KurtosisVeg2 MaxRef2 StdMeanRef2 Canopy70P	0.63	0.58	6.05	25.30
		MaxVeg1 P_Veg1_20 MaxRef1 MaxRef2 StdMeanRef2 Canopy70P	0.63	0.58	6.29	25.37

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.4.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 2,972 segments (0.318 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.4.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	MeanVeg1 RangeRef1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy20P Canopy90P	0.58	0.56	8.98	56.05
		MeanVeg1 RangeRef1 StdMeanRef1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy20P Canopy90P	0.58	0.56	9.00	55.84
		* MeanVeg1 ModeVeg2 P_Veg2_10 Canopy20P Canopy90P	0.57	0.55	9.13	56.49
		MeanVeg1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy20P Canopy90P	0.57	0.55	9.13	56.29
		MeanVeg1 StdMeanRef1 ModeVeg2 P_Veg2_10 Canopy20P Canopy90P	0.57	0.55	9.86	56.44
		MeanVeg1 RangeRef1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy20P	0.57	0.55	9.86	56.44
		MeanVeg1 StdMeanRef1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy20P Canopy90P	0.57	0.55	10.26	56.32
		MeanVeg1 RangeRef1 StdMeanRef1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy20P	0.57	0.55	10.80	56.43
		MeanVeg1 RangeRef1 ModeVeg2 P_Veg2_10 Canopy20P Canopy90P	0.57	0.55	10.90	56.66
		MeanVeg1 ModeVeg2 P_Veg2_10 MaxRef2 Canopy90P	0.56	0.54	10.97	56.87
	C	P_Veg1_40 MinRef2 StdRef2 Vegratio	0.57	0.54	5.00	50.92
		* P_Veg1_40 StdRef2 Vegratio	0.56	0.54	5.09	51.29
		P_Veg1_40 Vegratio	0.52	0.51	8.36	52.68
		P_Veg1_40 MinRef2 Vegratio	0.53	0.51	9.39	52.72
	A	StdMeanVeg1 P_Veg1_40 ModeVeg2 P_Veg2_10 StdRef2 Vegratio Canopy90P	0.57	0.56	8.48	55.55
		StdMeanVeg1 P_Veg1_40 ModeVeg2 P_Veg2_10 MinRef2 StdRef2 Vegratio	0.57	0.56	8.87	55.60
		* StdMeanVeg1 P_Veg1_40 ModeVeg2 P_Veg2_10 StdRef2 Vegratio	0.57	0.56	8.95	55.74
		StdMeanVeg1 P_Veg1_40 ModeVeg2 P_Veg2_10 MinRef2 StdRef2 Vegratio Canopy90P	0.58	0.56	9.00	55.49
		P_Veg1_40 ModeVeg2 P_Veg2_10 StdRef2 Vegratio	0.56	0.55	9.72	55.97
		StdMeanVeg1 P_Veg1_40 ModeVeg2 P_Veg2_10 StdRef2 Canopy90P	0.56	0.55	10.47	55.94
		P_Veg1_40 ModeVeg2 P_Veg2_10 StdRef2 Vegratio Canopy90P	0.56	0.55	10.60	55.96
		P_Veg1_40 ModeVeg2 P_Veg2_10 MinRef2 StdRef2 Vegratio	0.56	0.55	10.87	55.99
Biomass	D	* StdMeanVeg1 P_Veg1_50 ModeVeg2 P_Veg2_10	0.52	0.51	4.17	41.17
		41585 P_Veg1_50 ModeVeg2 P_Veg2_10	0.51	0.50	5.92	41.59
		StdMeanVeg1 P_Veg1_50 ModeVeg2 P_Veg2_10 Canopy80P	0.52	0.50	6.00	41.29
		P_Veg1_50 ModeVeg2 P_Veg2_10 Canopy80P	0.51	0.49	7.43	41.66

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.4.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	C	* P_Veg1_40 MaxRef1 MinRef2 Vegratio	0.54	0.51	4.61	18.58
		P_Veg1_40 MinRef2 StdRef2 Vegratio	0.53	0.51	5.54	18.69
		P_Veg1_40 MaxRef1 MinRef2 StdRef2 Vegratio	0.54	0.51	5.69	18.59
		P_Veg1_40 StdRef2 Vegratio	0.52	0.50	5.75	18.84
		P_Veg1_40 MaxRef1 ModeVeg2 MinRef2 Vegratio	0.54	0.51	6.10	18.64
		_Veg1_40 ModeVeg2 MinRef2 StdRef2 Vegratio	0.54	0.51	6.33	18.67
		P_Veg1_40 MaxRef1 Vegratio	0.51	0.49	6.56	18.94
		P_Veg1_40 Vegratio	0.50	0.49	6.76	19.08
		P_Veg1_40 ModeVeg2 StdRef2 Vegratio	0.52	0.50	6.89	18.86
	A	MeanVeg1 CVVeg1 MedianRef1 ModeVeg2 P_Veg2_30	0.61	0.60	5.02	38.06
		* MeanVeg1 MedianRef1 ModeVeg2 P_Veg2_30	0.60	0.60	6.99	38.33
		MeanVeg1 CVVeg1 MedianRef1 ModeVeg2 P_Veg2_30 Canopy80P	0.61	0.60	7.00	38.15
		MeanVeg1 MedianRef1 ModeVeg2 P_Veg2_30 Canopy80P	0.61	0.60	8.85	38.40

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.4.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 2,972 segments (0.318 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M)*. Selected models are shown with an *

Table G.4.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	MedianVeg1 RangeRef1 ModeVeg2 MaxRef2	0.56	0.54	3.93	60.27
		MedianVeg1 RangeRef1 MaxRef2	0.55	0.54	4.27	60.64
		MedianVeg1 ModeVeg2	0.54	0.53	4.31	60.93
		* MedianVeg1	0.53	0.53	4.34	61.21
		MedianVeg1 ModeVeg2 P_Veg2_25	0.55	0.54	4.51	60.71
		60.79840 MedianVeg1 RangeRef1 ModeVeg2	0.55	0.54	4.82	60.80
		MedianVeg1 ModeVeg2 P_Veg2_25 MaxRef2	0.56	0.54	4.84	60.52
		MedianVeg1 RangeRef1 ModeVeg2 P_Veg2_25 MaxRef2	0.57	0.55	4.89	60.25
		MedianVeg1 MaxRef2	0.54	0.53	5.56	61.27
		MinVeg1 MedianVeg1 RangeRef1 ModeVeg2 MaxRef2	0.56	0.54	5.58	60.45
		MedianVeg1 RangeRef1	0.54	0.53	5.62	61.29
		MedianVeg1 MeanRef1 RangeRef1 ModeVeg2 MaxRef2	0.56	0.54	5.88	60.53
		MinVeg1 MedianVeg1 RangeRef1 MaxRef2	0.55	0.54	5.89	60.82
		MedianVeg1 ModeVeg2 MaxRef2	0.54	0.53	5.95	61.11
	C	* MeanVeg1 StdVeg2 StdRef2 Vegratio Canopy80P	0.64	0.61	5.07	41.53
		MeanVeg1 StdRef2 Vegratio Canopy80P	0.62	0.59	5.86	42.27
		MeanVeg1 MaxRef1 StdVeg2 StdRef2 Vegratio Canopy80P	0.65	0.60	7.00	41.92
		MeanVeg1 MaxRef1 StdRef2 Vegratio Canopy80P	0.62	0.59	7.85	42.69
		MeanVeg1 StdRef2 Vegratio	0.59	0.56	8.89	43.85
		MeanVeg1 StdVeg2 StdRef2 Vegratio	0.60	0.57	9.06	43.55
	M	* P_Veg1_10 StdMeanRef1 MinRef2	0.47	0.43	4.91	54.63
		P_Veg1_10 MinRef2	0.44	0.42	5.44	55.47
		P_Veg1_10 MinRef1 StdMeanRef1 MinRef2	0.48	0.44	5.69	54.52
		P_Veg1_10 MinRef1 StdMeanRef1 MaxRef2 MinRef2	0.50	0.44	6.00	54.11
		P_Veg1_10 StdMeanRef1 MaxRef2 MinRef2	0.47	0.43	6.44	54.95
		P_Veg1_10 MaxRef2 MinRef2	0.45	0.41	6.87	55.74
		P_Veg1_10 MinRef1 MinRef2	0.44	0.40	7.39	56.03
		P_Veg1_10	0.39	0.37	8.26	57.46
Biomass	D	* P_Veg1_60 ModeVeg2 P_Veg2_10	0.55	0.53	2.77	42.73
		P_Veg1_60 ModeVeg2 P_Veg2_10 MaxRef2	0.55	0.53	4.01	42.78
		P_Veg1_60 ModeVeg2 P_Veg2_10 Canopy80P	0.55	0.53	4.77	42.93
		P_Veg1_60 ModeVeg2 P_Veg2_10 MaxRef2 Canopy80P	0.55	0.53	6.00	42.98
		P_Veg1_60 ModeVeg2	0.52	0.51	6.06	43.58
		P_Veg1_60	0.51	0.51	6.90	43.93

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.4.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	C	P_Veg1_30 StdMeanVeg2 StdRef2 Vegratio Canopy80P	0.63	0.59	6.76	12.76
		* P_Veg1_30 StdMeanVeg2 StdRef2 Vegratio	0.61	0.58	6.99	12.91
		P_Veg1_30 MinRef1 StdMeanVeg2 StdRef2 Vegratio Canopy80P	0.64	0.60	7.00	12.66
		P_Veg1_30 MinRef1 StdMeanVeg2 StdRef2 Vegratio	0.62	0.59	7.32	12.83
		P_Veg1_30 StdRef2 Vegratio Canopy80P	0.61	0.58	7.58	12.98
		P_Veg1_30 MinRef1 StdMeanVeg2 Vegratio	0.60	0.57	8.66	13.11
		P_Veg1_30 StdRef2 Vegratio	0.58	0.56	9.06	13.26
		P_Veg1_30 MinRef1 StdMeanVeg2 Vegratio Canopy80P	0.61	0.57	9.07	13.04
		P_Veg1_30 MinRef1 StdRef2 Vegratio Canopy80P	0.61	0.57	9.30	13.07
		P_Veg1_30 StdMeanVeg2 Vegratio	0.58	0.56	10.24	13.37
	M	* P_Veg2_60 Canopy70P Canopy90P	0.49	0.46	4.14	28.65
		P_Veg2_60 MinRef2 Canopy70P Canopy90P	0.51	0.47	4.41	28.42
		MedianRef1 P_Veg2_60 Canopy70P Canopy90P	0.50	0.46	4.86	28.56
		MedianRef1 P_Veg2_60 MinRef2 Canopy70P Canopy90P	0.52	0.47	5.02	28.29
		P_Veg1_20 P_Veg2_60 Canopy70P Canopy90P	0.50	0.45	5.66	28.81
		P_Veg1_20 P_Veg2_60 MinRef2 Canopy70P Canopy90P	0.52	0.46	5.90	28.59
		P_Veg2_60 Grnd1ratio Canopy70P Canopy90P	0.49	0.45	6.00	28.91
		P_Veg1_20 MedianRef1 P_Veg2_60 Canopy70P Canopy90P	0.51	0.46	6.06	28.63
		P_Veg1_20 MedianRef1 P_Veg2_60 MinRef2 Canopy70P Canopy90P	0.53	0.47	6.17	28.34
		P_Veg1_20 MedianRef1 MinRef2 Canopy70P Canopy90P	0.51	0.46	6.19	28.67
		P_Veg1_20 MedianRef1 Canopy70P Canopy90P	0.49	0.44	6.34	29.02
		P_Veg2_60 MinRef2 Grnd1ratio Canopy70P Canopy90P	0.51	0.46	6.39	28.73
		P_Veg1_20 MinRef2 Canopy70P Canopy90P	0.49	0.44	6.62	29.10
		P_Veg1_20 Canopy70P Canopy90P	0.46	0.43	6.63	29.40

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.5.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 1,473 segments (0.642 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.5.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_40 MaxRef1 StdMeanRef1 RangeVeg2 P_Veg2_10 MaxRef2 Canopy90P	0.55	0.53	8.00	57.75
		* P_Veg1_40 MaxRef1 RangeVeg2 P_Veg2_10 MaxRef2	0.54	0.52	8.80	58.35
		P_Veg1_40 MaxRef1 StdMeanRef1 P_Veg2_10 MaxRef2 Canopy90P	0.54	0.52	8.92	58.17
		P_Veg1_40 MaxRef1 RangeVeg2 P_Veg2_10 MaxRef2 Canopy90P	0.54	0.52	9.02	58.19
		P_Veg1_40 RangeVeg2 P_Veg2_10	0.52	0.51	9.45	58.90
		P_Veg1_40 MaxRef1 StdMeanRef1 RangeVeg2 P_Veg2_10 MaxRef2	0.54	0.52	9.51	58.30
		P_Veg1_40 RangeVeg2 P_Veg2_10 MaxRef2	0.53	0.51	9.84	58.78
		P_Veg1_40 MaxRef1 StdMeanRef1 MaxRef2 Canopy90P	0.53	0.52	9.87	58.58
	C	* P_Veg1_40 MedianRef1 ZeroNgrnd1ratio	0.53	0.51	3.11	52.84
		P_Veg1_40 MedianRef1 MinRef2 ZeroNgrnd1ratio	0.53	0.51	4.55	52.99
		P_Veg1_40 MedianRef1 MeanRef2 ZeroNgrnd1ratio	0.53	0.51	4.76	53.07
		P_Veg1_40 MedianRef1 MeanRef2 MinRef2 ZeroNgrnd1ratio	0.54	0.50	6.00	53.15
		P_Veg1_40 MedianRef1	0.50	0.48	6.29	54.29
		P_Veg1_40 ZeroNgrnd1ratio	0.50	0.48	6.32	54.30
		P_Veg1_40 MeanRef2 ZeroNgrnd1ratio	0.51	0.49	6.53	54.04
	A	StdMeanVeg1 P_Veg1_40 P_Veg2_10 MinRef2 StdRef2 ZeroNgrnd1ratio	0.56	0.54	7.82	56.44
		StdMeanVeg1 P_Veg1_40 P_Veg2_10 MinRef2 StdRef2 ZeroNgrnd1ratio Canopy90P	0.56	0.55	8.00	56.33
		StdMeanVeg1 P_Veg1_40 P_Veg2_10 StdRef2 ZeroNgrnd1ratio Canopy90P	0.56	0.54	8.31	56.51
		* StdMeanVeg1 P_Veg1_40 P_Veg2_10 StdRef2 ZeroNgrnd1ratio	0.55	0.54	9.04	56.73
		P_Veg1_40 P_Veg2_10 StdRef2 ZeroNgrnd1ratio	0.54	0.53	10.30	57.03
Biomass	D	P_Veg1_50 RangeVeg2 P_Veg2_10 ZeroNgrnd1ratio	0.50	0.48	5.00	42.20
		* P_Veg1_50 P_Veg2_10 ZeroNgrnd1ratio	0.49	0.48	5.06	42.36
		P_Veg1_50 RangeVeg2 P_Veg2_10	0.48	0.47	6.84	42.63
		P_Veg1_50 ZeroNgrnd1ratio	0.47	0.47	6.86	42.79
		P_Veg1_50 P_Veg2_10	0.47	0.46	8.36	43.01
	C	* P_Veg1_40 Vegratio	0.47	0.46	3.68	19.62
		P_Veg1_40 MinRef2 Vegratio	0.48	0.46	4.00	19.53
		P_Veg1_40	0.46	0.45	4.05	19.79
		P_Veg1_40 MinRef2	0.47	0.45	4.47	19.72

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.5.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass		* CVVeg1 MedianVeg1 MedianRef1 MinVeg2	0.59	0.58	5.78	39.01
		CVVeg1 MedianVeg1 MedianRef1 MinVeg2 P_Veg2_30	0.59	0.58	6.00	38.94
		CVVeg1 MedianVeg1 MedianRef1 P_Veg2_30	0.59	0.58	6.39	39.07
	A	MedianVeg1 MedianRef1 P_Veg2_30	0.58	0.58	8.08	39.31
		MedianVeg1 MedianRef1 MinVeg2 P_Veg2_30	0.59	0.58	8.52	39.26
		MedianVeg1 MedianRef1 MinVeg2	0.58	0.58	8.64	39.36
		CVVeg1 MedianVeg1 MedianRef1	0.58	0.57	9.43	39.43

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.5.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 1,473 segments (0.642 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.5.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* MedianVeg1 RangeRef1 RangeVeg2 MaxRef2	0.54	0.52	5.41	61.72
		CVVeg1 MedianVeg1 RangeRef1 RangeVeg2 MaxRef2	0.55	0.52	6.00	61.61
		CVVeg1 MedianVeg1 RangeRef1 RangeVeg2	0.54	0.52	6.26	61.97
		MedianVeg1 RangeRef1 RangeVeg2	0.53	0.51	6.59	62.34
		MedianVeg1 RangeVeg2	0.52	0.51	6.78	62.67
		CVVeg1 MedianVeg1 RangeVeg2	0.52	0.51	7.01	62.46
		CVVeg1 MedianVeg1	0.52	0.51	7.13	62.76
	C	* P_Veg1_30 StdMeanRef2 StdRef2 MedianRef2 ZeroNgrnd1ratio Canopy70P	0.70	0.67	7.00	38.24
		P_Veg1_30 StdRef2 MedianRef2 ZeroNgrnd1ratio Canopy70P	0.68	0.65	9.51	39.56
		P_Veg1_30 StdMeanRef2 StdRef2 ZeroNgrnd1ratio Canopy70P	0.66	0.63	12.03	40.48
		P_Veg1_30 StdRef2 ZeroNgrnd1ratio Canopy70P	0.65	0.62	12.54	40.97
	M	* MinRef1 StdMeanRef1 MaxRef2 MinRef2 Canopy60P	0.67	0.63	4.56	44.22
		MaxRef1 MinRef1 StdMeanRef1 MaxRef2 MinRef2 Canopy60P	0.67	0.62	6.12	44.50
		MinRef1 StdMeanRef1 MedianRef1 MaxRef2 MinRef2 Canopy60P	0.67	0.62	6.39	44.64
		MinRef1 StdMeanRef1 MaxRef2 Canopy60P	0.63	0.60	7.19	46.01
		MaxRef1 MinRef1 StdMeanRef1 MaxRef2 Canopy60P	0.64	0.60	7.83	45.85
Biomass	D	* P_Veg1_60 CVVeg2	0.51	0.50	2.88	44.21
		P_Veg1_60 CVVeg2 Canopy80P	0.51	0.50	4.00	44.23
		P_Veg1_60	0.48	0.48	6.35	45.10
		P_Veg1_60 Canopy80P	0.49	0.48	8.00	45.24
	C	* P_Veg1_25 StdMeanVeg2 CVRef2 StdRef2 ZeroNgrnd1ratio Canopy70P	0.66	0.62	7.00	12.35
		P_Veg1_25 StdMeanVeg2 CVRef2 StdRef2 ZeroNgrnd1ratio	0.63	0.59	9.83	12.81
		P_Veg1_25 StdMeanVeg2 StdRef2 ZeroNgrnd1ratio Canopy70P	0.62	0.59	9.88	12.82
		P_Veg1_25 StdMeanVeg2 StdRef2 ZeroNgrnd1ratio	0.60	0.57	11.25	13.08
		P_Veg1_25 StdMeanVeg2 ZeroNgrnd1ratio Canopy70P	0.60	0.57	11.47	13.12
		P_Veg1_25 CVRef2 StdRef2 ZeroNgrnd1ratio Canopy70P	0.61	0.57	11.92	13.06
	M	* ZeroNVeg3_5ratio Canopy60P Canopy90P	0.53	0.50	5.90	27.63
		P_Veg2_40 Canopy60P Canopy90P	0.53	0.49	6.04	27.67
		ZeroNVeg3_5ratio Canopy60P	0.50	0.48	6.34	28.01
		KurtosisVeg1 ZeroNVeg3_5ratio Canopy60P Canopy90P	0.54	0.50	6.45	27.51
		KurtosisVeg1 P_Veg2_40 Canopy60P Canopy90P	0.54	0.50	6.81	27.62
		Canopy60P	0.48	0.47	6.82	28.39
		KurtosisVeg1 ZeroNVeg3_5ratio Canopy60P	0.52	0.49	6.90	27.91

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.6.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 981 segments (0.964 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.6.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	MinVeg1 StdMeanVeg1 P_Veg1_10 P_Veg1_40 MinRef2 ZeroNgrnd1ratio	0.56	0.54	6.89	57.07
		MinVeg1 StdMeanVeg1 P_Veg1_10 P_Veg1_40 MaxRef1 ZeroNgrnd1ratio	0.56	0.54	6.92	57.08
		* StdMeanVeg1 P_Veg1_10 P_Veg1_40 MaxRef1 ZeroNgrnd1ratio	0.56	0.54	7.32	57.39
		MinVeg1 StdMeanVeg1 P_Veg1_10 P_Veg1_40 MaxRef1 MinRef2 ZeroNgrnd1ratio	0.57	0.55	7.46	56.98
		MinVeg1 P_Veg1_10 P_Veg1_40 ZeroNgrnd1ratio	0.55	0.53	7.67	57.68
		MinVeg1 StdMeanVeg1 P_Veg1_10 P_Veg1_40 ZeroNgrnd1ratio	0.55	0.54	7.89	57.51
		StdMeanVeg1 P_Veg1_10 P_Veg1_40 MaxRef1 MinRef2 ZeroNgrnd1ratio	0.56	0.54	8.00	57.32
		MinVeg1 P_Veg1_40 ZeroNgrnd1ratio	0.54	0.53	8.11	57.98
		P_Veg1_40 ZeroNgrnd1ratio	0.53	0.53	8.18	58.20
		P_Veg1_40 MaxRef1 ZeroNgrnd1ratio	0.54	0.53	8.31	58.02
	C	* MeanVeg1 P_Veg1_10 MedianRef1 MaxVeg2 ZeroNgrnd1ratio Canopy30P	0.59	0.55	6.25	50.30
		MeanVeg1 MedianRef1 MaxVeg2 MaxRef2 ZeroNgrnd1ratio Canopy30P	0.59	0.55	7.09	50.62
		MeanVeg1 P_Veg1_10 MedianRef1 MaxVeg2 MaxRef2 ZeroNgrnd1ratio Canopy30P	0.60	0.55	7.25	50.30
		MeanVeg1 P_Veg1_10 MedianRef1 MaxVeg2 P_Veg2_30 ZeroNgrnd1ratio Canopy30P	0.59	0.55	7.87	50.54
		MeanVeg1 MedianRef1 MaxVeg2 ZeroNgrnd1ratio Canopy30P	0.57	0.54	7.97	51.32
		80522 MeanVeg1 MedianRef1 MaxVeg2 P_Veg2_30 MaxRef2 ZeroNgrnd1ratio Canopy30P	0.59	0.55	8.57	50.81
		MeanVeg1 MedianRef1 MaxVeg2 P_Veg2_30 ZeroNgrnd1ratio Canopy30P	0.58	0.54	8.84	51.28
	A	MinVeg1 P_Veg1_40 MinRef1 MedianRef1 ZeroNgrnd1ratio	0.55	0.54	6.26	56.93
		* P_Veg1_40 MedianRef1 ZeroNgrnd1ratio	0.54	0.54	6.61	57.25
		MinVeg1 P_Veg1_40 MinRef1 MedianRef1 MinRef2 ZeroNgrnd1ratio	0.56	0.54	7.02	56.90
		MinVeg1 P_Veg1_40 MedianRef1 ZeroNgrnd1ratio	0.55	0.54	7.05	57.18
		P_Veg1_40 MinRef1 MedianRef1 ZeroNgrnd1ratio	0.55	0.54	7.25	57.20
		P_Veg1_40 MinRef1 MedianRef1 MinRef2 ZeroNgrnd1ratio	0.55	0.54	7.84	57.15
		MinVeg1 StdMeanVeg1 P_Veg1_40 MinRef1 MedianRef1 ZeroNgrnd1ratio	0.55	0.54	7.87	57.01
		StdMeanVeg1 P_Veg1_40 MedianRef1 ZeroNgrnd1ratio	0.54	0.54	7.87	57.29
Biomass	D	StdMeanVeg1 P_Veg1_40 MaxRef1 MinVeg2 ZeroNgrnd1ratio	0.55	0.53	5.81	40.42
		StdMeanVeg1 P_Veg1_40 MinRef2 ZeroNgrnd1ratio	0.54	0.53	6.53	40.68
		* P_Veg1_40 MaxRef1 ZeroNgrnd1ratio	0.53	0.52	6.54	40.83
		StdMeanVeg1 P_Veg1_40 MaxRef1 ZeroNgrnd1ratio	0.54	0.53	6.60	40.69
		StdMeanVeg1 P_Veg1_40 MaxRef1 MinRef2 ZeroNgrnd1ratio	0.55	0.53	6.69	40.55
		P_Veg1_40 MaxRef1 MinVeg2 ZeroNgrnd1ratio	0.54	0.53	6.79	40.72
		StdMeanVeg1 P_Veg1_40 MinVeg2 MinRef2 ZeroNgrnd1ratio	0.55	0.53	6.84	40.58

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.6.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	C	MeanVeg1 P_Veg1_10 MaxVeg2 MinVeg2 ZeroNgrnd1ratio Canopy30P	0.52	0.48	7.00	19.54
		* MeanVeg1 P_Veg1_10 MaxVeg2 MinVeg2 ZeroNgrnd1ratio	0.51	0.47	7.11	19.70
		MeanVeg1 P_Veg1_10 MinVeg2 ZeroNgrnd1ratio	0.49	0.46	7.75	19.93
		MeanVeg1 P_Veg1_10 MaxVeg2 ZeroNgrnd1ratio Canopy30P	0.50	0.46	8.53	19.90
		MeanVeg1 P_Veg1_10 MaxVeg2 ZeroNgrnd1ratio	0.48	0.45	8.61	20.05
		MeanVeg1 P_Veg1_10 ZeroNgrnd1ratio	0.46	0.44	9.54	20.30
		MeanVeg1 P_Veg1_10 MinVeg2 ZeroNgrnd1ratio Canopy30P	0.49	0.45	9.70	20.07
	A	CVVeg1 MinVeg1 StdMeanVeg1 P_Veg1_40 CVRef1 MinVeg2 MinRef2 Canopy70P	0.64	0.62	8.19	37.42
		CVVeg1 MinVeg1 StdMeanVeg1 P_Veg1_40 CVRef1 MinVeg2 MinRef2	0.63	0.62	8.83	37.57
		CVVeg1 MinVeg1 StdMeanVeg1 P_Veg1_40 CVRef1 MinRef2 Canopy70P	0.63	0.62	9.00	37.59
		* CVVeg1 StdMeanVeg1 P_Veg1_40 CVRef1 MinVeg2 MinRef2	0.63	0.62	9.33	37.71
		CVVeg1 StdMeanVeg1 P_Veg1_40 CVRef1 MinVeg2 MinRef2 Canopy70P	0.63	0.62	9.38	37.62

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.6.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 981 segments (0.964 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.6.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* P_Veg1_40 ZeroNgrnd1ratio	0.55	0.54	4.53	60.80
		MinVeg1 P_Veg1_40 ZeroNgrnd1ratio	0.56	0.55	4.58	60.53
		P_Veg1_40 CVVeg2 ZeroNgrnd1ratio	0.56	0.54	5.21	60.71
		MinVeg1 P_Veg1_40 ModeVeg2 ZeroNgrnd1ratio	0.56	0.55	5.49	60.50
		MinVeg1 P_Veg1_40 CVVeg2 ZeroNgrnd1ratio	0.56	0.55	5.58	60.53
		P_Veg1_40 CVVeg2	0.54	0.54	5.70	61.13
		P_Veg1_40 ModeVeg2 ZeroNgrnd1ratio	0.55	0.54	5.77	60.87
		MinVeg1 P_Veg1_40 CVVeg2	0.55	0.54	5.85	60.89
		P_Veg1_40 CVVeg2 ModeVeg2 ZeroNgrnd1ratio	0.56	0.54	5.94	60.64
		MinVeg1 P_Veg1_40 CVVeg2 ModeVeg2 ZeroNgrnd1ratio	0.57	0.55	6.00	60.36
	C	* MeanVeg1 StdVeg2 StdRef2 Vegratio Canopy80P	0.63	0.59	6.00	43.89
		MeanVeg1 StdVeg2 StdRef2 Vegratio	0.60	0.57	7.36	44.97
		MeanVeg1 StdRef2 Vegratio Canopy80P	0.59	0.56	8.25	45.38
		MeanVeg1 StdRef2 Vegratio	0.57	0.54	9.33	46.26
	M	* MeanRef1 P_Veg2_60 Canopy60P	0.51	0.48	3.75	51.96
		MeanRef1 Canopy60P	0.48	0.46	4.93	53.16
		MeanRef1 StdMeanRef1 P_Veg2_60 Canopy60P	0.52	0.48	5.00	52.10
		MeanRef1 StdMeanRef1 Canopy60P	0.49	0.46	6.10	53.28
		P_Veg2_60 Canopy60P	0.46	0.44	7.01	54.28
		StdMeanRef1 P_Veg2_60 Canopy60P	0.46	0.43	8.50	54.59
		Canopy60P	0.42	0.40	8.95	55.75
Biomass	D	* MinVeg1 MedianVeg1 ZeroNgrnd1ratio	0.55	0.54	2.00	42.71
		MinVeg1 MedianVeg1 CVVeg2 ZeroNgrnd1ratio	0.56	0.54	2.60	42.62
		MedianVeg1 ZeroNgrnd1ratio	0.54	0.53	2.98	43.13
		MedianVeg1 CVVeg2 ZeroNgrnd1ratio	0.55	0.53	3.08	42.94
		MinVeg1 StdMeanVeg1 MedianVeg1 CVVeg2	0.55	0.54	3.34	42.78
		MinVeg1 MedianVeg1 CVVeg2	0.55	0.53	3.49	43.02
		MinVeg1 StdMeanVeg1 MedianVeg1 ZeroNgrnd1ratio	0.55	0.54	3.89	42.90
		MinVeg1 MedianVeg1 ZeroNgrnd1ratio Vegratio	0.55	0.53	3.97	42.91
		MinVeg1 MedianVeg1 CVRef1 ZeroNgrnd1ratio	0.55	0.53	3.99	42.92
		StdMeanVeg1 MedianVeg1 CVVeg2	0.54	0.53	3.99	43.13
	C	* P_Veg1_30 P_Veg2_80 StdRef2 Vegratio	0.60	0.56	5.00	13.55
		P_Veg1_30 StdRef2 Vegratio	0.55	0.53	8.06	14.11
		P_Veg1_30 P_Veg2_80 Vegratio	0.53	0.50	11.19	14.53

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.6.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass		* P_Veg2_60 MinRef2 Canopy70P	0.58	0.55	3.02	26.54
		P_Veg2_60 MinRef2 Canopy60P	0.57	0.54	3.72	26.74
		P_Veg2_60 MinRef2 Canopy60P Canopy70P	0.59	0.55	4.10	26.55
		RangeVeg1 MinRef2 Canopy60P	0.57	0.54	4.13	26.86
		RangeVeg1 MinRef2 Canopy60P Canopy70P	0.58	0.54	4.62	26.71
	M	RangeVeg1 P_Veg2_60 MinRef2 Canopy70P	0.58	0.54	5.01	26.83
		RangeVeg1 MinRef2 Canopy70P	0.56	0.53	5.03	27.12
		MinRef2 Canopy60P	0.54	0.52	5.10	27.42
		MinRef2 Canopy60P Canopy70P	0.56	0.53	5.26	27.19
		MinRef2 Canopy70P	0.54	0.52	5.44	27.51
		RangeVeg1 P_Veg2_60 MinRef2 Canopy60P	0.57	0.53	5.56	26.99

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.7.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 749 segments (1.263 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.7.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* P_Veg1_40 MaxRef1 StdMeanRef2	0.54	0.53	6.19	58.28
		MinVeg1 P_Veg1_40 MaxRef1 MinRef1	0.55	0.53	6.21	58.07
		MinVeg1 P_Veg1_40 MaxRef1 StdMeanRef2	0.55	0.53	6.57	58.15
		MinVeg1 P_Veg1_10 P_Veg1_40 MaxRef1 MinRef1	0.55	0.53	6.77	57.98
		P_Veg1_40 MaxRef1 MinRef1	0.54	0.53	6.80	58.42
		MinVeg1 P_Veg1_40 MaxRef1 MinRef1 StdMeanRef2	0.55	0.53	6.87	58.00
	C	* MeanVeg1 P_Veg1_10 MedianRef1 ZeroNgrnd1ratio	0.57	0.55	5.00	51.07
		P_Veg1_10 MedianRef1 ZeroNgrnd1ratio	0.53	0.51	9.70	53.17
		MeanVeg1 MedianRef1 ZeroNgrnd1ratio	0.53	0.51	9.76	53.19
	A	* P_Veg1_40 MedianRef1 MinVeg2 StdMeanRef2 ZeroNgrnd1ratio	0.56	0.55	6.00	56.85
		P_Veg1_40 MedianRef1 StdMeanRef2 ZeroNgrnd1ratio	0.55	0.54	7.06	57.14
		P_Veg1_40 MedianRef1 ZeroNgrnd1ratio	0.54	0.54	8.20	57.44
		P_Veg1_40 MedianRef1 MinVeg2 StdMeanRef2	0.55	0.54	8.38	57.32
		P_Veg1_40 MedianRef1 MinVeg2 ZeroNgrnd1ratio	0.55	0.54	9.18	57.44
Biomass	D	MinVeg1 P_Veg1_40 MaxRef1 MinRef1 CVRef2	0.56	0.5390	5.06	40410
		* P_Veg1_40 MaxRef1 CVRef2	0.54	0.5316	5.10	40730
		P_Veg1_40 MaxRef1 MinRef1 CVRef2	0.55	0.5343	5.36	40615
		MinVeg1 P_Veg1_40 MaxRef1 CVRef2	0.55	0.5313	6.18	40744
		MinVeg1 P_Veg1_40 MaxRef1 MinRef1	0.55	0.5307	6.35	40772
		MinVeg1 P_Veg1_40 MaxRef1 MinRef1 StdRef1CVRef2	0.56	0.5356	7.00	40559
	C	* P_Veg1_10 MinVeg2 Vegratio Canopy50P	0.57	0.56	4.23	18.06
		P_Veg1_10 MedianRef1 MinVeg2 Vegratio Canopy50P	0.60	0.56	5.49	18.10
		P_Veg1_10 MinVeg2 SkewnessVeg2 Vegratio Canopy50P	0.59	0.56	5.64	18.12
		P_Veg1_10 Vegratio Canopy50P	0.56	0.54	6.87	18.55
		P_Veg1_10 MedianRef1 MinVeg2 SkewnessVeg2 Vegratio Canopy50P	0.59	0.56	7.00	18.17

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.7.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass		CVVeg1 P_Veg1_40 MaxRef1 MinVeg2 StdMeanRef2	0.63	0.62	6.88	37.87
		CVVeg1 P_Veg1_40 MeanRef1 MaxRef1 MinVeg2 StdMeanRef2	0.63	0.62	7.00	37.79
		CVVeg1 P_Veg1_40 MeanRef1 MaxRef1 MinVeg2	0.63	0.62	7.60	37.94
		* P_Veg1_40 MaxRef1 MinVeg2 StdMeanRef2	0.62	0.61	7.65	38.03
	A	CVVeg1 P_Veg1_40 MeanRef1 MaxRef1	0.62	0.61	7.85	38.05
		P_Veg1_40 MeanRef1 MaxRef1 MinVeg2 StdMeanRef2	0.62	0.61	8.41	38.01
		CVVeg1 P_Veg1_40 MaxRef1 StdMeanRef2	0.62	0.61	8.47	38.11
		CVVeg1 P_Veg1_40 MeanRef1 MaxRef1 StdMeanRef2	0.62	0.61	8.56	38.03
		P_Veg1_40 MaxRef1 StdMeanRef2	0.61	0.61	9.53	38.30

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.7.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 749 segments (1.263 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.7.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* CVVeg1 P_Veg1_40 MaxRef1	0.57	0.55	7.10	60.37
		CVVeg1 P_Veg1_40 MaxRef1 StdMeanRef2	0.57	0.55	7.85	60.31
		CVVeg1 P_Veg1_10 P_Veg1_40 MaxRef1	0.57	0.55	7.86	60.31
		CVVeg1 MinVeg1 P_Veg1_10 P_Veg1_40 MaxRef1	0.58	0.56	7.88	60.03
		MinVeg1 P_Veg1_40 MaxRef1 StdMeanRef2	0.57	0.55	7.90	60.32
		P_Veg1_40 MaxRef1 StdMeanRef2	0.56	0.55	7.93	60.61
	C	* MeanVeg1 StdVeg1 MinVeg2 ZeroNgrnd1ratio	0.62	0.59	5.00	44.46
		MeanVeg1 StdVeg1 ZeroNgrnd1ratio	0.60	0.57	6.21	45.51
	M	MedianRef1 ModeVeg2 StdMeanVeg2 Canopy60P	0.59	0.55	3.29	48.03
		* MedianRef1 ModeVeg2 Canopy60P	0.57	0.54	3.56	48.74
		MedianRef1 ModeVeg2 MedianVeg2 Canopy60P	0.58	0.55	3.58	48.19
		MedianRef1 ModeVeg2 StdMeanVeg2 MedianVeg2 Canopy60P	0.60	0.55	4.30	48.01
		MedianRef1 ModeVeg2 StdMeanVeg2 MinRef2 Canopy60P	0.59	0.54	5.03	48.43
		MedianRef1 ModeVeg2 MedianVeg2 CVRef2 Canopy60P	0.59	0.54	5.20	48.52
		MedianRef1 ModeVeg2 StdMeanVeg2 CVRef2 Canopy60P	0.59	0.54	5.21	48.53
		MedianRef1 ModeVeg2 MinRef2 Canopy60P	0.57	0.53	5.22	49.09
		MedianRef1 ModeVeg2 MedianVeg2 MinRef2 Canopy60P	0.59	0.54	5.46	48.67
		MedianRef1 ModeVeg2 CVRef2 Canopy60P	0.57	0.53	5.49	49.24
Biomass	D	* CVVeg1 P_Veg1_40 MaxRef1	0.56	0.55	4.91	42.36
		CVVeg1 MinVeg1 P_Veg1_40 MaxRef1	0.57	0.56	5.00	42.18
		MinVeg1 P_Veg1_40 MaxRef1	0.56	0.54	7.06	42.80
		P_Veg1_40 MaxRef1	0.55	0.54	7.35	43.06
		CVVeg1 MinVeg1 P_Veg1_40	0.55	0.54	7.37	42.87
		MinVeg1 P_Veg1_40	0.54	0.53	8.86	43.36

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.7.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	C	* P_Veg1_20 MinVeg2 StdRef2 Vegratio	0.61	0.57	4.598	13.58
		P_Veg1_20 MinVeg2 StdMeanVeg2 StdRef2 Vegratio	0.62	0.57	5.61	13.58
		P_Veg1_20 StdMeanVeg2 StdRef2 Vegratio	0.60	0.56	5.83	13.77
		P_Veg1_20 MinVeg2 Vegratio	0.58	0.55	6.05	13.94
		P_Veg1_20 MinRef1 MinVeg2 StdRef2 Vegratio	0.61	0.57	6.09	13.65
		P_Veg1_20 MinRef1 MinVeg2 StdMeanVeg2 Vegratio	0.61	0.57	6.22	13.67
		P_Veg1_20 MinRef1 StdMeanVeg2 Vegratio	0.59	0.56	6.32	13.84
		P_Veg1_20 MinRef1 MinVeg2 StdMeanVeg2 StdRef2 Vegratio	0.63	0.58	6.35	13.54
		P_Veg1_20 MinRef1 StdMeanVeg2 StdRef2 Vegratio	0.61	0.57	6.51	13.72
		P_Veg1_20 CVVeg2 MinVeg2 StdRef2 Vegratio	0.61	0.56	6.56	13.73
		P_Veg1_20 MinVeg2 StdMeanVeg2 Vegratio	0.59	0.55	6.66	13.89
		P_Veg1_20 MinRef1 MinVeg2 Vegratio	0.59	0.55	6.81	13.91
		P_Veg1_20 StdMeanVeg2 Vegratio	0.57	0.54	6.82	14.05
	M	* ModeVeg2 Canopy60P	0.56	0.54	1.81	26.11
		ModeVeg2 Canopy10P Canopy60P	0.57	0.54	3.21	26.22
		ModeVeg2 Canopy60P Canopy90P	0.56	0.53	3.58	26.32
		ModeVeg2 Canopy10P Canopy60P Canopy90P	0.57	0.53	5.00	26.45
		Canopy60P	0.50	0.49	6.21	27.58

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.8.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 502 segments (1.885 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.8.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_30 MinRef1 MedianRef1 MaxRef2	0.58	0.57	3.97	56.36
		* P_Veg1_30 MinRef1 MaxRef2	0.57	0.56	4.10	56.62
		P_Veg1_30 MinRef1 MedianRef1	0.57	0.56	4.66	56.75
		P_Veg1_30 MinRef1	0.56	0.56	4.98	57.05
	C	* P_Veg1_10 MeanRef1	0.47	0.46	3.00	52.34
		MeanRef1	0.38	0.37	12.82	56.47
	A	* P_Veg1_40 MinRef1 MedianRef1 ZeroNgrnd1ratio	0.58	0.57	4.15	54.79
		P_Veg1_40 MinRef1 MedianRef1 MeanRef2 ZeroNgrnd1ratio	0.58	0.57	6.00	54.91
		P_Veg1_40 MinRef1 MedianRef1 MeanRef2	0.57	0.56	9.90	55.62
Biomass	D	RangeVeg1 P_Veg1_40 MinRef1 StdMeanVeg2 ZeroNgrnd1ratio	0.59	0.57	6.84	39.43
		* P_Veg1_40 MinRef1 StdMeanVeg2 ZeroNgrnd1ratio	0.58	0.56	6.99	39.61
		RangeVeg1 P_Veg1_40 MinRef1 ModeVeg2 StdMeanVeg2 ZeroNgrnd1ratio	0.59	0.57	7.00	39.29
		P_Veg1_40 MinRef1 ModeVeg2 StdMeanVeg2 ZeroNgrnd1ratio	0.58	0.57	7.14	39.47
		RangeVeg1 P_Veg1_40 MinRef1 ModeVeg2 ZeroNgrnd1ratio	0.58	0.56	7.89	39.59
		RangeVeg1 P_Veg1_40 MinRef1 ZeroNgrnd1ratio	0.57	0.56	8.19	39.80
		P_Veg1_40 MinRef1 ModeVeg2 ZeroNgrnd1ratio	0.57	0.56	8.26	39.81
		P_Veg1_40 MinRef1 ZeroNgrnd1ratio	0.57	0.56	8.61	40.02
	C	* MedianRef1 MedianVeg2 ZeroNVeg2ratio Canopy30P Canopy50P Canopy70P	0.61	0.57	6.73	15.64
		MedianRef1 SkewnessVeg2 MedianVeg2 ZeroNVeg2ratio Canopy30P Canopy50P Canopy70P	0.62	0.57	7.00	15.54
		MedianRef1 SkewnessVeg2 MedianVeg2 ZeroNVeg2ratio Canopy30P Canopy50P	0.60	0.56	7.77	15.77
		MedianRef1 MedianVeg2 ZeroNVeg2ratio Canopy30P Canopy50P	0.59	0.55	7.97	15.92
		P_Veg1_40 MedianRef1 MedianVeg2 ZeroNVeg2ratio Canopy30P Canopy50P Canopy70P	0.61	0.56	8.71	15.76
	A	CVVeg1 P_Veg1_40 MeanRef1 MinRef1 MinVeg2 MaxRef2	0.64	0.63	8.30	37.32
		CVVeg1 P_Veg1_40 MeanRef1 MinRef1 MinVeg2 MaxRef2 MedianRef2	0.65	0.63	8.40	37.23
		* CVVeg1 P_Veg1_40 MeanRef1 MinRef1 MaxRef2	0.64	0.63	8.73	37.46
		CVVeg1 P_Veg1_40 MeanRef1 MinRef1 MaxRef2 MedianRef2	0.64	0.63	8.85	37.37
		CVVeg1 P_Veg1_40 MeanRef1 MinRef1 ModeVeg2 MaxRef2	0.64	0.63	8.91	37.38
		CVVeg1 P_Veg1_40 MeanRef1 MinRef1 MinVeg2 ModeVeg2 MaxRef2	0.65	0.63	8.94	37.28
		CVVeg1 P_Veg1_40 MeanRef1 MinRef1 ModeVeg2 MaxRef2 MedianRef2	0.65	0.63	8.98	37.29
		CVVeg1 P_Veg1_40 MeanRef1 MinRef1 MinVeg2 ModeVeg2 MaxRef2 MedianRef2	0.65	0.63	9.00	37.19
		CVVeg1 P_Veg1_40 MinRef1 ModeVeg2 MaxRef2 MedianRef2	0.64	0.63	9.15	37.40
		CVVeg1 P_Veg1_40 MinRef1 MaxRef2 MedianRef2	0.64	0.63	9.20	37.50
		CVVeg1 P_Veg1_40 MinRef1 MinVeg2 MaxRef2 MedianRef2	0.64	0.63	9.23	37.41

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.8.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 502 segments (1.885 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.8.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* P_Veg1_30 MinRef1	0.59	0.59	4.15	58.10
		P_Veg1_30 MinRef1 MedianRef1	0.60	0.59	4.25	57.84
		P_Veg1_30 MinRef1 MaxRef2	0.60	0.59	4.45	57.90
		P_Veg1_30 MinRef1 MedianRef1 MaxRef2	0.61	0.59	5.00	57.76
		P_Veg1_30 MinRef1 Canopy70P	0.60	0.58	5.24	58.13
		P_Veg1_30 MinRef1 MaxRef2 Canopy70P	0.60	0.59	5.36	57.87
		P_Veg1_30 MinRef1 MedianRef1 Canopy70P	0.60	0.59	5.41	57.89
		P_Veg1_30	0.58	0.57	5.84	58.87
	C	* MeanVeg1 StdVeg2 MinRef2 StdRef2 Vegratio Canopy30P Canopy50P Canopy70P	0.72	0.66	9.00	40.13
		MeanVeg1 StdVeg2 StdRef2 Vegratio Canopy30P Canopy50P Canopy70P	0.67	0.61	14.44	43.24
		MeanVeg1 MinRef2 StdRef2 Vegratio Canopy30P Canopy50P Canopy70P	0.67	0.61	14.74	43.38
		MeanVeg1 StdVeg2 MinRef2 StdRef2 Vegratio Canopy30P Canopy70P	0.66	0.60	16.21	44.05
		MeanVeg1 StdRef2 Vegratio Canopy30P Canopy50P Canopy70P	0.64	0.59	16.31	44.45
	M	* MinRef1 MedianRef1 MinRef2 Canopy50P Canopy90P	0.54	0.49	6.32	50.05
		MinRef1 MedianRef1 MinRef2 ZeroNgrnd1ratio Canopy50P Canopy90P	0.56	0.49	7.00	49.85
		MedianRef1 MinRef2 ZeroNgrnd1ratio Canopy50P Canopy90P	0.53	0.47	7.78	50.95
		MinRef1 MedianRef1 Canopy50P Canopy90P	0.50	0.45	8.12	51.71
		MinRef1 MedianRef1 Canopy50P	0.48	0.44	8.25	52.32
		MinRef1 MedianRef1 MinRef2 Canopy50P	0.50	0.45	8.57	51.98
		MinRef1 MedianRef1 ZeroNgrnd1ratio Canopy50P	0.50	0.45	8.62	52.01
		MedianRef1 ZeroNgrnd1ratio Canopy50P Canopy90P	0.49	0.45	8.69	52.05
Biomass	D	P_Veg1_40 MinRef1 StdMeanVeg2 ZeroNgrnd1ratio	0.60	0.58	5.31	41.12
		P_Veg1_40 MinRef1 ModeVeg2 StdMeanVeg2 ZeroNgrnd1ratio	0.61	0.59	5.49	40.94
		P_Veg1_40 MinRef1 ModeVeg2 ZeroNgrnd1ratio	0.60	0.58	5.74	41.21
		* P_Veg1_40 MinRef1 ZeroNgrnd1ratio	0.59	0.58	6.06	41.48
		P_Veg1_40 MinRef1 StdMeanVeg2 ZeroNgrnd1ratio Canopy70P	0.60	0.58	6.74	41.21
		P_Veg1_40 ModeVeg2 ZeroNgrnd1ratio	0.59	0.57	6.74	41.63
		P_Veg1_40 MinRef1 ModeVeg2 ZeroNgrnd1ratio Canopy70P	0.60	0.58	6.74	41.21
		P_Veg1_40 MinRef1 ZeroNgrnd1ratio Canopy70P	0.59	0.58	6.86	41.45

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.8.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	C	* P_Veg1_30 StdRef2 Vegratio	0.57	0.54	4.64	13.96
		P_Veg1_30 P_Veg2_80 StdRef2 Vegratio	0.59	0.55	4.65	13.81
		P_Veg1_30 P_Veg2_30 P_Veg2_80 StdRef2 Vegratio	0.60	0.55	6.00	13.86
		P_Veg1_30 P_Veg2_30 StdRef2 Vegratio	0.58	0.54	6.14	14.05
		P_Veg1_30 P_Veg2_30 P_Veg2_80 Vegratio	0.54	0.50	9.56	14.58
		P_Veg1_30 P_Veg2_80 Vegratio	0.52	0.49	9.99	14.78
	M	* MinRef2 ZeroNgrnd1ratio Canopy60P	0.50	0.46	4.45	28.42
		KurtosisVeg1 MinRef2 ZeroNgrnd1ratio Canopy60P	0.51	0.46	5.52	28.45
		ZeroNgrnd1ratio Canopy60P	0.46	0.43	5.87	29.20
		KurtosisVeg1 StdMeanVeg1 MinRef2 ZeroNgrnd1ratio Canopy60P	0.53	0.47	6.00	28.27
		StdMeanVeg1 MinRef2 ZeroNgrnd1ratio Canopy60P	0.50	0.45	6.25	28.70

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.9.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 374 segments (2.530 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.9.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_50 StdRef1 Canopy90P	0.54	0.53	4.00	56.49
		* P_Veg1_50 Canopy90P	0.53	0.52	4.75	56.92
		P_Veg1_50 StdRef1	0.53	0.52	5.90	57.20
		P_Veg1_50	0.52	0.51	6.16	57.50
	C	P_Veg1_40 MeanRef1 MaxRef1 ModeVeg2 RangeVeg2 StdRef2 Canopy30P Canopy90P	0.59	0.54	9.00	47.42
		* P_Veg1_40 MeanRef1 MaxRef1 RangeVeg2 StdRef2 Canopy30P Canopy90P	0.58	0.53	9.23	47.92
		P_Veg1_40 MeanRef1 RangeVeg2 StdRef2 Canopy30P Canopy90P	0.55	0.50	11.22	49.09
		P_Veg1_40 MeanRef1 RangeVeg2 Canopy30P Canopy90P	0.53	0.49	11.62	49.59
	A	* CVVeg1 P_Veg1_40 MedianRef1 Canopy90P	0.55	0.54	5.04	55.26
		CVVeg1 MinVeg1 P_Veg1_40 MedianRef1 Canopy90P	0.55	0.54	5.83	55.23
		CVVeg1 P_Veg1_40 MedianRef1 MaxVeg2 Canopy90P	0.55	0.54	6.51	55.34
		CVVeg1 P_Veg1_40 MaxRef1 MedianRef1 Canopy90P	0.55	0.54	6.60	55.35
		CVVeg1 MinVeg1 P_Veg1_40 MaxRef1 MedianRef1 Canopy90P	0.55	0.54	6.64	55.20
Biomass	D	MedianVeg1 ModeVeg2 CVRef2 Canopy90P	0.53	0.51	5.71	41.11
		MedianVeg1 MinRef1 ModeVeg2 CVRef2 Canopy90P	0.54	0.52	6.00	40.98
		* MedianVeg1 CVRef2 Canopy90P	0.52	0.51	6.02	41.34
		MedianVeg1 CVRef2	0.51	0.50	6.70	41.64
		MedianVeg1 Canopy90P	0.51	0.50	7.00	41.69
	C	* P_Veg1_40 RangeRef1 StdVeg2 Canopy80P	0.44	0.40	3.48	18.01
		P_Veg1_40 MeanRef1 RangeRef1 StdVeg2 Canopy80P	0.45	0.40	4.25	17.97
		P_Veg1_40 RangeRef1	0.39	0.37	4.71	18.49
		P_Veg1_40 RangeRef1 StdVeg2	0.41	0.38	4.74	18.35
	A	CVVeg1 P_Veg1_50 MeanRef1 P_Veg2_20 Canopy80P	0.63	0.62	5.11	37.18
		* CVVeg1 P_Veg1_50 MeanRef1 Canopy80P	0.63	0.62	6.17	37.40
		CVVeg1 SkewnessVeg1 P_Veg1_50 MeanRef1 P_Veg2_20 Canopy80P	0.63	0.62	7.00	37.27
		CVVeg1 SkewnessVeg1 P_Veg1_50 MeanRef1 Canopy80P	0.63	0.62	7.37	37.42
		P_Veg1_50 MeanRef1 P_Veg2_20 Canopy80P	0.62	0.61	7.75	37.56

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.9.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 374 segments (2.530 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.9.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg1_50 RangeRef1 CVVeg2 MaxRef2 Canopy90P	0.59	0.56	5.72	59.71
		P_Veg1_50 RangeRef1 CVVeg2 MaxRef2	0.58	0.56	5.72	60.07
		* P_Veg1_50 CVVeg2 MaxRef2	0.57	0.55	5.80	60.45
		P_Veg1_50 MaxRef2 MedianRef2	0.57	0.55	5.97	60.50
	C	* KurtosisVeg1 P_Veg1_25 RangeRef1 P_Veg2_70 ZeroNgrnd1ratio Canopy30P	0.67	0.62	6.18	40.45
		KurtosisVeg1 P_Veg1_25 RangeRef1 P_Veg2_70 StdRef2 ZeroNgrnd1ratio Canopy30P	0.68	0.62	7.12	40.40
		KurtosisVeg1 P_Veg1_25 RangeRef1 ZeroNgrnd1ratio Canopy30P	0.64	0.60	7.57	41.67
		KurtosisVeg1 P_Veg1_25 P_Veg2_70 StdRef2 ZeroNgrnd1ratio Canopy30P	0.66	0.60	7.68	41.23
	M	* SkewnessVeg2 Canopy60P Canopy90P	0.55	0.53	4.00	42.60
		SkewnessVeg2 Canopy90P	0.53	0.51	4.93	43.49
		Canopy60P Canopy90P	0.49	0.47	8.61	45.13
Biomass	D	* MedianVeg1 MedianRef2	0.53	0.52	5.93	44.62
		MedianVeg1 MedianRef2 Canopy90P	0.54	0.52	6.35	44.48
		MedianVeg1 MaxRef2 MedianRef2	0.54	0.52	6.38	44.48
		MedianVeg1 MaxRef1 MaxRef2 MedianRef2	0.55	0.53	6.50	44.26
		MedianVeg1 MaxRef1 MaxRef2 MedianRef2 Canopy90P	0.56	0.53	6.70	44.06
		MedianVeg1 MaxRef1 MaxRef2 Canopy90P	0.55	0.53	6.91	44.37
		MedianVeg1 ModeVeg2 MedianRef2 Canopy90P	0.55	0.53	6.93	44.38
		MedianVeg1 ModeVeg2 MedianRef2	0.54	0.52	6.95	44.63
	C	P_Veg1_25 MinVeg2 P_Veg2_70 ZeroNgrnd1ratio Canopy40P	0.60	0.55	6.00	12.77
		* P_Veg1_25 MinVeg2 P_Veg2_70 ZeroNgrnd1ratio	0.58	0.54	6.37	12.98
		P_Veg1_25 P_Veg2_70 ZeroNgrnd1ratio	0.55	0.52	6.97	13.21
		P_Veg1_25 P_Veg2_70 ZeroNgrnd1ratio Canopy40P	0.57	0.53	7.39	13.14
	M	SkewnessVeg1 MinRef1 MedianRef1 Canopy80P	0.70	0.67	4.23	19.66
		SkewnessVeg1 P_Veg1_10 MinRef1 MedianRef1 Canopy80P	0.71	0.67	5.03	19.61
		* SkewnessVeg1 MinRef1 Canopy80P	0.68	0.66	5.29	20.12
		SkewnessVeg1 P_Veg1_10 MinRef1 Canopy80P	0.69	0.66	5.99	20.06
		SkewnessVeg1 MinRef1 MedianRef1 Canopy60P Canopy80P	0.70	0.67	6.22	19.89

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.10.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 240 segments (3.942 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.10.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* P_Veg1_60 StdMeanRef1 RangeRef2 ZeroNVeg2ratio Canopy90P	0.58	0.55	5.76	55.31
		P_Veg1_60 StdMeanRef1 MedianRef1 RangeRef2 ZeroNVeg2ratio Canopy90P	0.58	0.55	7.00	55.39
		P_Veg1_60 StdMeanRef1 ZeroNVeg2ratio Canopy90P	0.56	0.54	7.00	56.02
		P_Veg1_60 ZeroNVeg2ratio Canopy90P	0.55	0.53	7.75	56.56
	C	* RangeVeg1 P_Veg1_40 RangeVeg2 Canopy90P	0.58	0.55	4.76	48.33
		RangeVeg1 P_Veg1_40 ModeVeg2 RangeVeg2 Canopy90P	0.59	0.55	5.74	48.32
		RangeVeg1 P_Veg1_40 RangeVeg2 P_Veg2_25 Canopy90P	0.58	0.54	6.64	48.75
		RangeVeg1 P_Veg1_40 ModeVeg2 RangeVeg2 P_Veg2_25 Canopy90P	0.60	0.55	7.00	48.44
		P_Veg1_40 RangeVeg2 Canopy90P	0.55	0.52	7.04	49.81
	A	* P_Veg1_40 MedianRef1 ZeroNgrnd1ratio Canopy90P	0.53	0.52	6.03	56.23
		P_Veg1_40 StdMeanRef1 MedianRef1ZeroNgrnd1ratio Canopy90P	0.54	0.52	6.04	56.04
		P_Veg1_40 StdMeanRef1 MedianRef1 P_Veg2_20 ZeroNgrnd1ratio Canopy90P	0.54	0.52	7.00	56.03
		P_Veg1_40 MedianRef1 P_Veg2_20 ZeroNgrnd1ratio Canopy90P	0.53	0.52	7.72	56.37
Biomass	D	P_Veg1_60 MaxVeg2 P_Veg2_10 ZeroNVeg2ratio Canopy90P	0.58	0.56	6.18	39.47
		* P_Veg1_60 MaxVeg2 ZeroNVeg2ratio Canopy90P	0.57	0.55	6.20	39.70
		P_Veg1_60 MaxVeg2 P_Veg2_10 RangeRef2 ZeroNVeg2ratio Canopy90P	0.59	0.56	6.90	39.40
	C	P_Veg1_30 RangeRef1 RangeVeg2 Canopy90P	0.51	0.47	5.19	17.63
		* P_Veg1_30 RangeRef1 RangeVeg2	0.49	0.46	5.36	17.83
		P_Veg1_30 RangeVeg2 MaxRef2 Canopy90P	0.50	0.46	5.67	17.72
		P_Veg1_30 RangeRef1 RangeVeg2 MaxRef2 Canopy90P	0.52	0.47	6.00	17.60
	A	StdMeanRef1 CVVeg2 P_Veg2_75 MinRef2 ZeroNgrnd3_5ratio Canopy90P	0.63	0.61	7.00	36.86
		* StdMeanRef1 CVVeg2 P_Veg2_75 ZeroNgrnd3_5ratio Canopy90P	0.62	0.61	8.14	37.13
		StdMeanRef1 P_Veg2_75 MinRef2 ZeroNgrnd3_5ratio Canopy90P	0.61	0.60	11.29	37.54
		CVVeg2 P_Veg2_75 ZeroNgrnd3_5ratio Canopy90P	0.60	0.59	11.61	37.70
		StdMeanRef1 P_Veg2_75 ZeroNgrnd3_5ratio Canopy90P	0.60	0.59	11.98	37.74

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.10.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 240 segments (3.942 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.10.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* P_Veg1_60 ZeroNgrnd2ratio	0.61	0.60	2.17	58.79
		P_Veg1_60 KurtosisVeg2 ZeroNgrnd2ratio	0.61	0.59	4.00	59.19
		P_Veg1_60	0.57	0.56	6.39	61.25
		P_Veg1_60 KurtosisVeg2	0.58	0.56	7.29	61.24
	C	MinVeg1 P_Veg1_25 ZeroNgrnd2ratio Canopy30P	0.70	0.66	5.46	40.70
		* P_Veg1_25 ZeroNgrnd2ratio Canopy30P	0.68	0.65	5.52	41.37
		P_Veg1_25 RangeVeg2 ZeroNgrnd2ratio Canopy30P	0.70	0.65	5.66	40.84
		P_Veg1_25 ZeroNgrnd2ratio	0.65	0.63	5.96	42.22
		MinVeg1 P_Veg1_25 RangeVeg2 ZeroNgrnd2ratio Canopy30P	0.71	0.66	6.00	40.39
		MinVeg1 P_Veg1_25 ZeroNgrnd2ratio	0.67	0.64	6.05	41.71
		P_Veg1_25 RangeVeg2 Canopy30P	0.67	0.64	6.27	41.85
		P_Veg1_25 RangeVeg2 ZeroNgrnd2ratio	0.66	0.63	6.85	42.21
	M	* MinVeg1 KurtosisVeg2 Canopy10P Canopy90P	0.58	0.54	3.66	43.87
		MinVeg1 MinRef1 KurtosisVeg2 Canopy10P Canopy90P	0.58	0.53	5.10	44.08
		MinVeg1 KurtosisVeg2 Canopy10P Canopy60P Canopy90P	0.58	0.53	5.15	44.11
		MinVeg1 P_Veg1_30 KurtosisVeg2 Canopy10P Canopy90P	0.58	0.53	5.63	44.35
		MinVeg1 CVVeg2 KurtosisVeg2 Canopy10P Canopy90P	0.58	0.53	5.66	44.37
		MinVeg1 MinRef1 CVVeg2 Canopy60P Canopy90P	0.57	0.53	5.96	44.52
Biomass	D	* P_Veg1_60 ZeroNVeg2ratio	0.59	0.57	5.77	43.81
		P_Veg1_60 MaxVeg2 ZeroNVeg2ratio Canopy90P	0.61	0.58	5.83	43.15
		P_Veg1_60 ZeroNVeg2ratio Canopy90P	0.60	0.58	5.89	43.52
		P_Veg1_60 MinRef2 ZeroNVeg2ratio	0.59	0.57	6.41	43.70
		P_Veg1_60 MaxVeg2 ModeVeg2 ZeroNVeg2ratio Canopy90P	0.62	0.59	6.42	43.00
		P_Veg1_60 MaxVeg2 ZeroNVeg2ratio	0.59	0.57	6.93	43.88
	C	* P_Veg1_25 MedianRef1 StdRef2 ZeroNgrnd3_5ratio	0.67	0.63	5.70	12.06
		P_Veg1_25 MedianRef1 StdRef2 MedianRef2 ZeroNgrnd3_5ratio	0.69	0.63	6.00	11.92
		P_Veg1_25 ZeroNgrnd3_5ratio	0.62	0.60	6.05	12.48
		P_Veg1_25 MedianRef1 ZeroNgrnd3_5ratio	0.64	0.61	6.08	12.31
		P_Veg1_25 StdRef2 MedianRef2 ZeroNgrnd3_5ratio	0.66	0.62	6.10	12.14
		P_Veg1_25 MedianRef2 ZeroNgrnd3_5ratio	0.64	0.61	6.39	12.37
	M	MeanVeg1 MinVeg1 RangeRef1 Vegratio Canopy70P	0.67	0.63	6.00	21.74
		* MeanVeg1 MinVeg1 RangeRef1 Canopy70P	0.65	0.62	6.81	22.17

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.11.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 168 segments (5.632 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.11.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* P_Veg2_70 ZeroNgrnd3_5ratio Canopy80P	0.61	0.59	2.62	51.15
		P_Veg2_70 MinRef2 ZeroNgrnd3_5ratio Canopy80P	0.62	0.59	3.39	51.05
		P_Veg2_10 P_Veg2_70 ZeroNgrnd3_5ratio Canopy80P	0.61	0.59	4.32	51.41
		P_Veg2_70 ZeroNgrnd3_5ratio Canopy20P Canopy80P	0.61	0.59	4.58	51.50
		P_Veg2_10 P_Veg2_70 MinRef2 ZeroNgrnd3_5ratio Canopy80P	0.62	0.59	5.05	51.30
		P_Veg2_70 MinRef2 ZeroNgrnd3_5ratio Canopy20P Canopy80P	0.62	0.59	5.38	51.42
		P_Veg2_70 ZeroNgrnd3_5ratio	0.58	0.57	5.78	52.66
		P_Veg2_70 MinRef2 ZeroNgrnd3_5ratio	0.59	0.57	5.97	52.38
	C	* ModeVeg1 P_Veg1_40 RangeVeg2 StdRef2	0.69	0.66	6.02	38.03
		ModeVeg1 P_Veg1_40 RangeVeg2 MedianVeg2 StdRef2	0.70	0.66	6.89	37.98
		ModeVeg1 P_Veg1_40 RangeVeg2 MedianVeg2 StdRef2 ZeroNgrnd3_5ratio	0.72	0.67	7.00	37.51
		ModeVeg1 P_Veg1_40 RangeVeg2 StdRef2 ZeroNgrnd3_5ratio	0.69	0.65	7.82	38.47
	A	* P_Veg1_40 StdMeanVeg2	0.55	0.54	2.96	52.16
		P_Veg1_40 StdMeanVeg2 StdRef2	0.56	0.54	3.19	51.98
		P_Veg1_40	0.54	0.53	4.17	52.67
		P_Veg1_40 StdRef2	0.54	0.54	4.39	52.49
		P_Veg1_40 StdMeanVeg2 ZeroNgrnd3_5ratio	0.55	0.54	4.89	52.38
Biomass	D	MinVeg1 P_Veg2_10 P_Veg2_75 CVRef2	0.61	0.59	5.72	37.09
		P_Veg2_10 P_Veg2_75 ZeroNVeg2ratio Canopy70P	0.61	0.59	5.94	37.14
		* P_Veg2_75 ZeroNVeg2ratio Canopy70P	0.60	0.58	5.96	37.41
		MinVeg1 P_Veg2_10 P_Veg2_75 CVRef2 ZeroNVeg2ratio	0.62	0.59	6.21	36.95
		MinVeg1 P_Veg2_10 P_Veg2_75 ZeroNVeg2ratio Canopy70P	0.62	0.59	6.28	36.97
		MinVeg1 P_Veg2_10 P_Veg2_75 ZeroNVeg2ratio	0.61	0.58	6.32	37.24
		MinVeg1 P_Veg2_10 P_Veg2_75 CVRef2 Canopy20P	0.62	0.59	6.36	36.99
		MinVeg1 P_Veg2_10 P_Veg2_75 CVRef2 Canopy70P	0.62	0.59	6.57	37.05
		MinVeg1 P_Veg2_10 P_Veg2_75 CVRef2 ZeroNVeg2ratio Canopy70P	0.63	0.59	6.58	36.78
		P_Veg2_10 P_Veg2_75 CVRef2 ZeroNVeg2ratio Canopy70P	0.61	0.59	6.87	37.13
		P_Veg2_10 P_Veg2_75 ZeroNVeg2ratio	0.59	0.57	6.93	37.66
	C	* P_Veg1_40 MaxVeg2	0.55	0.52	1.93	15.98
		P_Veg1_40 MaxVeg2 Canopy80P	0.57	0.53	2.51	15.88
		P_Veg1_40 MaxVeg2 ModeVeg2	0.55	0.52	3.39	16.07
		P_Veg1_40	0.50	0.49	3.69	16.56
		P_Veg1_40 MaxVeg2 P_Veg2_40	0.55	0.51	3.87	16.18
		P_Veg1_40 Canopy80P	0.52	0.50	3.94	16.41

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.11.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	A	* CVRef1 CVVeg2 P_Veg2_75 ZeroNVeg3_5ratio Canopy80P	0.68	0.66	7.57	33.14
		CVRef1 CVVeg2 P_Veg2_10 P_Veg2_75 ZeroNVeg3_5ratio Canopy80P	0.68	0.66	7.73	33.02
		CVRef1 CVVeg2 P_Veg2_10 P_Veg2_30 P_Veg2_75 ZeroNVeg3_5ratio Canopy80P	0.69	0.67	8.00	32.91
		CVRef1 P_Veg2_75 ZeroNVeg3_5ratio Canopy80P	0.67	0.66	8.58	33.44
		CVRef1 P_Veg2_10 P_Veg2_75 ZeroNVeg3_5ratio Canopy80P	0.67	0.66	8.72	33.31
		CVVeg2 P_Veg2_75 ZeroNVeg3_5ratio Canopy80P	0.67	0.66	8.81	33.47

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.11.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 168 segments (5.632 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.11.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	P_Veg2_70 ZeroNgrnd3_5ratio	0.65	0.63	1.56	54.80
		MaxVeg2 P_Veg2_70 ZeroNgrnd3_5ratio	0.66	0.63	2.48	54.73
		* P_Veg2_70	0.63	0.62	2.55	55.98
		MaxVeg2 P_Veg2_70	0.64	0.62	3.04	55.67
		MinRef1 P_Veg2_70 ZeroNgrnd3_5ratio	0.65	0.63	3.24	55.20
		ModeVeg2 P_Veg2_70 ZeroNgrnd3_5ratio	0.65	0.63	3.50	55.36
	C	* P_Veg1_40 CVVeg2	0.52	0.47	1.82	52.13
		P_Veg1_40	0.45	0.43	2.48	54.34
		KurtosisVeg1 P_Veg1_40	0.50	0.45	2.62	53.22
		KurtosisVeg1 P_Veg1_40 CVVeg2	0.54	0.46	3.19	52.58
		P_Veg1_40 P_Veg2_40	0.48	0.43	3.35	54.21
		KurtosisVeg1 P_Veg1_40 P_Veg2_40	0.53	0.45	3.59	53.16
		P_Veg1_40 CVVeg2 P_Veg2_40	0.52	0.45	3.75	53.38
		KurtosisVeg1 P_Veg1_40 CVVeg2 P_Veg2_40	0.54	0.44	5.00	53.74
	M	ModeVeg1 MinRef1 MinVeg2 StdMeanRef2 ZeroNgrnd1ratio Canopy10P Canopy50P Canopy80P	0.80	0.75	9.00	27.48
		* ModeVeg1 MinRef1 StdMeanRef2 ZeroNgrnd1ratio Canopy10P Canopy50P Canopy80P	0.78	0.74	9.35	28.02
		ModeVeg1 MinRef1 MinVeg2 StdMeanRef2 Canopy10P Canopy50P Canopy80P	0.78	0.73	10.01	28.28
		ModeVeg1 MinRef1 ZeroNgrnd1ratio Canopy10P Canopy50P Canopy80P	0.77	0.73	10.35	28.76
Biomass	D	* P_Veg2_75 StdMeanRef2 ZeroNgrnd3_5ratio	0.65	0.62	5.32	39.48
		ModeVeg2 P_Veg2_75 StdMeanRef2 ZeroNgrnd3_5ratio	0.66	0.63	5.34	39.07
		ModeVeg2 P_Veg2_75 StdMeanRef2 ZeroNgrnd3_5ratio Canopy90P	0.68	0.64	5.63	38.76
		P_Veg2_75 ZeroNgrnd3_5ratio	0.63	0.61	5.65	40.01
		P_Veg2_75 ZeroNgrnd3_5ratio Canopy90P	0.64	0.62	6.00	39.76
		P_Veg2_75 StdMeanRef2 ZeroNgrnd3_5ratio Canopy90P	0.65	0.62	6.35	39.50
		ModeVeg2 P_Veg2_75 StdMeanRef2	0.64	0.62	6.40	39.93
		ModeVeg2 P_Veg2_75 ZeroNgrnd3_5ratio Canopy90P	0.65	0.62	6.54	39.59
	C	ModeVeg1 P_Veg1_30 RangeVeg2	0.57	0.50	4.79	14.45
		* ModeVeg1 RangeVeg1 P_Veg1_30 RangeVeg2	0.61	0.52	5.00	14.16
		P_Veg1_30 RangeVeg2	0.51	0.46	5.48	15.01
		RangeVeg1 P_Veg1_30 RangeVeg2	0.55	0.47	5.93	14.86

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.11.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass		KurtosisVeg1 MinVeg1 P_Veg1_10 RangeRef1 ZeroNVeg3_5ratio Canopy80P	0.83	0.80	7.77	15.89
		KurtosisVeg1 MinVeg1 ModeVeg1 P_Veg1_10 RangeRef1 ZeroNVeg3_5ratio Canopy80P	0.83	0.80	8.66	15.87
		* MinVeg1 P_Veg1_10 RangeRef1 ZeroNVeg3_5ratio Canopy80P	0.81	0.79	8.83	16.32
	M	KurtosisVeg1 MinVeg1 ModeVeg1 P_Veg1_10 RangeRef1 ZeroNVeg3_5ratio Canopy60P Canopy80P	0.84	0.80	9.01	15.72
		KurtosisVeg1 MinVeg1 P_Veg1_10 RangeRef1 ZeroNVeg3_5ratio Canopy60P Canopy80P	0.83	0.80	9.05	15.96
		MinVeg1 ModeVeg1 P_Veg1_10 RangeRef1 ZeroNVeg3_5ratio Canopy60P Canopy80P	0.83	0.80	9.38	16.03

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.12.1 2-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 167 Appomattox forest stands (5.666 ha/segment) across forest types (*deciduous = D; coniferous = C; all segments/types = A*). Selected models are shown with an *

Table G.12.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	MaxVeg1 MedianVeg1 MaxRef1 RangeVeg2 StdMeanVeg2 ZeroNVeg2ratio Canopy10P Canopy80P	8.66	0.46	8.90	62.44
		MaxVeg1 MedianVeg1 MaxRef1 MedianRef1 RangeVeg2 StdMeanVeg2 ZeroNVeg2ratio Canopy10P Canopy80P	9.09	0.46	9.75	62.35
		* MedianVeg1 MaxRef1 RangeVeg2 StdMeanVeg2 ZeroNVeg2ratio Canopy10P Canopy80P	9.78	0.44	9.82	63.56
		MaxVeg1 MedianVeg1 MaxRef1 RangeVeg2 ZeroNVeg2ratio Canopy10P Canopy80P	0.50	0.44	10.00	63.67
	C	* MaxRef1 MedianRef1 MinRef2	0.53	0.48	1.87	55.61
		P_Veg1_30 MaxRef1 MedianRef1 MinRef2	0.56	0.49	2.37	55.02
		P_Veg1_30 MaxRef1 MedianRef1	0.51	0.47	2.87	56.60
		P_Veg1_30 MedianRef1	0.48	0.44	3.13	57.82
		P_Veg1_30 MaxRef1 MedianRef1 MinRef2 ZeroNVeg2ratio	0.57	0.49	3.47	55.03
		MaxRef1 MedianRef1 KurtosisVeg2 MinRef2	0.53	0.47	3.64	56.33
		MaxRef1 MedianRef1 MinRef2 ZeroNVeg2ratio	0.53	0.47	3.68	56.36
		MaxRef1 StdRef1 MedianRef1 MinRef2	0.53	0.47	3.70	56.39
		MaxRef1 MedianRef1 MinRef2 StdRef2	0.53	0.47	3.82	56.51
	A	MedianVeg1 CVRef1 RangeRef1 MedianRef1 P_Veg2_90 StdMeanRef2 ZeroNgrnd1ratio Canopy90P	0.48	0.43	9.95	61.96
		CVVeg1 MedianVeg1 CVRef1 RangeRef1 MedianRef1 StdMeanRef2 ZeroNgrnd1ratio Canopy90P	0.48	0.43	9.96	61.96
		CVVeg1 MedianVeg1 CVRef1 RangeRef1 MedianRef1 P_Veg2_90 StdMeanRef2 ZeroNgrnd1ratio Canopy90P	0.49	0.44	10.00	61.62
		* MedianVeg1 CVRef1 RangeRef1 MedianRef1 P_Veg2_90 StdMeanRef2 ZeroNgrnd1ratio	0.46	0.42	10.12	62.36
		MedianVeg1 CVRef1 RangeRef1 MedianRef1 P_Veg2_90 ZeroNgrnd1ratio	0.45	0.41	10.24	62.74
		CVVeg1 MedianVeg1 CVRef1 RangeRef1 MedianRef1 P_Veg2_90 StdMeanRef2 ZeroNgrnd1ratio	0.47	0.43	10.47	62.14
Biomass	D	* MedianVeg1 P_Veg2_10 ZeroNVeg2ratio Canopy10P	0.46	0.43	5.00	45.84
		MedianVeg1 ZeroNVeg2ratio Canopy10P	0.42	0.39	7.57	47.25
	C	P_Veg1_30 Vegratio	0.42	0.38	3.62	19.65
		* P_Veg1_30 MedianRef1 Vegratio	0.45	0.40	4.00	19.45
		P_Veg1_30	0.37	0.35	4.41	20.17
		P_Veg1_30 MedianRef1	0.40	0.36	4.81	20.01

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.12.1		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Biomass	A	* RangeRef1 StdRef1 MedianRef1 StdMeanVeg2 _Veg2_75 MedianRef2 ZeroNgrnd1ratio Canopy90P	0.51	0.46	8.66	41.18
		RangeRef1 StdRef1 MedianRef1 StdMeanVeg2 P_Veg2_75 MedianRef2 ZeroNVeg3_5ratio ZeroNgrnd1ratio Canopy90P	0.52	0.46	9.09	41.04
		RangeRef1 StdRef1 MedianRef1 StdMeanVeg2 P_Veg2_75 MedianRef2 ZeroNVeg3_5ratio Canopy90P	0.50	0.45	9.78	41.45

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table G.12.2 3-class lidar distributional volume (m^3/ha) and biomass (kg/ha) models for 167 Appomattox forest stands (5.666 ha/segment) across forest types (*deciduous = D; coniferous = C; mixed = M*). Selected models are shown with an *

Table G.12.2		Candidate models	R^2	Adjusted R^2	Cp	RMSE m^3/ha or Mg/ha
Volume	D	* P_Veg1_50 P_Veg2_10 ZeroNgrnd2ratio Canopy10P	0.51	0.46	5.96	68.16
		P_Veg1_50 MaxRef1 P_Veg2_10 ZeroNgrnd2ratio Canopy10P	0.53	0.47	6.17	67.48
		P_Veg1_50 MaxRef1 MedianRef1 RangeVeg2 ZeroNgrnd2ratio Canopy10P	0.55	0.47	6.85	67.18
	C	* P_Veg1_25 MaxRef1 MaxVeg2 Canopy80P	0.78	0.73	5.00	40.08
		P_Veg1_25 MaxVeg2 Canopy80P	0.73	0.68	7.03	43.32
	M	* MedianVeg1 MinRef1 Canopy90P	0.61	0.57	2.08	46.68
		MedianVeg1 MinRef1 MinRef2 Canopy90P	0.63	0.57	3.07	46.60
		MedianVeg1 MinRef1 StdMeanRef2 Canopy90P	0.62	0.55	3.94	47.46
		MedianVeg1 MinRef1 CVRef2 Canopy90P	0.61	0.55	4.06	47.58
		MinRef1 MinRef2 Canopy90P	0.58	0.53	4.17	48.66
		MedianVeg1 StdMeanRef2 Canopy90	0.57	0.52	4.81	49.25
Biomass	D	* MedianVeg1 P_Veg2_10 ZeroNgrnd2ratio Canopy10P	0.51	0.46	5.00	48.61
		MedianVeg1 ZeroNgrnd2ratio Canopy10P	0.43	0.38	9.07	51.67
	C	* MaxRef1 StdVeg2 P_Veg2_30 StdMeanRef2	0.75	0.70	3.36	12.56
		MaxRef1 StdVeg2 StdMeanRef2	0.72	0.67	3.67	13.09
		P_Veg1_20 MaxRef1 StdVeg2 StdMeanRef2	0.73	0.66	5.10	13.25
		P_Veg1_20 MaxRef1 StdVeg2 P_Veg2_30 StdMeanRef2	0.76	0.68	5.21	12.88
		MaxRef1 StdVeg2 P_Veg2_30 StdMeanRef2	0.76	0.68	5.28	12.91
		MaxRef1 StdVeg2 StdMeanRef2 StdRef2	0.72	0.66	5.47	13.40
	M	* MeanVeg1 StdMeanRef1 MinVeg2 P_Veg2_20 Canopy90P	0.73	0.678	6.00	20.29
		MeanVeg1 StdMeanRef1 MinVeg2 Canopy90P	0.66	0.607	10.53	22.42

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Appendix H
Field-Measured vs. Predicted Value Plots for All Segmentation Results

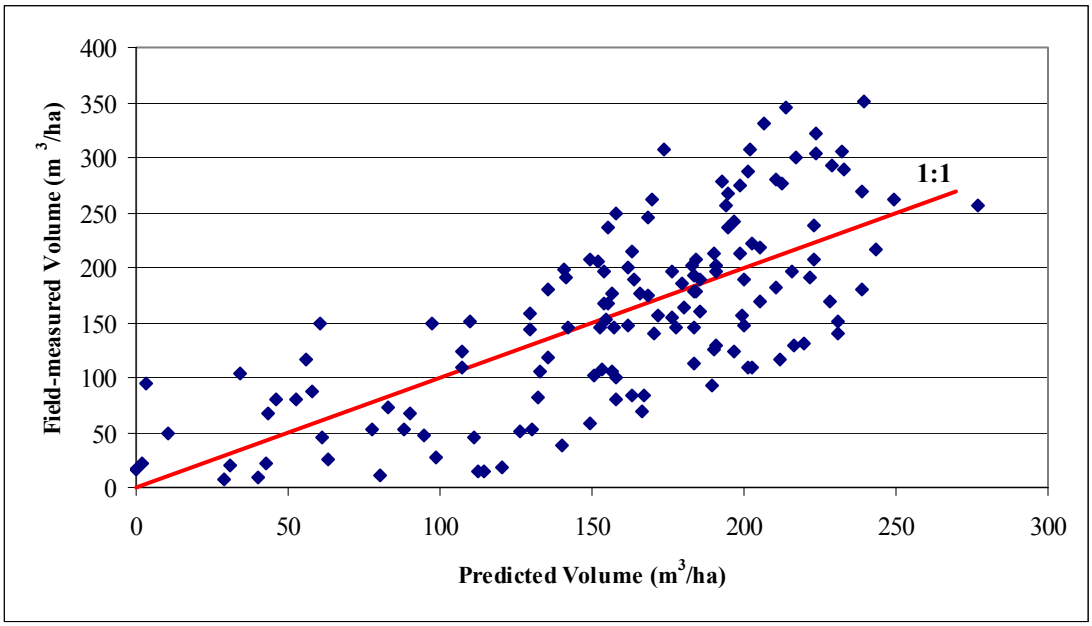


Figure H.1 2-class volume model (0.035 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.51$)

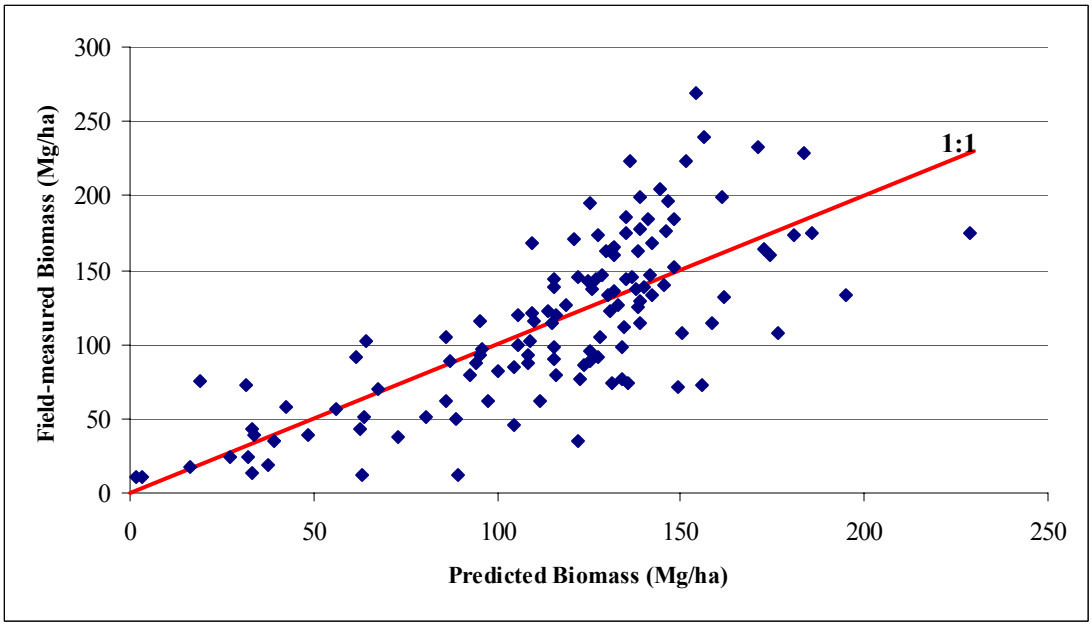


Figure H.2 2-class biomass model (0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.54$)

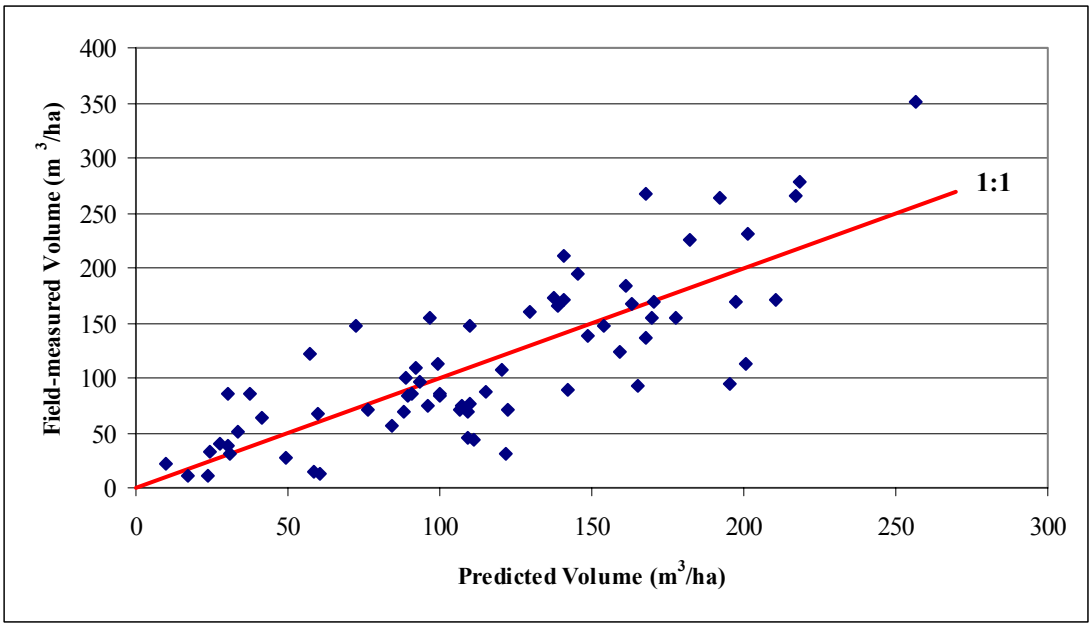


Figure H.3 2-class volume model (0.035 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.62$)

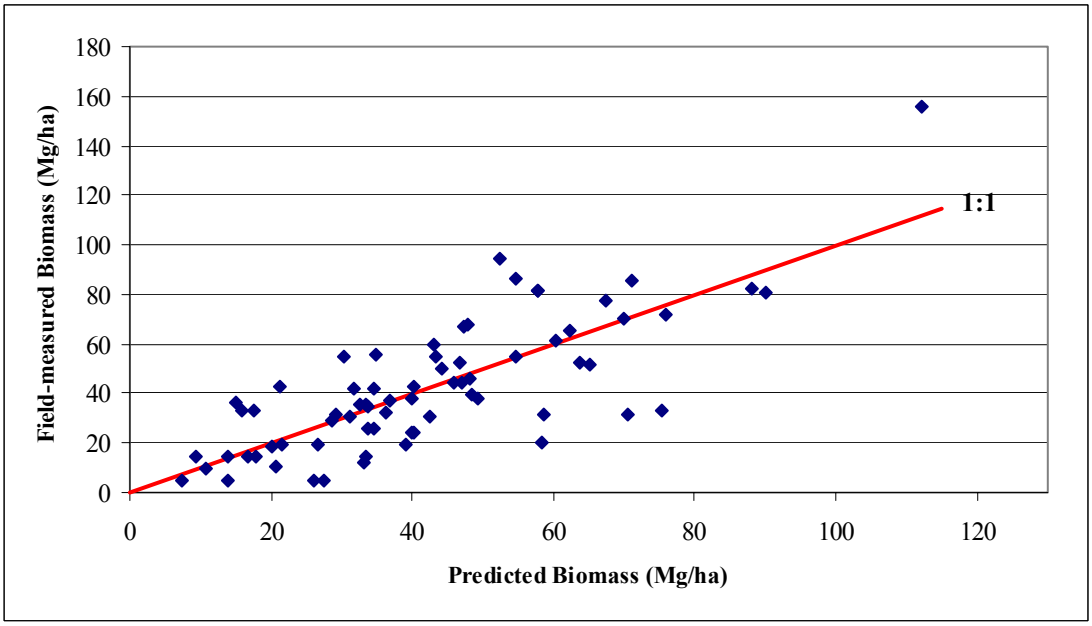


Figure H.4 2-class biomass model (0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.57$)

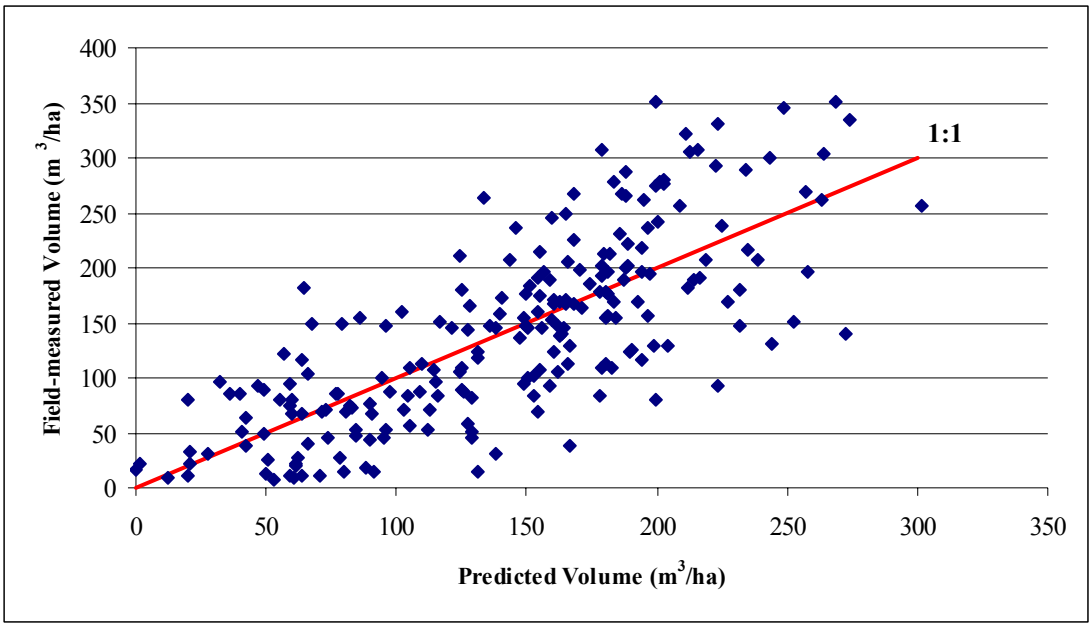


Figure H.5 2-class volume model (0.035 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.59$)

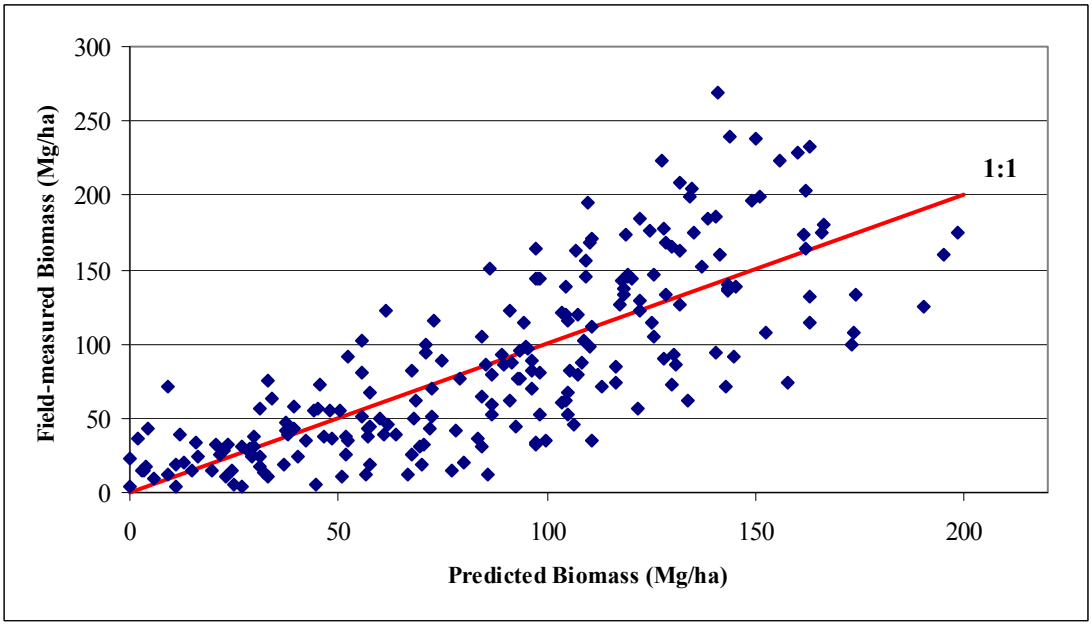


Figure H.6 2-class biomass model (0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.60$)

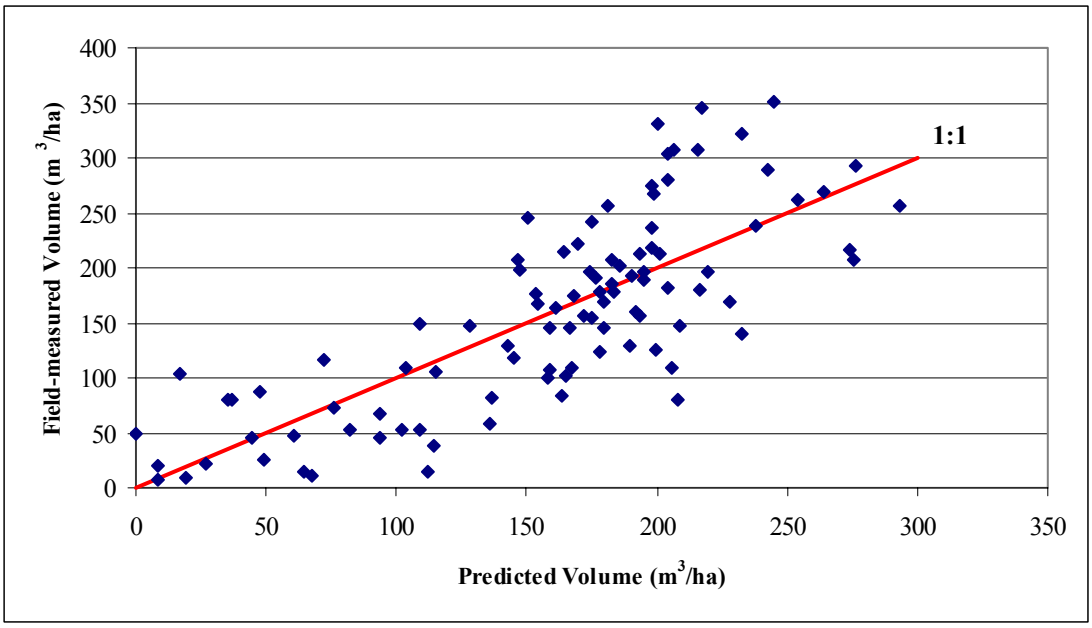


Figure H.7 3-class volume model (0.035 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.59$)

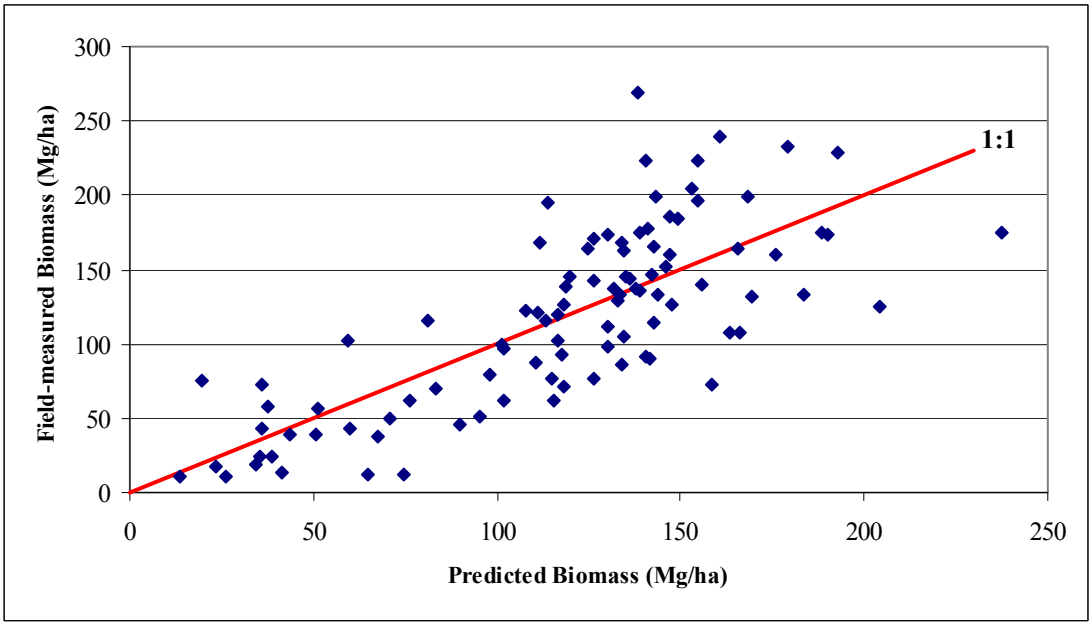


Figure H.8 3-class biomass model (0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.56$)

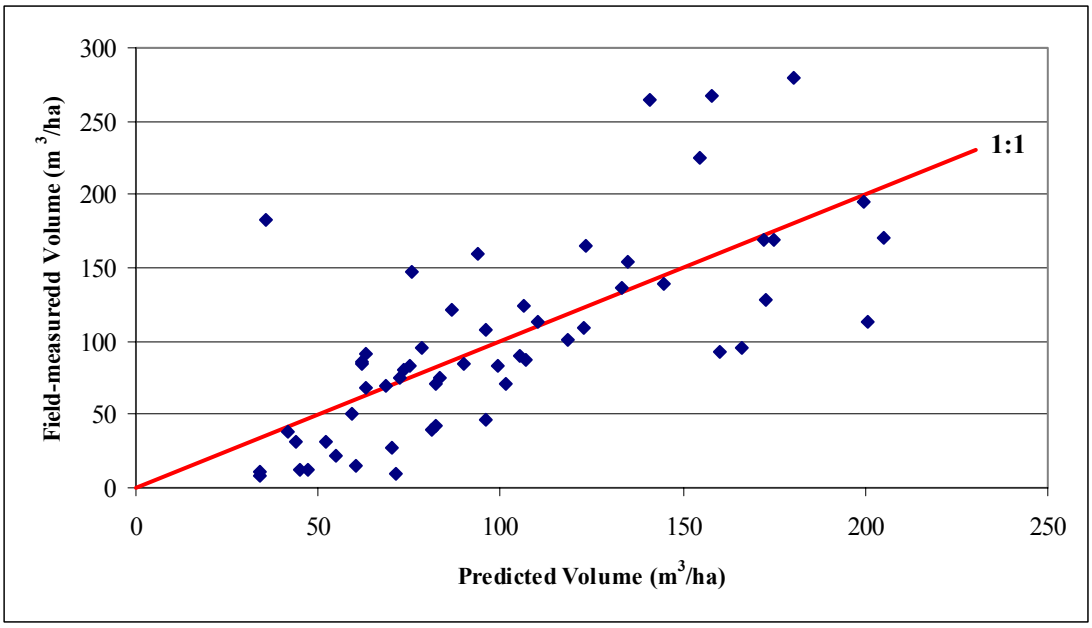


Figure H.9 3-class volume model (0.035 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.47$)

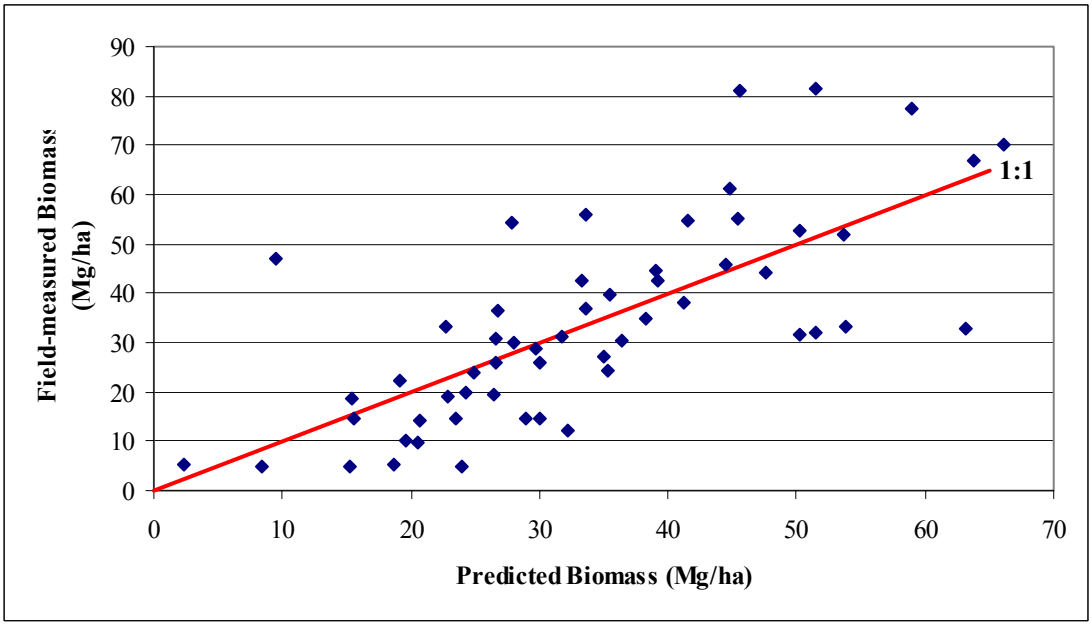


Figure H.10 3-class biomass model (0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.50$)

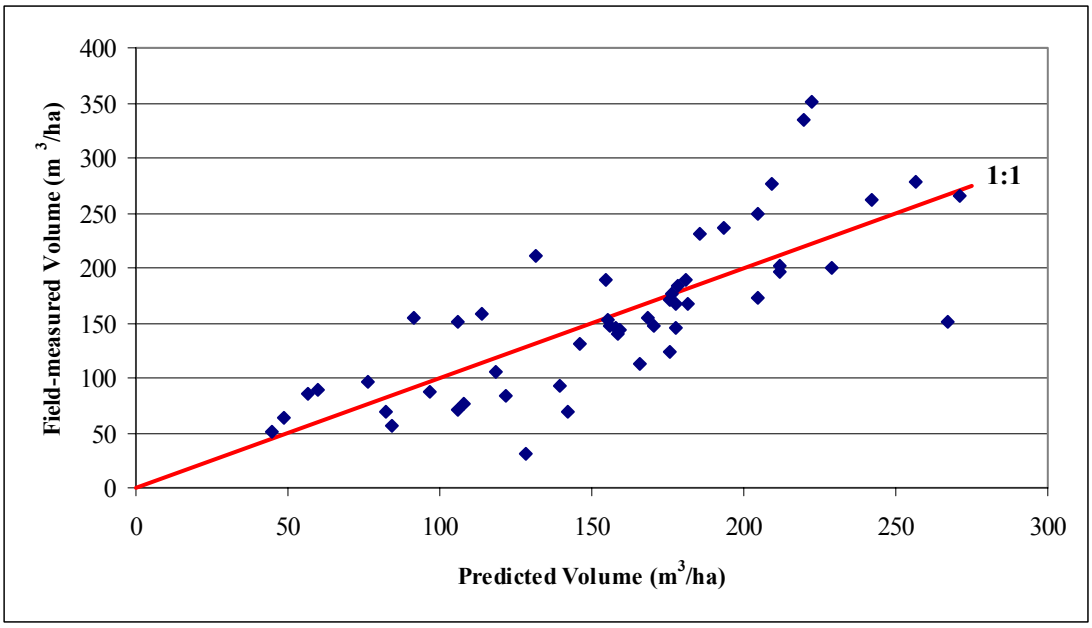


Figure H.11 3-class volume model (0.035 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.56$)

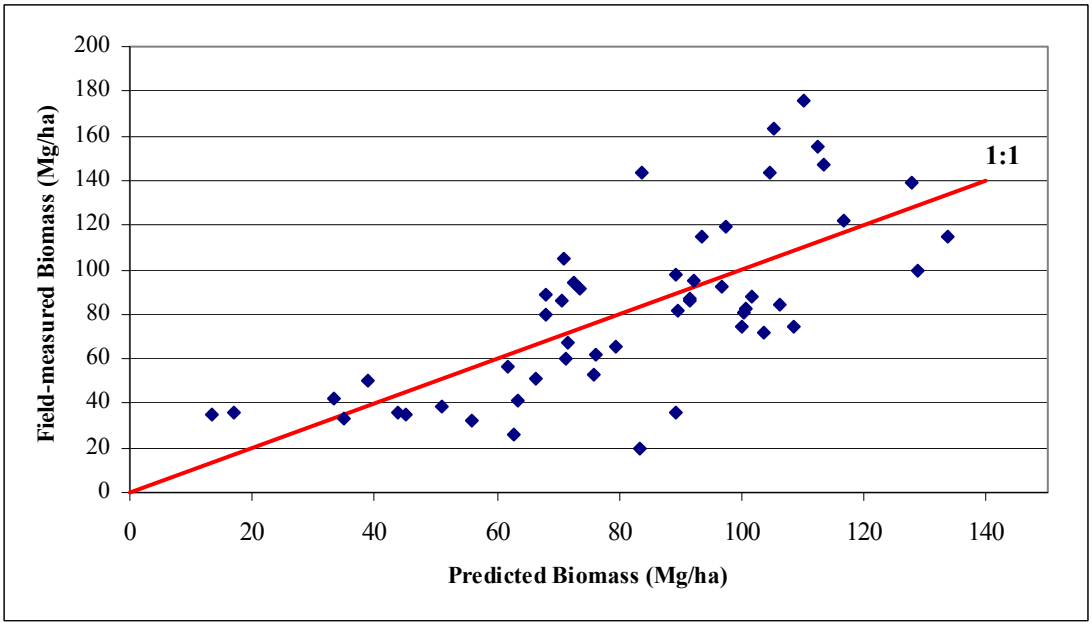


Figure H.12 3-class biomass model (0.035 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.48$)

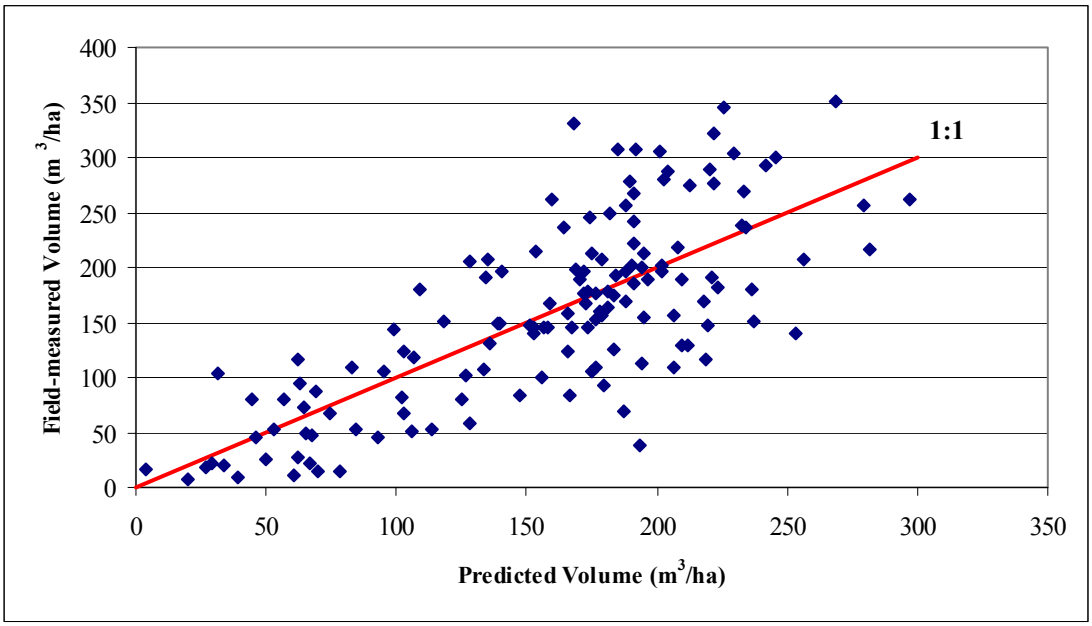


Figure H.13 2-class volume model (0.091 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.55$)

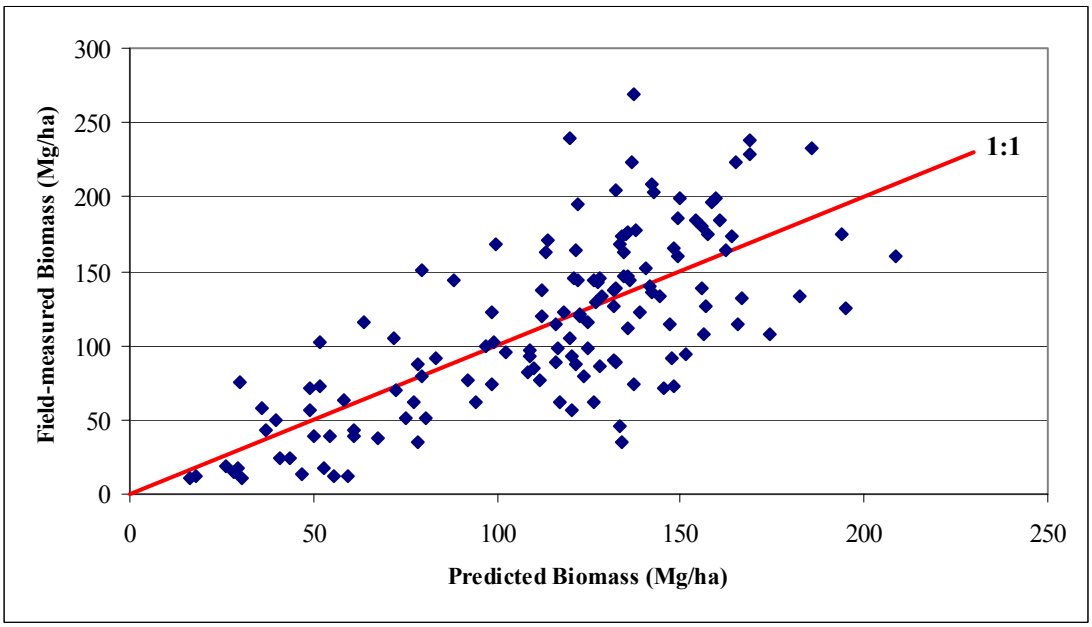


Figure H.14 2-class biomass model (0.091 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.51$)

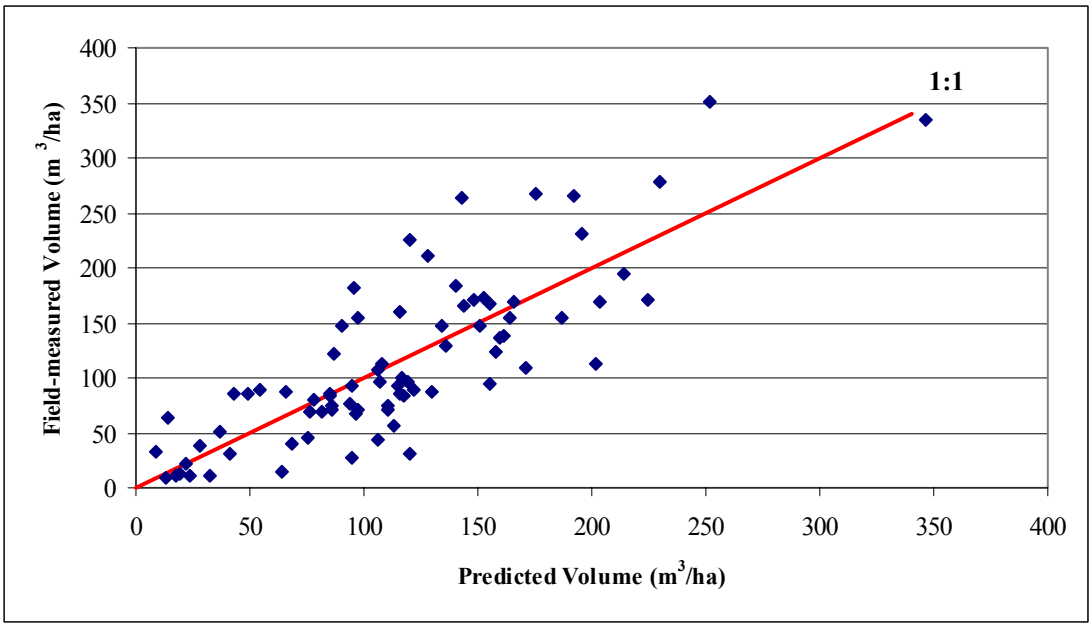


Figure H.15 2-class volume model (0.091 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.64$)

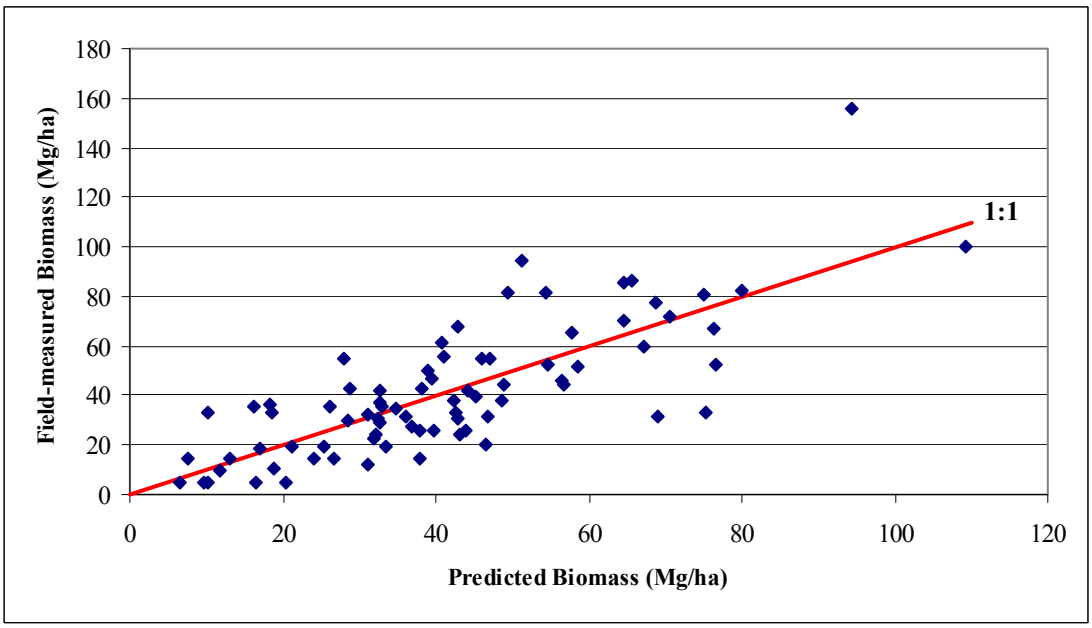


Figure H.16 2-class biomass model (0.091 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.59$)

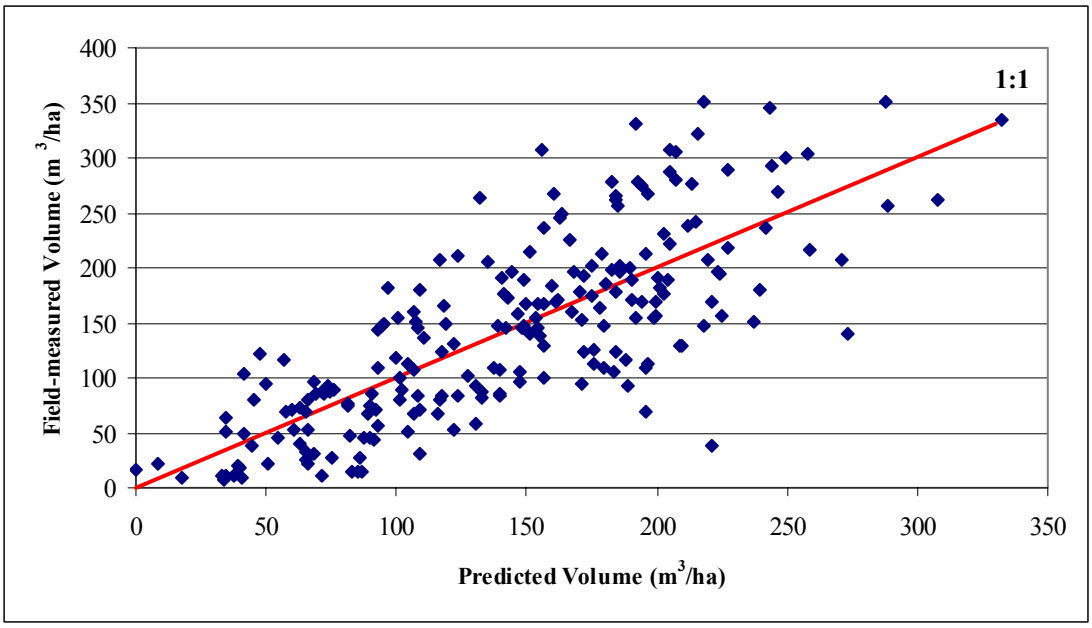


Figure H.17 2-class volume model (0.091 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.59$)

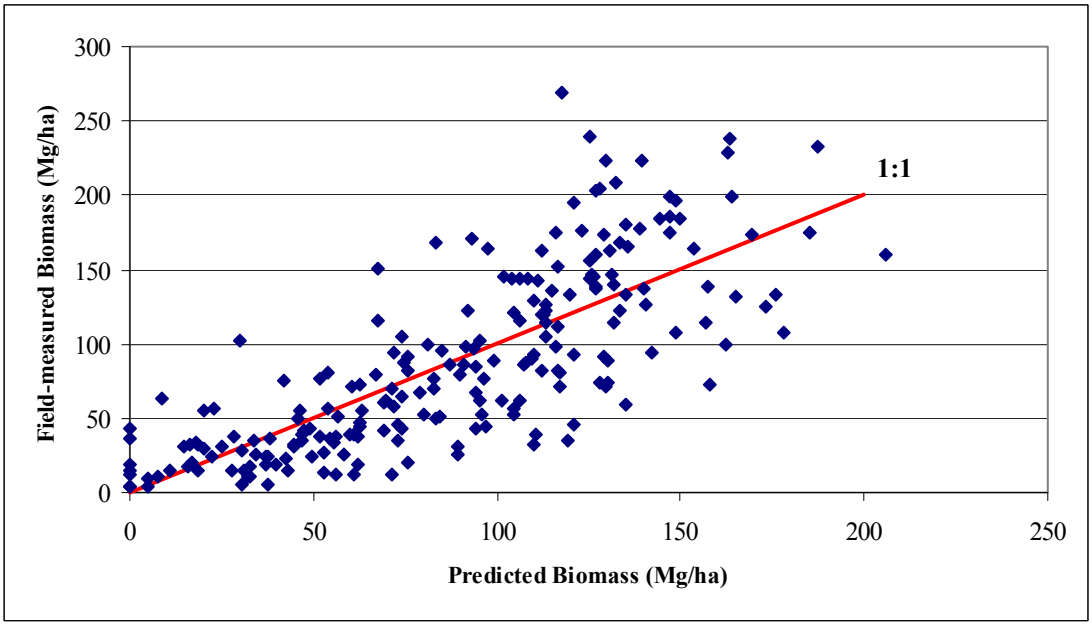


Figure H.18 2-class biomass model (0.091 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.59$)

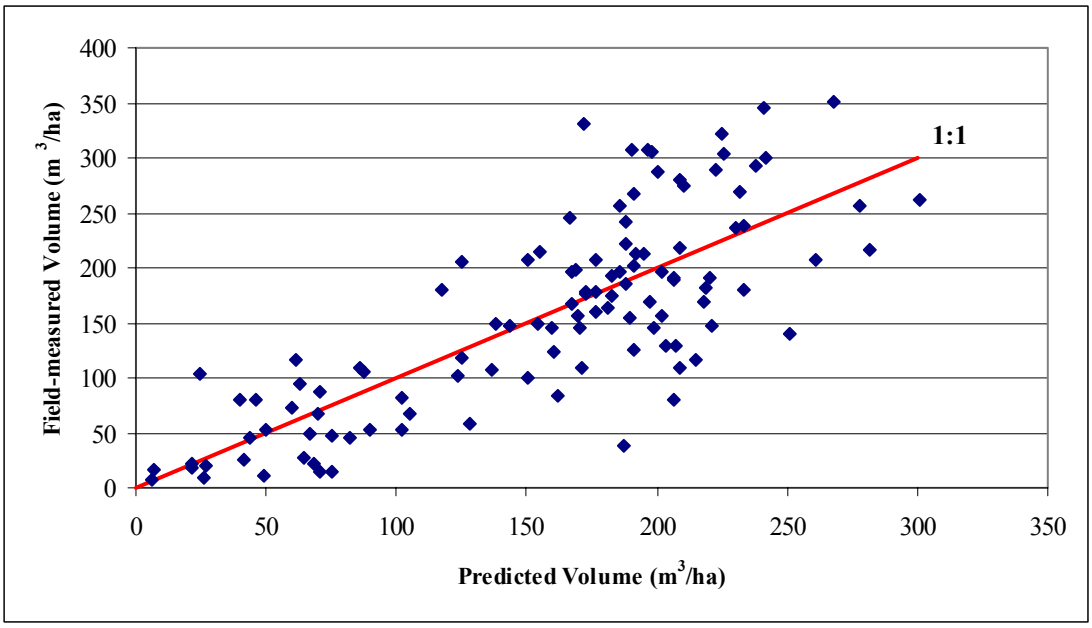


Figure H.19 3-class volume model (0.091 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.60$)

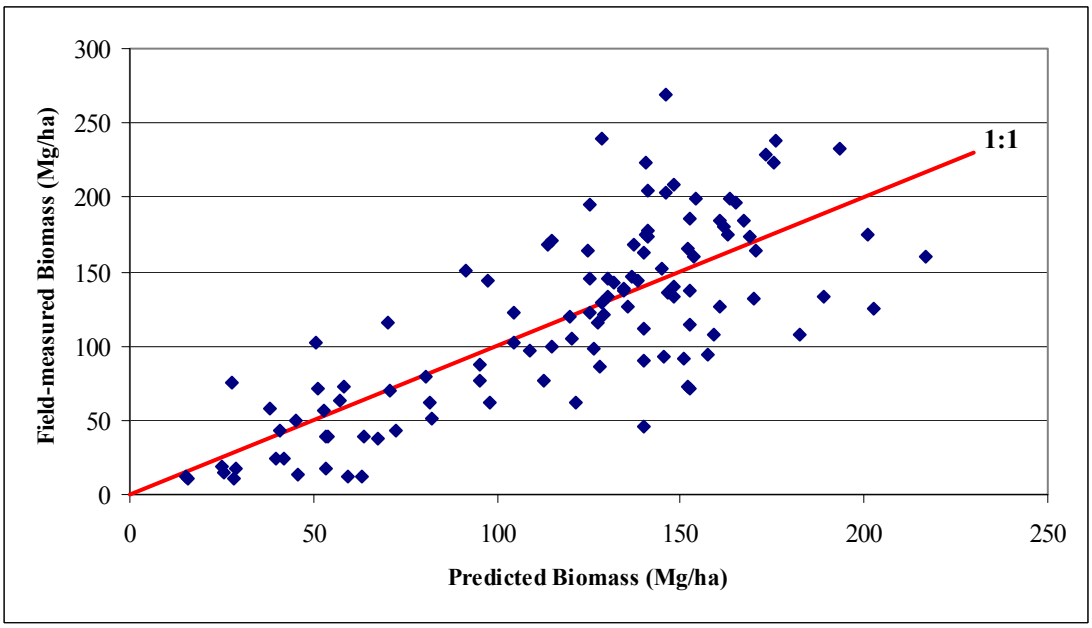


Figure H.20 3-class biomass model (0.091 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.57$)

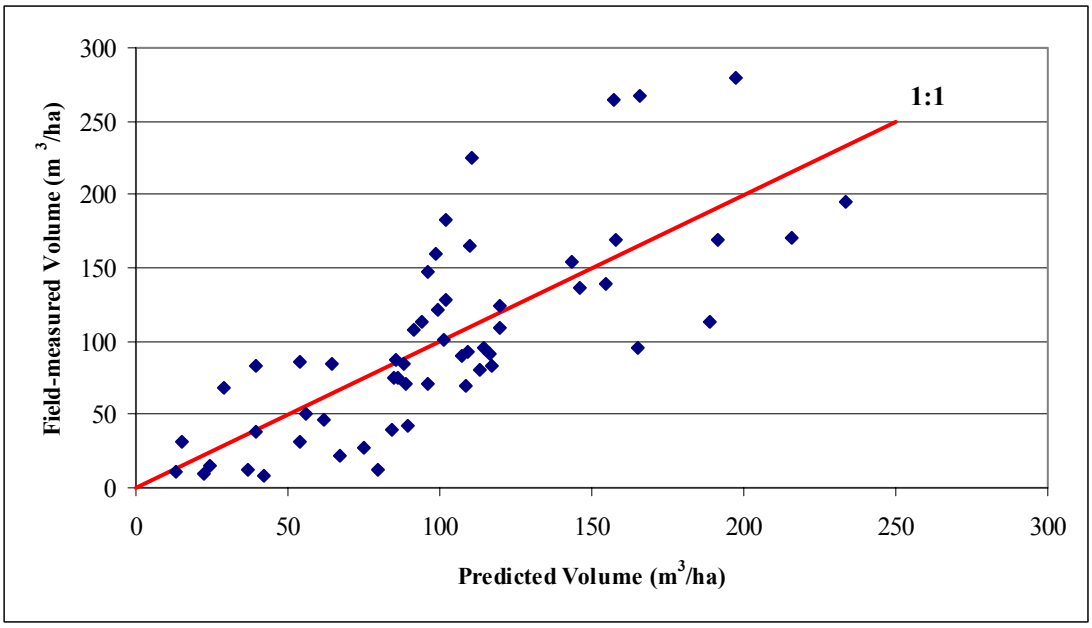


Figure H.21 3-class volume model (0.091 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.54$)

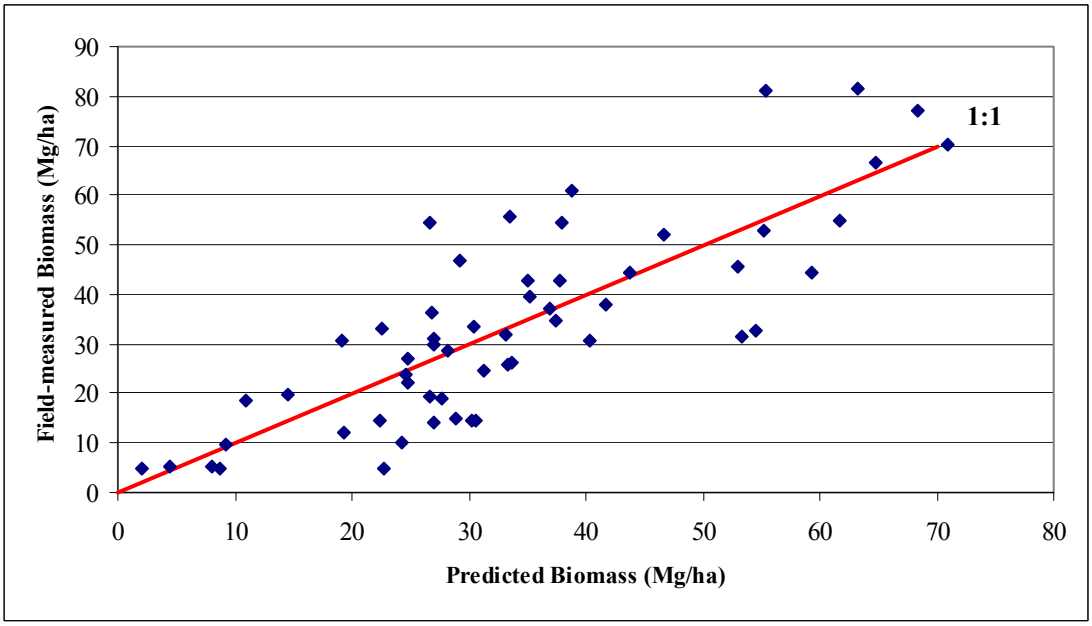


Figure H.22 3-class biomass model (0.091 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.62$)

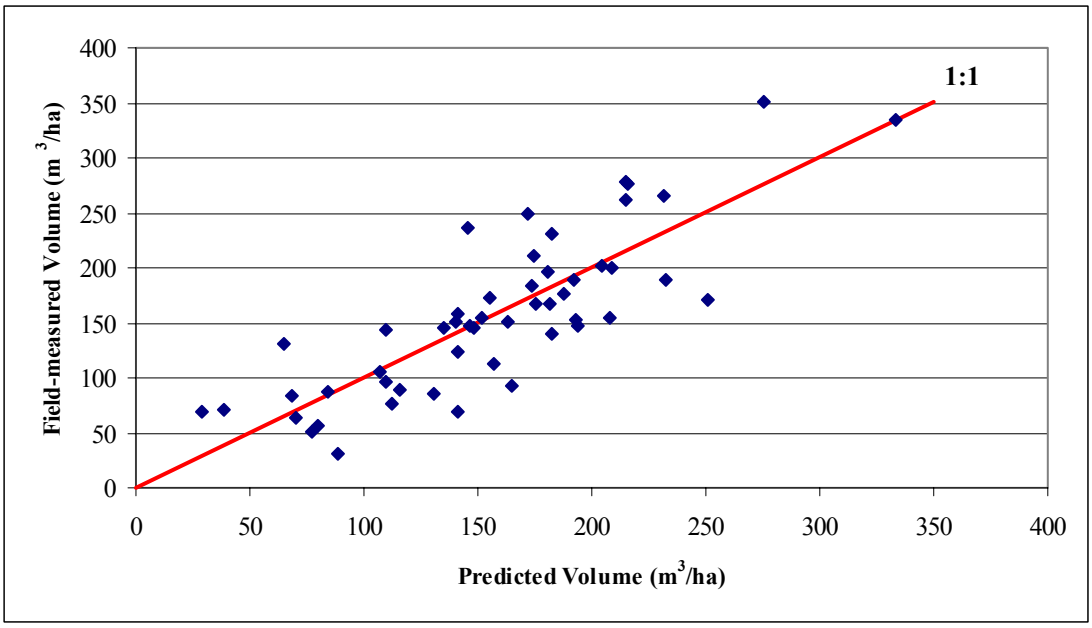


Figure H.23 3-class volume model (0.091 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.63$)

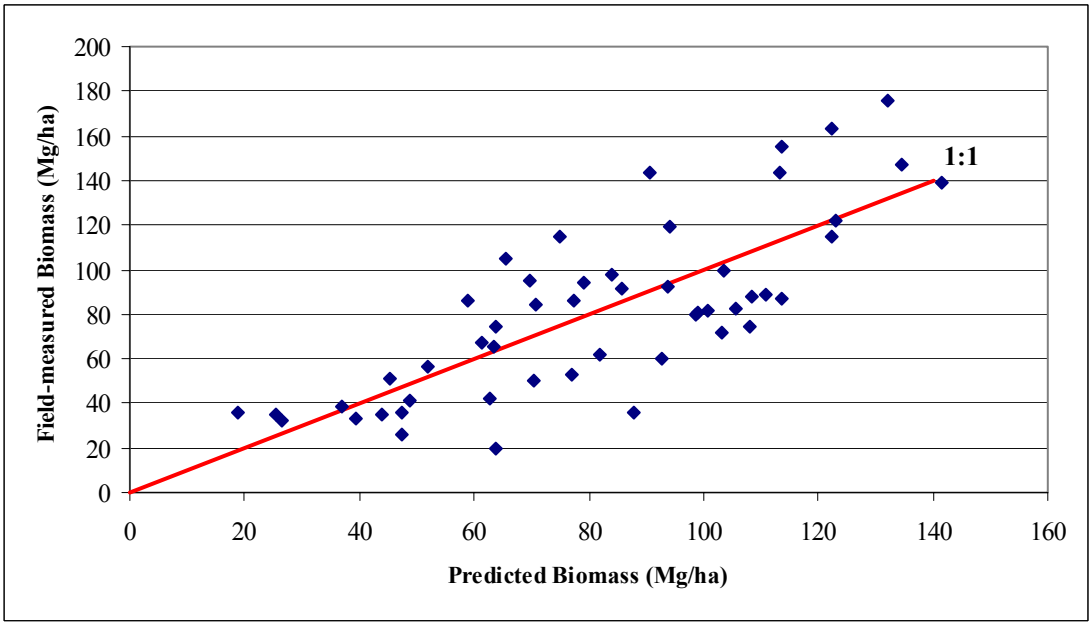


Figure H.24 3-class biomass model (0.091 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.55$)

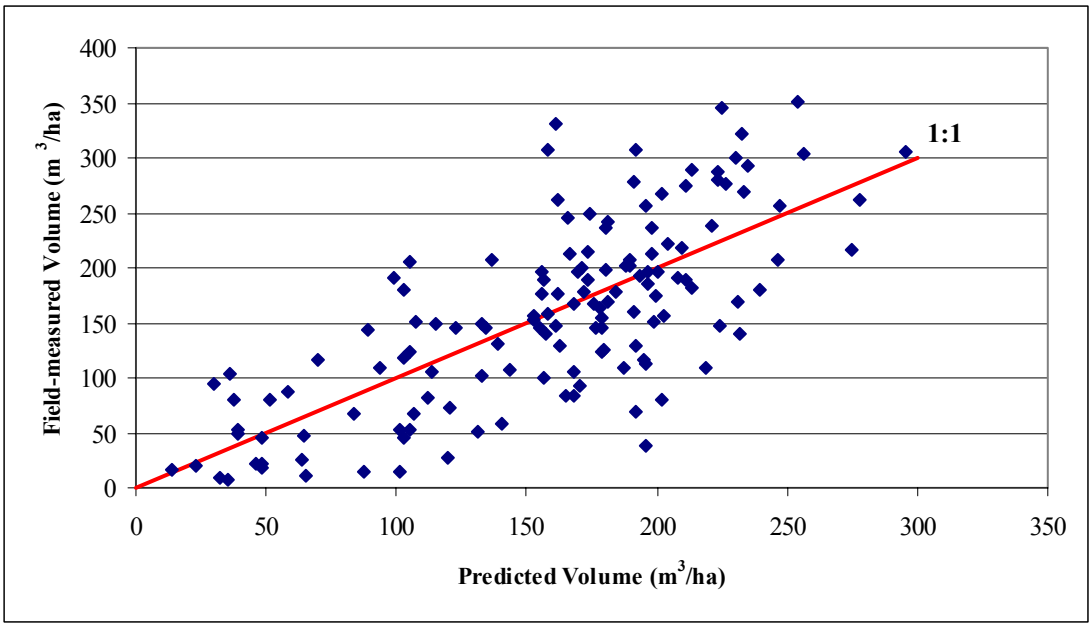


Figure H.25 2-class volume model (0.141 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.52$)

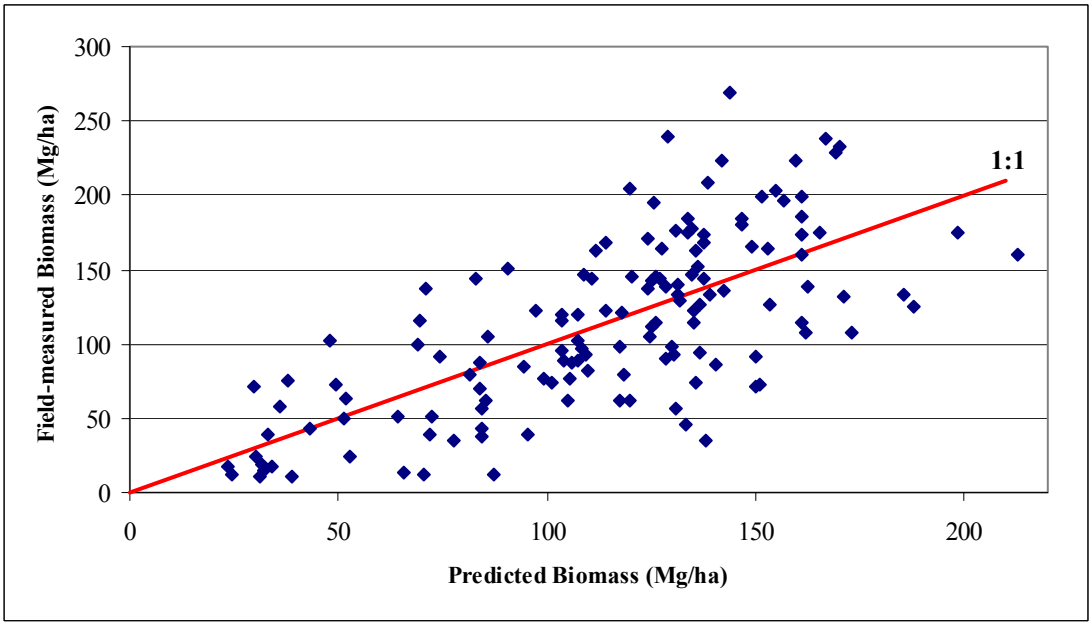


Figure H.26 2-class biomass model (0.141 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.48$)

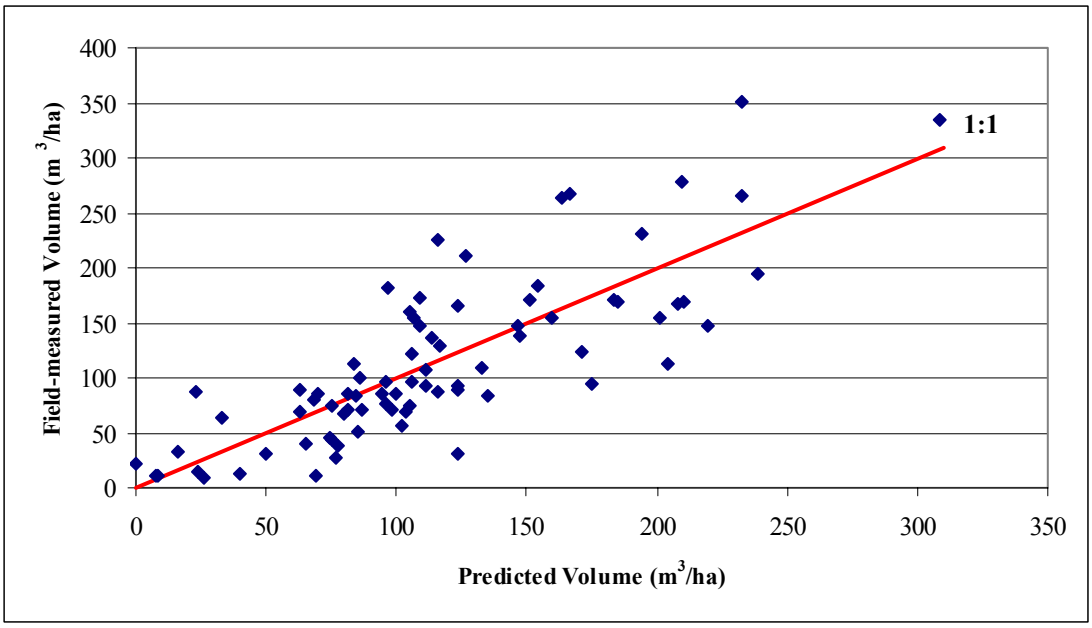


Figure H.27 2-class volume model (0.141 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.62$)

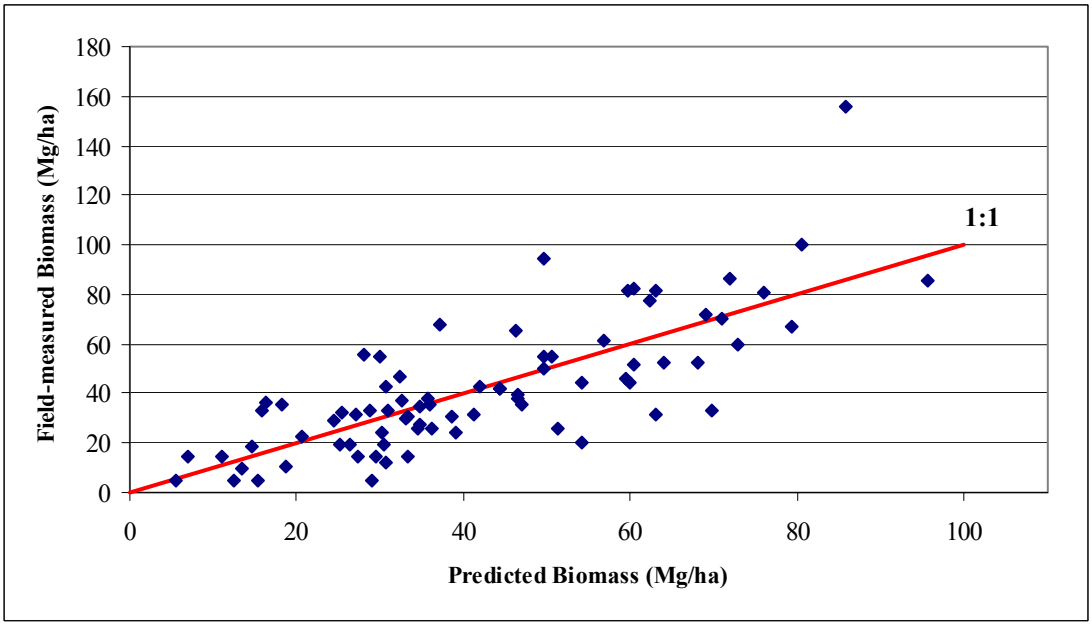


Figure H.28 2-class biomass model (0.141 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.57$)

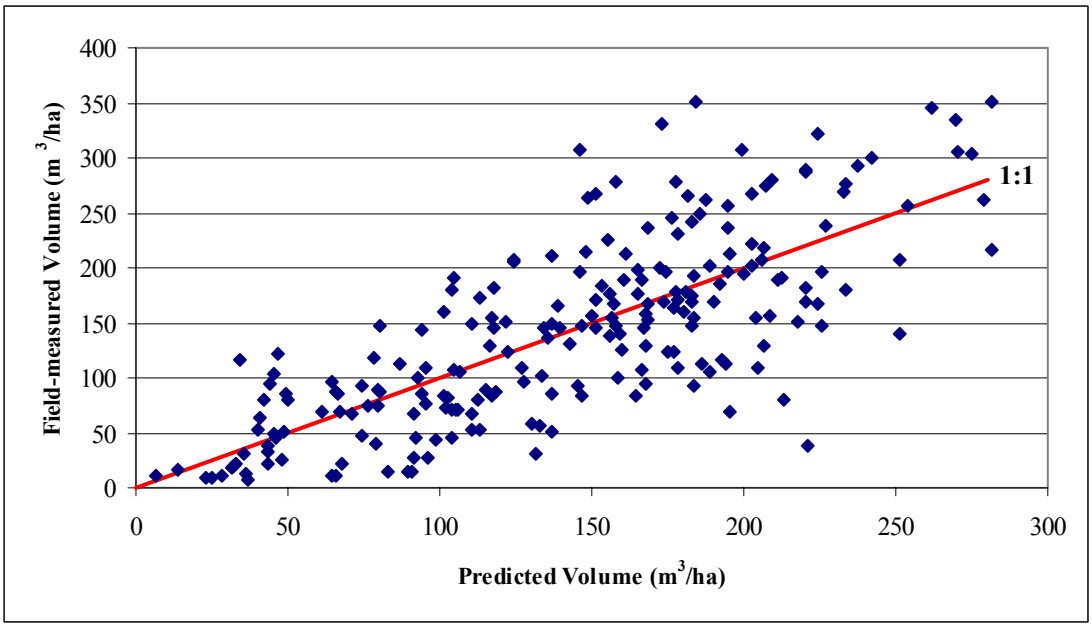


Figure H.29 2-class volume model (0.141 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.56$)

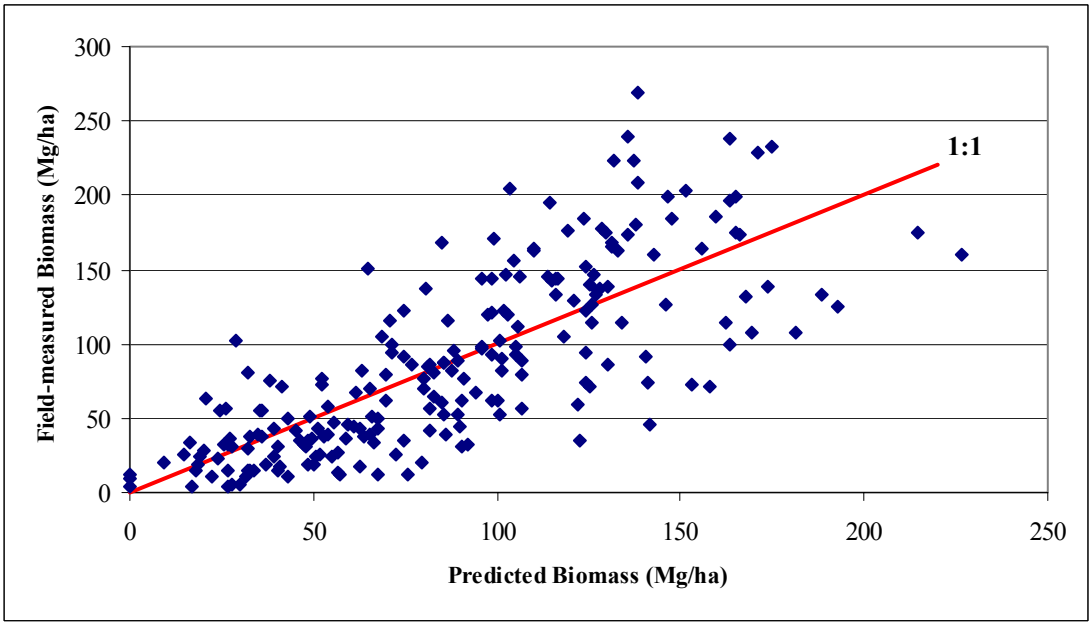


Figure H.30 2-class biomass model (0.141 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.58$)

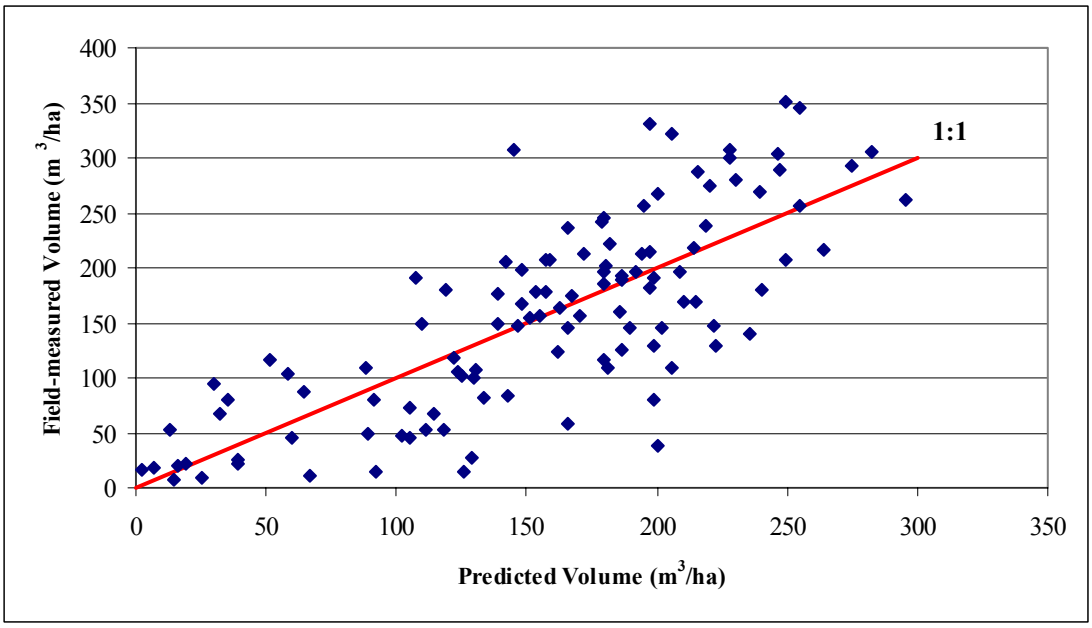


Figure H.31 3-class volume model (0.141 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.58$)

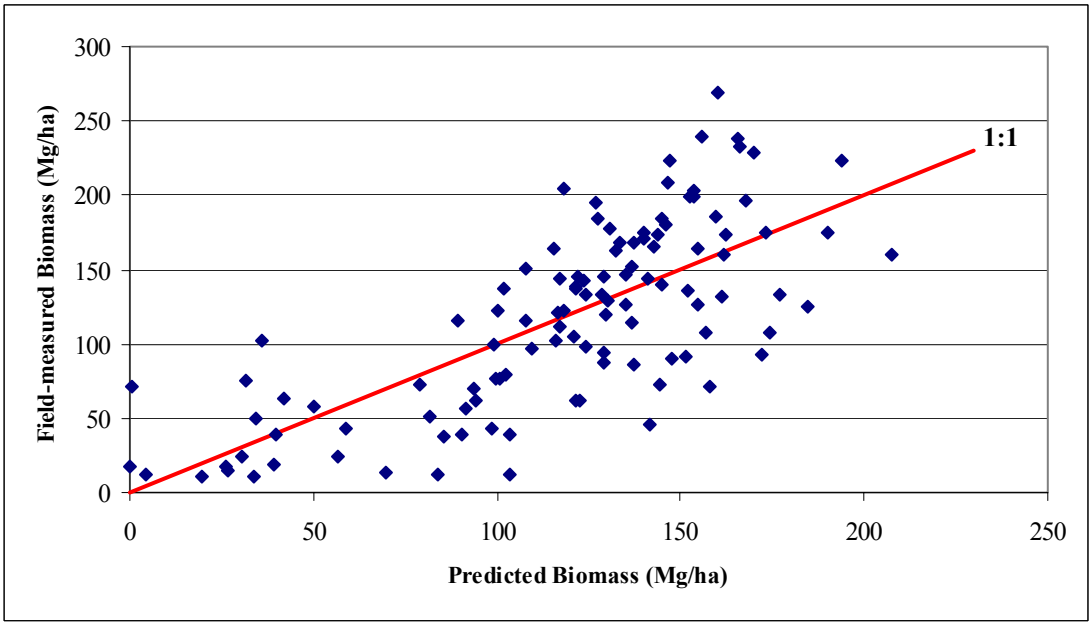


Figure H.32 3-class biomass model (0.141 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.52$)

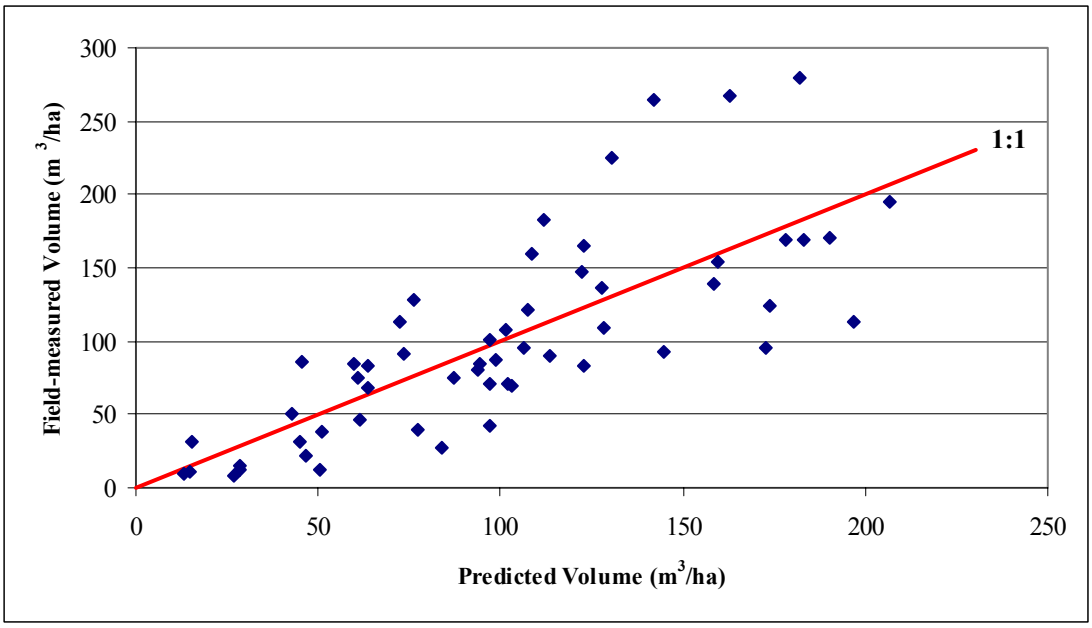


Figure H.33 3-class volume model (0.141 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.56$)

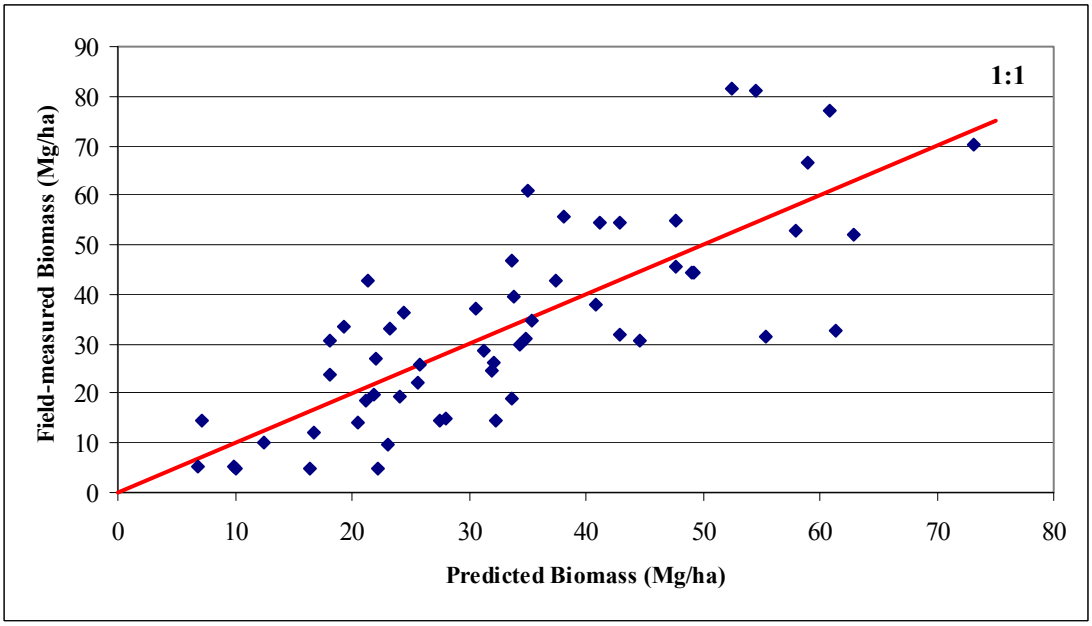


Figure H.34 3-class biomass model (0.141 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.59$)

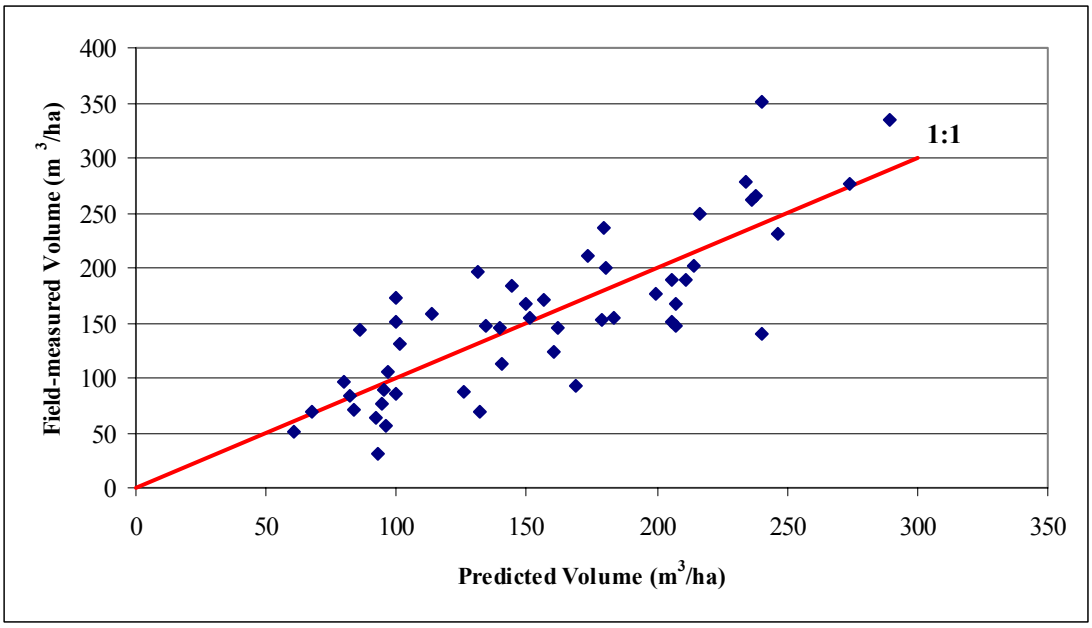


Figure H.35 3-class volume model (0.141 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.62$)

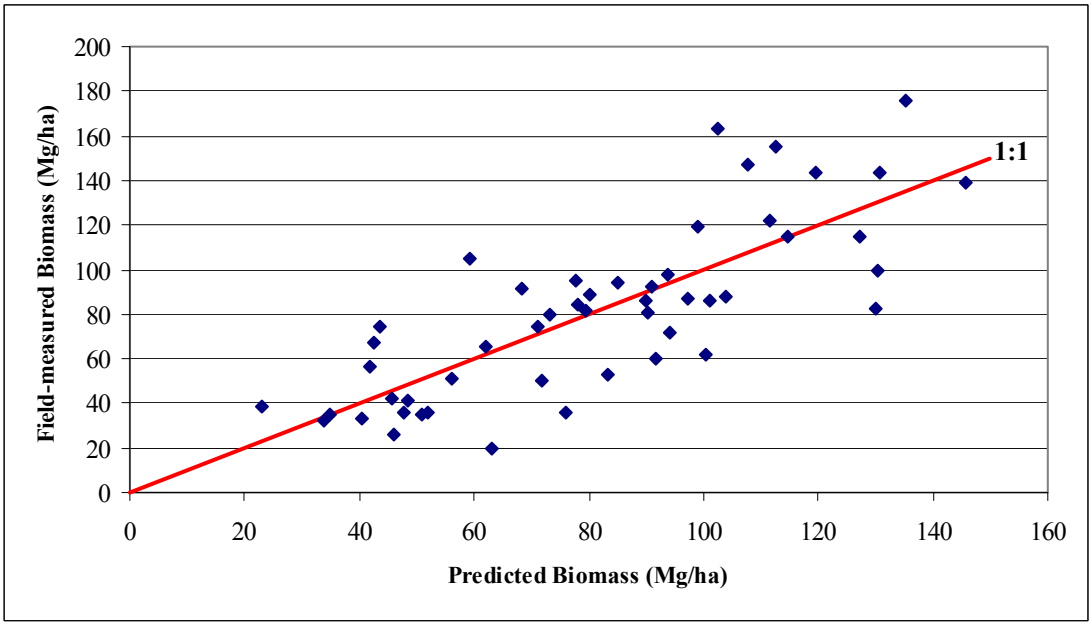


Figure H.36 3-class biomass model (0.141 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.58$)

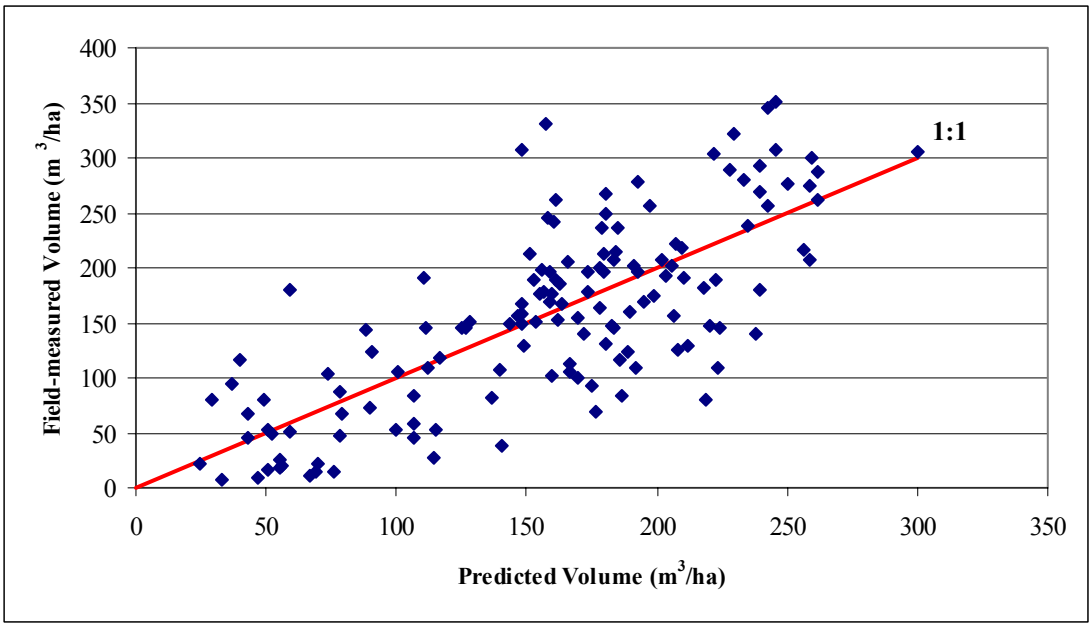


Figure H.37 2-class volume model (0.318 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.55$)

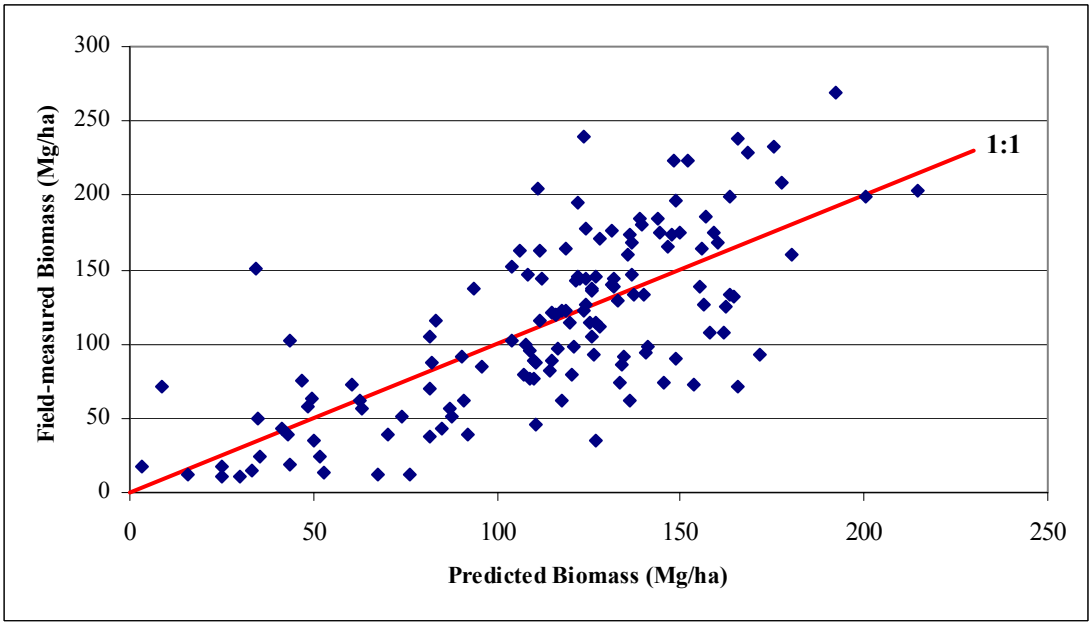


Figure H.38 2-class biomass model (0.318 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.51$)

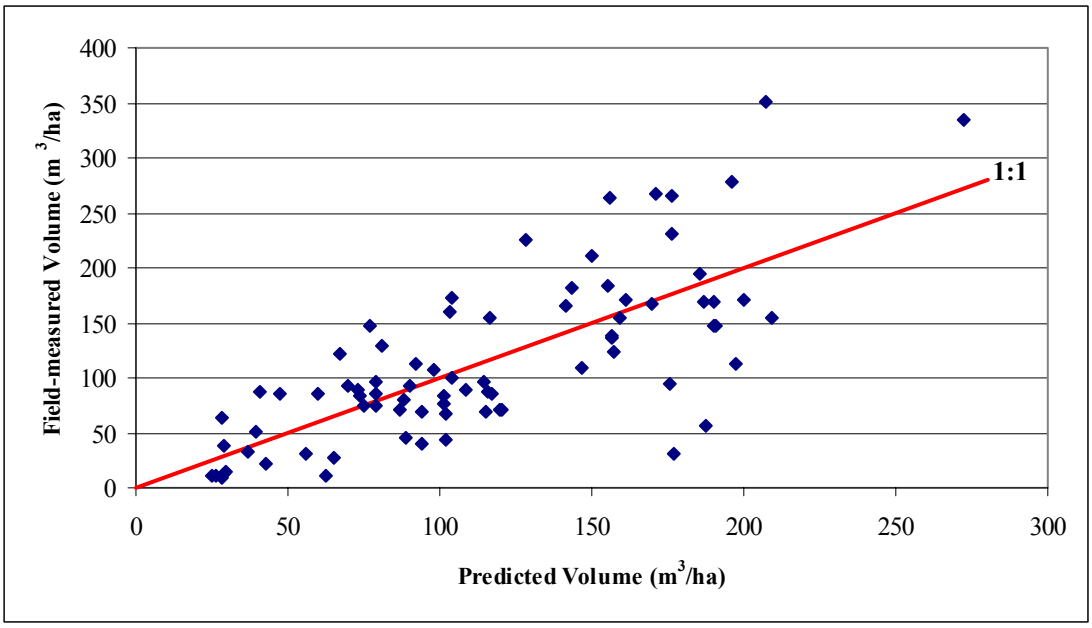


Figure H.39 2-class volume model (0.318 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.54$)

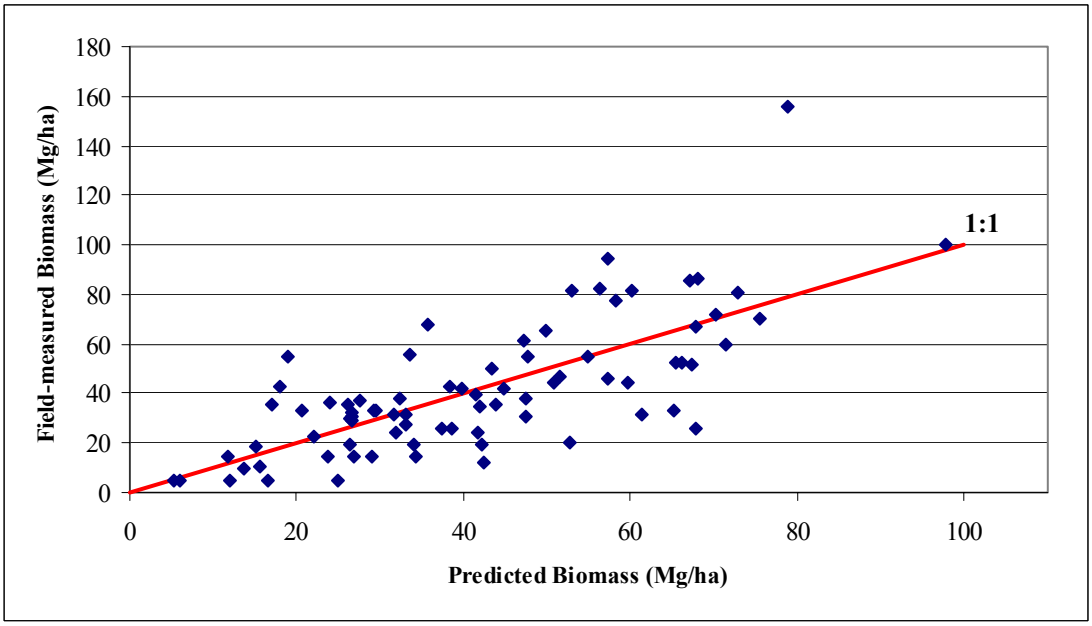


Figure H.40 2-class biomass model (0.318 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.51$)

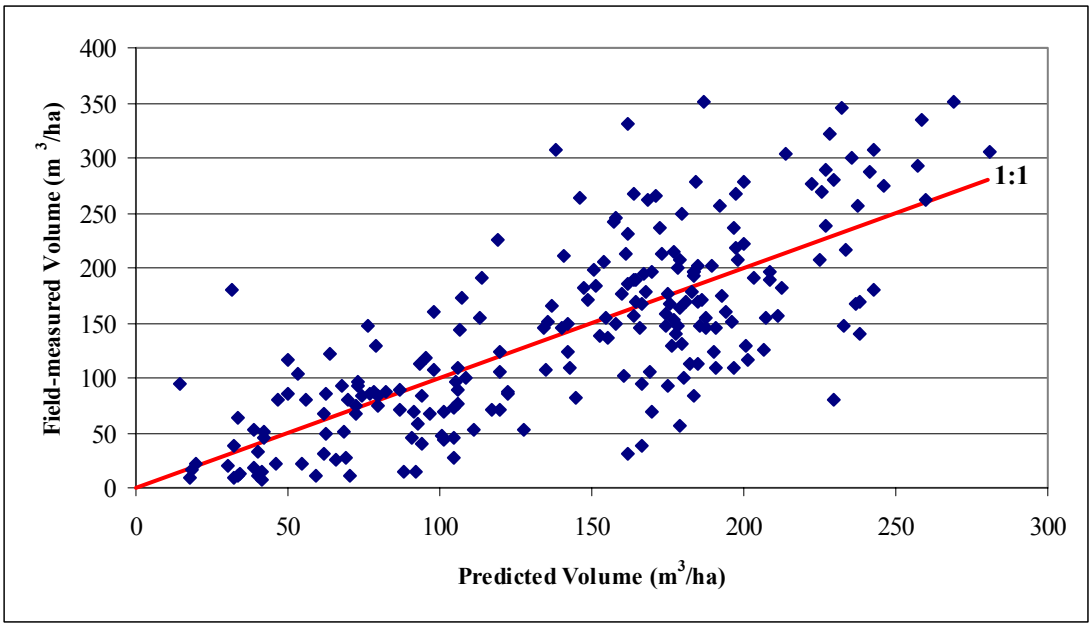


Figure H.41 2-class volume model (0.318 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.56$)

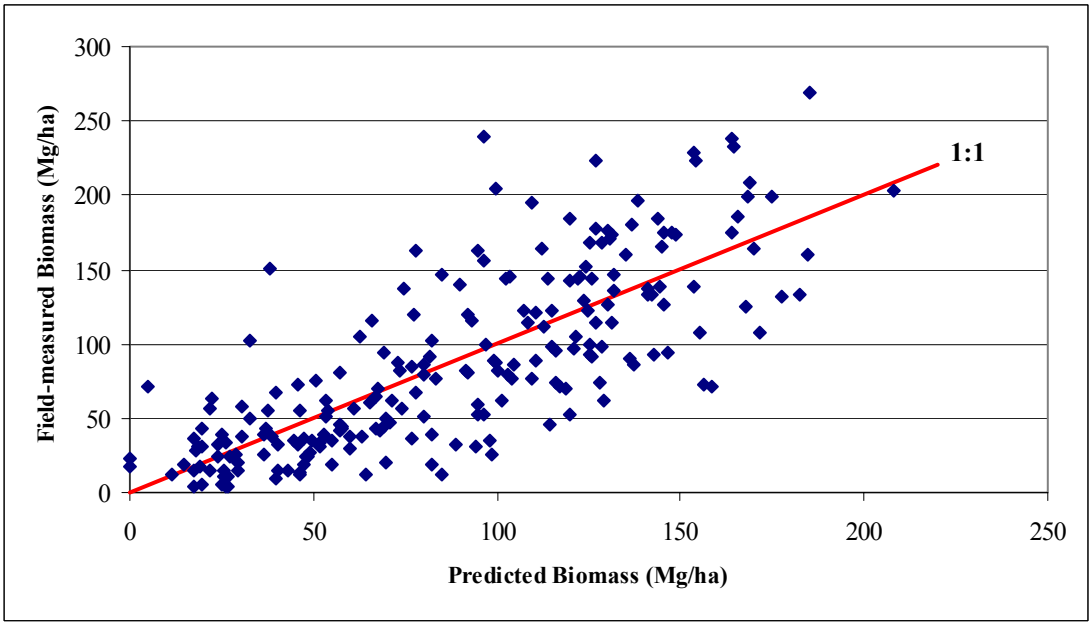


Figure H.42 2-class biomass model (0.318 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.60$)

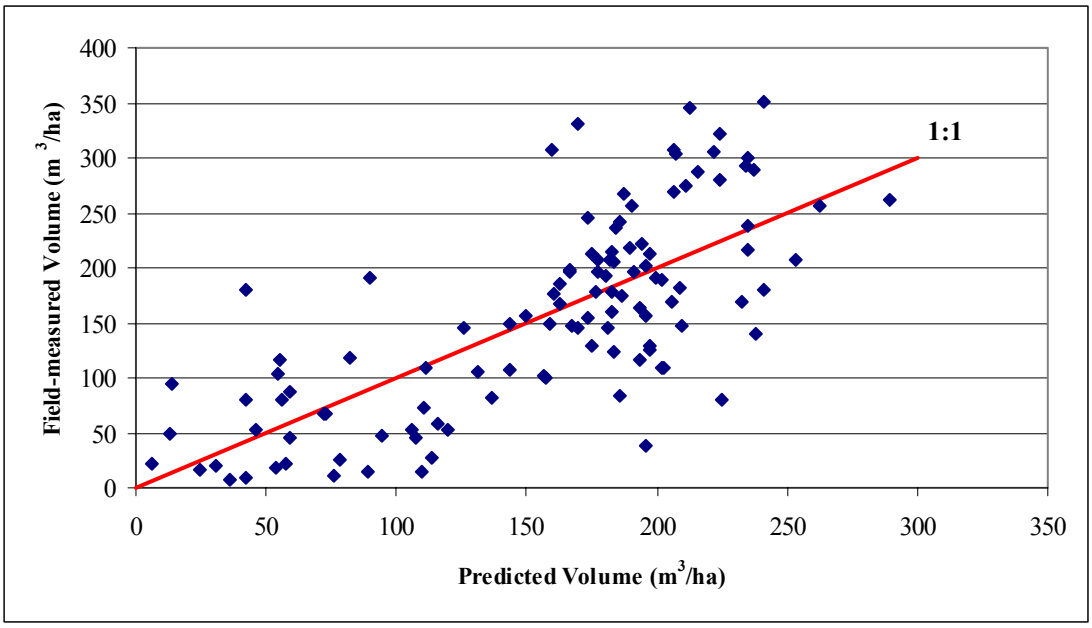


Figure H.43 3-class volume model (0.318 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.53$)

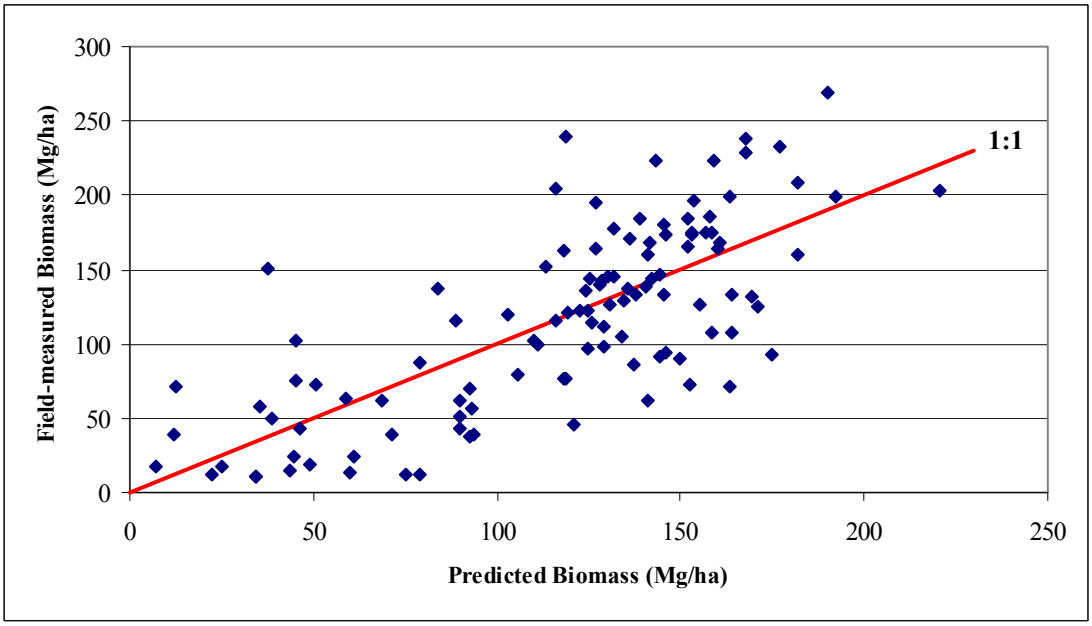


Figure H.44 3-class biomass model (0.318 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.53$)

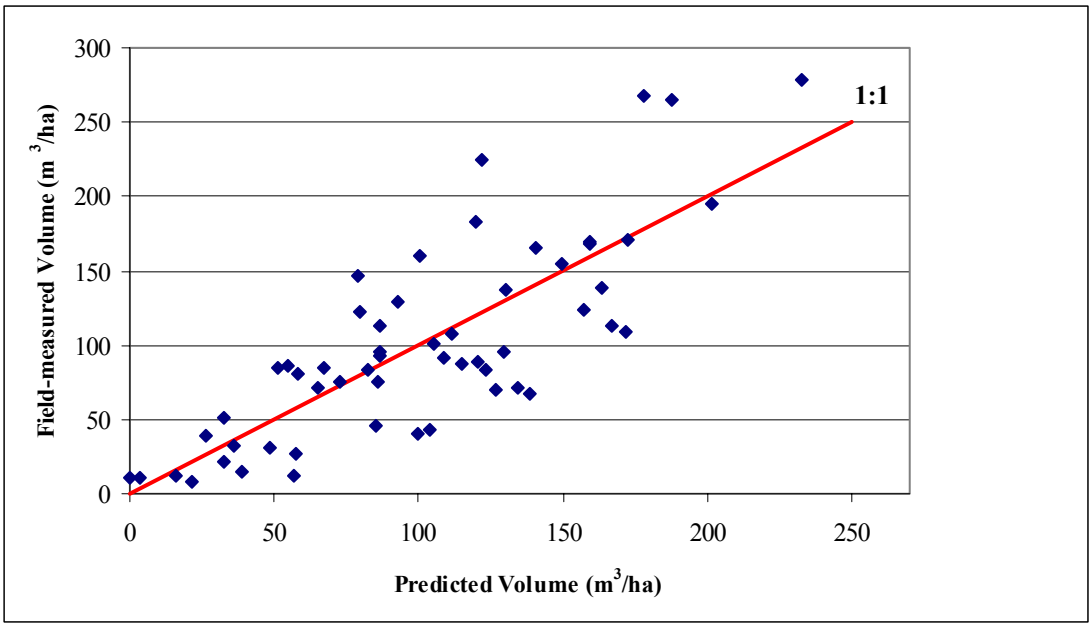


Figure H.45 3-class volume model (0.318 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.61$)

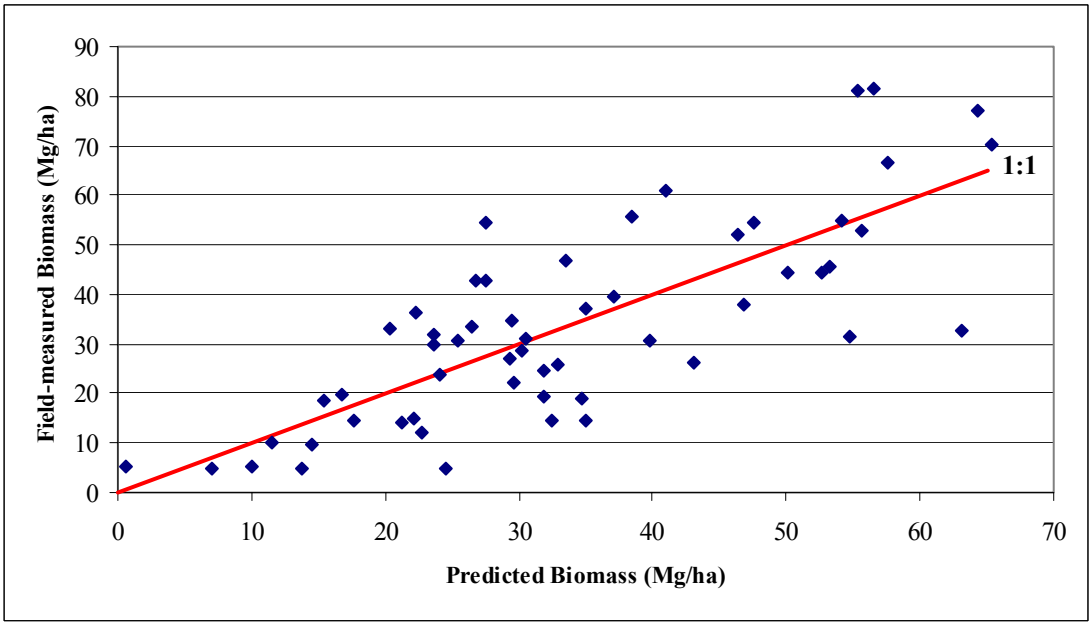


Figure H.46 3-class biomass model (0.318 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.58$)

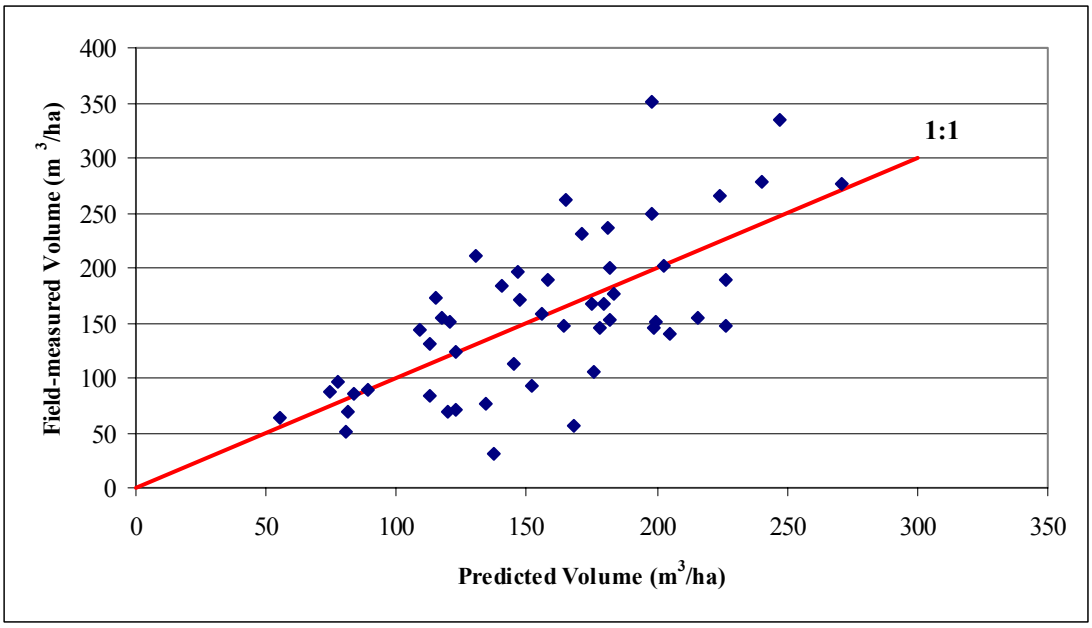


Figure H.47 3-class volume model (0.318 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.43$)

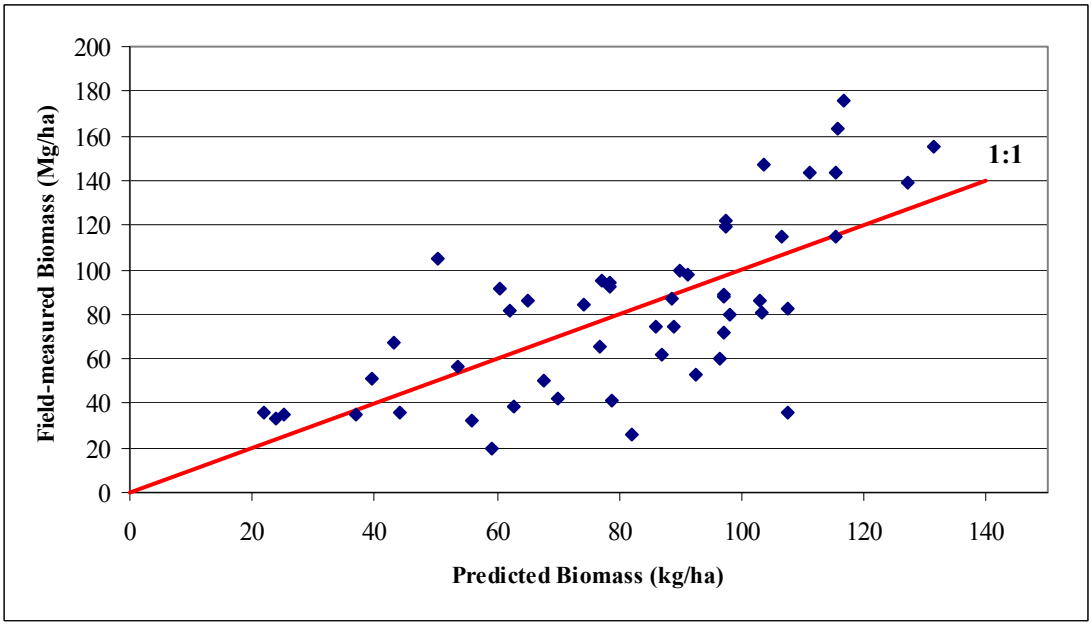


Figure H.48 3-class biomass model (0.318 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.46$)

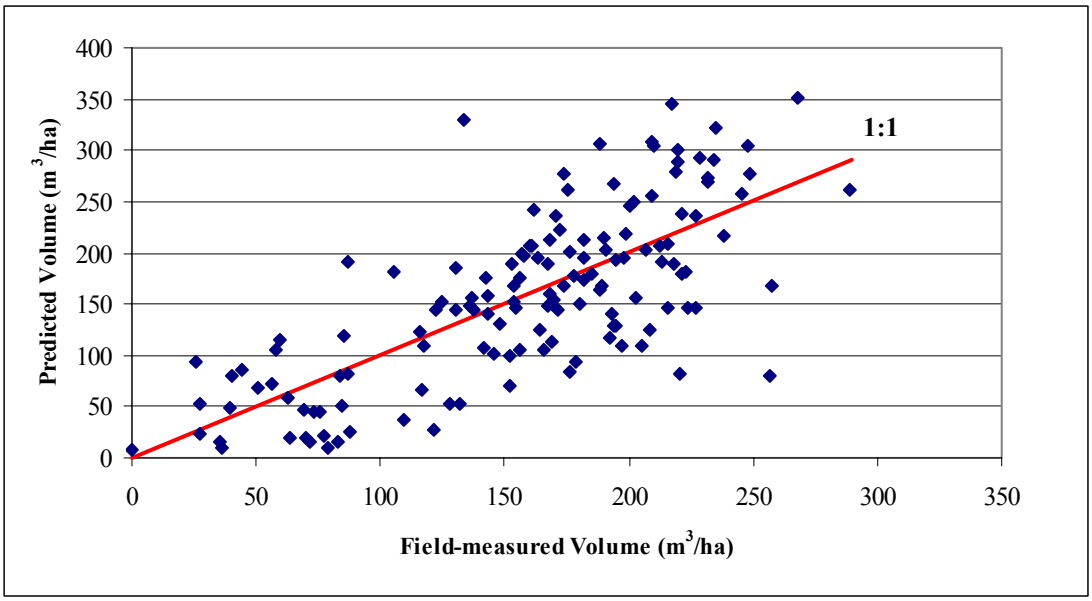


Figure H.49 2-class volume model (0.642 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.52$)

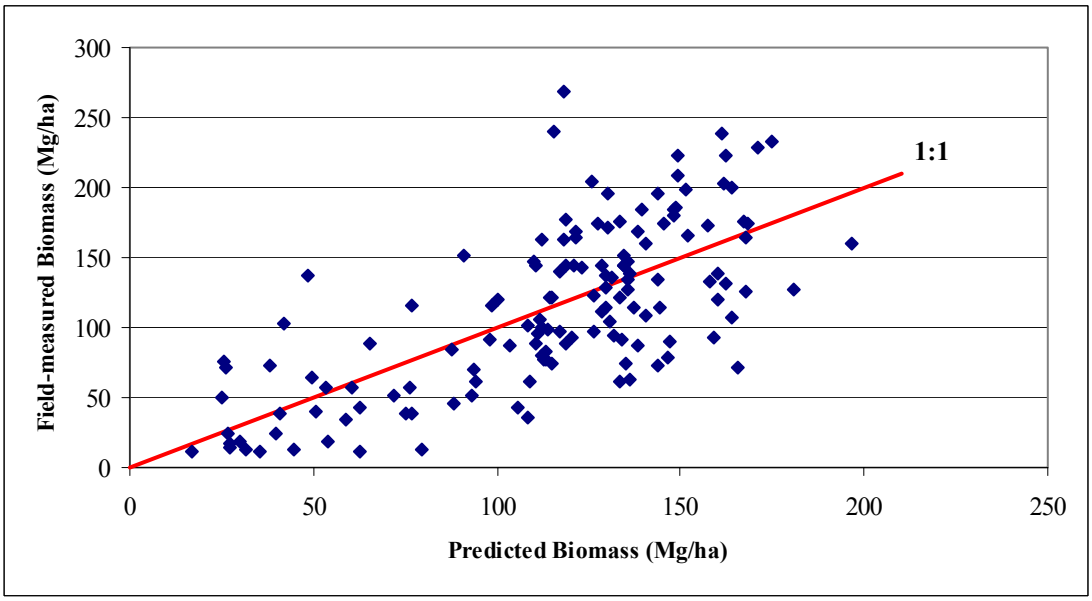


Figure H.50 2-class biomass model (0.642 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.48$)

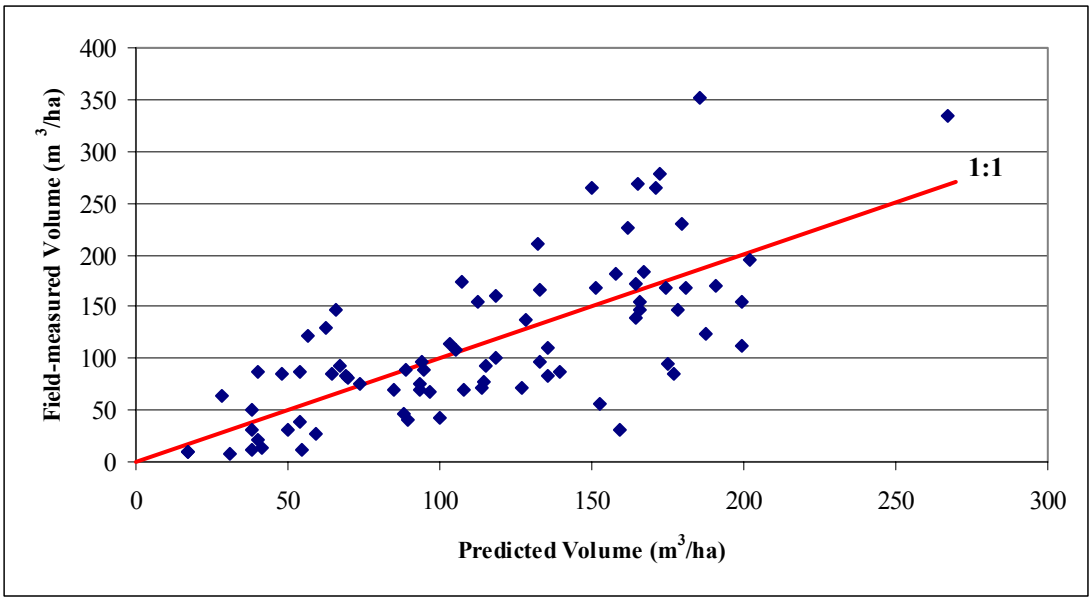


Figure H.51 2-class volume model (0.642 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.51$)

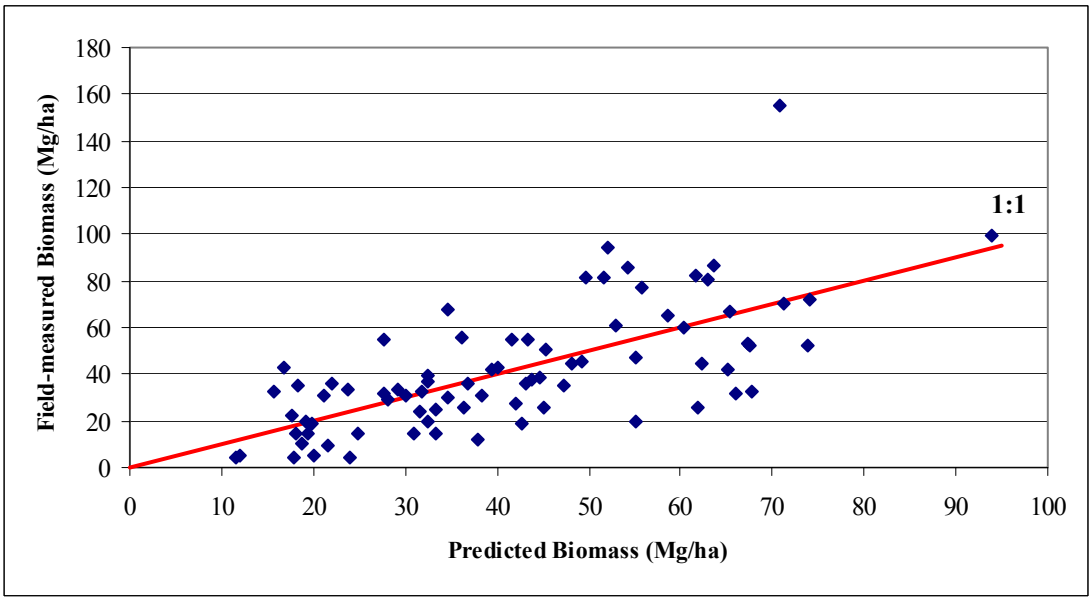


Figure H.52 2-class biomass model (0.642 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.46$)

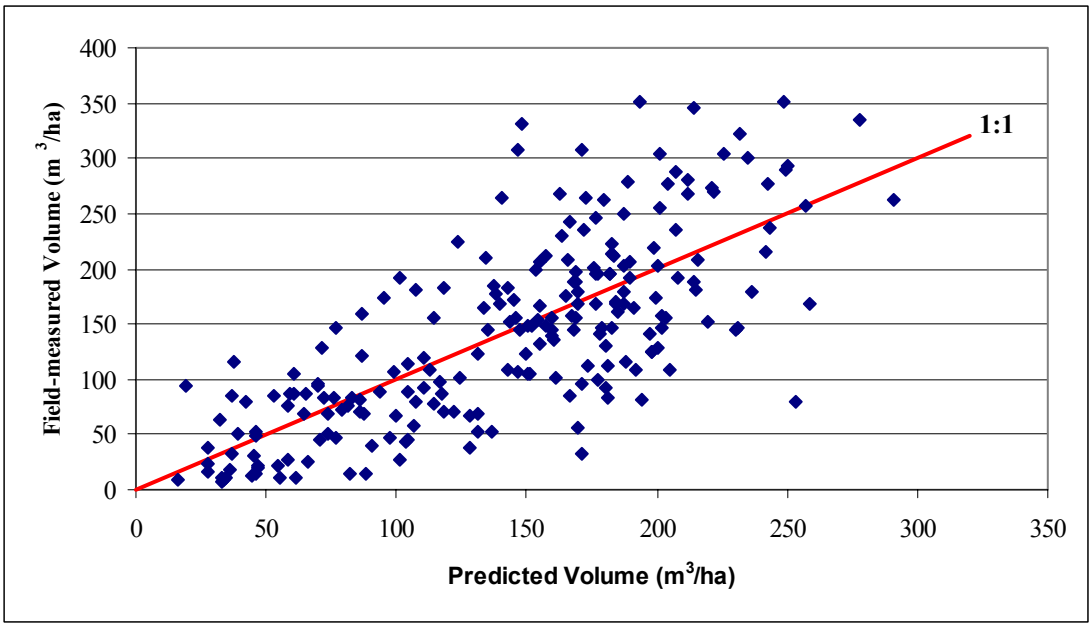


Figure H.53 2-class volume model (0.642 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.54$)

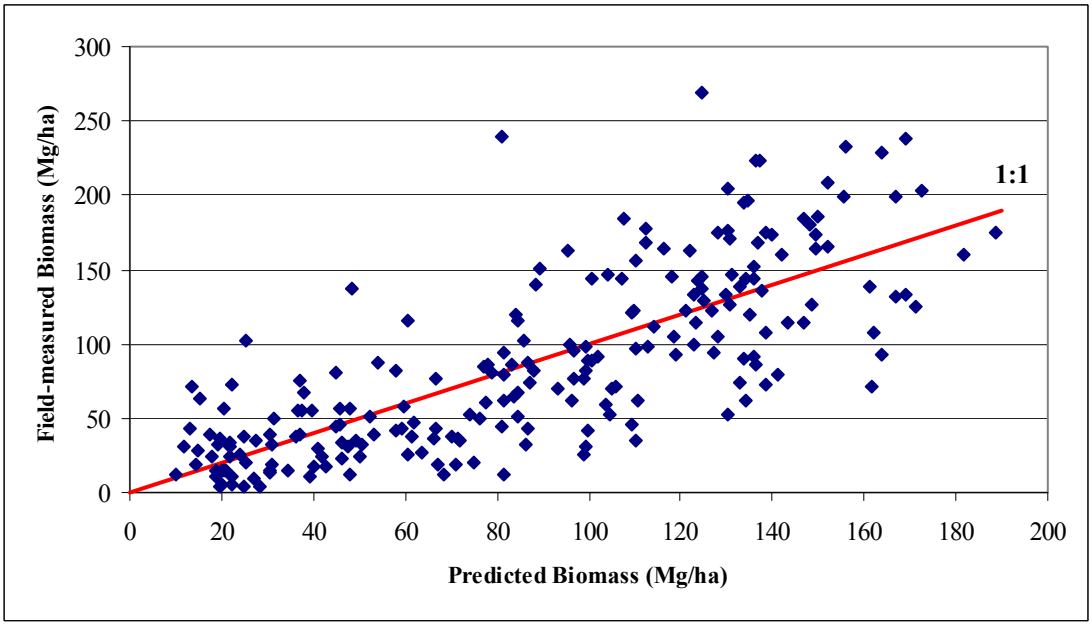


Figure H.54 2-class biomass model (0.642 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.58$)

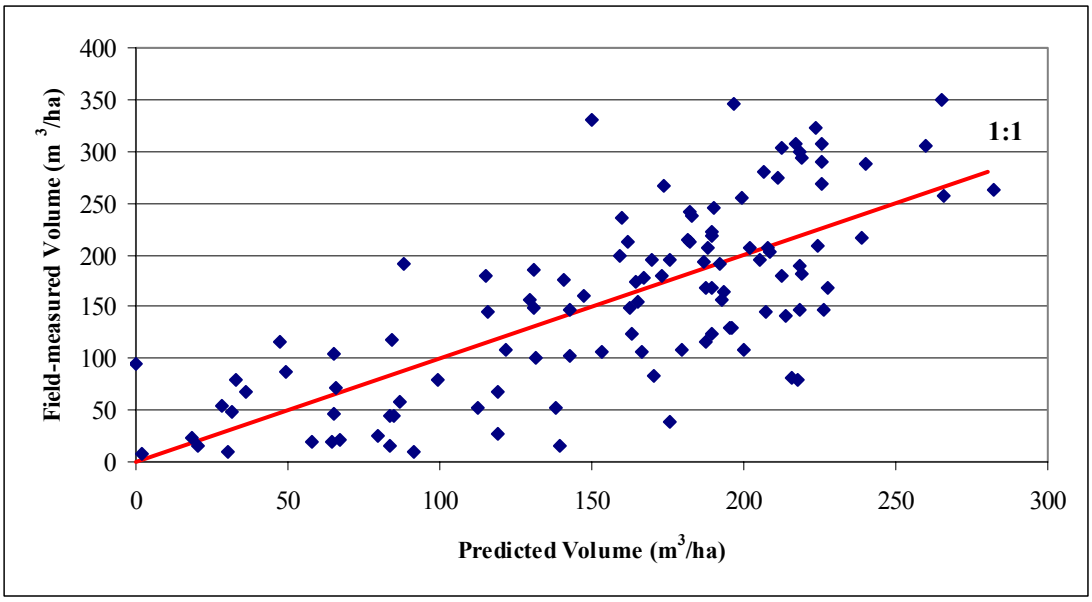


Figure H.55 3-class volume model (0.642 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.52$)

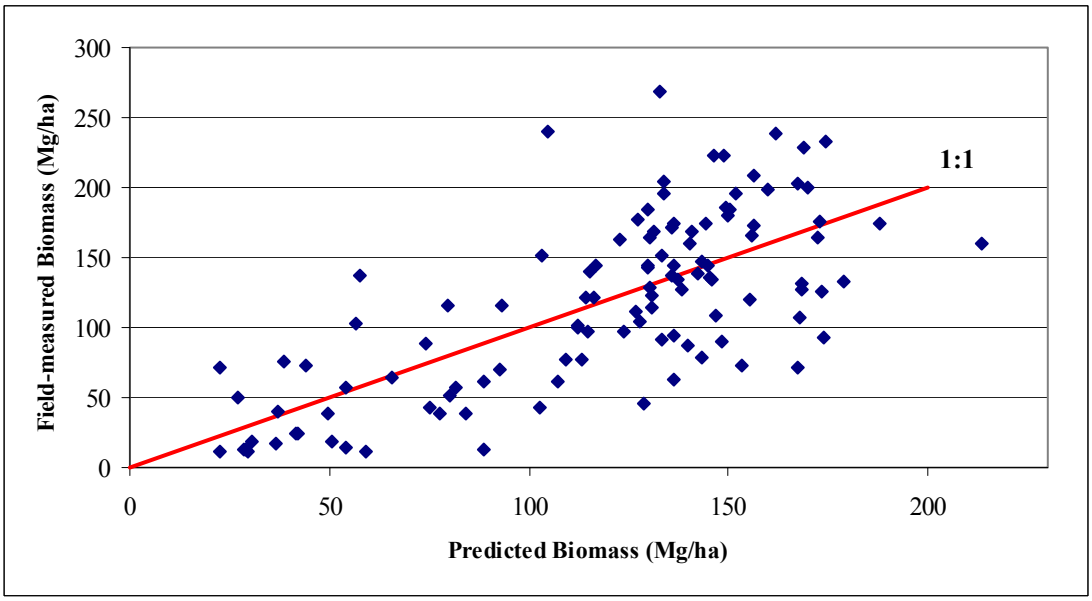


Figure H.56 3-class biomass model (0.642 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.50$)

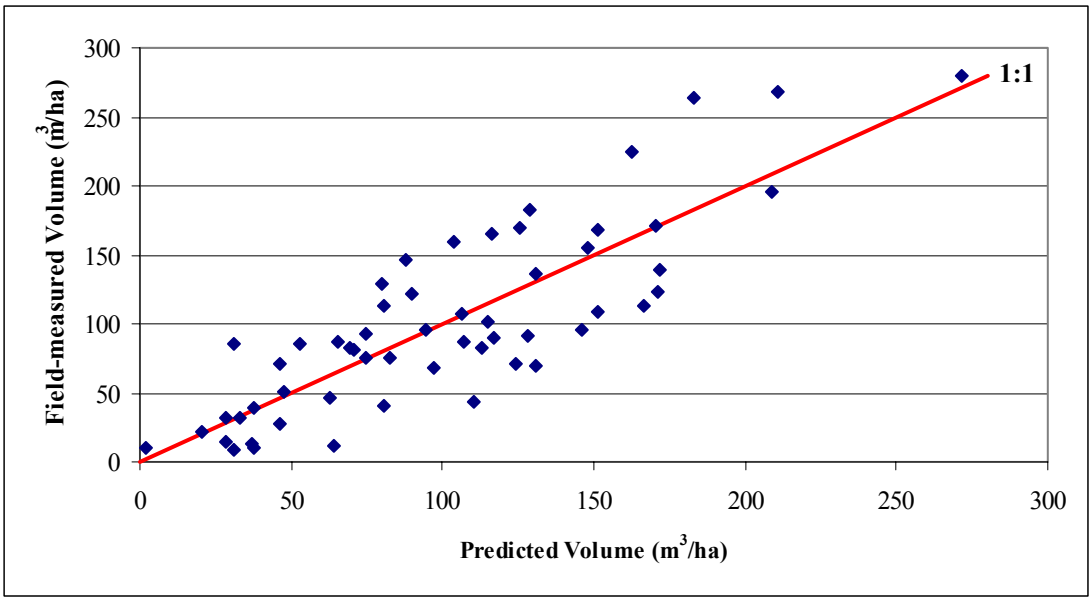


Figure H.57 3-class volume model (0.642 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.67$)

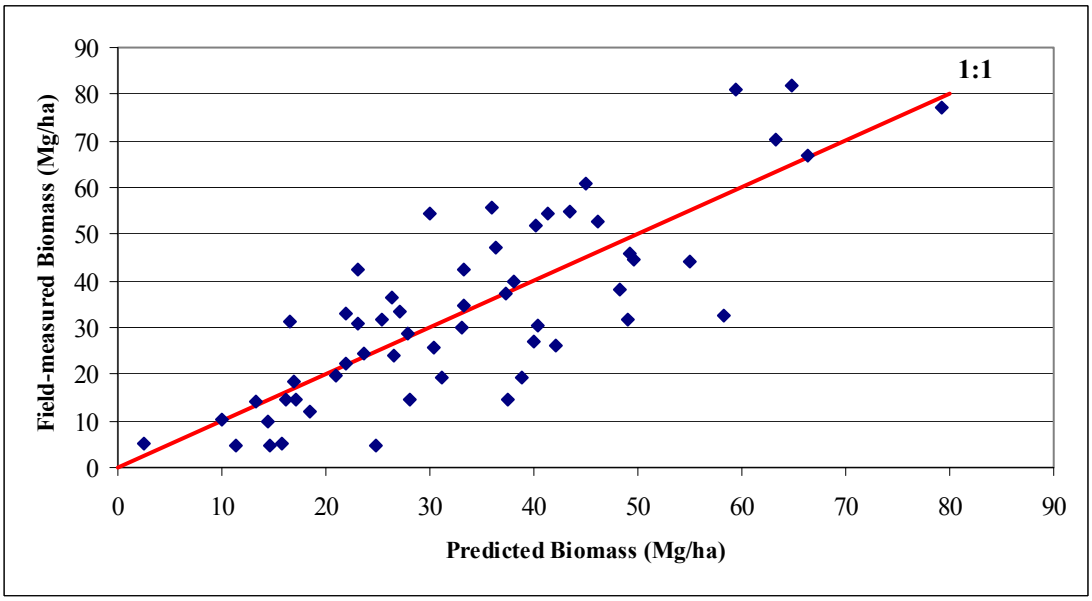


Figure H.58 3-class biomass model (0.642 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.62$)

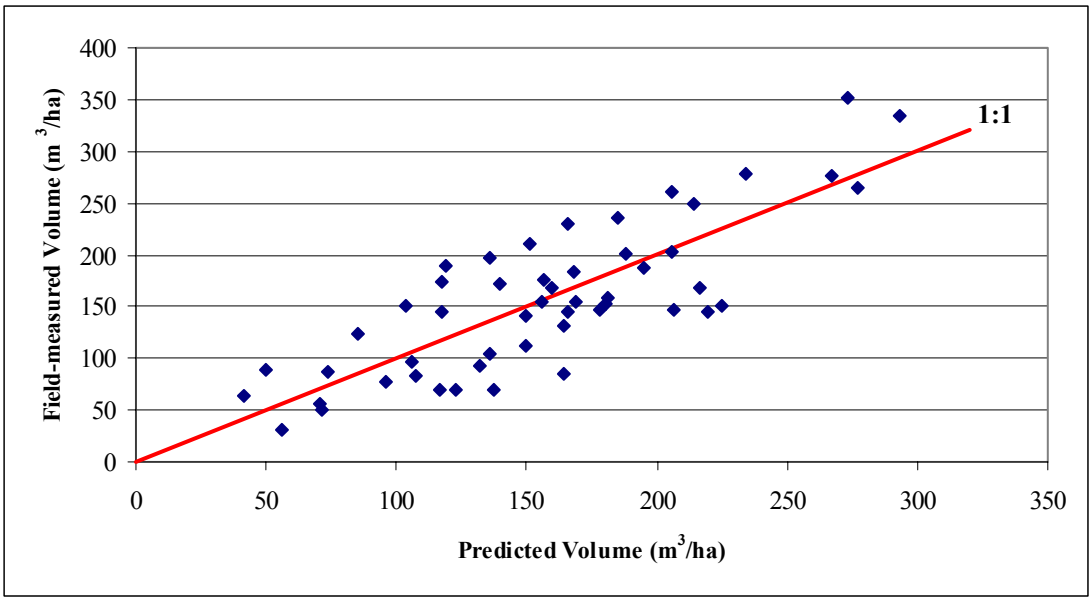


Figure H.59 3-class volume model (0.642 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.63$)

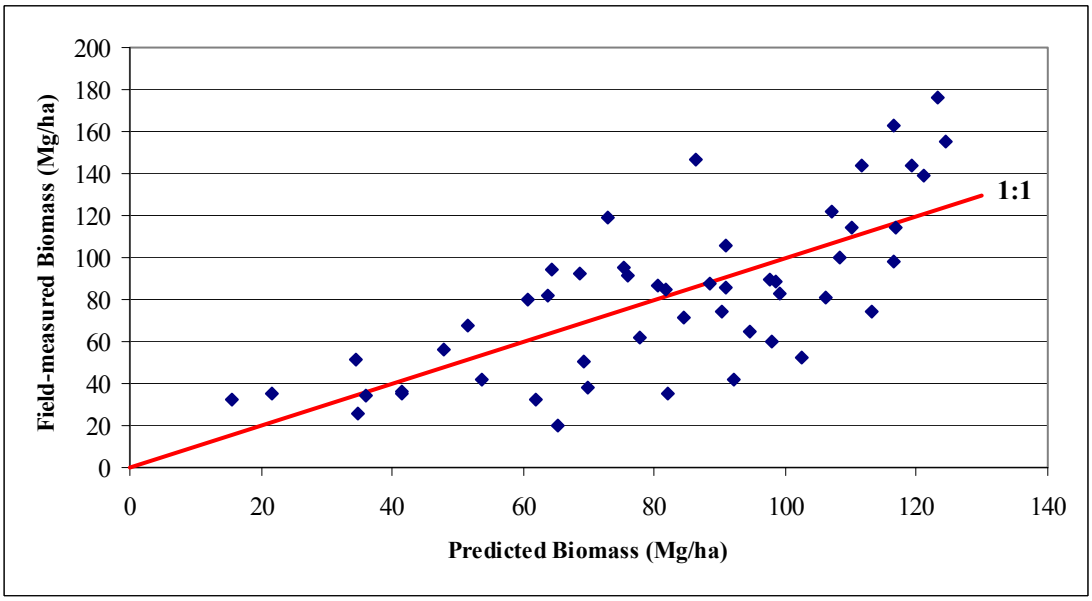


Figure H.60 3-class biomass model (0.642 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.50$)

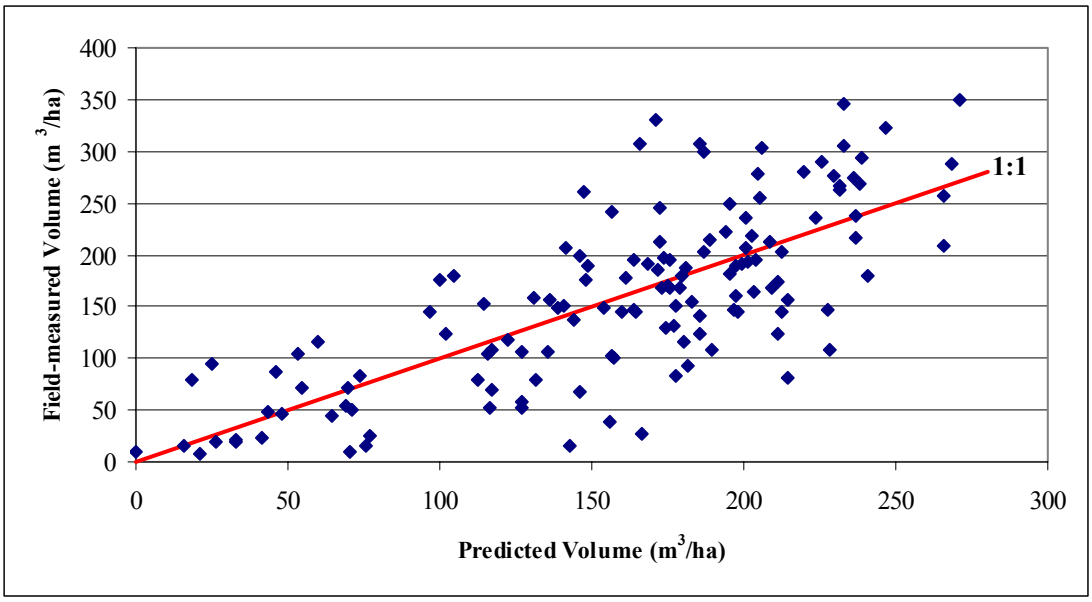


Figure H.61 2-class volume model (0.964 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.54$)

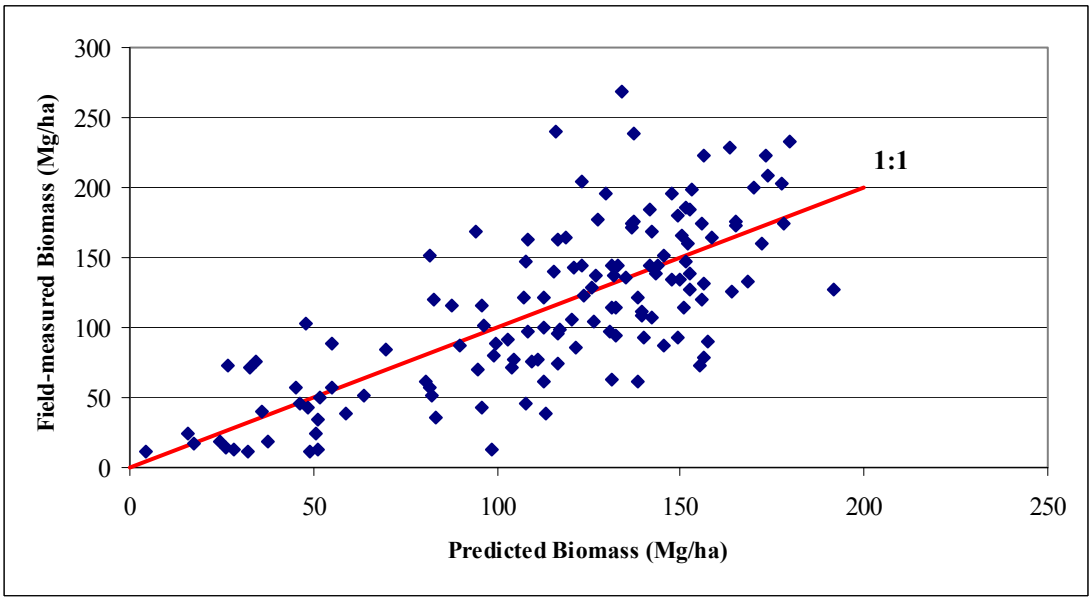


Figure H.62 2-class biomass model (0.964 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.52$)

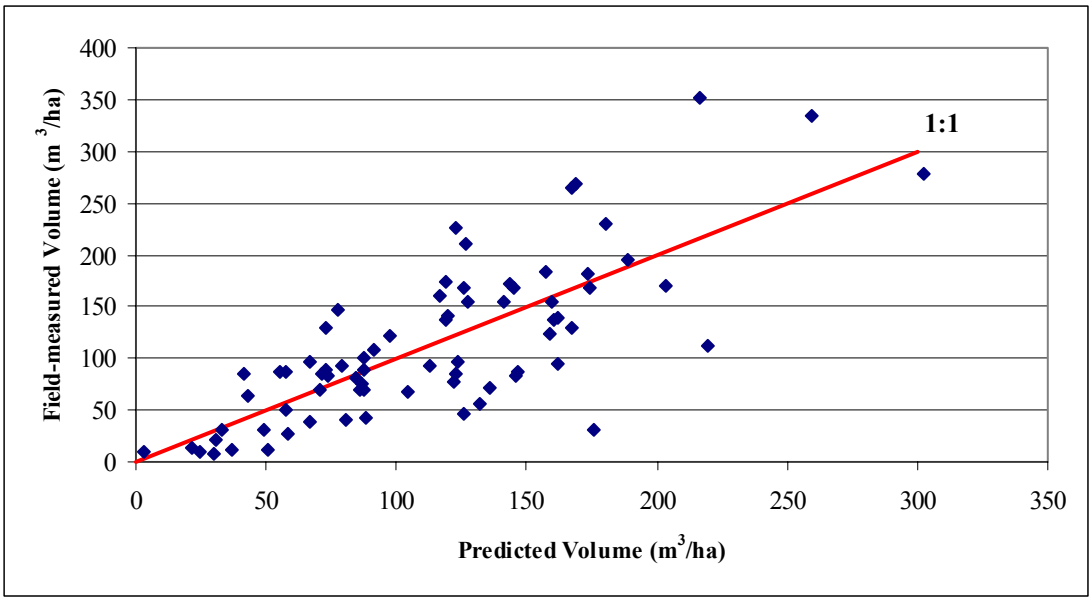


Figure H.63 2-class volume model (0.964 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.55$)

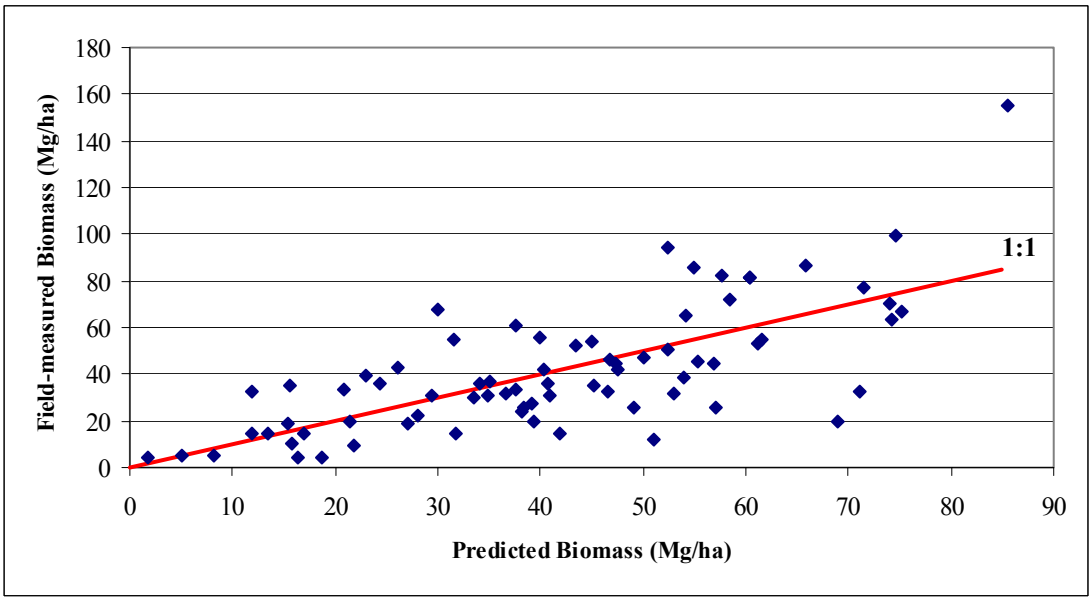


Figure H.64 2-class biomass model (0.964 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.47$)

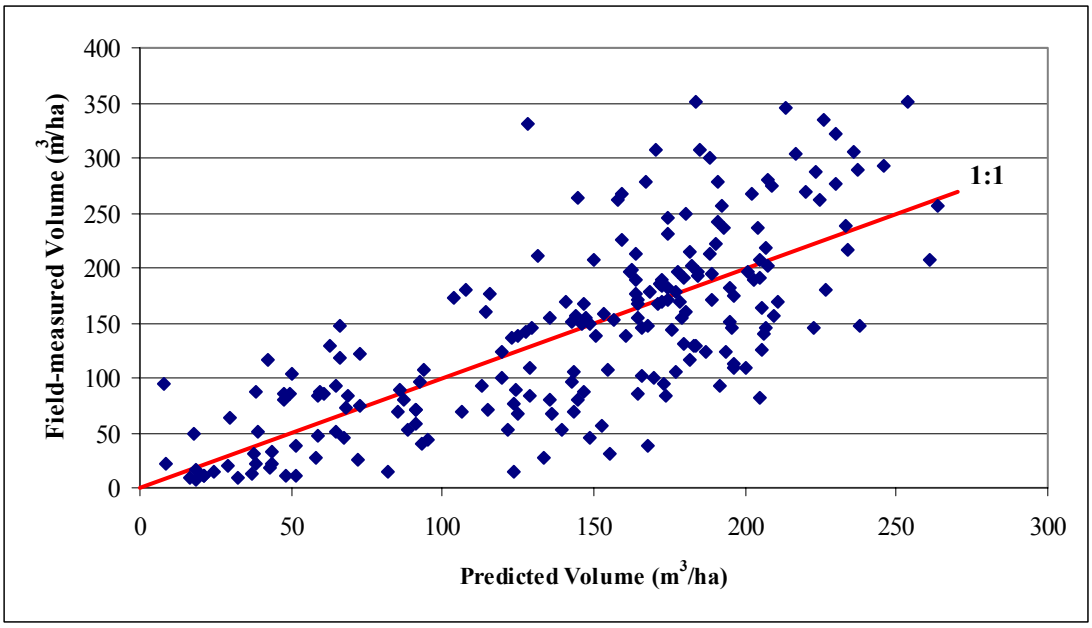


Figure H.65 2-class volume model (0.964 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.54$)

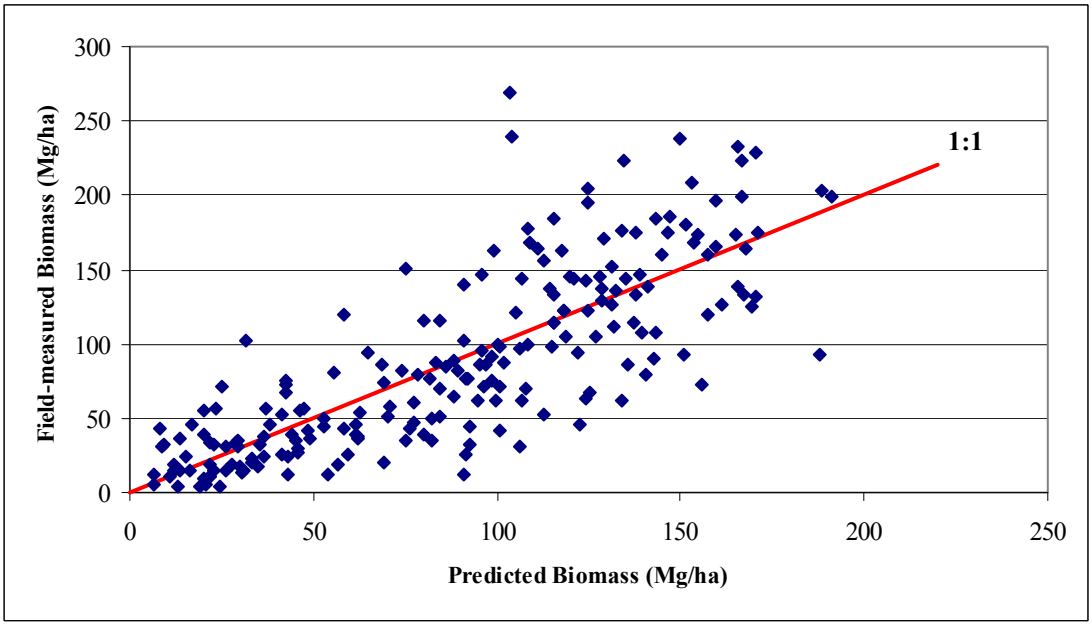


Figure H.66 2-class biomass model (0.964 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.62$)

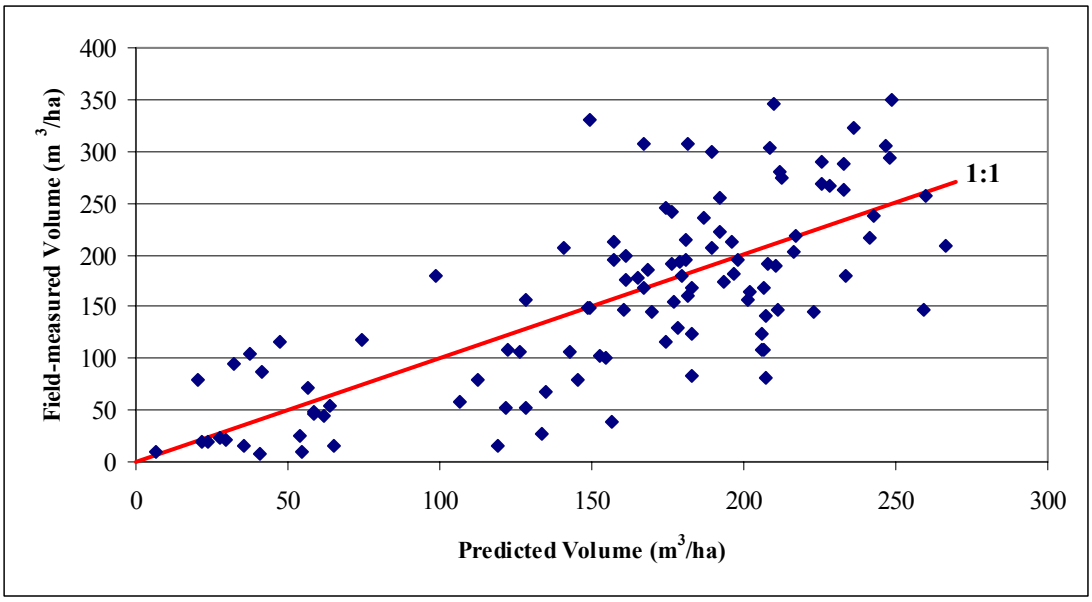


Figure H.67 3-class volume model (0.964 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.54$)

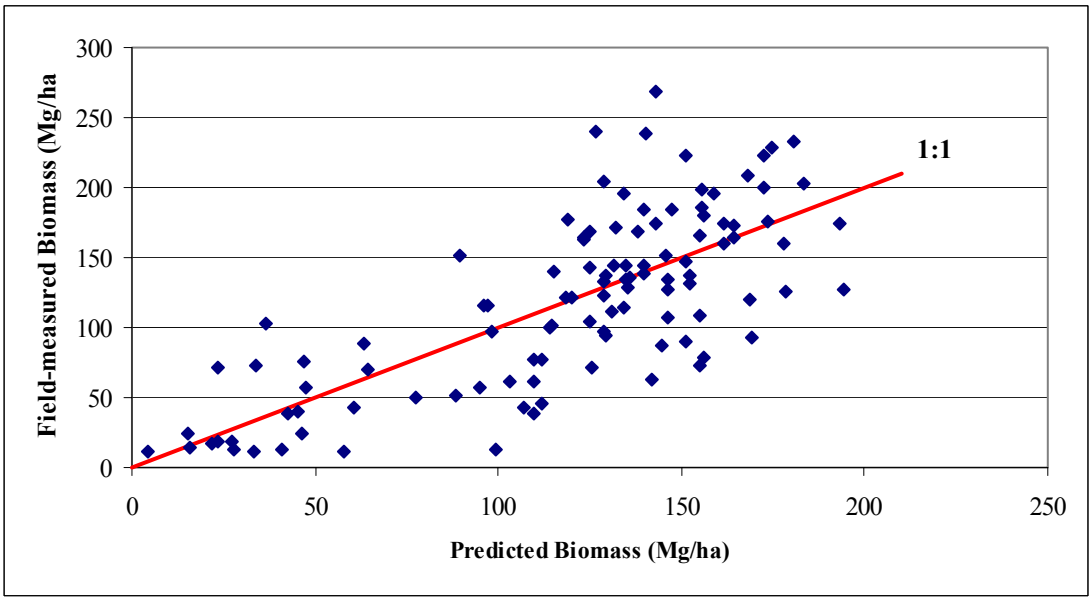


Figure H.68 3-class biomass model (0.964 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.54$)

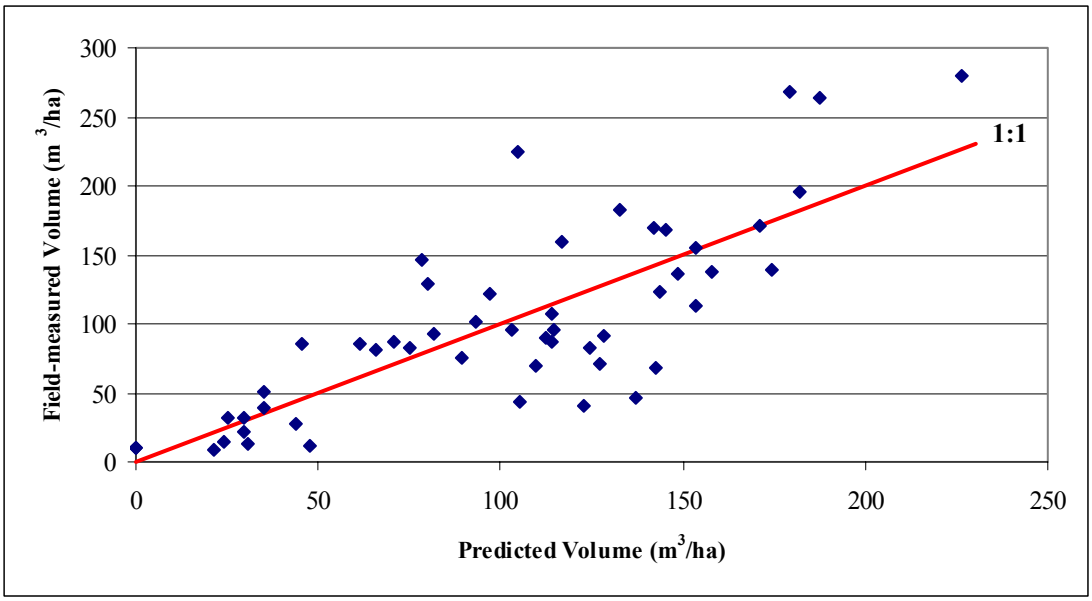


Figure H.69 3-class volume model (0.964 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.59$)

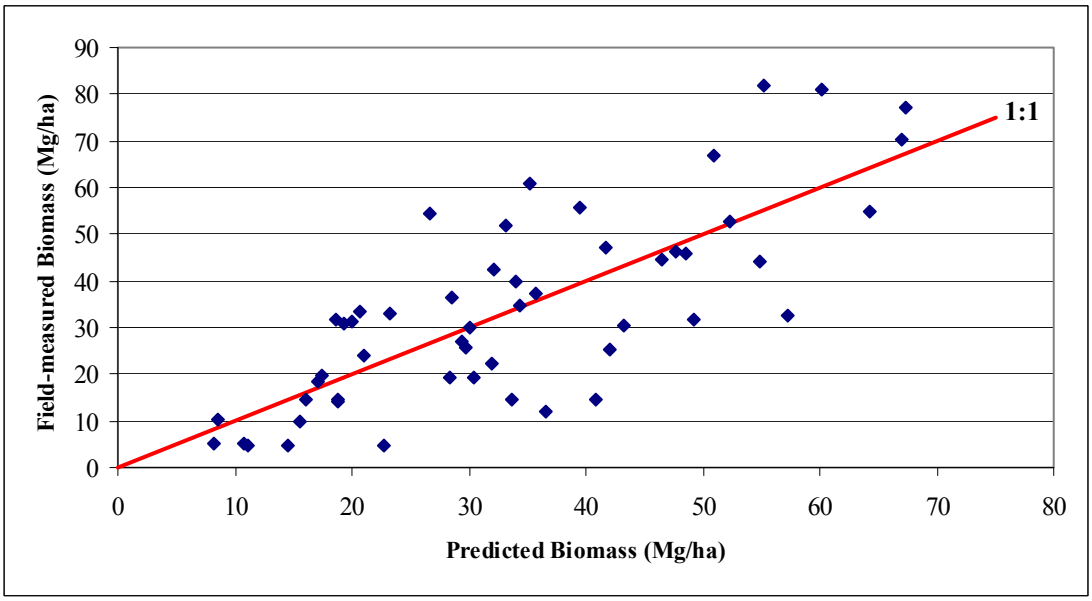


Figure H.70 3-class biomass model (0.964 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.56$)

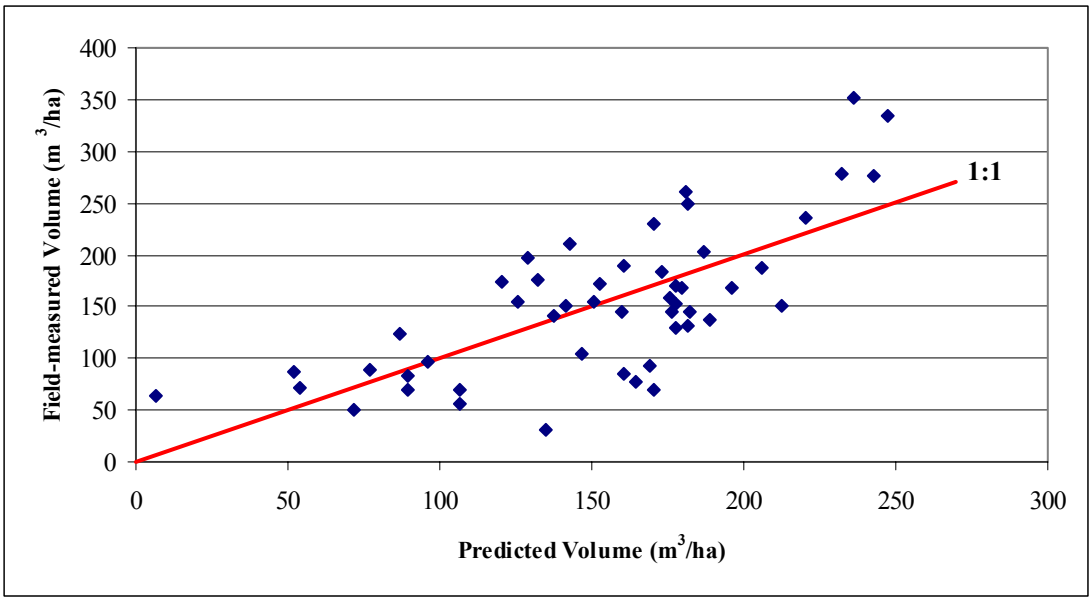


Figure H.71 3-class volume model (0.964 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.48$)

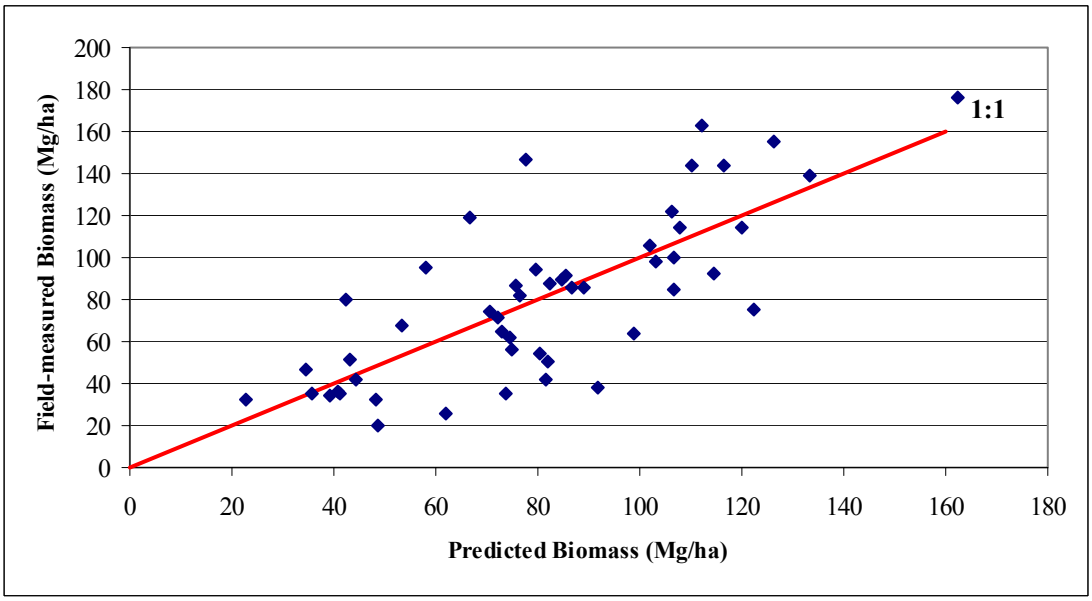


Figure H.72 3-class biomass model (0.964 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.55$)

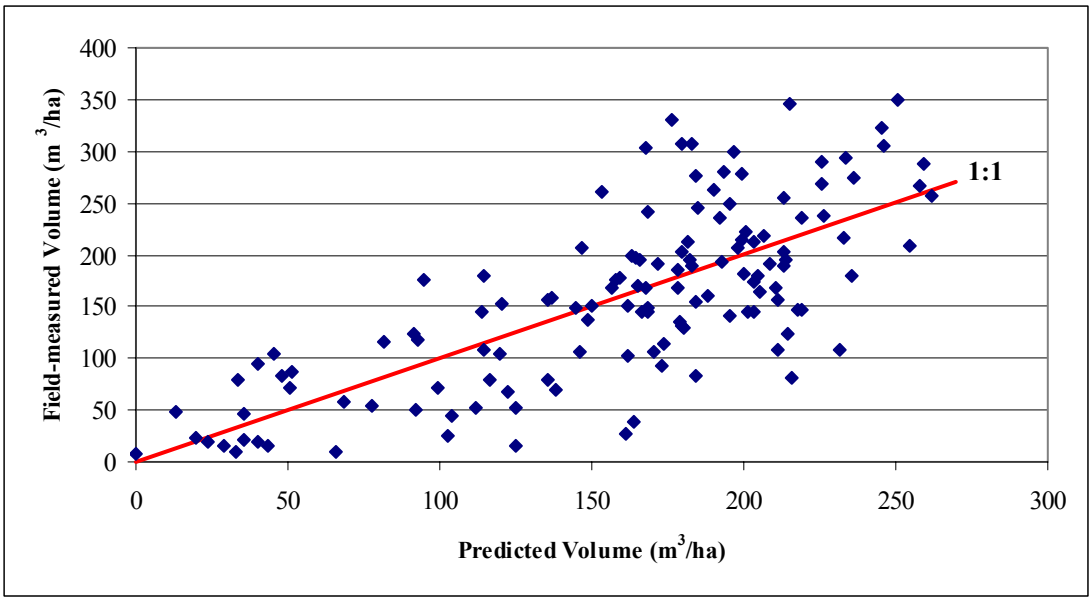


Figure H.73 2-class volume model (1.264 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.53$)

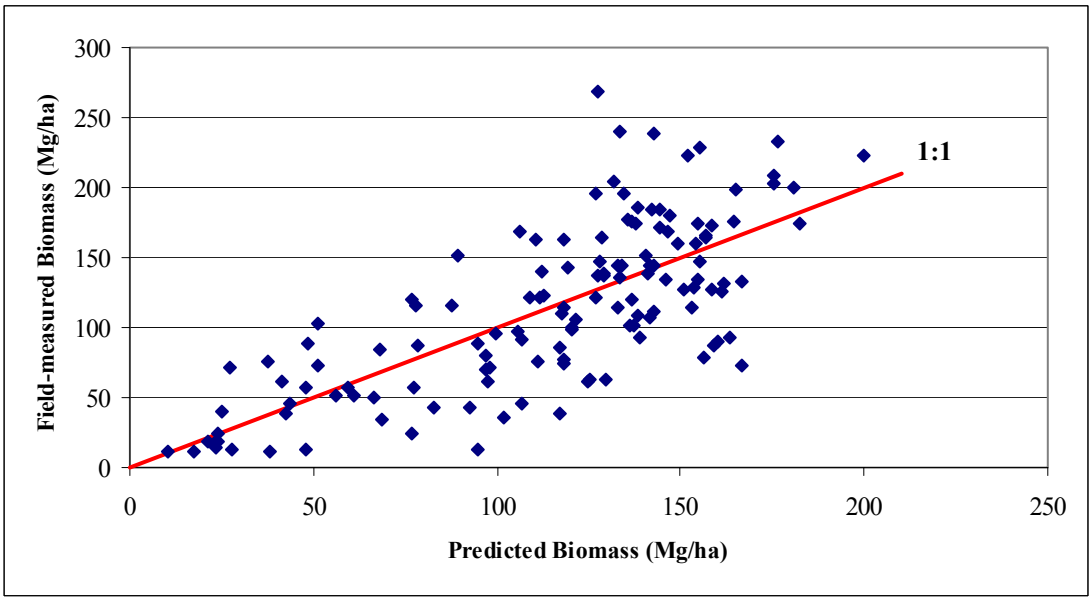


Figure H.74 2-class biomass model (1.264 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.53$)

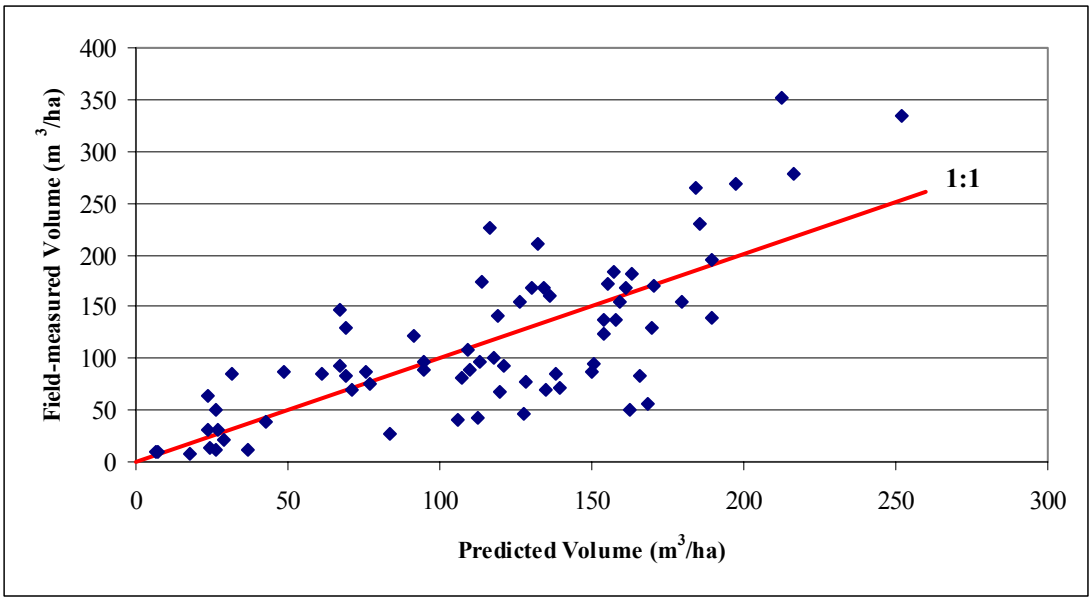


Figure H.75 2-class volume model (1.264 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.55$)

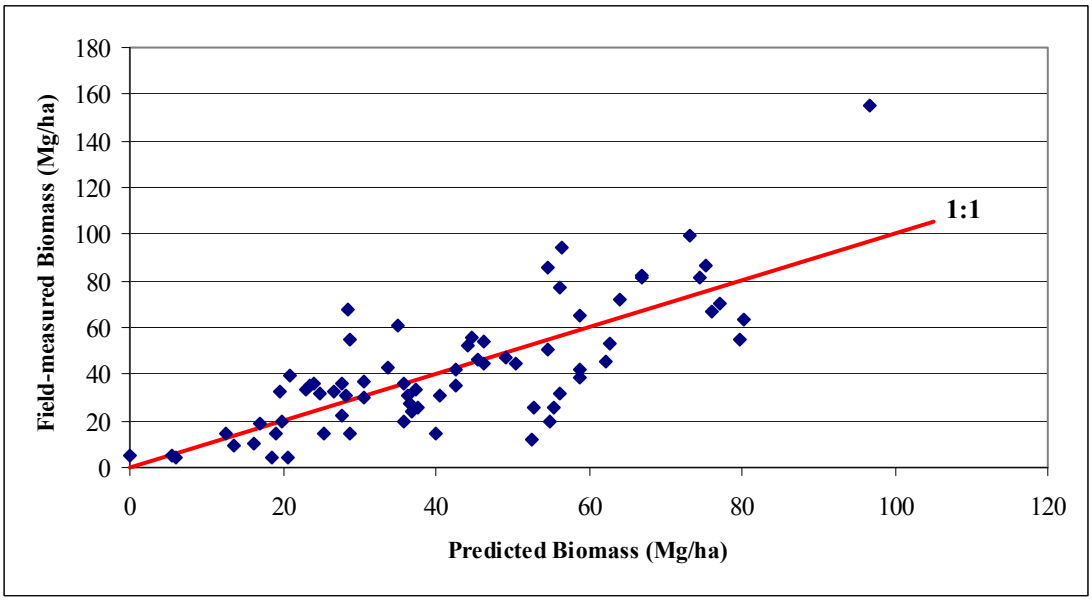


Figure H.76 2-class biomass model (1.264 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.56$)

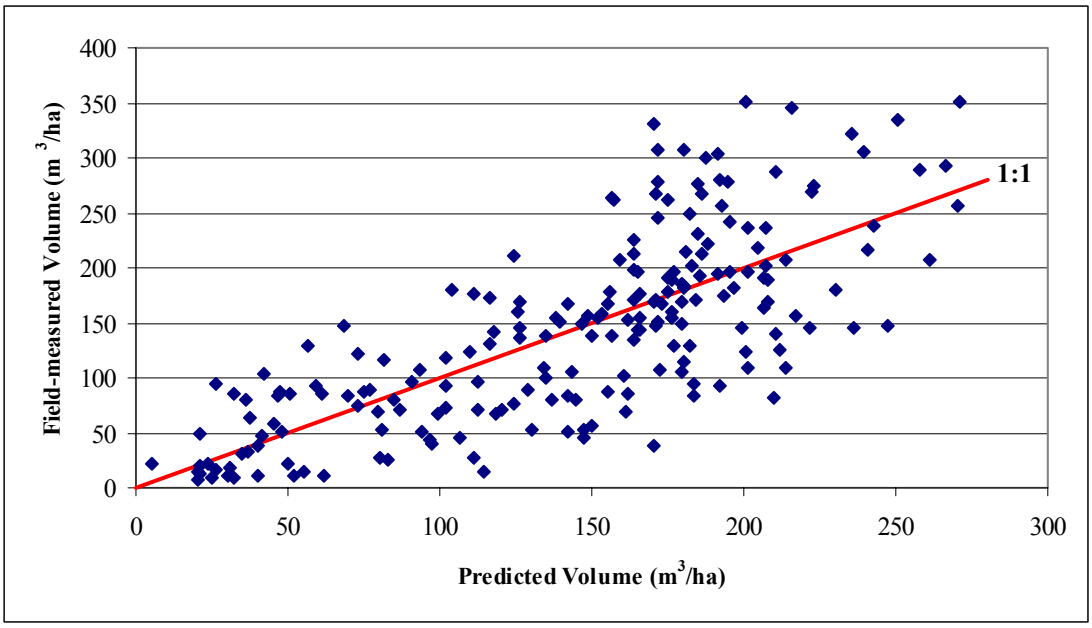


Figure H.77 2-class volume model (1.264 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.55$)

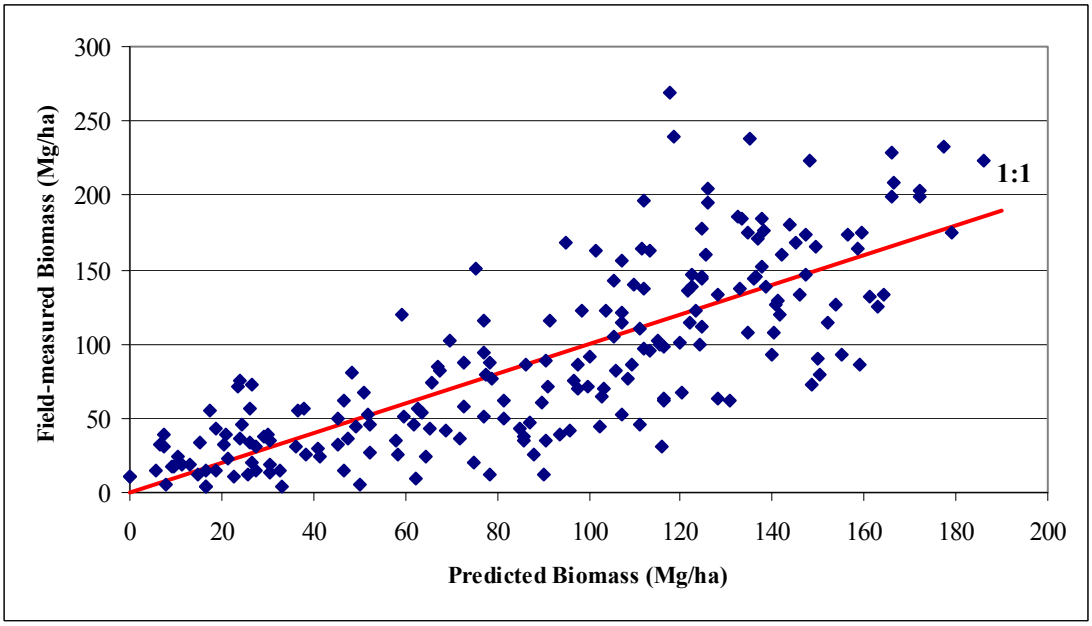


Figure H.78 2-class biomass model (1.264 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.61$)

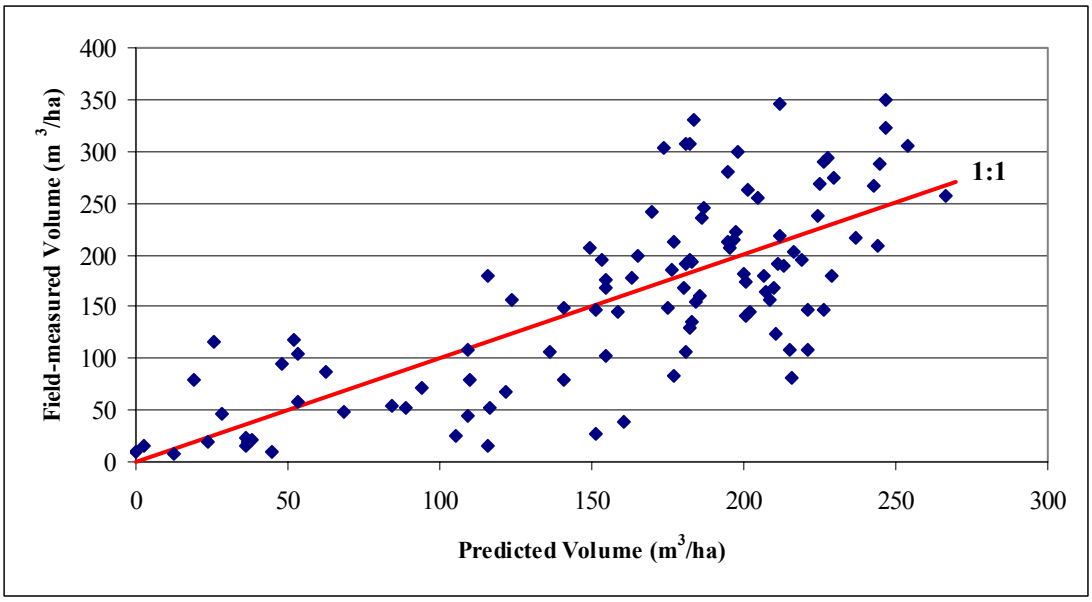


Figure H.79 3-class volume model (1.264 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.55$)

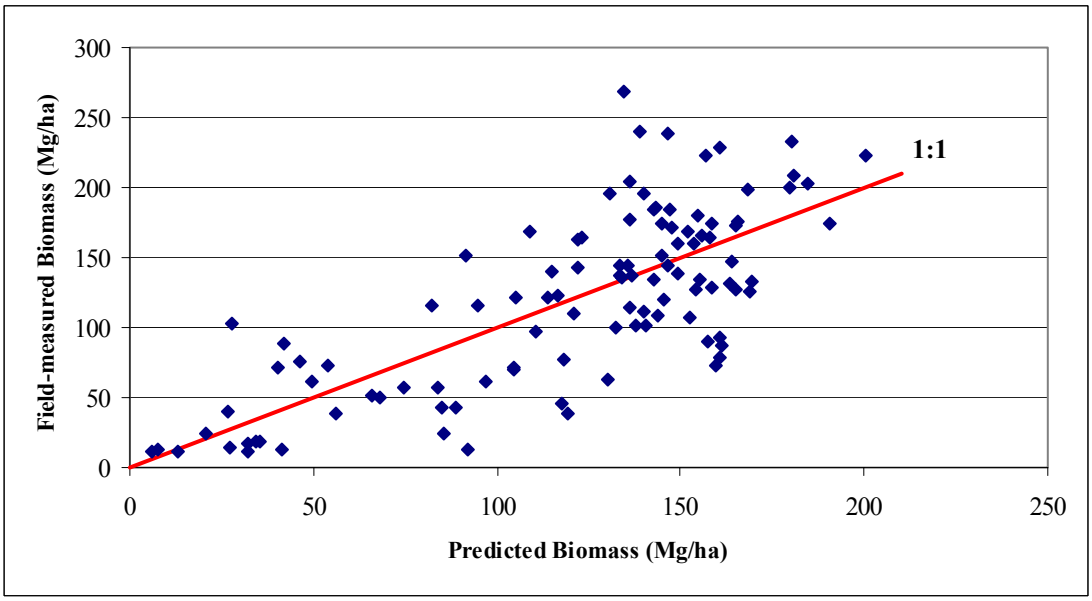


Figure H.80 3-class biomass model (1.264 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.55$)

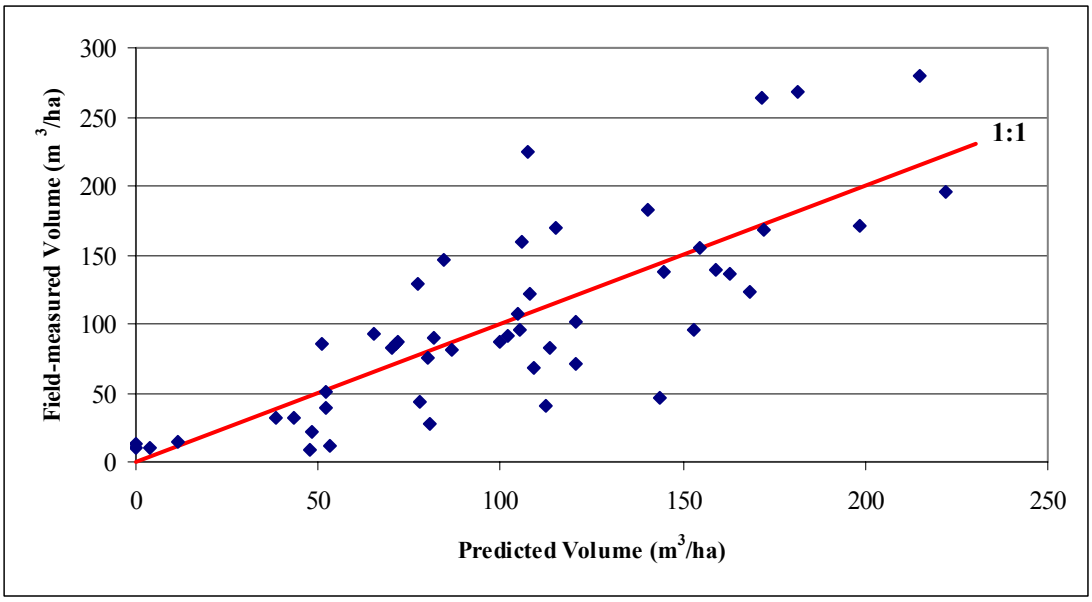


Figure H.81 3-class volume model (1.264 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.59$)

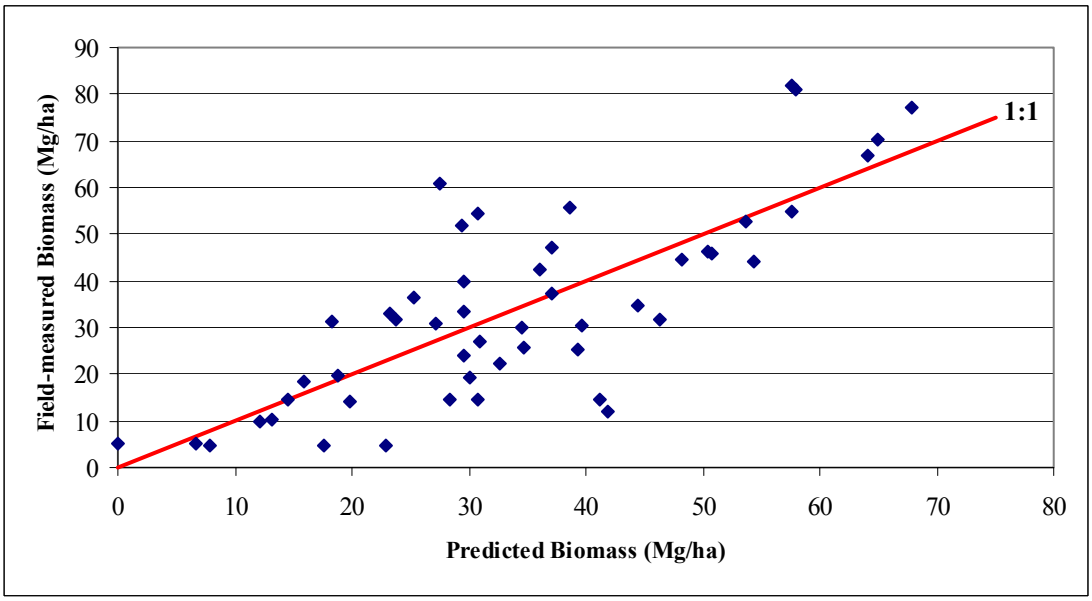


Figure H.82 3-class biomass model (1.264 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.57$)

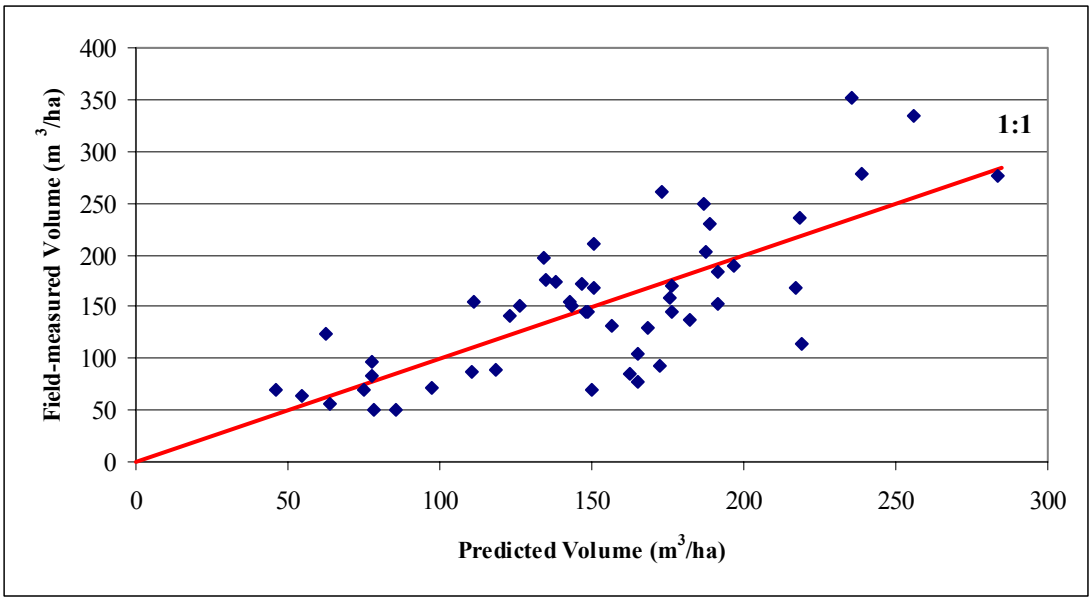


Figure H.83 3-class volume model (1.264 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.54$)

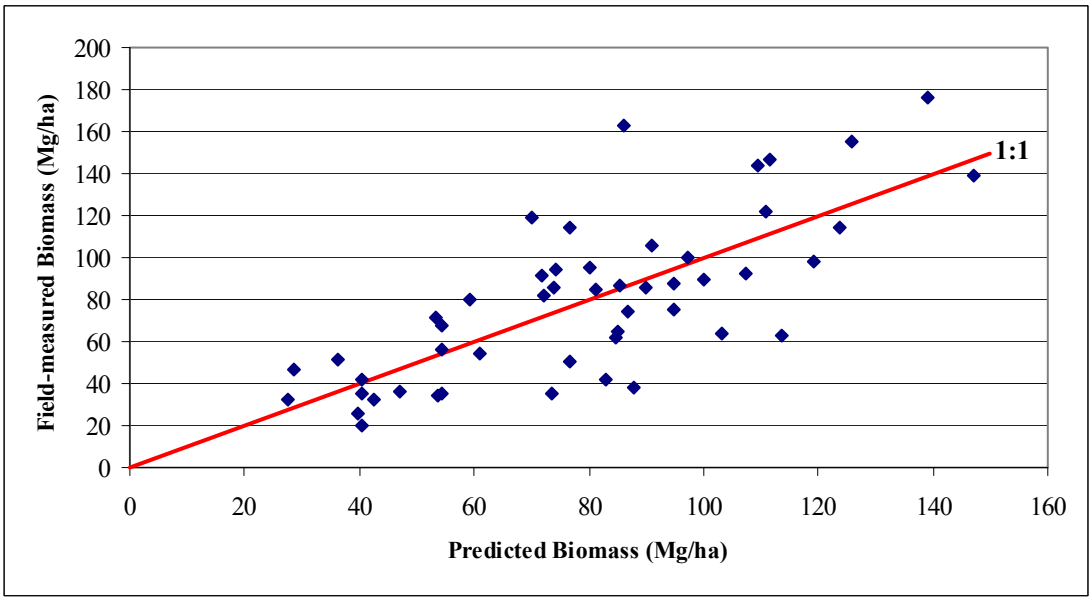


Figure H.84 3-class biomass model (1.264 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.54$)

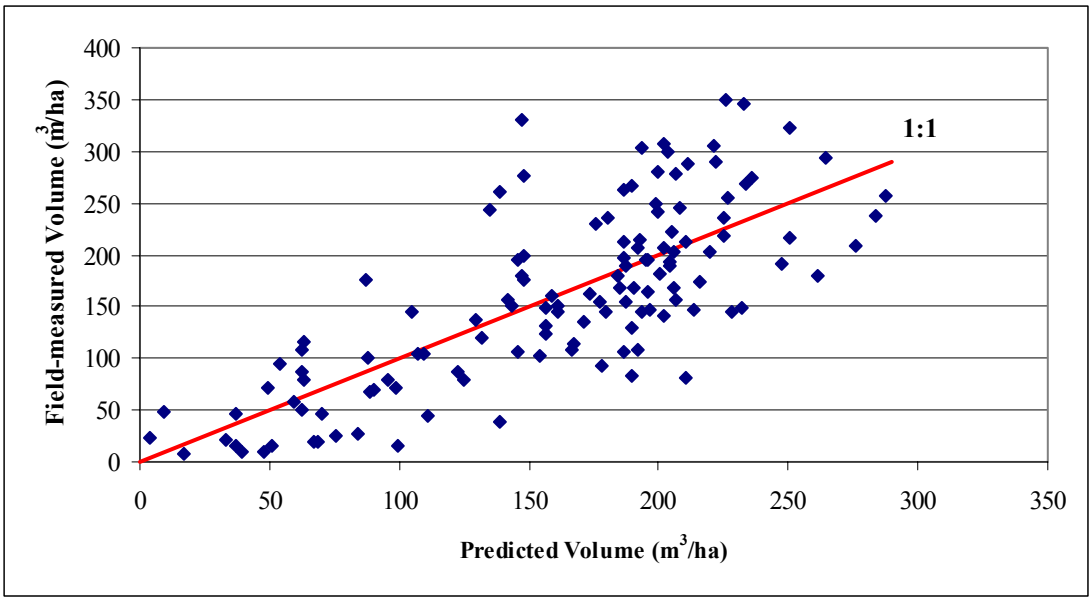


Figure H.85 2-class volume model (1.885 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.56$)

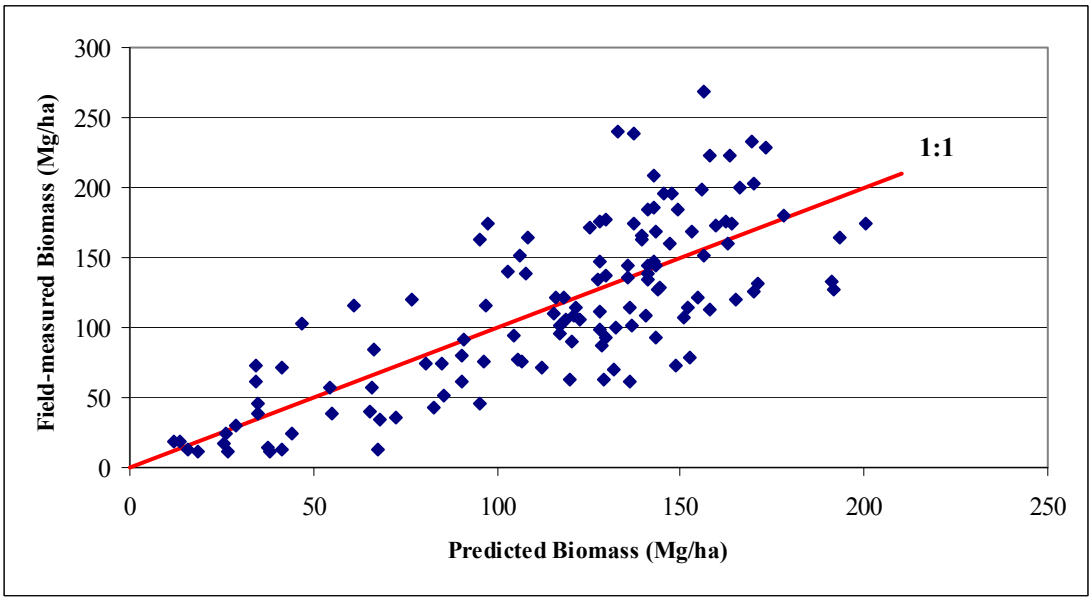


Figure H.86 2-class biomass model (1.885 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.56$)

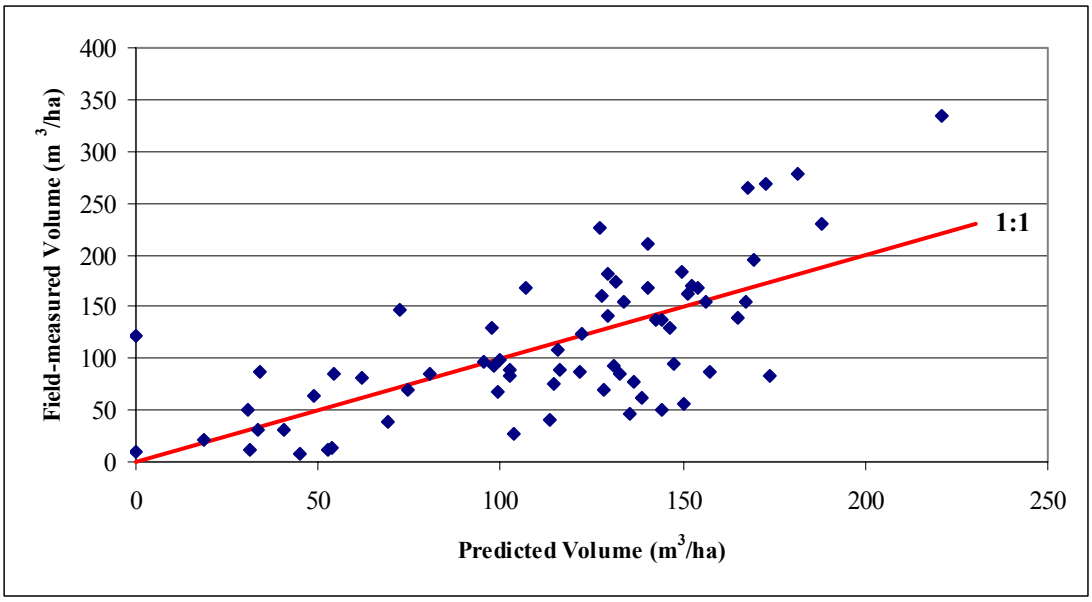


Figure H.87 2-class volume model (1.885 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.46$)

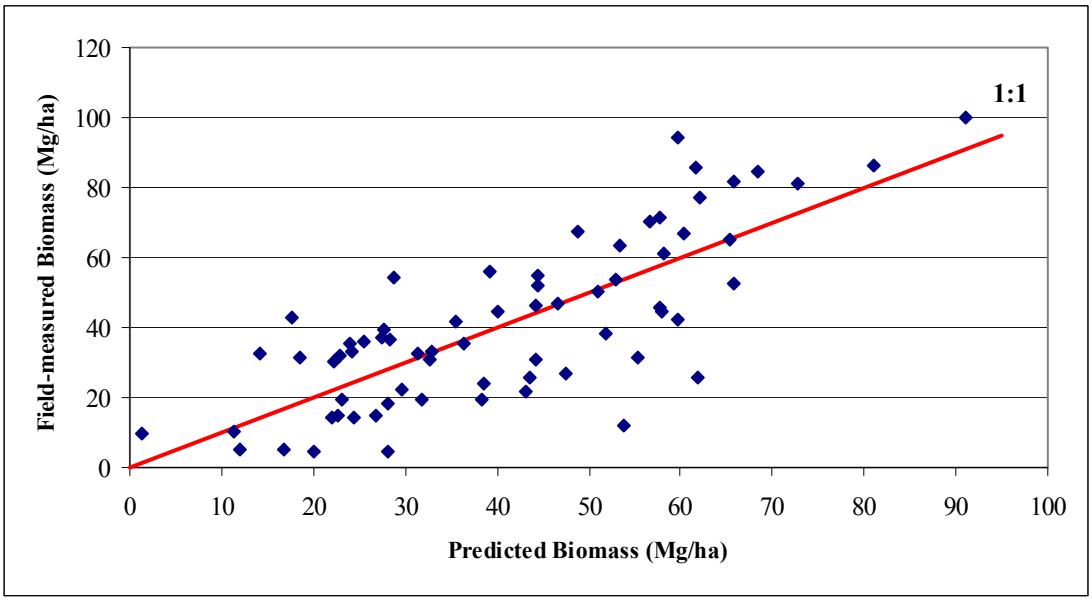


Figure H.88 2-class biomass model (1.885 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.57$)

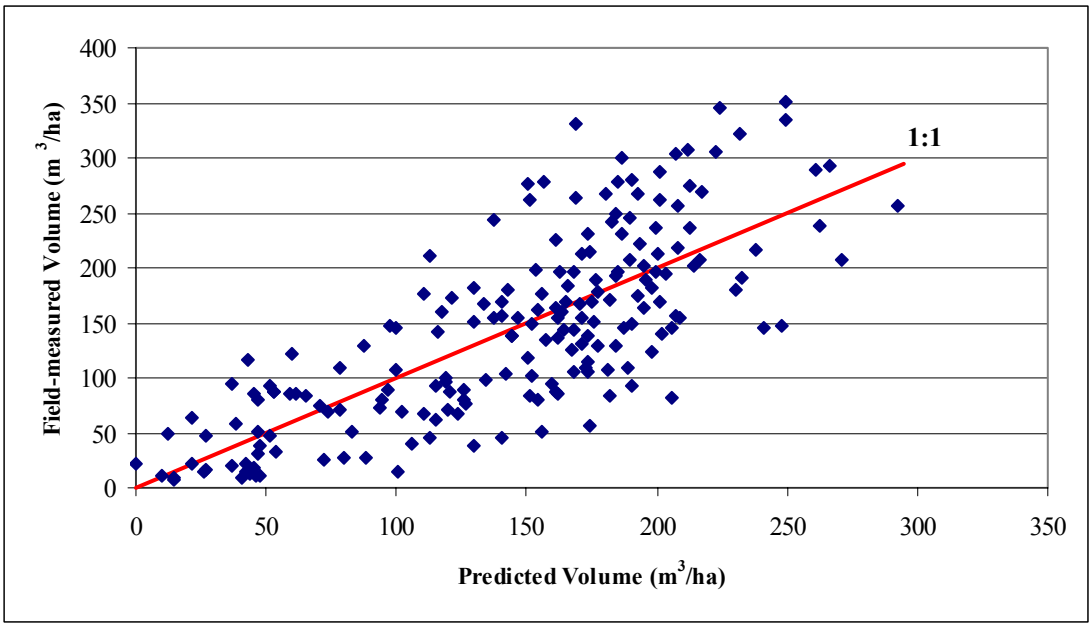


Figure H.89 2-class volume model (1.885 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.57$)

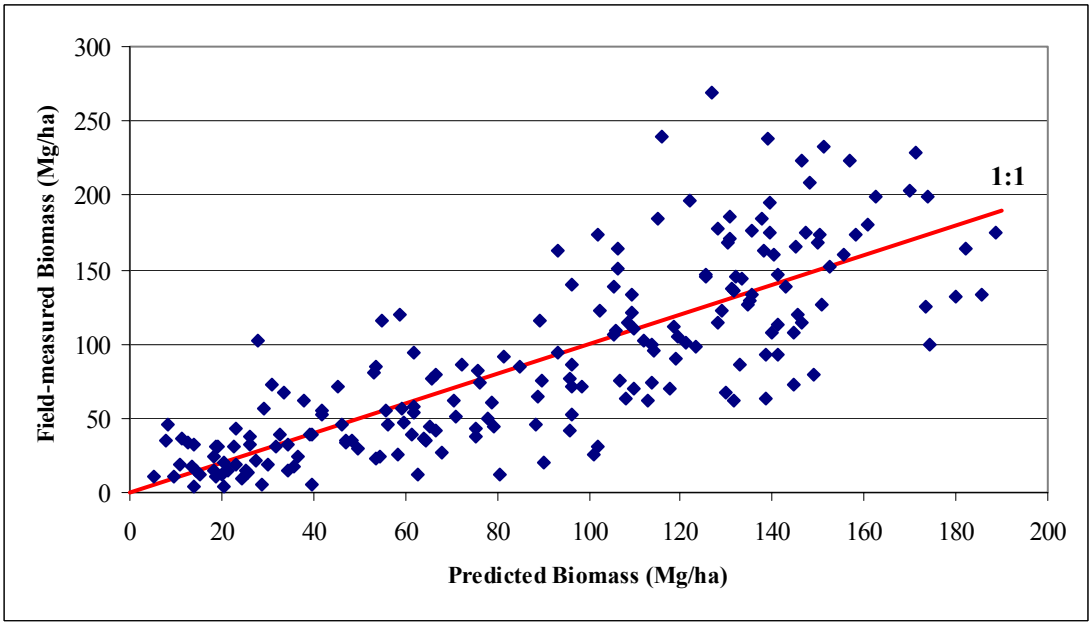


Figure H.90 2-class biomass model (1.885 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.63$)

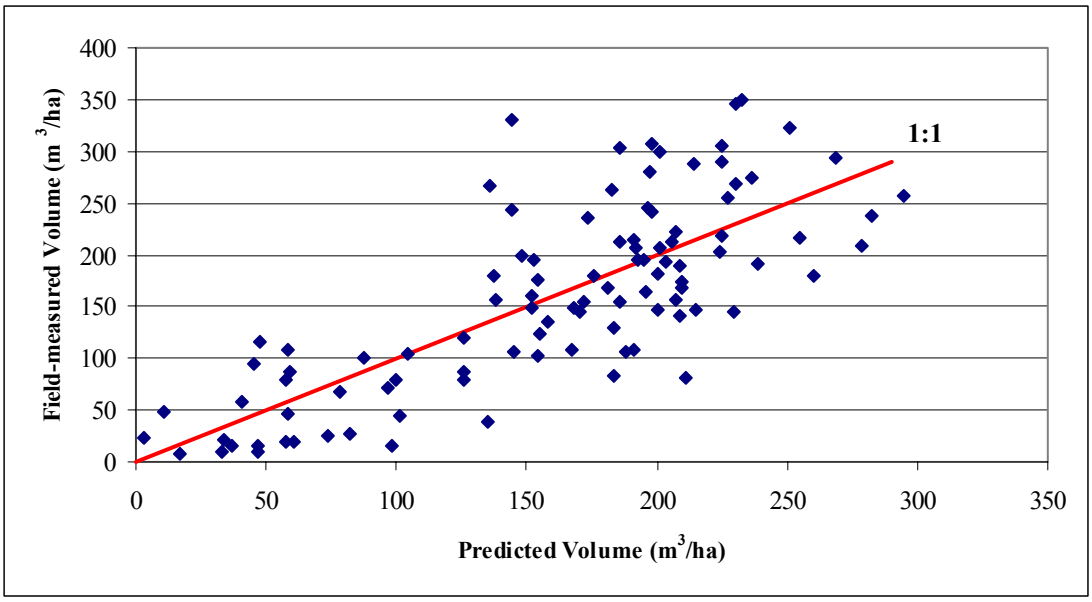


Figure H.91 3-class volume model (1.885 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.59$)

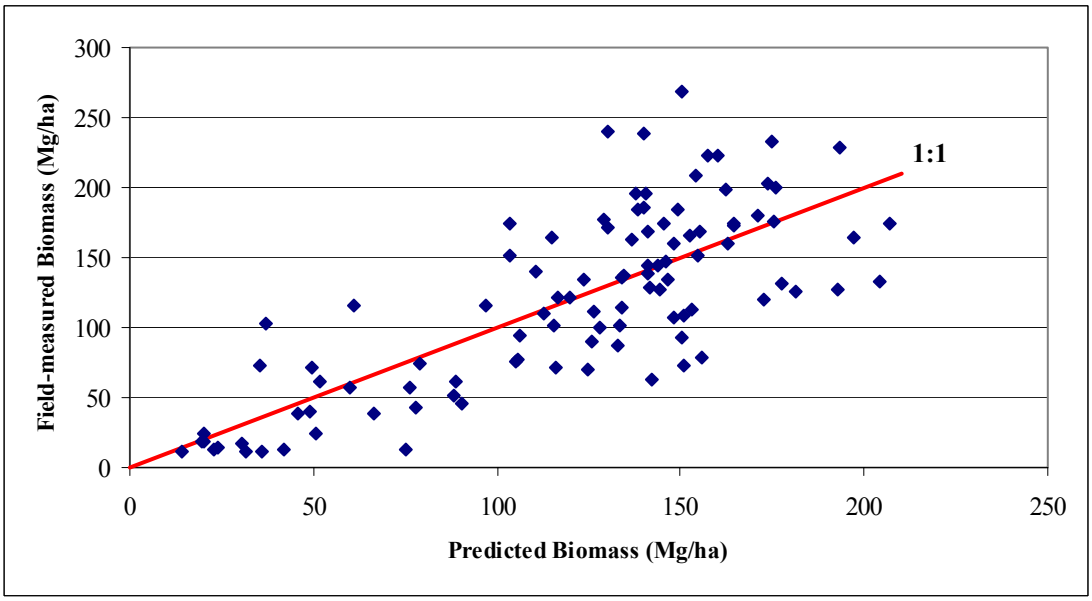


Figure H.92 3-class biomass model (1.885 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.58$)

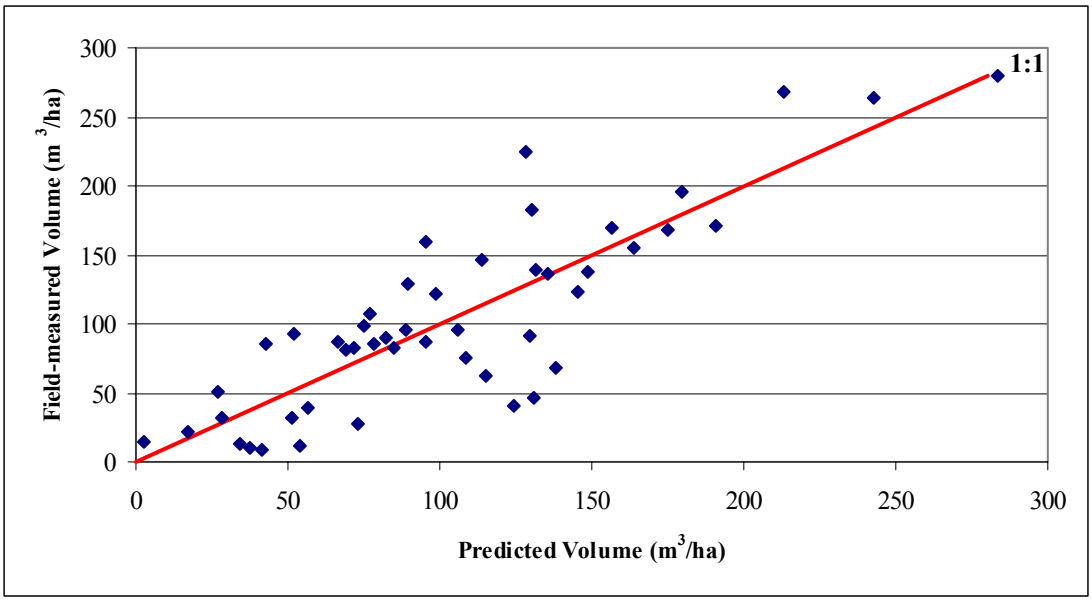


Figure H.93 3-class volume model (1.885 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.66$)

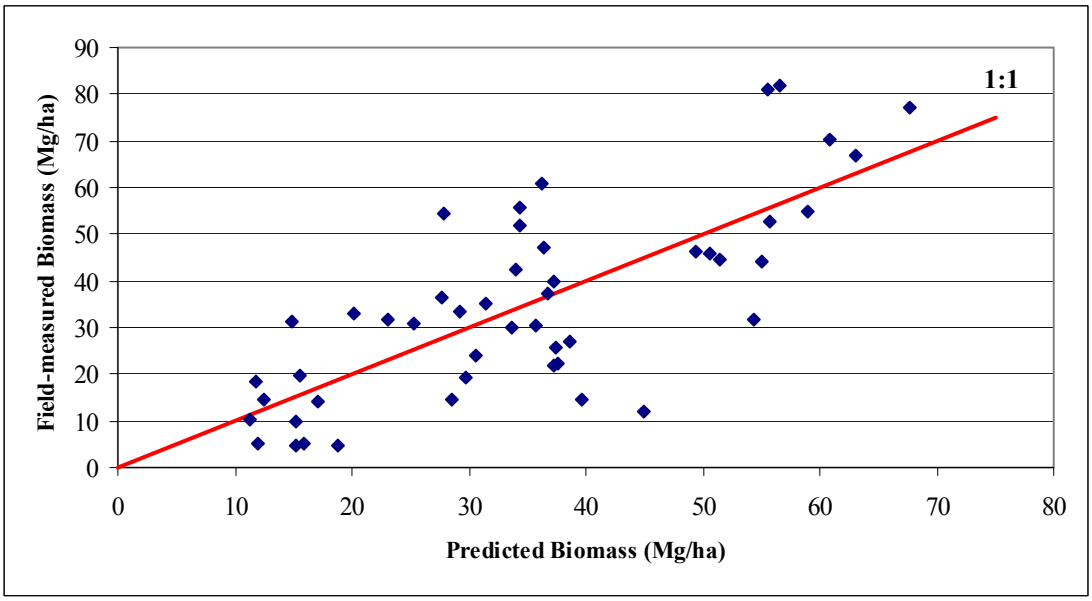


Figure H.94 3-class biomass model (1.885 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.54$)

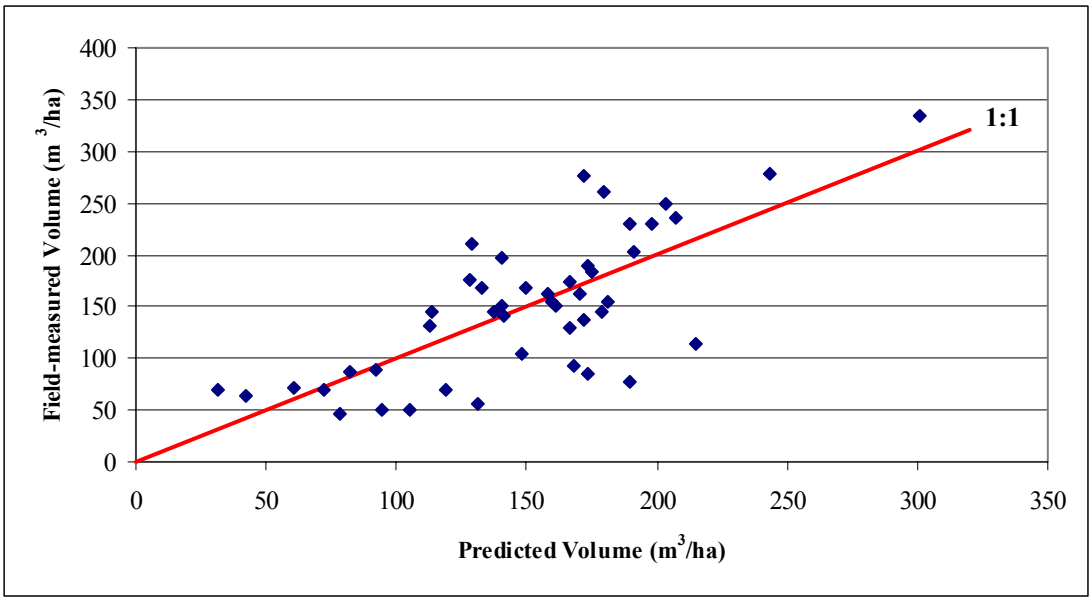


Figure H.95 3-class volume model (1.885 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.48$)

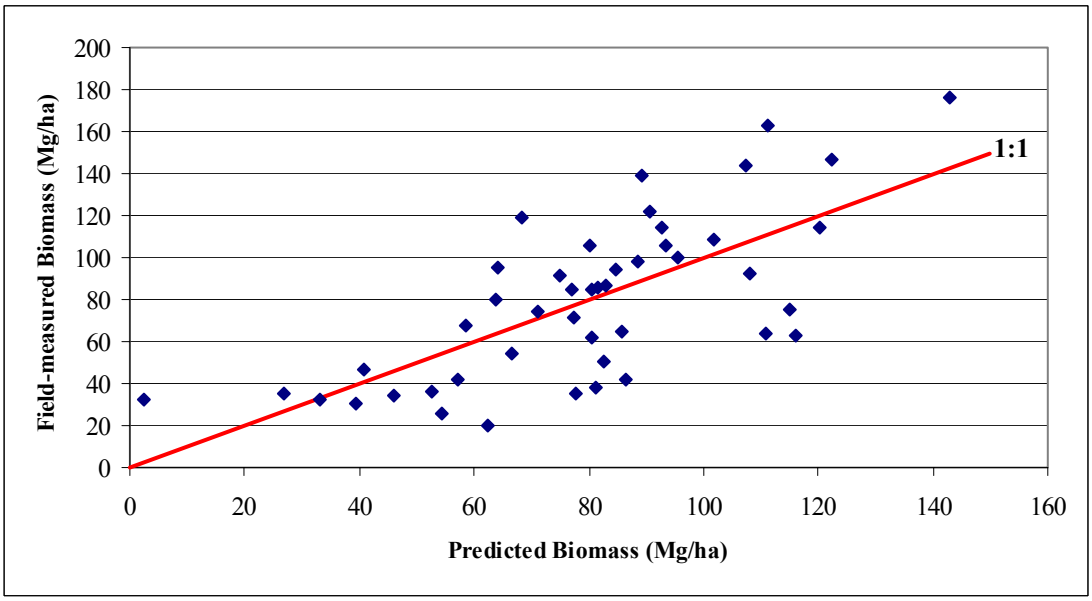


Figure H.96 3-class biomass model (1.885 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.46$)

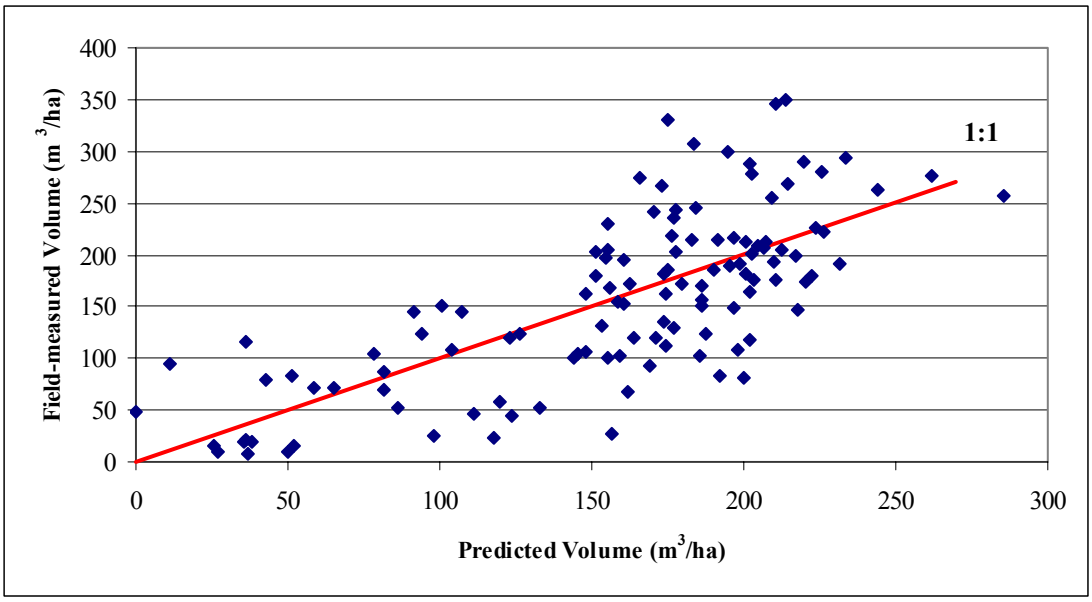


Figure H.97 2-class volume model (2.53 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.52$)

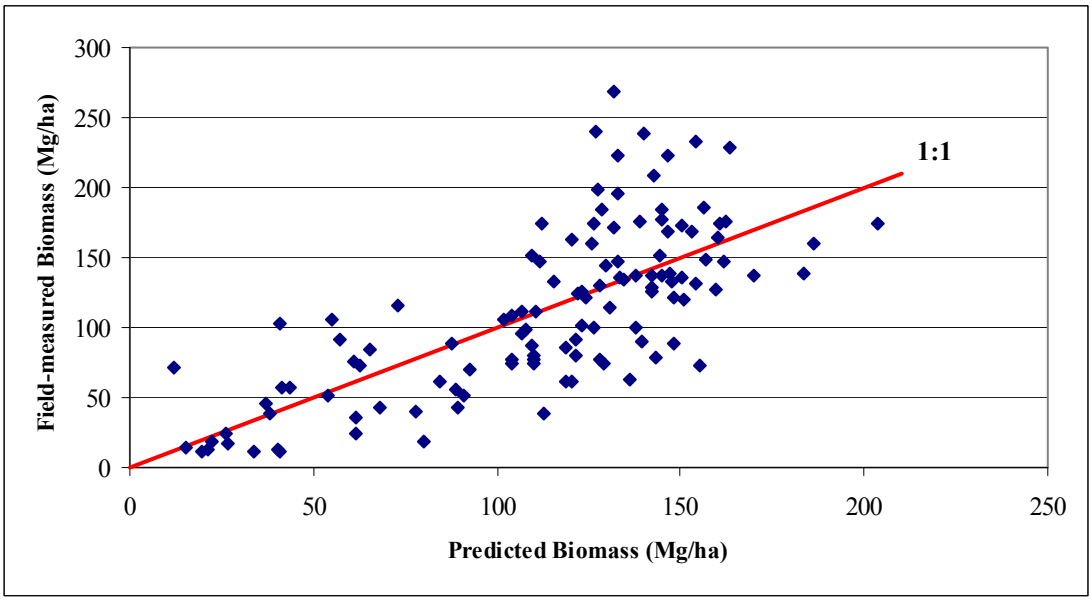


Figure H.98 2-class biomass model (2.53 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.51$)

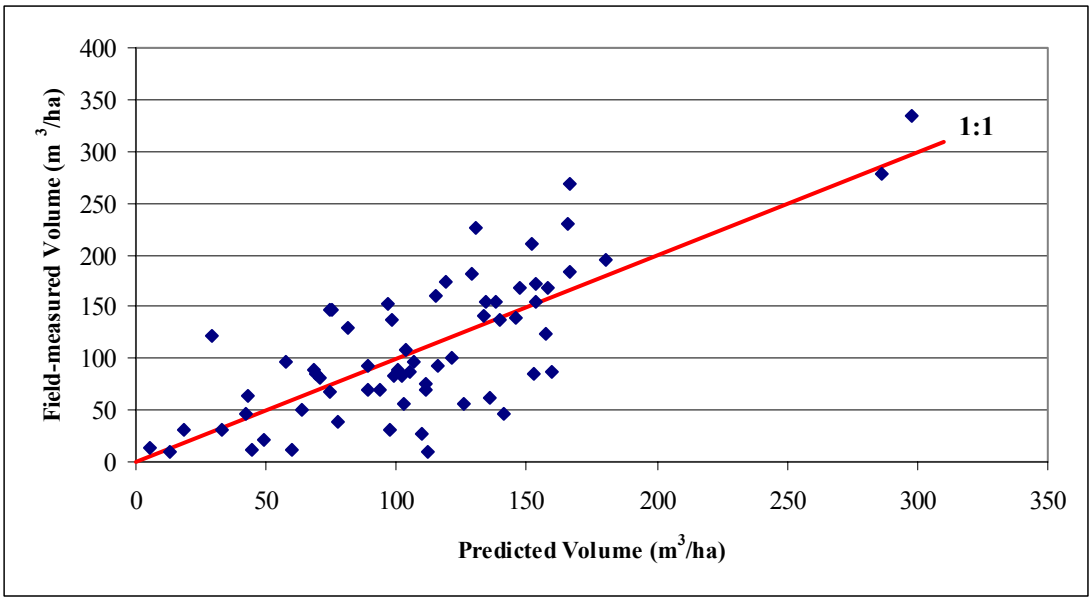


Figure H.99 2-class volume model (2.53 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.53$)

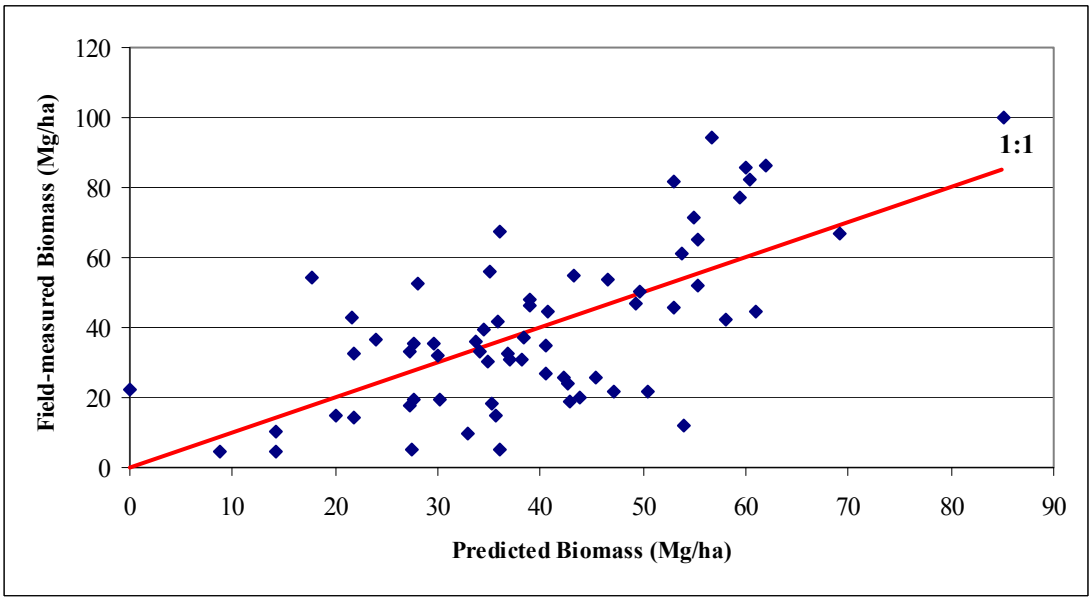


Figure H.100 2-class biomass model (2.53 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.40$)

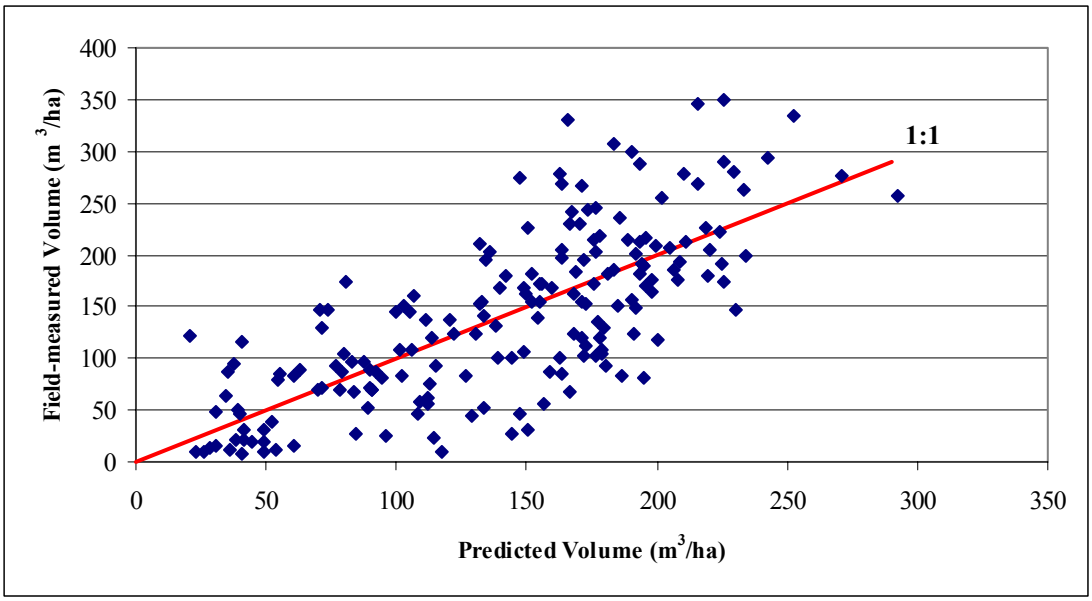


Figure H.101 2-class volume model (2.53 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.54$)

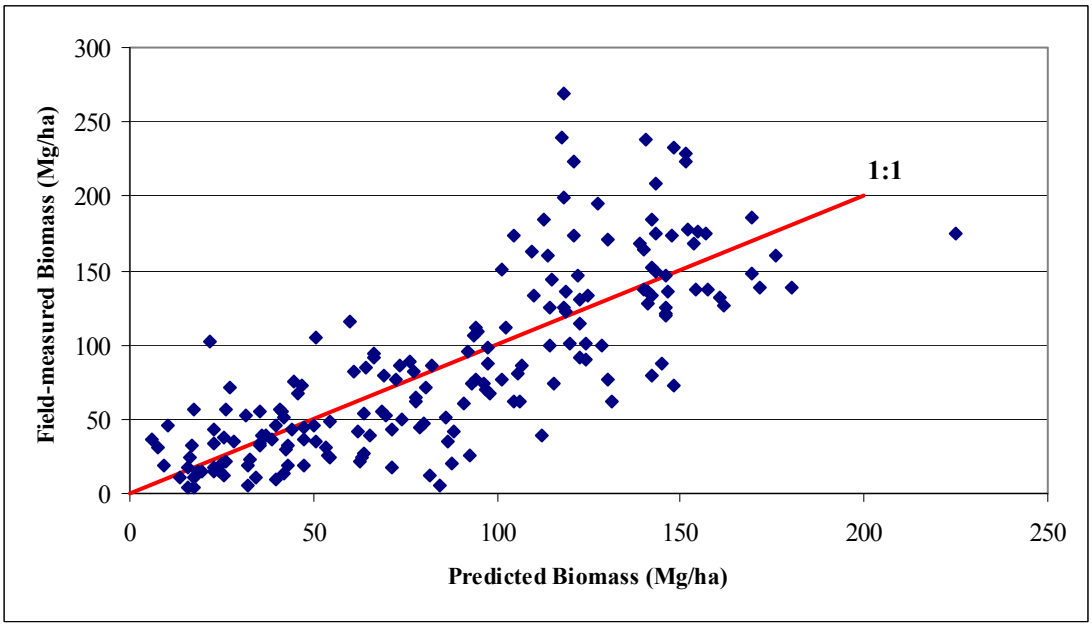


Figure H.102 2-class biomass model (2.53 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.62$)

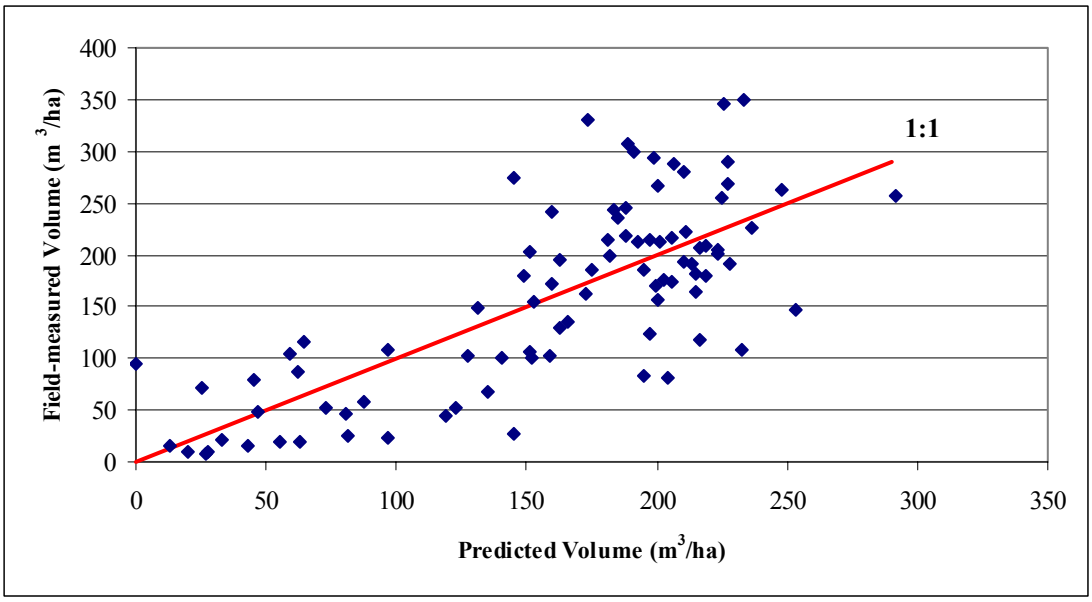


Figure H.103 3-class volume model (2.53 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.55$)

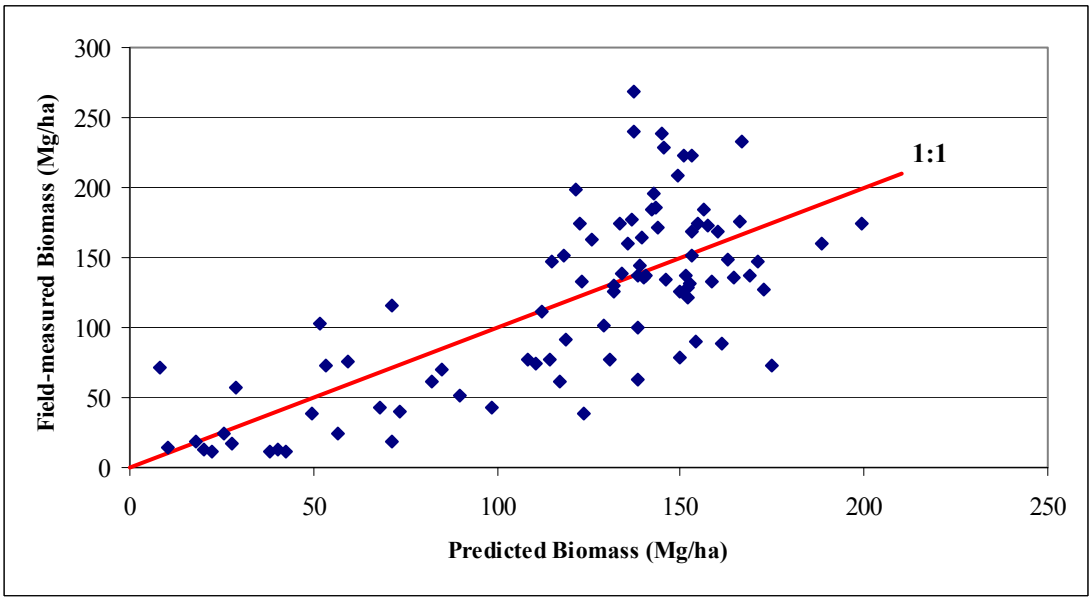


Figure H.104 3-class biomass model (2.53 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.52$)

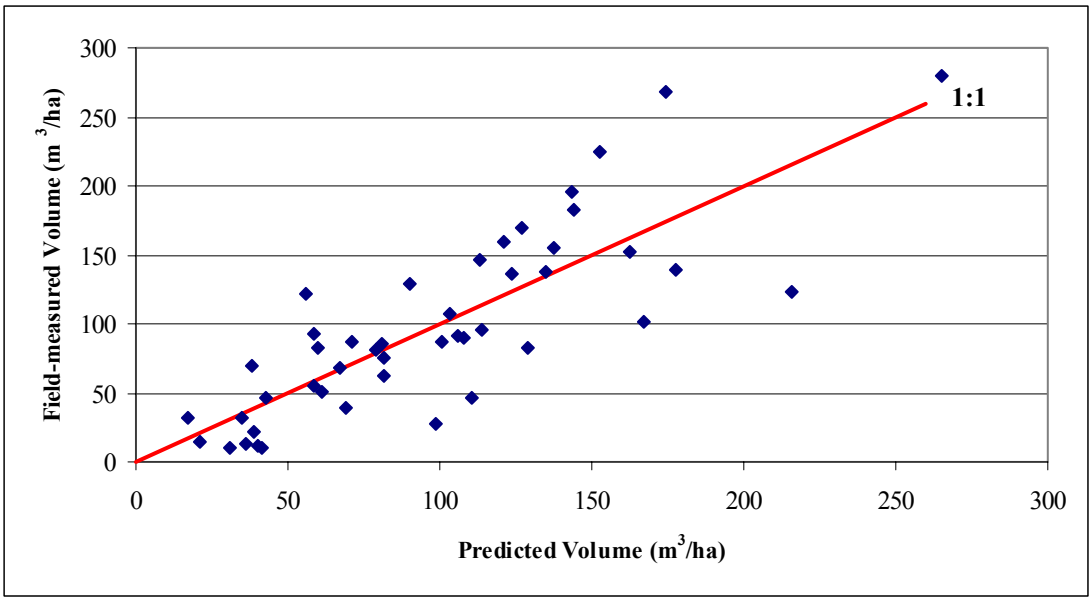


Figure H.105 3-class volume model (2.53 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.62$)

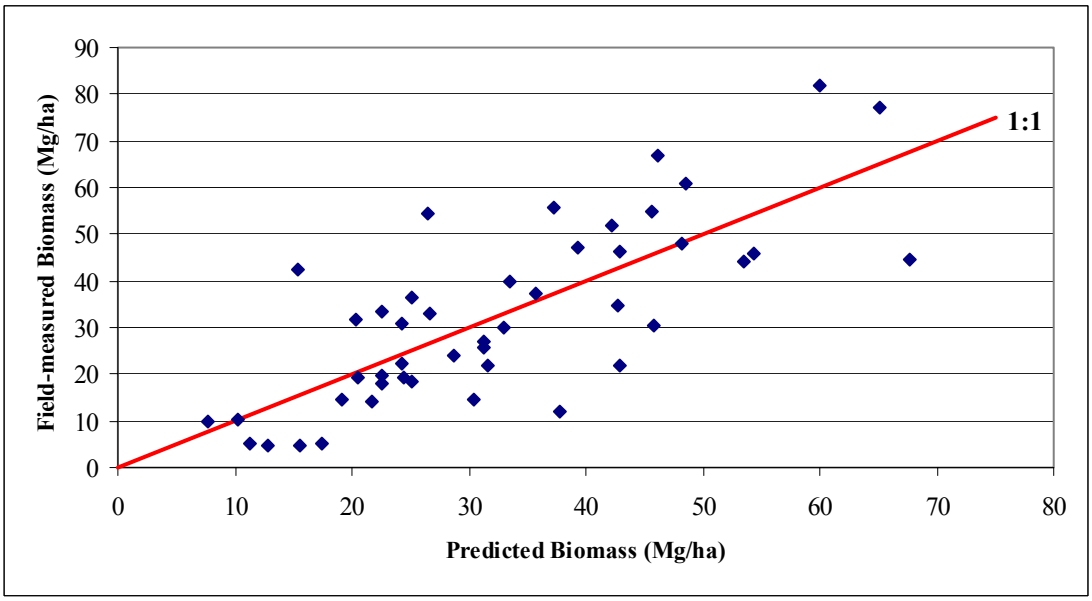


Figure H.106 3-class biomass model (2.53 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.54$)

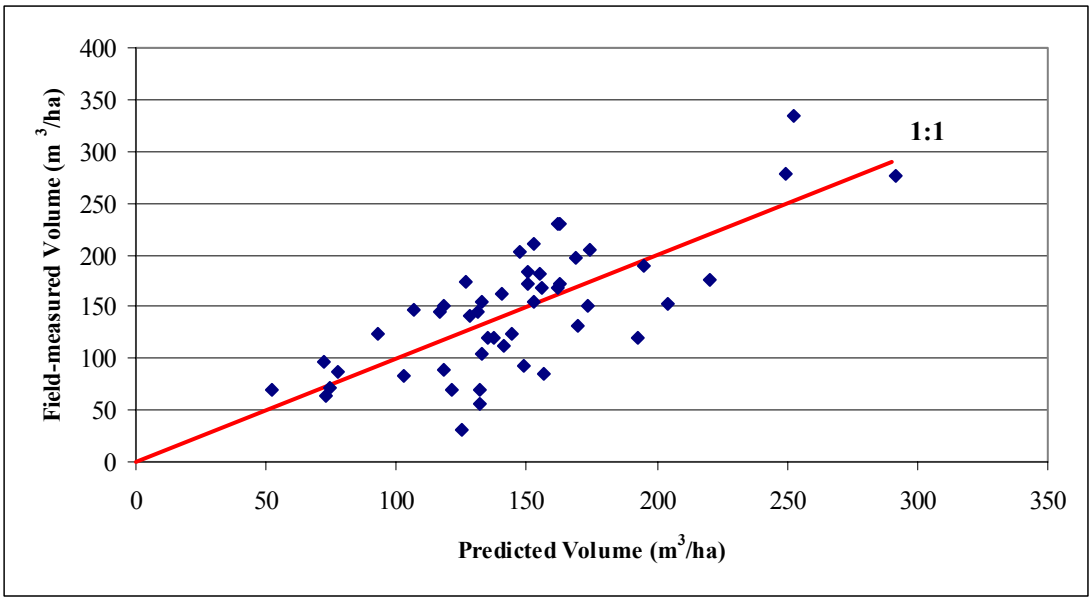


Figure H.107 3-class volume model (2.53 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.53$)

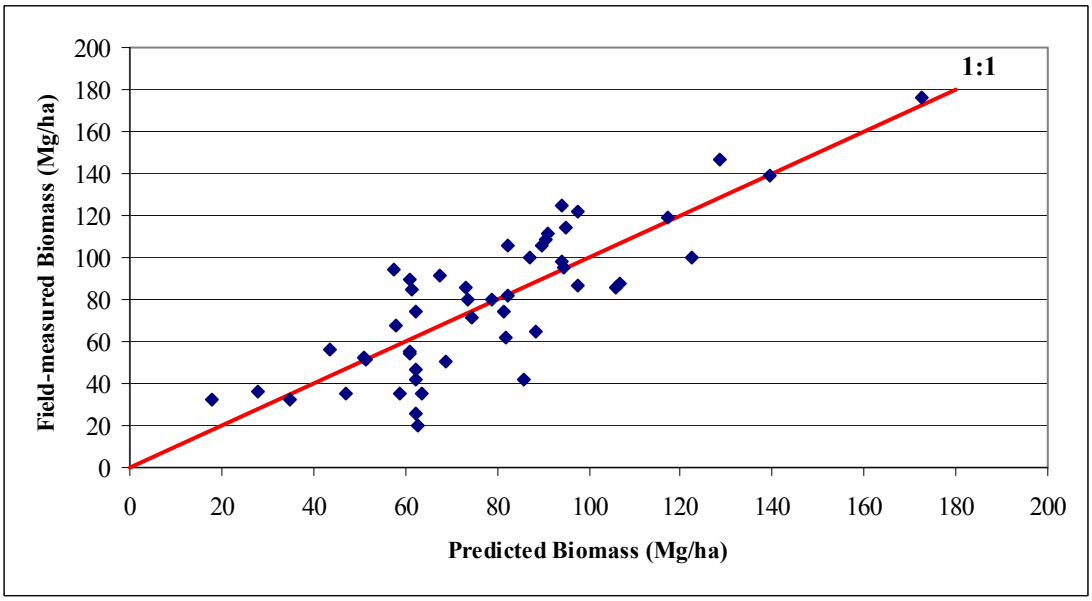


Figure H.108 3-class biomass model (2.53 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.66$)

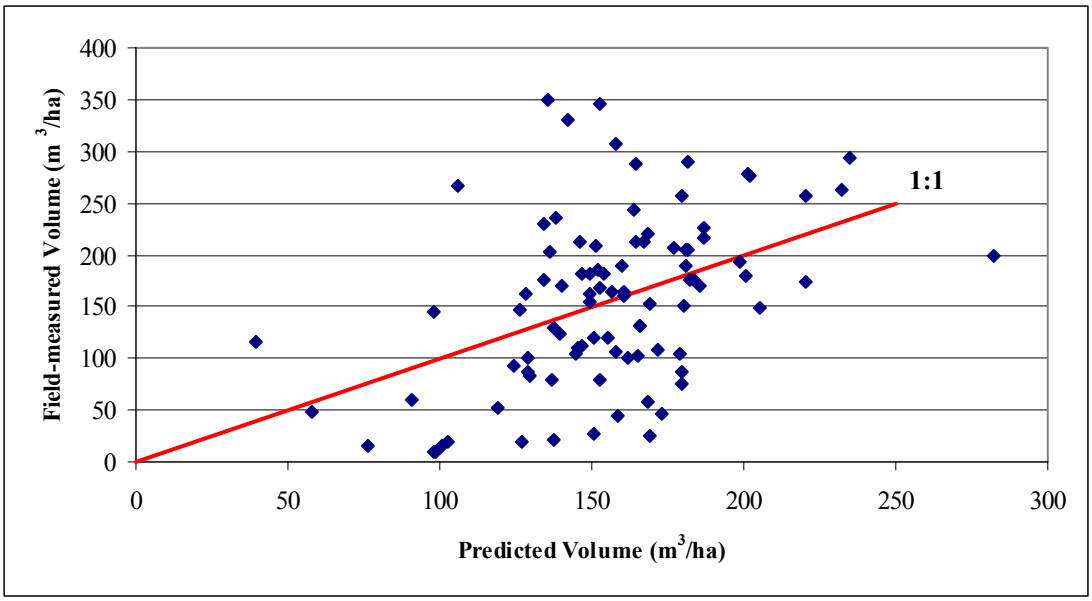


Figure H.109 2-class volume model (3.942 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.55$)

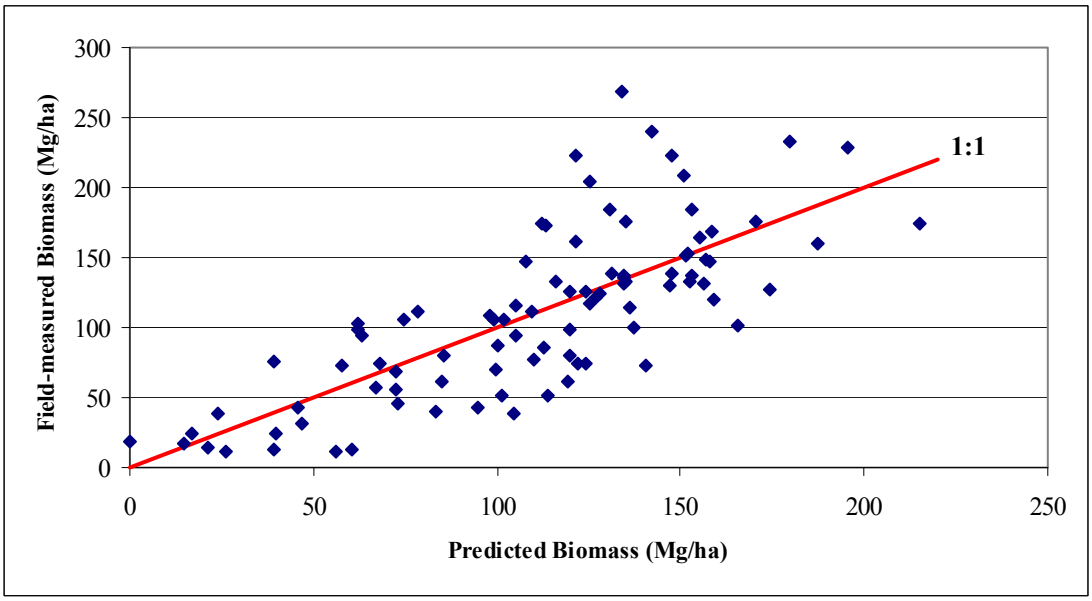


Figure H.110 2-class biomass model (3.942 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.55$)

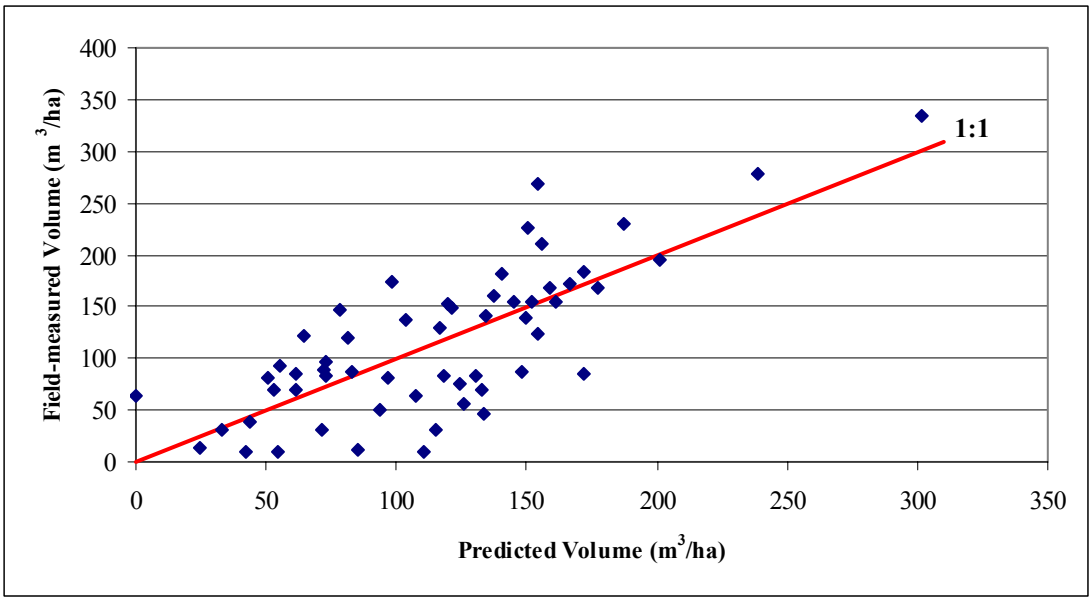


Figure H.111 2-class volume model (3.942 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.55$)

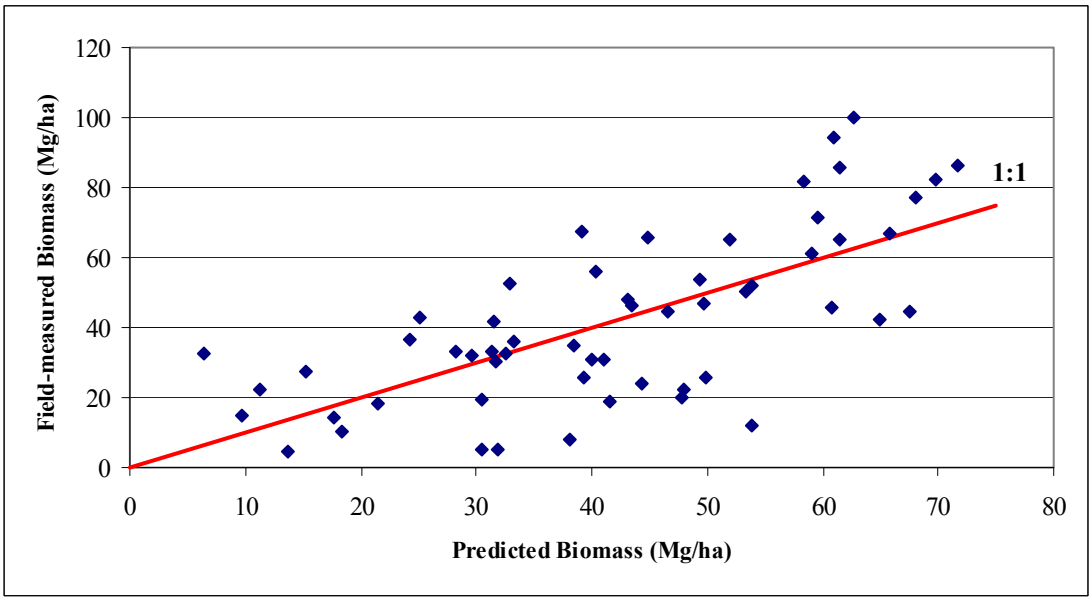


Figure H.112 2-class biomass model (3.942 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.46$)

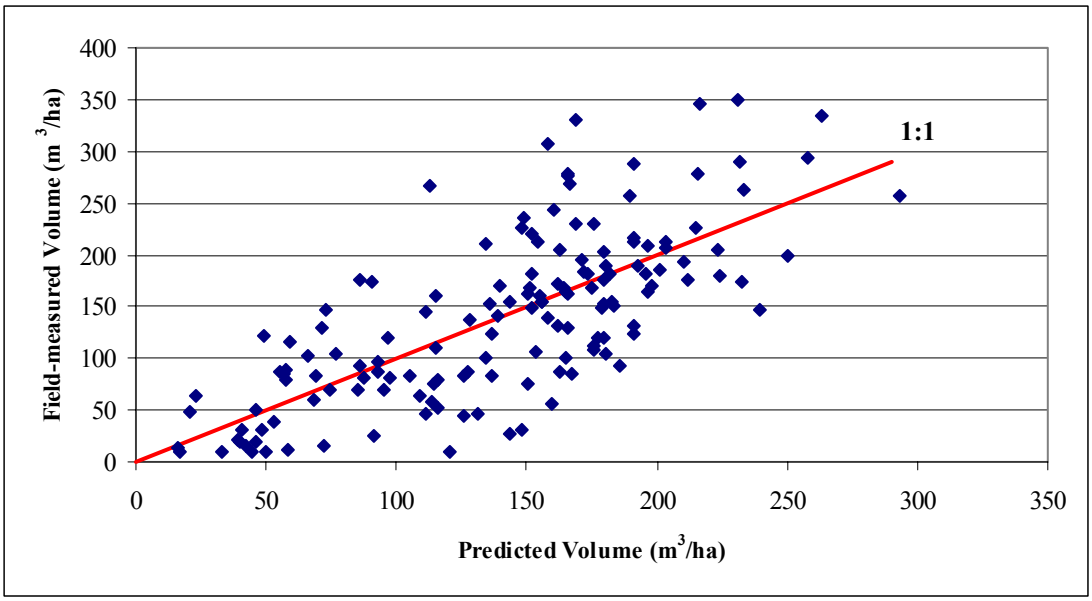


Figure H.113 2-class volume model (3.942 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.52$)

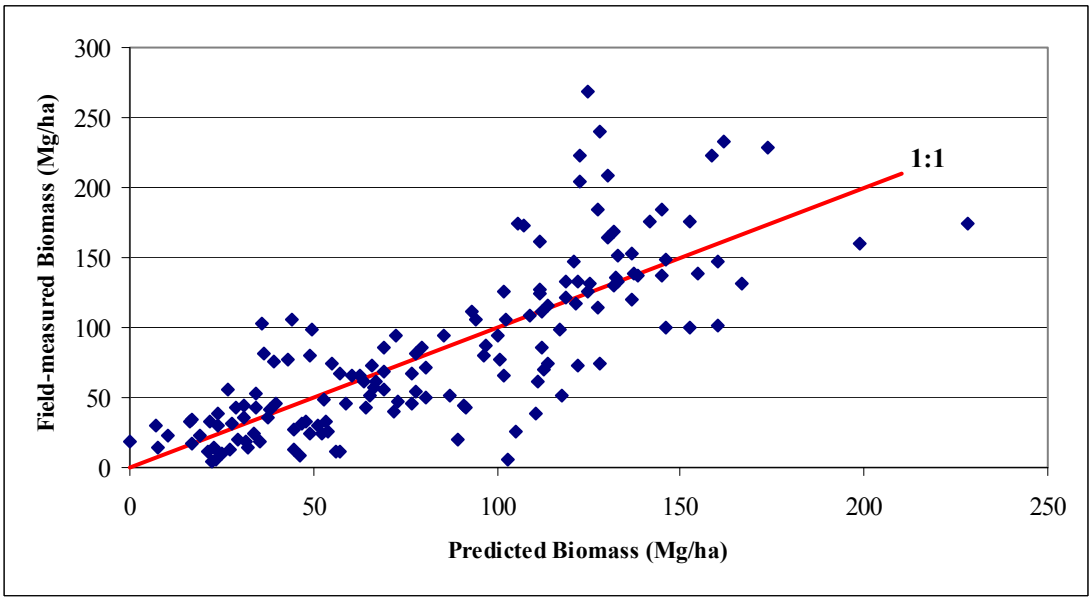


Figure H.114 2-class biomass model (3.942 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.61$)

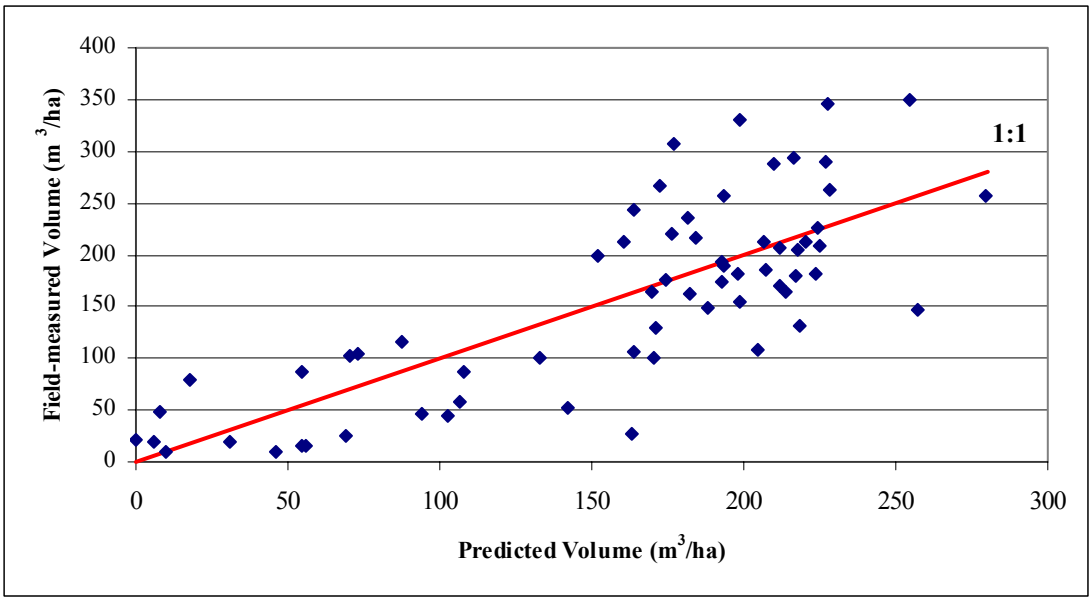


Figure H.115 3-class volume model (3.942 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.6$)

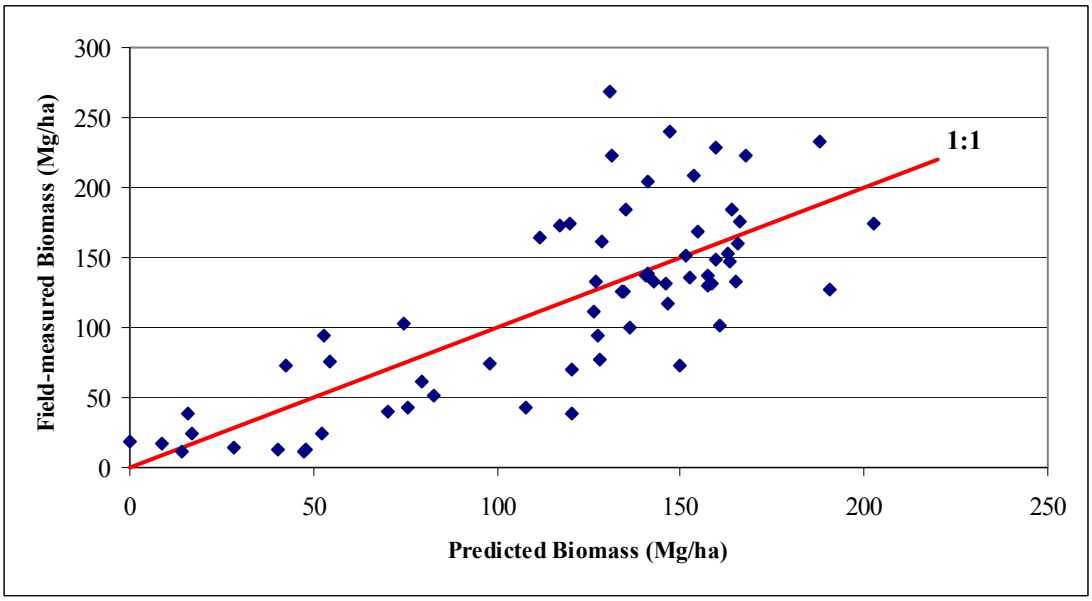


Figure H.116 3-class biomass model (3.942 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.57$)

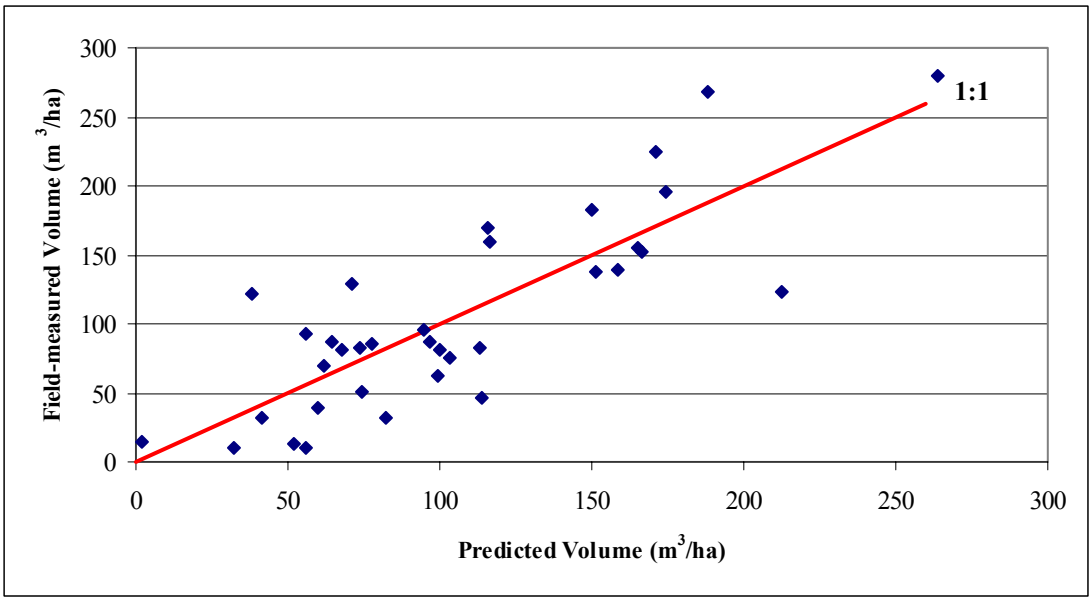


Figure H.117 3-class volume model (3.942 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.65$)

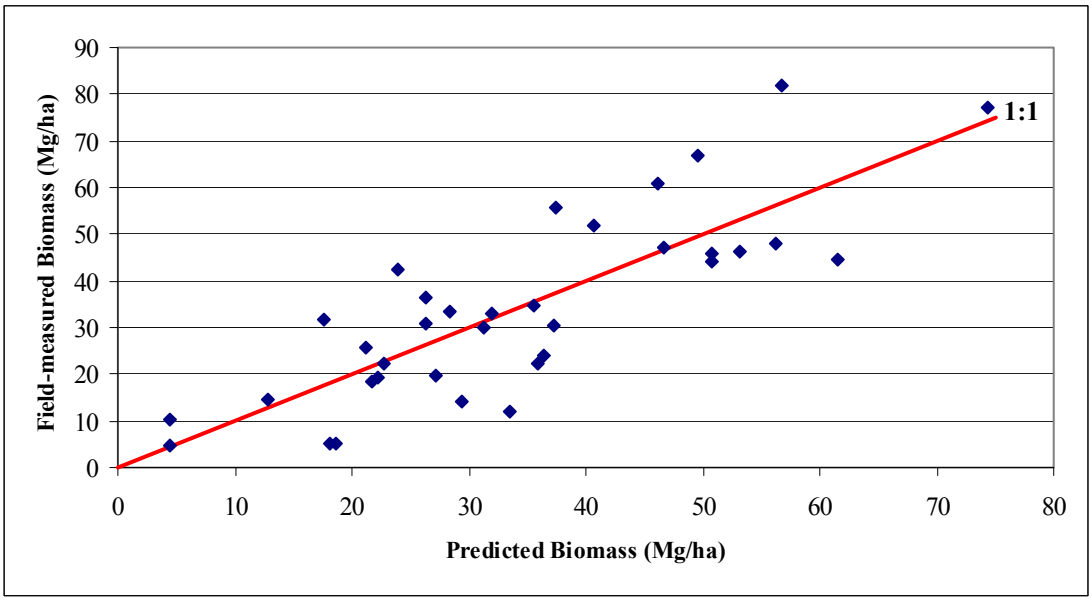


Figure H.118 3-class biomass model (3.942 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.63$)

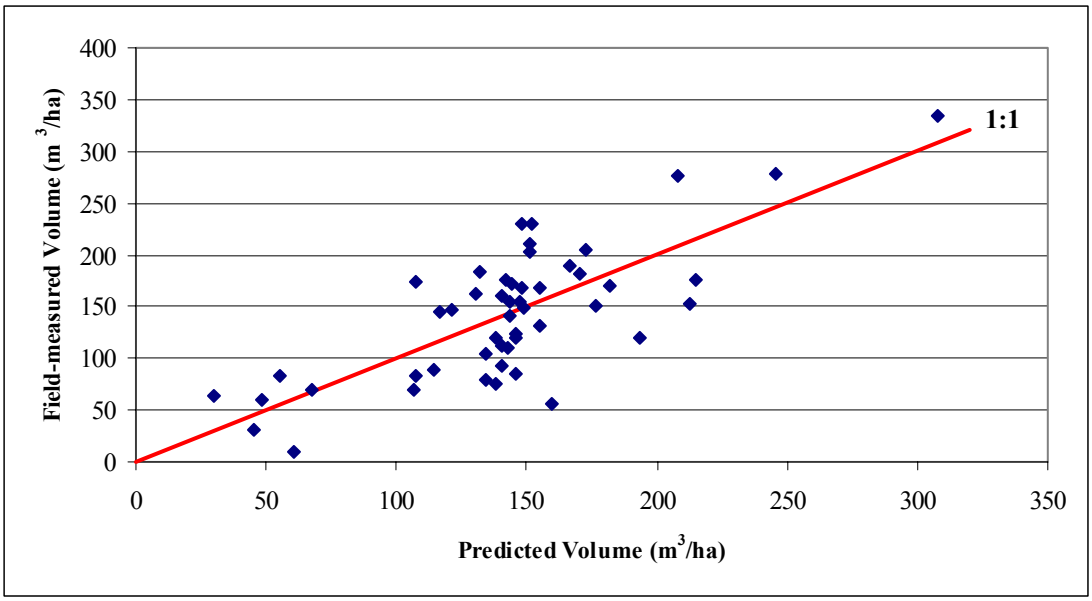


Figure H.119 3-class volume model (3.942 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.54$)

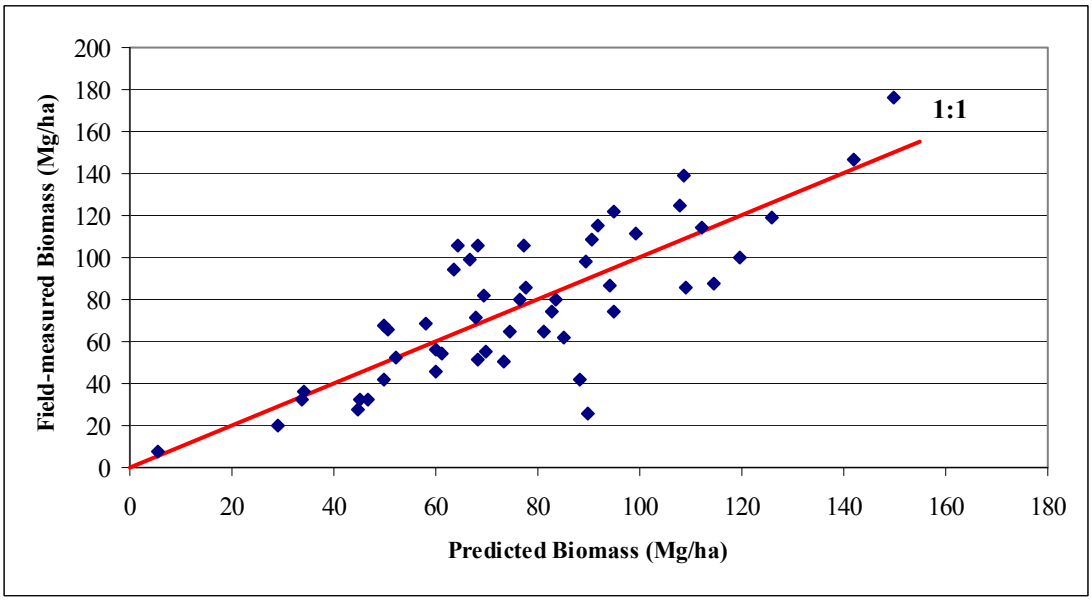


Figure H.120 3-class biomass model (3.942 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.62$)

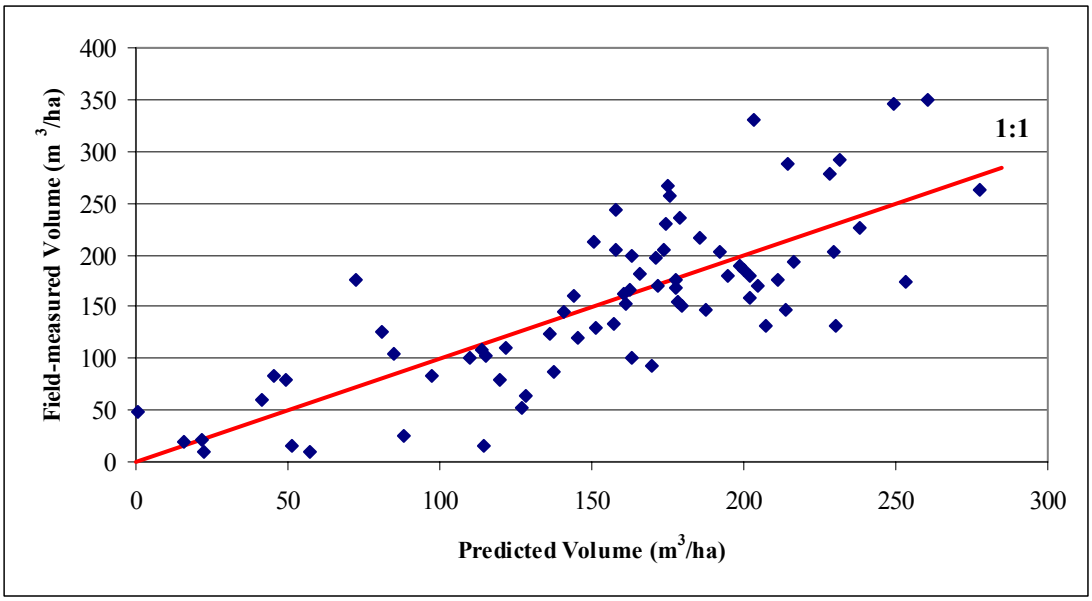


Figure H.121 2-class volume model (5.632 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.59$)

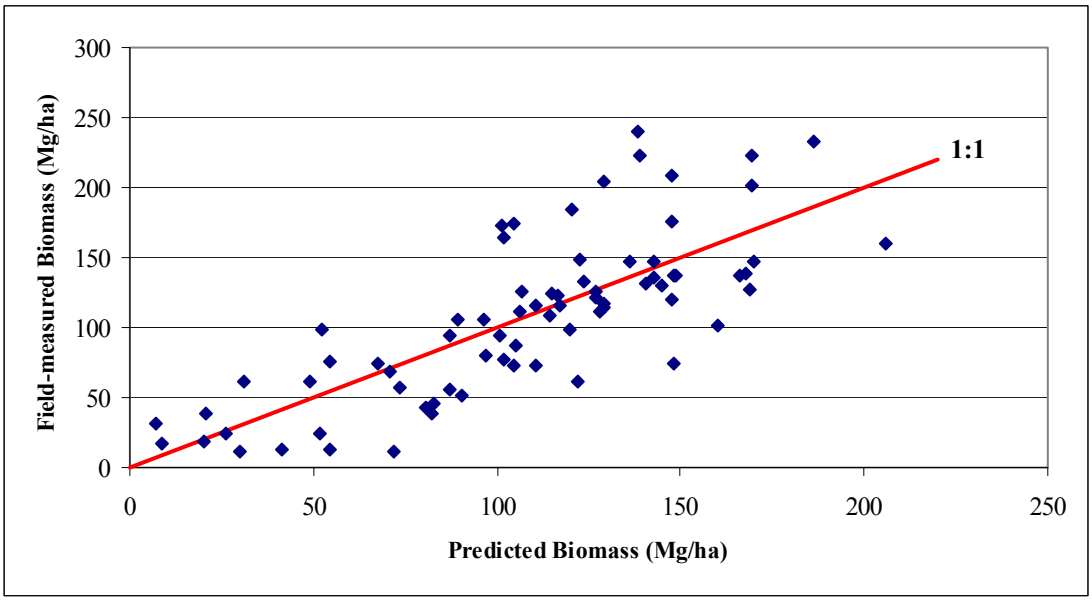


Figure H.122 2-class biomass model (5.632 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.58$)

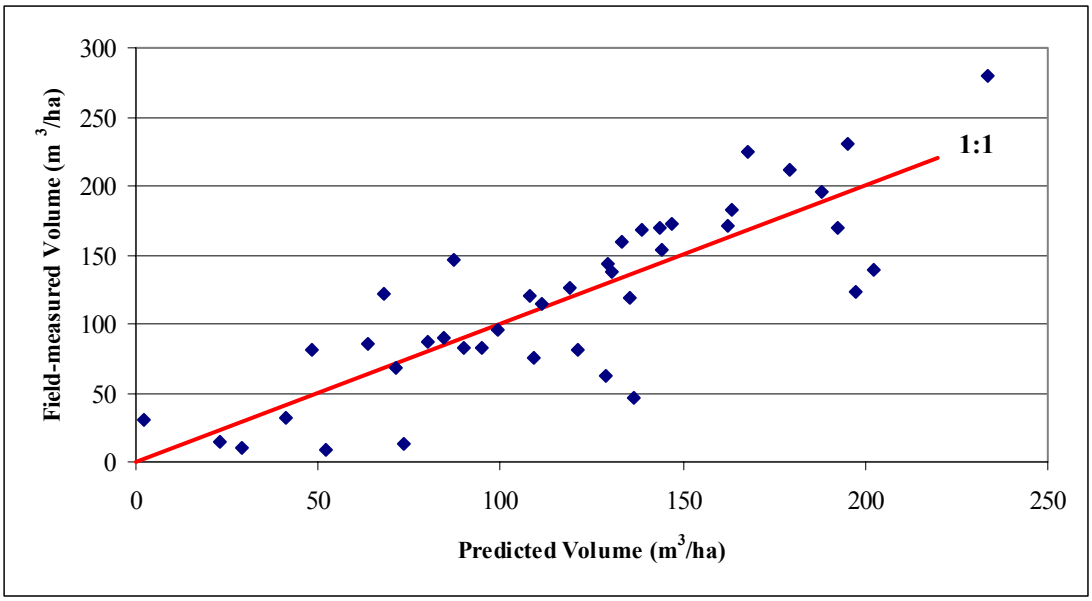


Figure H.123 2-class volume model (5.632 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.66$)

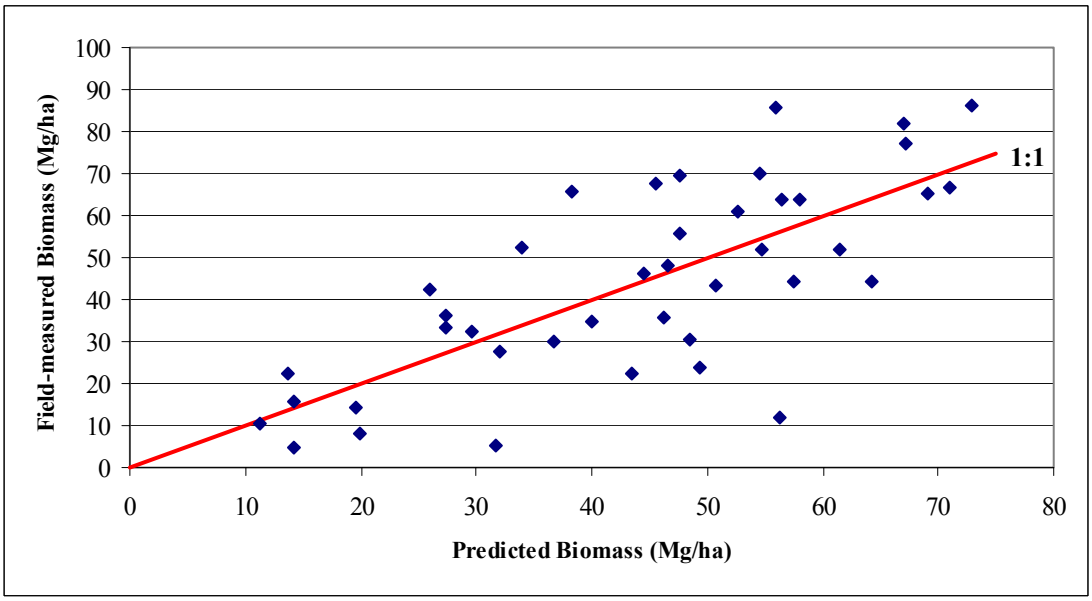


Figure H.124 2-class biomass model (5.632 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.52$)

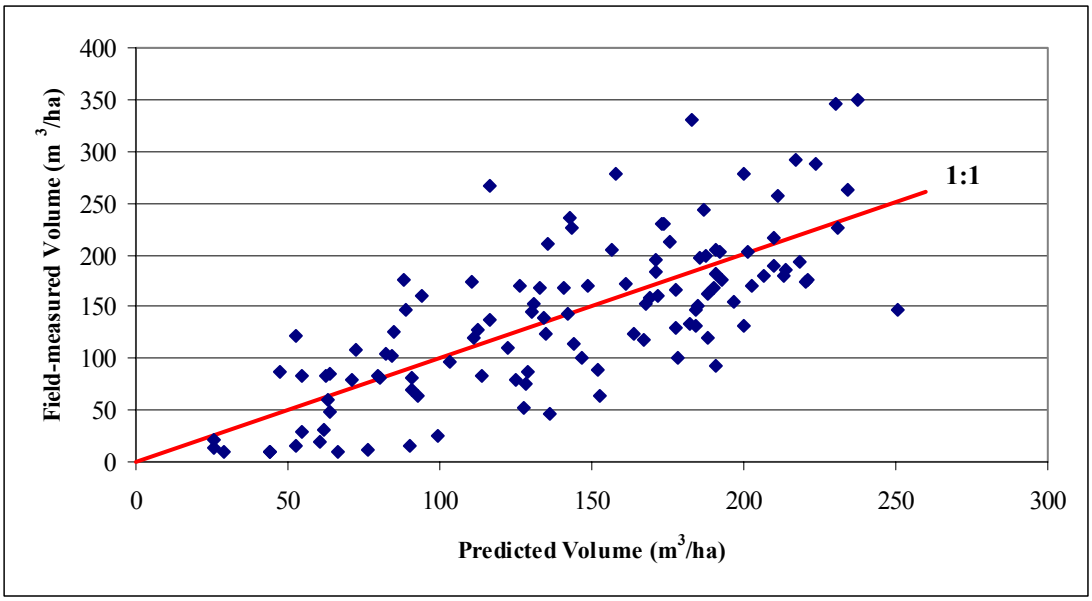


Figure H.125 2-class volume model (5.632 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.54$)

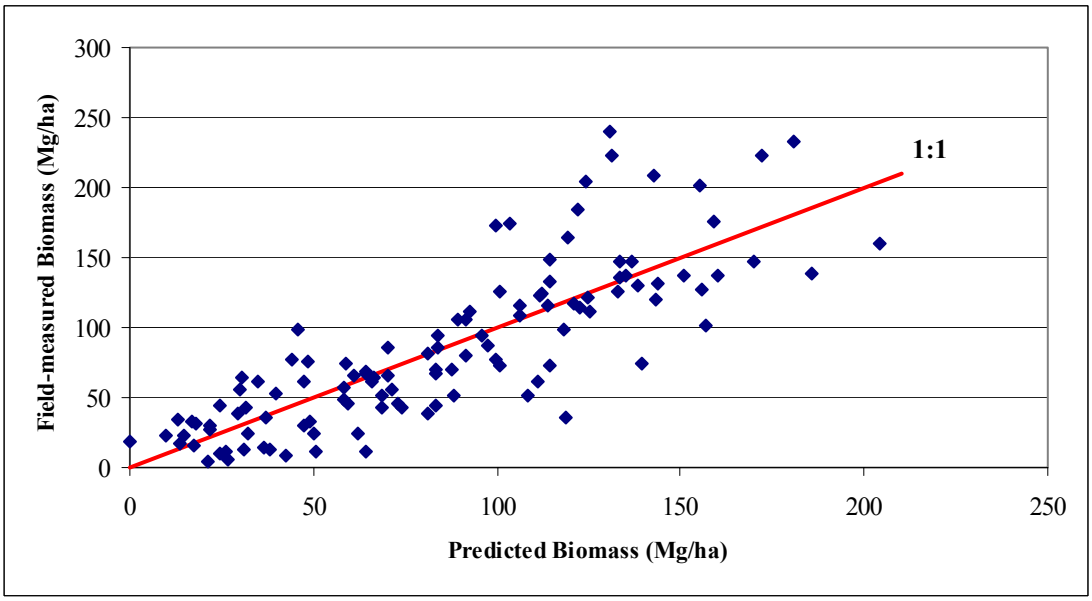


Figure H.126 2-class biomass model (5.632 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.66$)

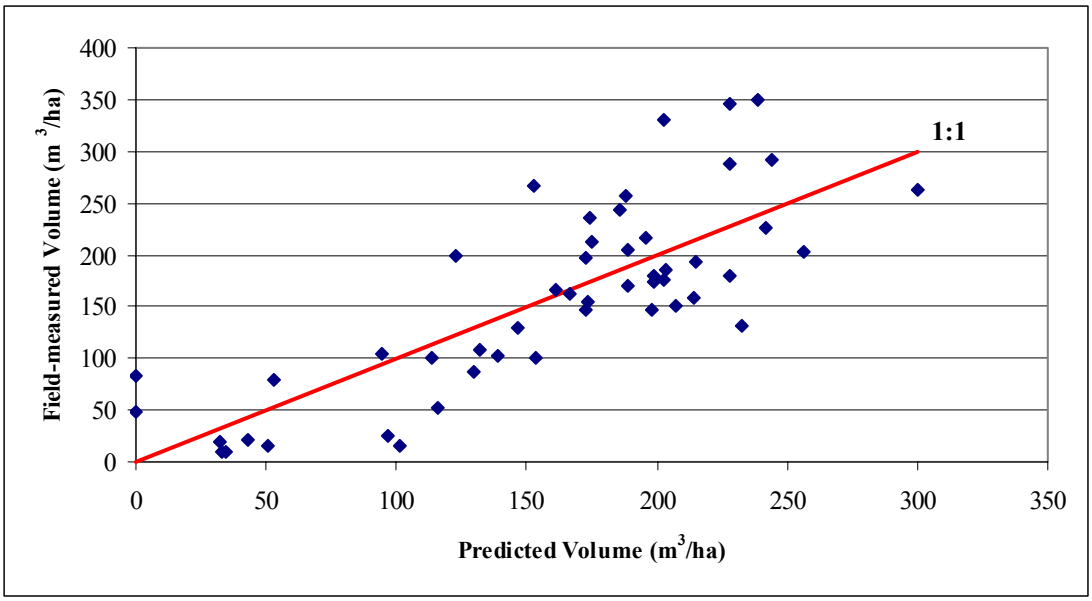


Figure H.127 3-class volume model (5.632 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.62$)

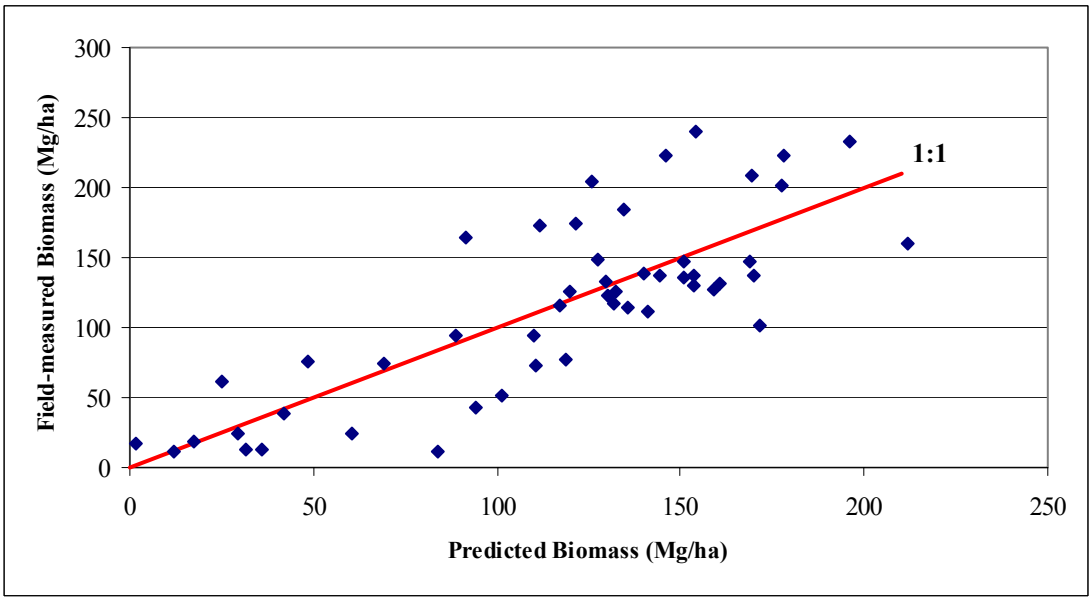


Figure H.128 3-class biomass model (5.632 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.62$)

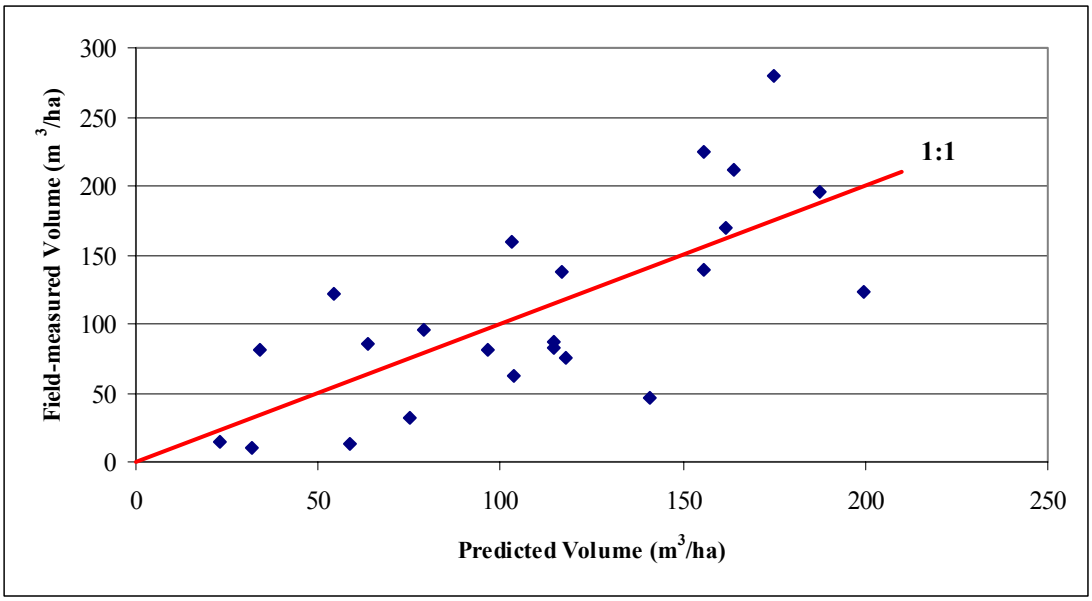


Figure H.129 3-class volume model (5.632 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.47$)

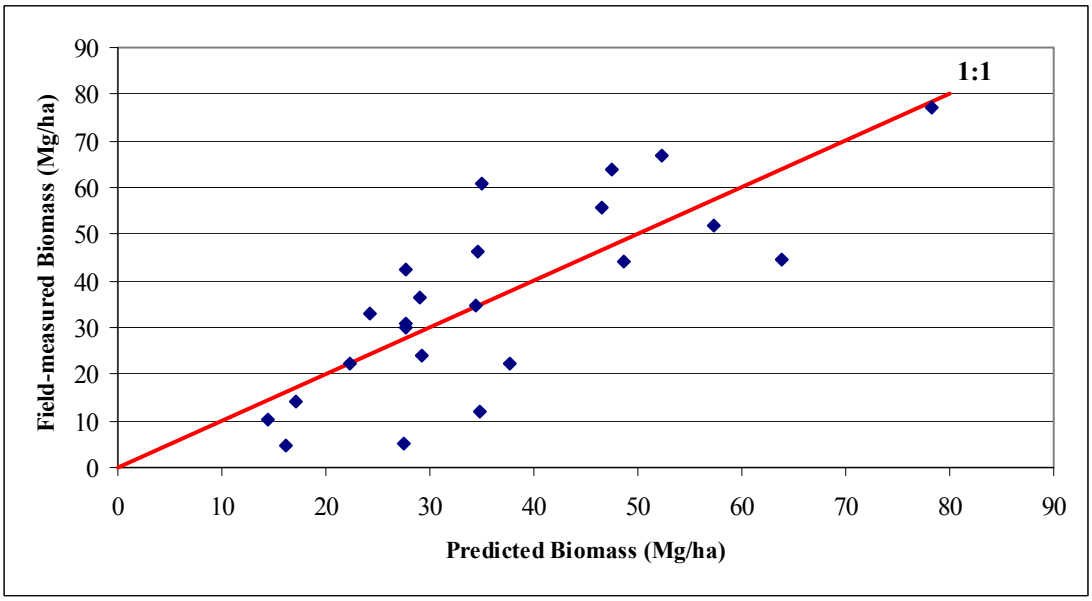


Figure H.130 3-class biomass model (5.632 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.52$)

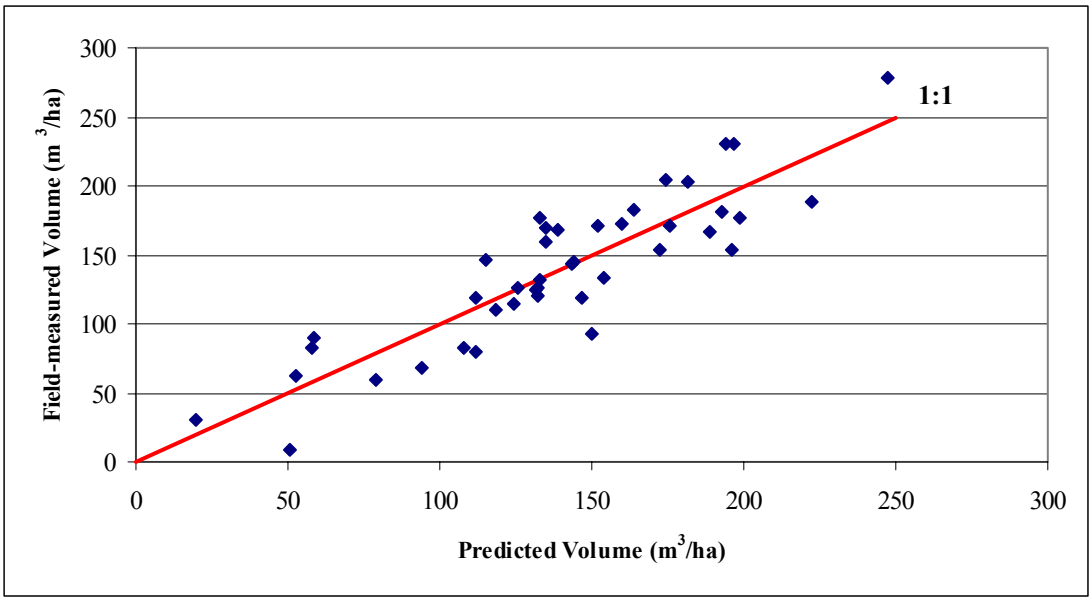


Figure H.131 3-class volume model (5.632 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.74$)

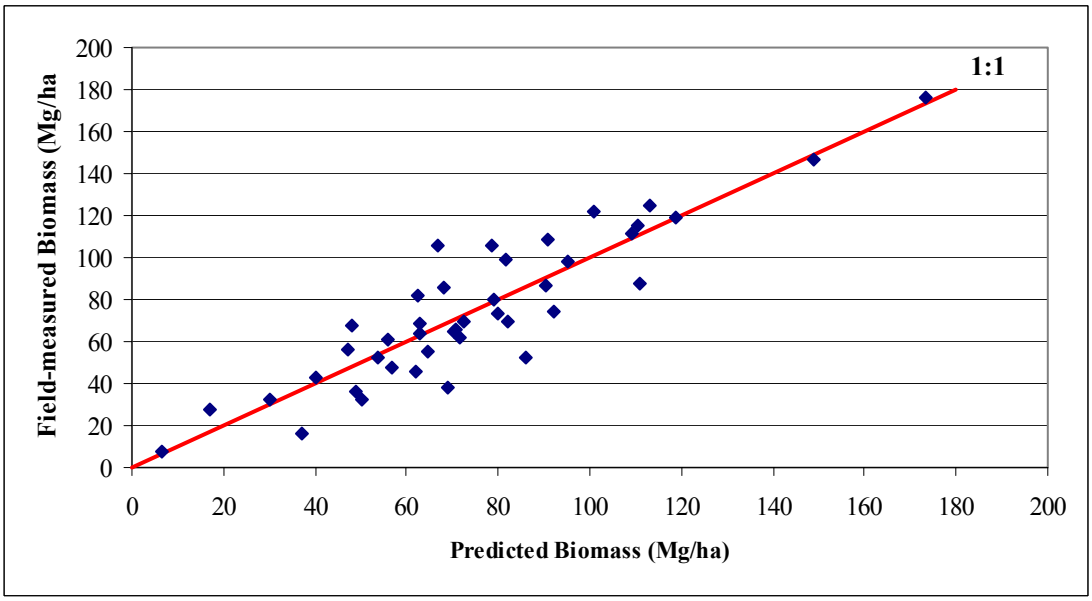


Figure H.132 3-class biomass model (5.632 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.79$)

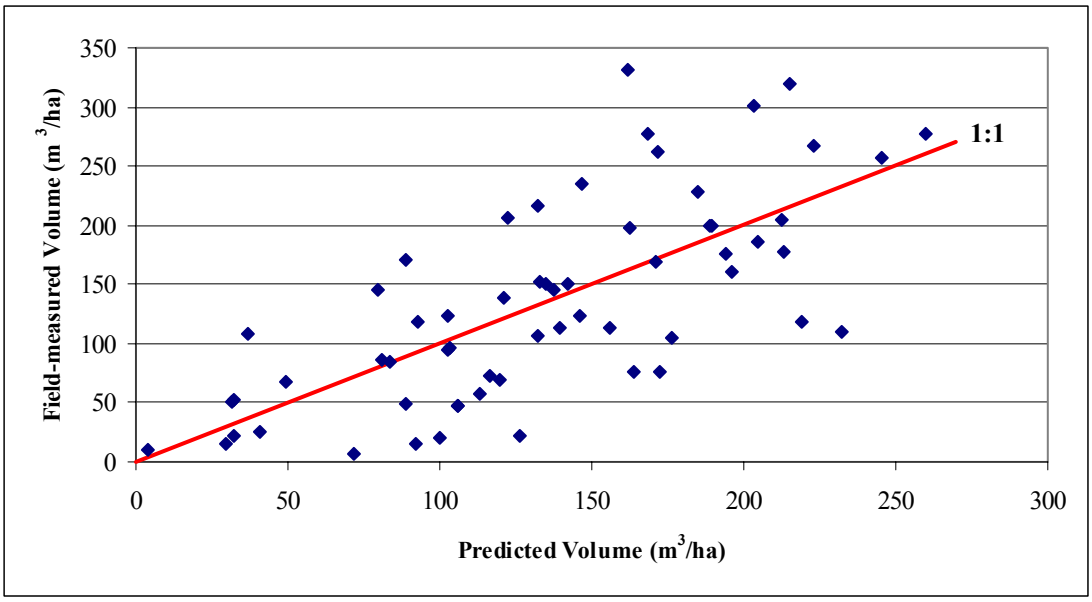


Figure H.133 2-class volume model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.44$)

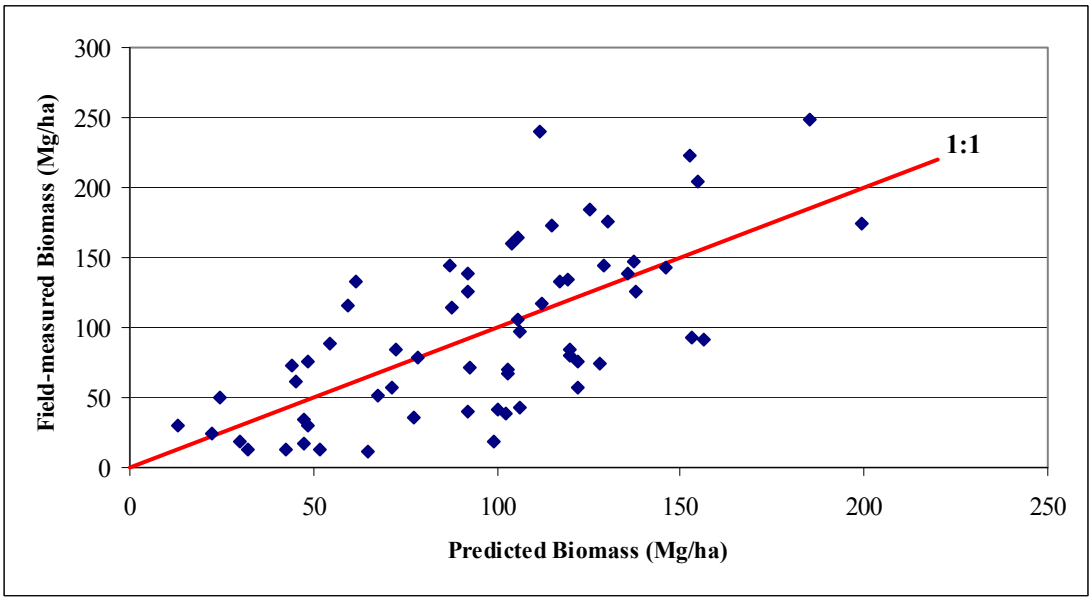


Figure H.134 2-class biomass model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.43$)

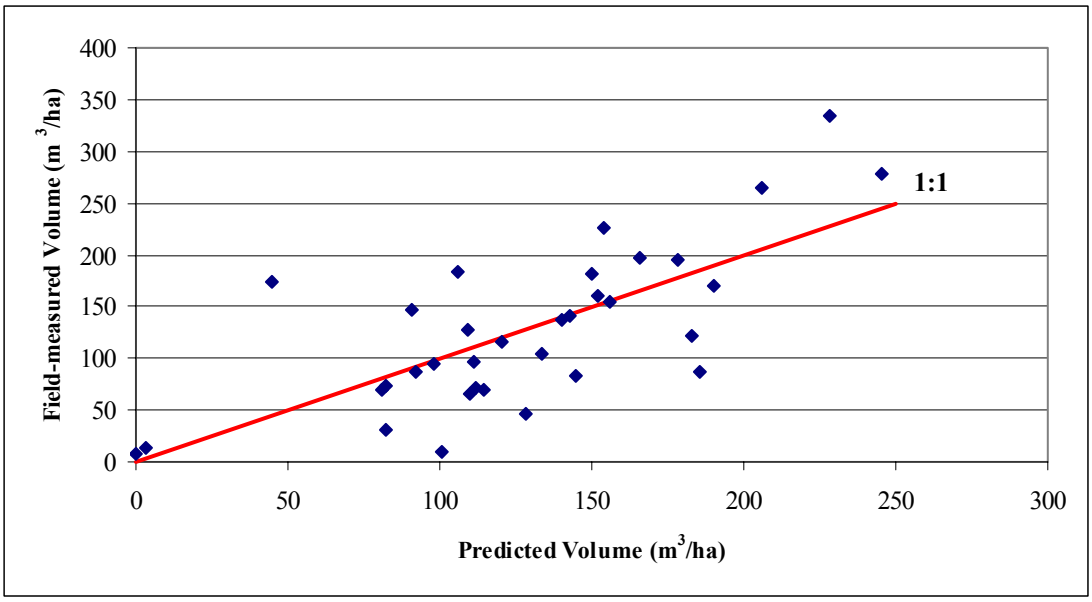


Figure H.135 2-class volume model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.48$)

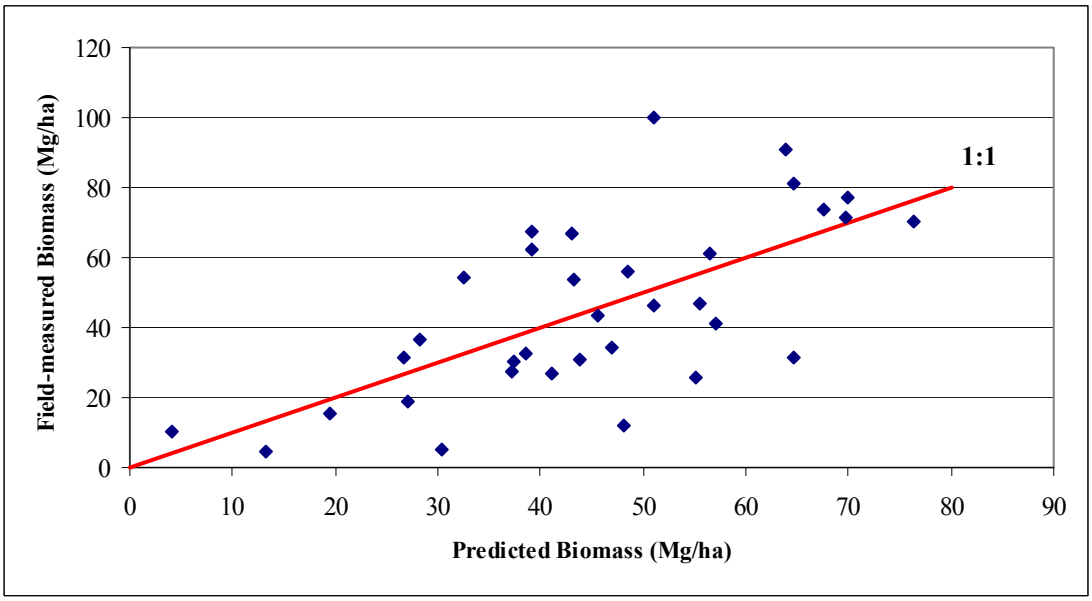


Figure H.136 2-class biomass model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.4$)

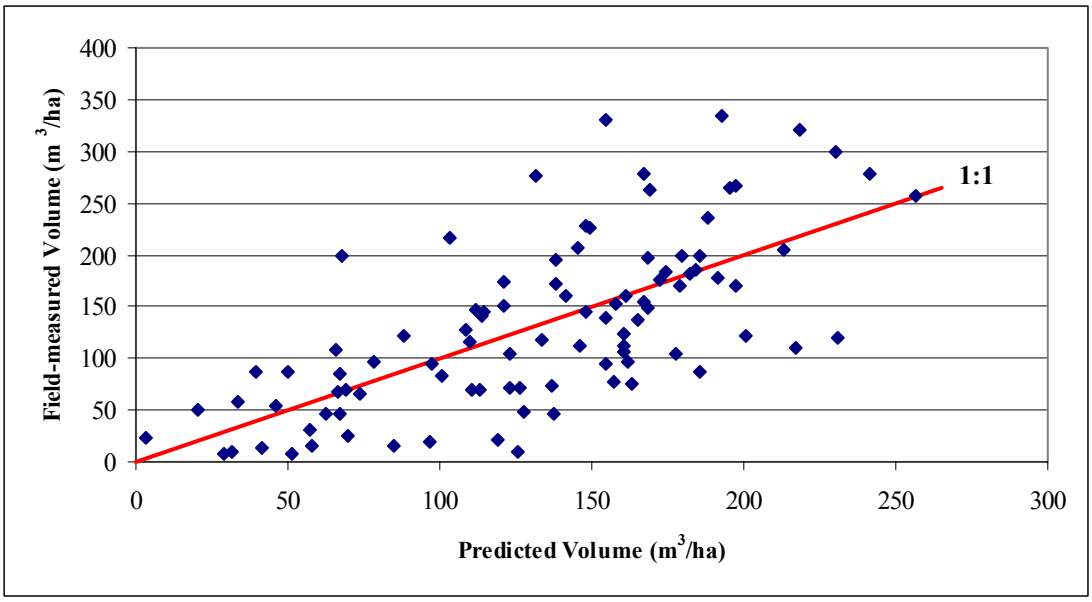


Figure H.137 2-class volume model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted volume/ha values for all plots (adjusted $R^2 = 0.42$)

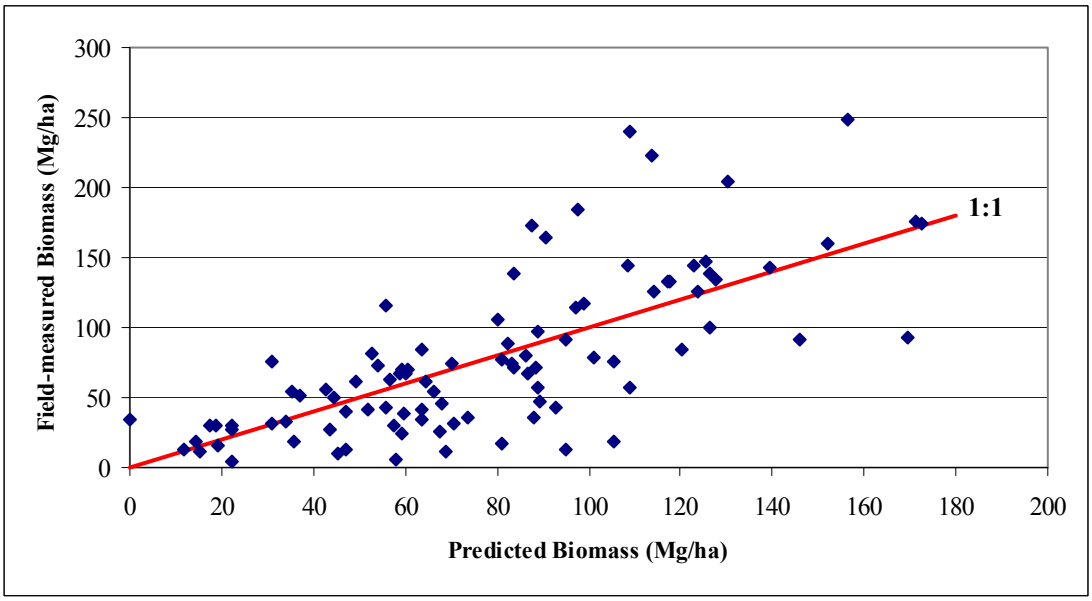


Figure H.138 2-class biomass model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for all plots (adjusted $R^2 = 0.46$)

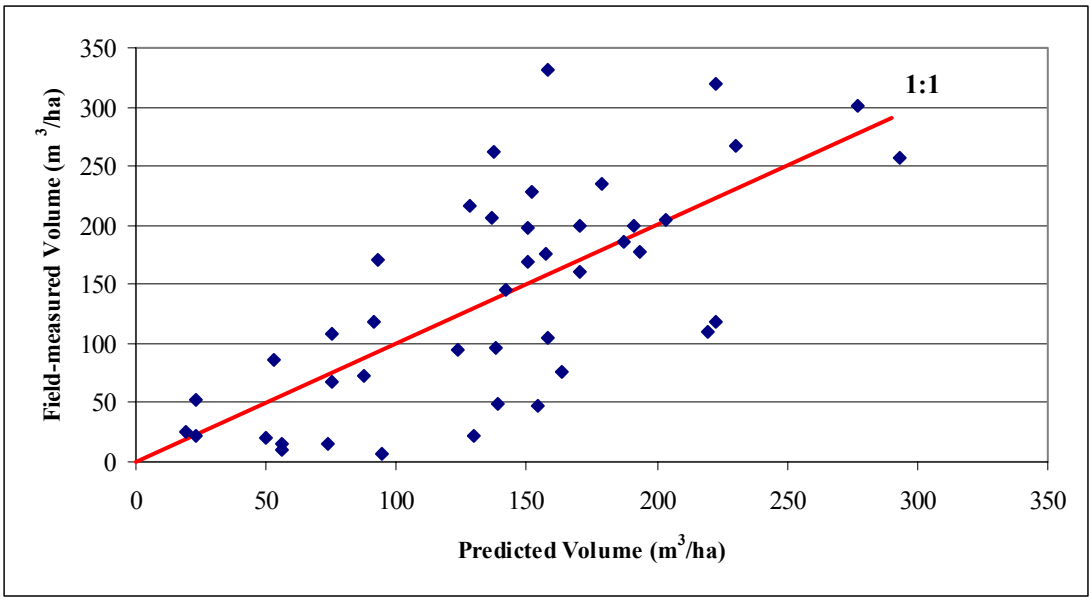


Figure H.139 3-class volume model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted volume/ha values for deciduous plots (adjusted $R^2 = 0.46$)

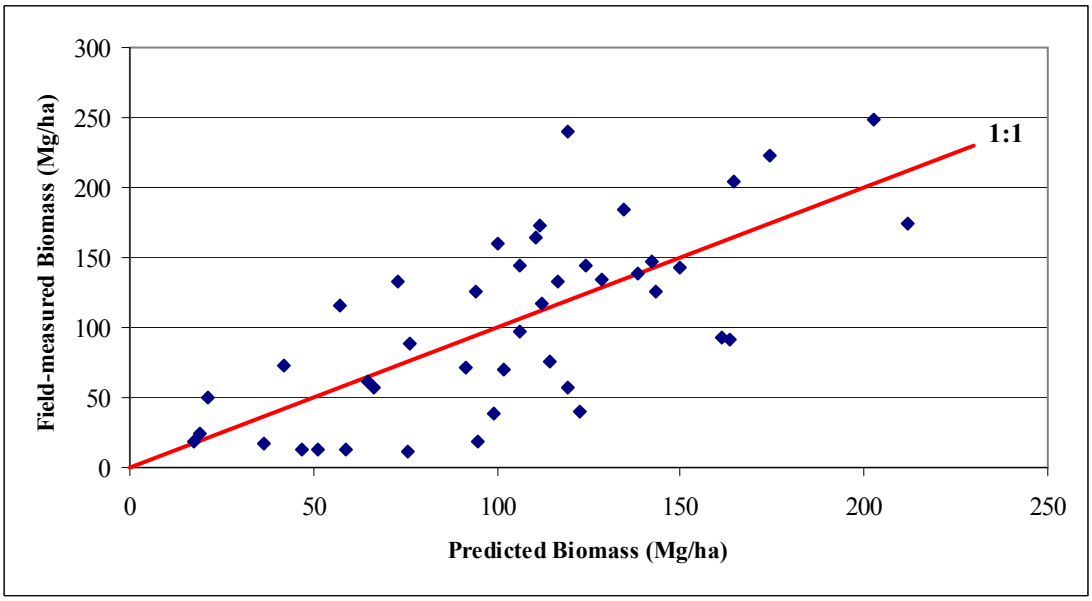


Figure H.140 3-class biomass model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for deciduous plots (adjusted $R^2 = 0.46$)

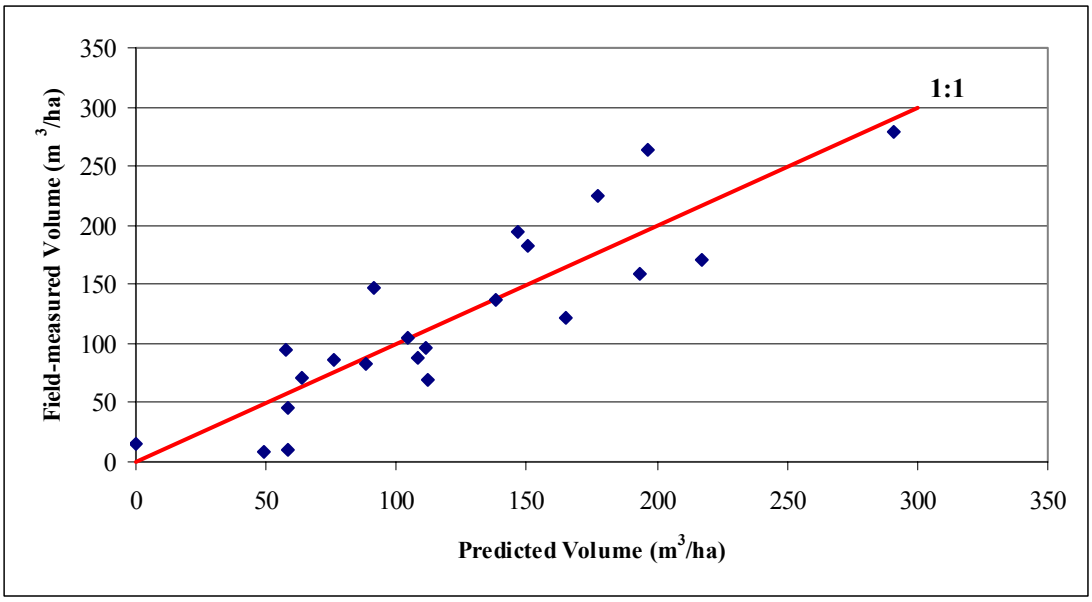


Figure H.141 3-class volume model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted volume/ha values for coniferous plots (adjusted $R^2 = 0.73$)

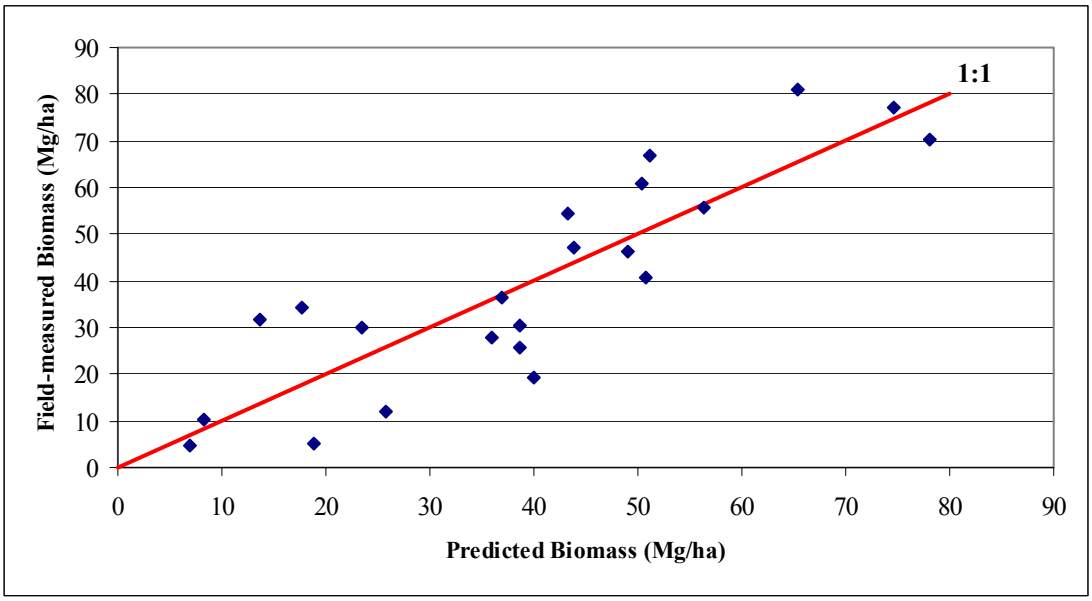


Figure H.142 3-class biomass model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for coniferous plots (adjusted $R^2 = 0.7$)

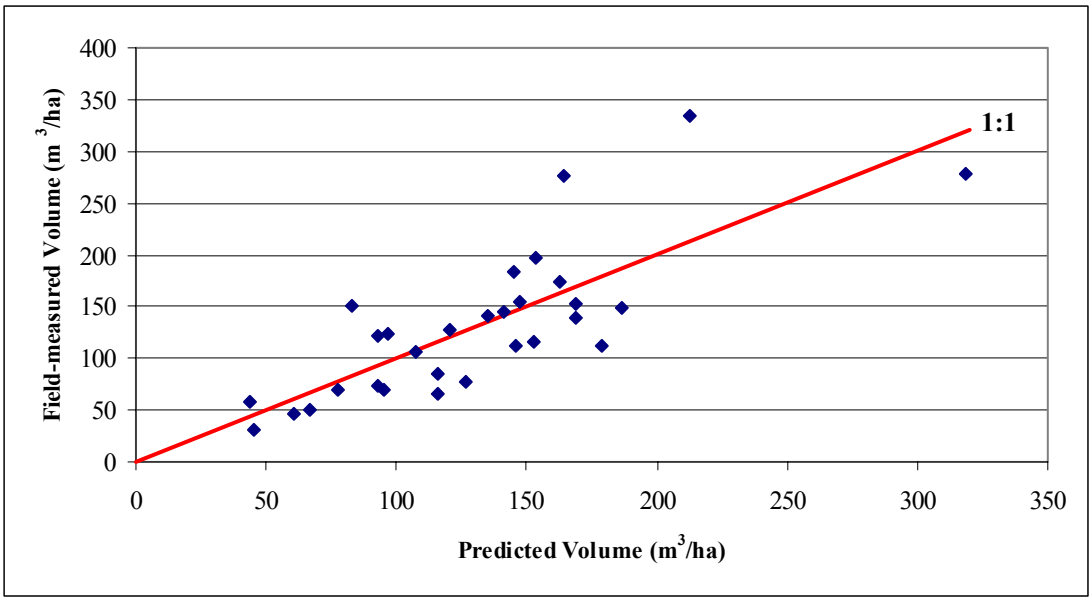


Figure H.143 3-class volume model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted volume/ha values for mixed plots (adjusted $R^2 = 0.57$)

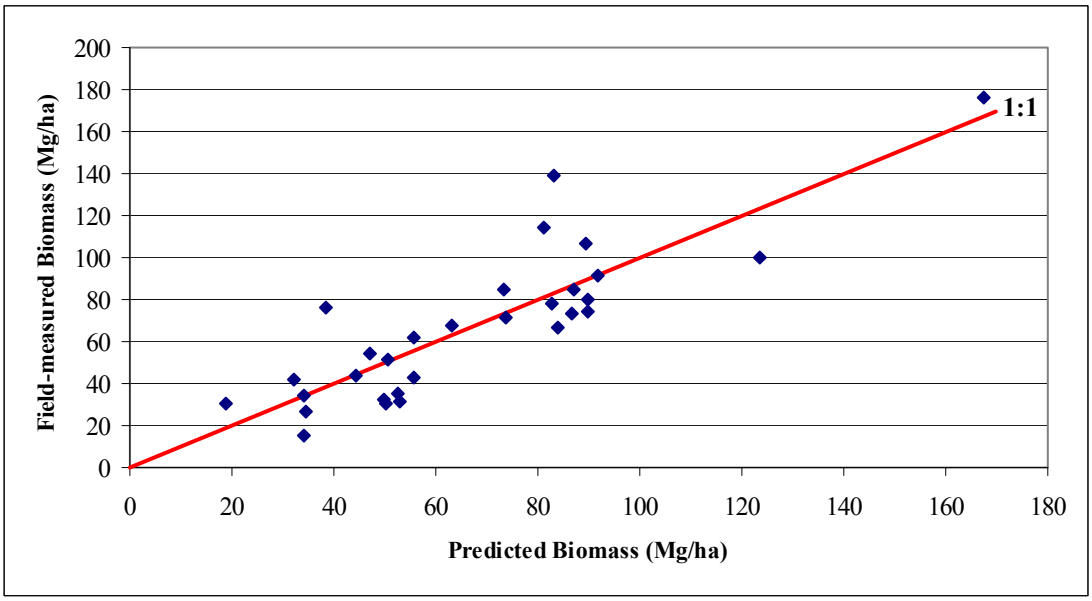


Figure H.144 3-class biomass model (Appomattox stands; 5.666 ha/segment): Field-measured vs. predicted biomass/ha values and residuals for mixed plots (adjusted $R^2 = 0.68$)

Appendix I
Lidar (Distributional) Discriminant Functions for All Model Types and Segmentation Applications

Table I.1 Discriminant functions for all classifications across segmentations applications: *2-class Deciduous-Coniferous (D/C); 3-class Deciduous-Coniferous-Mixed (D/C/M)*

Table I.1	Classification		Discriminant functions
27,050 segments (0.035 ha/segment)	2-class	D	-220.28186 + 2.21116 MaxVeg1 + 0.13773 MedianRef1 + 25.21254 Grnd2ratio + 0.1626 MinRef2 -0.3293 StdMeanRef2
		C	-196.74731 + 1.90057 MaxVeg1 + 0.13087 MedianRef1 + 36.76810 Grnd2ratio + 0.15728 MinRef2 -0.33499 StdMeanRef2
	3-class	D	-354.96818 + 0.19351 MedianRef1 + 1.31272 KurtosisVeg2 + 0.17134 ModeVeg2 - 40.34476 Canopy40P - 3.80619 MeanVeg2 + 303.72094 Veg2ratio
		C	-322.21780 + 0.18577 MedianRef1 + 1.53278 KurtosisVeg2 + 0.39403 ModeVeg2 - 47.10051 Canopy40P - 4.82535 MeanVeg2 + 293.93371 Veg2ratio
		M	-337.73848 + 0.18968 MedianRef1 + 1.32760 KurtosisVeg2 + 0.08721 ModeVeg2 -42.47079 Canopy40P - 3.98111 MeanVeg2 + 296.50557 Veg2ratio
10,352 segments (0.091 ha/segment)	2-class	D	-885.27272 + 6.91706 StdVeg2 + 0.14369 MedianRef1 + 397.4252 Grnd2ratio - 42.75664 Canopy50P + 1.22708 MinRef2 - 0.43471 StdMeanRef2 + 972.39027 Vegratio
		C	-872.63655 + 5.56663 StdVeg2 + 0.13598 MedianRef1 + 406.91115 Grnd2ratio -45.32381 Canopy50P + 1.26961 MinRef2 - 0.46685 StdMeanRef2 + 968.50784 Vegratio
	3-class	D	-978.38175 - 1.04833 StdVeg2 + 0.19634 MedianRef1 + 356.06717 Veg2ratio + 0.58814 StdRef2 - 21.09056 P_Veg2_10 - 103.72083 Canopy40P + 1.94068 MinRef2
		C	-943.74801 - 2.76251 StdVeg2 + 0.18672 MedianRef1 + 345.8106 Veg2ratio + 0.57779 StdRef2 - 21.10671 P_Veg2_10 - 106.91789 Canopy40P + 1.97784 MinRef2
		M	-965.59002 - 1.62510 StdVeg2 + 0.19289 MedianRef1 + 353.26261 Veg2ratio + 0.58116 StdRef2 - 21.83907 P_Veg2_10 - 106.31814 Canopy40P + 1.96162 MinRef2
6,687 segments (0.141 ha/segment)	2-class	D	-6126 + 14.49859 StdVeg2 + 0.14064 MedianRef1 + 59.01074 Veg2ratio + 10.91229 P_Veg2_30 + 3452 MinVeg2 + 5.11071 RangeRef2 + 27.84699 SkewnessVeg1
		C	-6043 + 13.05867 StdVeg2 + 0.13053 MedianRef1 + 47.68549 Veg2ratio + 11.25797 P_Veg2_30 + 3401 MinVeg2 + 5.09188 RangeRef2 + 27.4459 SkewnessVeg1
	3-class	D	-6623 + 24.0105 StdVeg2 + 281.42507 Veg2ratio + 0.03683 StdRef2 + 0.40472 MeanRef1 + 61.43605 P_Veg2_10 - 5.91934 P_Veg2_40 - 11.19655 Canopy70P - 21.83891 Canopy40P + 85.61175 MinVeg1 + 5.07971 RangeRef1
		C	-6503 + 21.84371 StdVeg2 + 265.02779 Veg2ratio + 0.01403 StdRef2 + 0.39181 MeanRef1 + 61.33555 P_Veg2_10 - 5.85239 P_Veg2_40 - 12.7986 Canopy70P - 21.53449 Canopy40P + 85.32023 MinVeg1 + 5.05860 RangeRef1
		M	-6561 + 23.25828 StdVeg2 + 275.47242 Veg2ratio + 0.0303 StdRef2 + 0.39886 MeanRef1 + 60.03777 P_Veg2_10 - 5.47192 P_Veg2_40 - 7.87319 Canopy70P - 24.92840 Canopy40P + 84.5413 MinVeg1 + 5.06374 RangeRef1
2,972 segments	2-class	D	-27677 + 74.9432 StdVeg2 - 0.04385 MedianRef1 - 2370 Grnd1ratio + 20.53036 MaxRef1 -0.07138 CVVeg1
		C	-27520 + 73.61387 StdVeg2 - 0.0531 MedianRef1 - 2331 Grnd1ratio + 20.48244 MaxRef1 - 0.09185 CVVeg1

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table I.1	Classification		Discriminant functions
(0.318 ha/segment)	3-class	D	-36037 + 28.13332 StdVeg2 - 0.39646 MedianRef1 - 2753 Grnd1ratio + 27.09630 MaxRef1 + 7.77237 StdMeanRef2 - 244.02115 Canopy80P + 67.63603 P_Veg2_10 + 56.74712 P_Veg2_40
		C	-35790 + 26.29979 StdVeg2 - 0.40687 MedianRef1 - 2713 Grnd1ratio + 27.01657 MaxRef1 + 7.69979 StdMeanRef2 - 243.80392 Canopy80P + 67.08428 P_Veg2_10 + 56.90142 P_Veg2_40
		M	-35963 + 27.58242 StdVeg2 -0.40372 MedianRef1 - 2737 Grnd1ratio + 27.07397 MaxRef1 + 7.74828 StdMeanRef2 - 238.18718 Canopy80P + 65.84744 P_Veg2_10 + 57.32129 P_Veg2_40
1,473 segments (0.642 ha/segment)	2-class	D	-7025 + 16.83780 StdVeg2 + 0.09654 MedianRef1 + 183.56649 ZeroNVeg2ratio + 30.84950 P_Veg2_30 + 5.86204 RangeRef1 - 0.52111 StdRef1 - 110.85139 Canopy80P
		C	-6936 +15.40834 StdVeg2 + 0.08617 MedianRef1 + 170.76793 ZeroNVeg2ratio + 31.09848 P_Veg2_30 + 5.84131 RangeRef1 - 0.51489 StdRef1 - 112.64084 Canopy80P
	3-class	D	-6195 + 25.06841 StdVeg2 + 0.17951 MedianRef1 + 5.10192 RangeRef1 -1957 ZeroNgrnd1ratio -296.83566 Canopy80P + 25.51799 SkewnessVeg1
		C	-6072 + 23.40229 StdVeg2 + 0.16857 MedianRef1 + 5.06255 RangeRef1 -1904 ZeroNgrnd1ratio -293.81630 Canopy80P + 24.93907 SkewnessVeg1
		M	-6164 + 24.80065 StdVeg2 + 0.17372 MedianRef1 + 5.09289RangeRef1 -1947 ZeroNgrnd1ratio -290.35414 Canopy80P + 25.24283 SkewnessVeg1
981 segments (0.964 ha/segment)	2-class	D	-11643 + 0.28117 MedianRef1 - 2222 ZeroNgrnd1ratio + 9.37966 RangeRef1 + 3197 StdMeanVeg2 + 0.56248 CVVeg2 + 2.65610 P_Veg2_90 + 3.56756 KurtosisVeg2
		C	-11479 + 0.26774 MedianRef1 - 2176 ZeroNgrnd1ratio + 9.32886 RangeRef1 + 3173 StdMeanVeg2 + 0.50876 CVVeg2 + 2.20126 P_Veg2_90 + 3.68584 KurtosisVeg2
	3-class	D	-851.43321 + 12.80738 StdVeg2 + 0.32042 MedianRef1+ 2.47033 MinRef1 -109.05780 ZeroNgrnd1ratio -4.29535 StdMeanRef2 + 70.91896 Canopy70P + 170.53690 ZeroNgrnd3_5ratio
		C	-846.32749 + 11.43988 StdVeg2 + 0.30860 MedianRef1 + 2.55924 MinRef1 -79.49820 ZeroNgrnd1ratio -4.44356 StdMeanRef2 + 72.39397 Canopy70P + 173.58616 ZeroNgrnd3_5ratio
		M	-833.09123 + 12.63929 StdVeg2 + 0.31339 MedianRef1 + 2.45733 MinRef1 -108.51113 ZeroNgrnd1ratio -4.33300 StdMeanRef2 + 74.73918 Canopy70P + 173.58843 ZeroNgrnd3_5ratio
749 segments (1.263 ha/segment)	2-class	D	-23298 - 66.25723 StdVeg2 + 0.18084 MedianRef1 - 885.03315 ZeroNgrnd1ratio - 1.55360 RangeRef1 + 22.63183 RangeVeg2 + 114.73402 Canopy20P + 245.21367 MinVeg1 + 18.52694 MaxRef2
		C	-23226 - 68.32787 StdVeg2 + 0.16873 MedianRef1 - 849.35617 ZeroNgrnd1ratio - 1.60951 RangeRef1 + 22.88804 RangeVeg2 + 111.28676 Canopy20P + 243.40661 MinVeg1 + 18.56196 MaxRef2
	3-class	D	-591.27546 + 10.41905 StdVeg2 + 0.24117 MedianRef1 + 1.71612 MinRef1 -273.44123 ZeroNgrnd1ratio -6.81183 P_Veg2_40 + 41.94105 ZeroNVeg3_5ratio
		C	-565.28167 + 8.74800 StdVeg2 + 0.22534 MedianRef1 + 1.78327 MinRef1 -249.25129 ZeroNgrnd1ratio -6.46401 P_Veg2_40 + 37.00133 ZeroNVeg3_5ratio
		M	-558.03491 + 9.66630 StdVeg2 + 0.23262 MedianRef1 + 1.68843 MinRef1 -262.09922 ZeroNgrnd1ratio -6.31780 P_Veg2_40 + 38.71754 ZeroNVeg3_5ratio

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table I.1	Classification		Discriminant functions
502 segments (1.885 ha/segment)	2-class	D	-10005 + 3.23738 StdVeg2 + 0.42628 MeanRef1 - 524.84239 ZeroNgrnd1ratio - 0.92909 StdRef2 + 7.60943 RangeVeg2 + 17.39611StdMeanRef2 - 1.75116 CVVeg1 - 9683 MinVeg2 + 8.19979 RangeRef1
		C	-9898 + 0.58529 StdVeg2 + 0.40759 MeanRef1 -489.97923 ZeroNgrnd1ratio -0.94604 StdRef2 + 7.87731 RangeVeg2 + 17.65906 StdMeanRef2 -1.78822 CVVeg1 -10055 MinVeg2 + 8.17949 RangeRef1
	3-class	D	-1059 + 10.29183 StdVeg2 + 0.16242 MeanRef1 + 2.56514 MinRef1 + 6.04805 P_Veg2_40 -424.80633 ZeroNgrnd1ratio + 1.12216 StdRef2 -99.51822 Canopy40P
		C	-1018 + 7.53512 StdVeg2 + 0.14517 MeanRef1 + 2.65830 MinRef1 + 6.20028 P_Veg2_40 -387.92146 ZeroNgrnd1ratio + 1.09585 StdRef2 -99.57357 Canopy40P
		M	-1028 + 8.97121 StdVeg2 + 0.15268 MeanRef1 2.56161 MinRef1 + 6.52509 P_Veg2_40 -413.68104 ZeroNgrnd1ratio + 1.11741 StdRef2 -99.40553 Canopy40P
374 segments (2.530 ha/segment)	2-class	D	-525.61126 + 18.48719 StdVeg2 + 0.41204 MeanRef1 + 85.88747 Canopy60P -25.63819 SkewnessVeg1 -6260 MinVeg2 - 0.73779 ModeVeg2
		C	-492.78326 + 17.23640 StdVeg2 + 0.39828 MeanRef1 + 89.30145 Canopy60P -24.97310 SkewnessVeg1 -7234 MinVeg2 - 0.23200 ModeVeg2
	3-class	D	-1586 + 18.70761 StdVeg2 + 0.22732 MeanRef1 + 13.40286 P_Veg1_10 + 2.60885 ModeVeg1 + 1.59183 StdRef2 - 196.60456 ZeroNgrnd2ratio -3.17305 ModeVeg2 + 11133 MinVeg2 + 1184 Canopy80P
		C	-1532 + 17.07606 StdVeg2 + 0.21593 MeanRef1 + 13.34382 P_Veg1_10 + 2.43820 ModeVeg1 + 1.55605 StdRef2 - 185.29991 ZeroNgrnd2ratio -2.12810 ModeVeg2 + 9694 MinVeg2 + 1193 Canopy80P
		M	-1570 + 18.02585 StdVeg2 + 0.21550 MeanRef1 + 13.64770 P_Veg1_10 + 2.46869 ModeVeg1 + 1.59052 StdRef2 - 193.04406 ZeroNgrnd2ratio -2.94849 ModeVeg2 + 10951 MinVeg2 + 1199 Canopy80P
240 segments (3.942 ha/segment)	2-class	D	-476.67360 + 16.62460 StdVeg2 + 0.41419 MeanRef1 -1.67260 RangeVeg2 + 1.69327 ModeVeg1 -217.99366 MinVeg1 - 102.13900 Canopy10P
		C	-434.03492 + 14.69419 StdVeg2 + 0.39656 MeanRef1 -1.48759 RangeVeg2 + 1.61290 ModeVeg1 -196.82809 MinVeg1 - 97.96007 Canopy10P
	3-class	D	-543.81149 + 6.83221 StdVeg2 + 0.23904 MeanRef1 + 7.61120 P_Veg2_10 -4.54951 P_Veg2_40 -5270 MinVeg2 -1.22781 StdMeanRef2 + 0.23332 MedianRef1 -90.49263 Canopy30P
		C	-495.86816 + 4.76171 StdVeg2 + 0.23640 MeanRef1 + 10.98595 P_Veg2_10 -4.77159 P_Veg2_40 -7636 MinVeg2 - 0.99212 StdMeanRef2 + 0.21941 MedianRef1 -91.40219 Canopy30P
		M	-518.03748 + 5.62093 StdVeg2 + 0.22780 MeanRef1 + 5.97844 P_Veg2_10 -4.15472 P_Veg2_40 -5089 MinVeg2 -1.22143 StdMeanRef2 + 0.23497 MedianRef1 -90.88664 Canopy30P
168 segments (5.632 ha/segment)	2-class	D	-20324 -83.30648 StdVeg2 + 0.01879 MedianRef1 + 88.55351 StdMeanRef1 + 16.15333 MaxRef1 + 13.76938 MaxVeg2 + 29.22759 ModeVeg2 -1.30544 RangeRef2 -280.85498 Canopy10P -204.89673 Canopy30P
		C	-20465 -85.72738 StdVeg2 -0.0003548 MedianRef1 + 89.83191 StdMeanRef1 + 16.22372 MaxRef1 + 14.03081 MaxVeg2 + 30.10724 ModeVeg2 -1.30565 RangeRef2 -277.56861 Canopy10P -206.95669 Canopy30P
	3-class	D	-1168 + 22.44972 StdVeg2 + 0.21955 MeanRef1 + 1.75425 StdRef2 + 8.52345 StdMeanRef1 + 17.83418 P_Veg1_10 + 7.71016 P_Veg2_40 + 149.67922 Canopy60P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table I.1	Classification		Discriminant functions
		C	-1102 + 20.01174 StdVeg2 + 0.21122 MeanRef1 + 1.70913 StdRef2 + 9.41433 StdMeanRef1 + 17.62814 P_Veg1_10 + 7.56637 P_Veg2_40 + 150.70204 Canopy60P
		M	-1138 + 21.19001 StdVeg2 + 0.20935 MeanRef1 + 1.75203 StdRef2 + 8.39062 StdMeanRef1 + 17.83999 P_Veg1_10 + 7.96522 P_Veg2_40 + 149.73814 Canopy60P
167 Appomattox forest stands (5.666 ha/segment)	2-class	D	-586.74222 + 0.46828 MedianRef1 + 0.78547 P_Veg2_80 -19.69664 P_Veg2_10 + 24.66836 Canopy60P -128.86603 Canopy10P
		C	-542.52477 + 0.45096 MedianRef1 + 0.37310 P_Veg2_80 -17.45435 P_Veg2_10 + 26.48977 Canopy60P -127.20950 Canopy10P
	3-class	D	-18597 -0.92292 MedianRef1+ 2.01373 MeanRef1 + 168.60523 P_Veg2_10 -13.56087 P_Veg2_80 + 28.04245 P_Veg1_80 -18.37901 MaxVeg2 + 12.87310 MaxRef1 + 4.67738 CVRef2 + 104.39083 ZeroNgrnd2ratio -173.32683 Canopy30P
		C	-18452 -0.95033 MedianRef1 + 2.02294 MeanRef1 + 173.48129 P_Veg2_10 -14.46028 P_Veg2_80 + 28.02604 P_Veg1_80 -18.25467 MaxVeg2 + 12.84133 MaxRef1 + 4.60070 CVRef2 + 101.01297 ZeroNgrnd2ratio -177.14513 Canopy30P
		M	-18417 -0.89293 MedianRef1 + 1.96387 MeanRef1 + 169.05846 P_Veg2_10 -13.75264 P_Veg2_80 + 27.39573 P_Veg1_80 -18.15894 MaxVeg2 + 12.81487 MaxRef1 + 4.90972 CVRef2 + 117.42060 ZeroNgrnd2ratio -172.24469 Canopy30P

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Appendix J
Canopy Height Model (Distributional) Discriminant Functions for All Model Types and Segmentation Applications

Table J.1 Discriminant functions for all classifications across segmentations applications: *2-class Deciduous-Coniferous (D/C); 3-class Deciduous-Coniferous-Mixed (D/C/M)*

Table J.1	Classification		Discriminant functions
27,050 segments (0.035 ha/segment)	2-class	D	-16.02088 + 4.18111 Canopy30P + 1.62393 P_Veg1_90 + 22.93282 Canopy60P -0.23539 RangeVeg1
		C	-10.13985 + 1.97358 Canopy30P + 1.20264 P_Veg1_90 + 19.49061 Canopy60P -0.10695 RangeVeg1
	3-class	D	-15.63599 + 1.31037 MaxVeg1 + 16.74440 Canopy60P
		C	-9.04940 + 0.96154 MaxVeg1 + 14.28681 Canopy60P
		M	-13.68514 + 1.22747 MaxVeg1 + 15.55567 Canopy60P
10,352 segments (0.091 ha/segment)	2-class	D	-9.48891 + 20.54109 Canopy30P + 44.75997 StdMeanVeg1 + 0.97840 P_Veg1_25 + 0.00517 CVVeg1
		C	-4.64922 + 13.27147 Canopy30P + 28.27693 StdMeanVeg1 + 0.67252 P_Veg1_25 + 0.01862 CVVeg1
	3-class	D	-11.02532 + 24.80700 Canopy30P + 28.22601 StdMeanVeg1 -0.00168 CVVeg1 + 1.07590 P_Veg1_60
		C	-4.91824 + 14.98880 Canopy30P + 13.78656 StdMeanVeg1 + 0.01872 CVVeg1 + 0.70758 P_Veg1_60
		M	-8.43425 + 19.11389 Canopy30P + 25.45987 StdMeanVeg1 + 0.00739 CVVeg1 + 0.93577 P_Veg1_60
6,687 segments (0.141 ha/segment)	2-class	D	-5.48812 -2.18589 Canopy20P + 40.36917 StdMeanVeg1 + 10.73661 Canopy60P + 1.02411 MinVeg1
		C	-4.07808 -4.14500 Canopy20P + 20.93507 StdMeanVeg1 + 11.77489 Canopy60P + 0.85857 MinVeg1
	3-class	D	-291.37783 + 2.57087 MaxVeg1 -5.68564 Canopy30P + 551.63322 Canopy90P
		C	-288.20630 + 2.20359 MaxVeg1 -10.06656 Canopy30P + 556.05046 Canopy90P
		M	-300.98766 + 2.51541 MaxVeg1 -8.49446 Canopy30P + 563.18661 Canopy90P
2,972 segments (0.318 ha/segment)	2-class	D	-3.51991 + 78.23869 StdMeanVeg1 + 4.19205 Canopy40P + 0.41452 MinVeg1
		C	-1.91652 + 46.29248 StdMeanVeg1 + 4.98637 Canopy40P + 0.33364 MinVeg1
	3-class	D	-24.90124 + 1.16742 MaxVeg1 -8.10690 StdMeanVeg1 -4.53093 Canopy40P + 1.11668 MinVeg1 + 32.77116 Canopy70P
		C	-21.23171 + 0.94357 MaxVeg1 -38.93434 StdMeanVeg1 -6.87550 Canopy40P + 1.23663 MinVeg1 + 36.93112 Canopy70P
		M	-25.75553 + 1.14759 MaxVeg1 -17.64346 StdMeanVeg1 -8.15406 Canopy40P + 1.06801 MinVeg1 + 36.96250 Canopy70P
1,473 segments (0.642 ha/segment)	2-class	D	-144.97677 -1.40571 MaxVeg1 + 52.42084 Canopy30P + 2.98673 MinVeg1 + 0.04887 CVVeg1 + 259.98032 Canopy80P + 6.30428 P_Veg1_30
		C	-140.99422 -1.55879 MaxVeg1 + 48.86259 Canopy30P + 3.21234 MinVeg1 + 0.04300 CVVeg1 + 262.80629 Canopy80P + 6.06250 P_Veg1_30
	3-class	D	-70.23616 + 119.85016 Canopy80P -7.19107 Canopy30P + 1.39643 MaxVeg1
		C	-67.98716 + 124.89660 Canopy80P -9.33713 Canopy30P + 1.09733 MaxVeg1
		M	-75.60618 + 127.17879 Canopy80P -9.77994 Canopy30P + 1.36008 MaxVeg1
981 segments (0.964 ha/segment)	2-class	D	-17.37863 + 1.53670 P_Veg1_90 + 26.51122 Canopy30P
		C	-10.42261 + 1.16259 P_Veg1_90 + 22.35694 Canopy30P
	3-class	D	-40.28102 + 2.21214 P_Veg1_90 + 15.87742 Canopy30P + 51.56988 Canopy70P -72.59962 StdMeanVeg1

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Table J.1	Classification		Discriminant functions
		C	-35.90474 + 1.86408 P_Veg1_90 + 11.04038 Canopy30P + 55.21969 Canopy70P -93.83738 StdMeanVeg1
		M	-42.75061 + 2.19049 P_Veg1_90 + 12.61167 Canopy30P + 57.36826 Canopy70P -98.88499 StdMeanVeg1
749 segments (1.263 ha/segment)	2-class	D	-168.91697 + 2.80160 P_Veg1_90 + 12.14562 Canopy20P + 0.43014 KurtosisVeg1 + 302.61449 Canopy80P
		C	-166.00525 + 2.46299 P_Veg1_90 + 7.36234 Canopy20P + 0.42948 KurtosisVeg1 + 306.00914 Canopy80P
	3-class	D	-28.92439 -1.25878 Canopy30P + 33.74289 Canopy70P + 1.17837 RangeVeg1 + 45.26000 StdMeanVeg1
		C	-26.51909 -4.18662 Canopy30P + 40.99007 Canopy70P + 0.86926 RangeVeg1 + 22.84735 StdMeanVeg1
		M	-31.91150 -3.45877 Canopy30P + 38.88901 Canopy70P + 1.18212 RangeVeg1 + 24.10937 StdMeanVeg1
502 segments (1.885 ha/segment)	2-class	D	-16.59886 + 1.65916 P_Veg1_90 + 32.78734 Canopy20P -0.01796 KurtosisVeg1 + 0.25038 MinVeg1 -112.42463 StdMeanVeg1
		C	-8.91154 + 1.20723 P_Veg1_90 + 25.88956 Canopy20P -0.01408 KurtosisVeg1 + 0.59259 MinVeg1 -89.76819 StdMeanVeg1
	3-class	D	-22.26192 + 1.06685 RangeVeg1 + 27.77242 Canopy20P + 0.72344 P_Veg1_60
		C	-11.68940 + 0.88986 RangeVeg1 + 19.36670 Canopy20P + 0.30929 P_Veg1_60
		M	-19.56962 + 1.17080 RangeVeg1 + 21.65176 Canopy20P + 0.40540 P_Veg1_60
374 segments (2.530 ha/segment)	2-class	D	-23.37206 + 1.74206 P_Veg1_90 + 3.91169 Canopy20P + 33.56530 Canopy40P -2.17339 SkewnessVeg1
		C	-17.83918 + 1.43622 P_Veg1_90 -0.29689 Canopy20P + 33.77598 Canopy40P -1.95373 SkewnessVeg1
	3-class	D	-504.09233 + 2.15980 P_Veg1_90 + 53.38467 Canopy30P -152.51264 Canopy60P + 1103 Canopy80P
		C	-501.06936 + 1.78409 P_Veg1_90 + 47.49819 Canopy30P -147.35499 Canopy60P + 1104 Canopy80P
		M	-512.95921 + 1.99352 P_Veg1_90 + 50.63491 Canopy30P -153.46304 Canopy60P + 1117 Canopy80P
240 segments (3.942 ha/segment)	2-class	D	-28.23360 + 35.41092 Canopy20P + 40.28714 StdMeanVeg1 + 1.26598 MaxVeg1 + 1.40925 P_Veg1_20 0.05796 StdVeg1
		C	-20.91936 + 28.10582 Canopy20P + 61.35986 StdMeanVeg1 + 1.37534 MaxVeg1 + 0.89122 P_Veg1_20 -1.20078 StdVeg1
	3-class	D	-87.64630 + 2.98511 P_Veg1_90 + 31.17477 Canopy20P + 127.69137 Canopy70P
		C	-81.37748 + 2.50435 P_Veg1_90 + 23.19332 Canopy20P + 131.04823 Canopy70P
		M	-84.03956 + 2.77933 P_Veg1_90 + 26.51731 Canopy20P + 128.84286 Canopy70P
168 segments (5.632 ha/segment)	2-class	D	-28.72889 + 13.87552 Canopy20P + 2.19963 MaxVeg1 + 0.19320 KurtosisVeg1 + 242.36834 StdMeanVeg1 -2.94237 StdVeg1
		C	-23.19582 + 13.41069 Canopy20P + 2.06047 MaxVeg1 + 0.17908 KurtosisVeg1 + 249.99577 StdMeanVeg1 -3.36415 StdVeg1
	3-class	D	-37871 -2.26247 RangeVeg1 + 233.42974 Canopy20P -1321 Canopy60P + 77015 Canopy90P
		C	-37886 -2.74698 RangeVeg1 + 225.35621 Canopy20P -1310 Canopy60P + 77036 Canopy90P
		M	-37991 -2.39128 RangeVeg1 + 230.28219 Canopy20P -1319 Canopy60P + 77138 Canopy90P
167 Appomattox forest stands (5.666 ha/segment)	2-class	D	-35.59318 + 1.17893 StdVeg1 + 0.18297 CVVeg1 + 1.82289 P_Veg1_20 + 1.40863 RangeVeg1
		C	-29.14895 + 0.36201 StdVeg1 + 0.15296 CVVeg1 + 1.57128 P_Veg1_20 + 1.44358 RangeVeg1
	3-class	D	-38.48022 + 2.14595 StdVeg1 + 0.19293 CVVeg1 + 1.74255 P_Veg1_25 + 0.21081 KurtosisVeg1 + 1.28327 RangeVeg1

Veg = Vegetation lidar hit; Grnd = Ground lidar hit; Ref = Reflectance associated with lidar hit; Veg1, 2, or 3_5 = 1st, 2nd, or grouped 3rd through 5th returns; P_..._10-90 = Percentiles; CV = Coefficient of variation; StdMean = Standard error of the mean; Std = Standard deviation; Canopy10-90 = Canopy cover percentiles; N..ratio = Vegetation or ground hits as a ratio of return totals; Vegratio = Vegetation hits as a ratio of total hits

Appendix K
Microsoft C++ Code for Between- and Within Segment Variance Calculation

```
// August 1, 2004 - For per-segment CHM between- and within-segment variance calculations
// Within and between variance calculation for eCognition output file with 4 input columns
// Columns 1 = ID 2-3 = Mean and Std. dev of layers 4 = Other eCognition vars., e.g., Area, Shape, etc.
// Outputs the within and between variances and between/within ratios
// 1 set of input columns: 1 CHM; Reads columns 1,2,3, then skips 4,5 and reads 6, ignores rest of line
// CAN HANDLE UP TO 299 INPUT FILES!!!!

#include <fstream.h>            //File streams
#include <string.h>             //Text string use
#include <math.h>               //For mathematical operators
#include <iomanip.h>            //For input/output manipulation
#include <iostream.h>           //For file input/output
#include <stdio.h>
#include <process.h>
#include <malloc.h>
#include <memory.h>
#include <string.h>             // For string manipulation
#include <stdlib.h>             // For text manipulation

// Global variables
const int NumColumns = 4;                                   //58 input columns; 51 calc columns
const int NumRows = 290000;                                 //250,000 for smallest of resolutions and scale parameters; 250,000 segments safe

double Segment_array[NumRows][NumColumns];                  //Declare array to hold input

int Rows = 0;                                               //Acts as counter and rownumber for reading in files
int Segment_elements = 0;                                   //Counter for # elements per each segment or row

double Design_mean = 0.0;                                   //Variable to hold design mean final for each file-run
double Design_w_var = 0.0;                                  //Variable to hold design within variance for each file-run
double Design_b_var = 0.0;                                  //Variable to hold design between variance final for each file-run
double Ratio = 0.0;                                         //Array to hold ratio of between and within variances
double Mi_total = 0.0;                                      //Variable for # elemens per segment

char CharNumRuns[5];                                        //Variable arrays that hold the character conversion for # runs requested
char CharNumRuns2[5];                                       //3 steps
```

```
char CharNumRuns3[5];

// BEGIN MAIN PROGRAM
void main () {

        char InputFileName[50];                 //Input file
        char InputFileNameTemp[50];             //Temp input file to hold root input file name; copied to read each run
        char OutputFileName[50];                //Output file

        int NumFiles = 0;                       //# files to use for calculation
        double Resolution = 0;                  //Resolution segment pixels

        ifstream inFile;                        //Input file stream
        ofstream outFile;                       //Output file stream


        //INITIALIZE THE MAIN ARRAY SO THAT NO "FALSE" VALUES REMAIN
        for (int CleanRows = 0; CleanRows < NumRows; CleanRows++)
        {
                for (int CleanColumns =0; CleanColumns < NumColumns; CleanColumns++)
                        Segment_array[CleanRows][CleanColumns] = 0.0;
        }

        //USER INPUT SECTION: IN AND OUT FILENAMES, RESOLUTION, # FILES TO RUN
        cout << endl << endl;
        cout << "*****************************************************" << endl;
        cout << endl;
        cout << "INPUT SECTION:  Please specify the following file names:" << endl;
        cout << "_____________________________________________________" << endl;
        cout << endl << endl;
        cout << endl;
        cout << "Input BASE file name (max. 50 chars; EXCLUDE extension): ";
        cin  >> InputFileName;
        cout << "Output file name (max. 50 chars; include extension):     ";
        cin  >> OutputFileName;
        cout << "Segmentation input resolution in meters:                 ";
        cin  >> Resolution;
        cout << "Number of input files to use for variance calculation:   ";
        cin  >> NumFiles;
        cout << endl;
```

```
cout << "*****************************************************************" << endl;
cout << endl << endl;
cout << "Thank you!  Please wait for your output file...." << endl;
cout << endl;

//OPEN OUTPUT FILE
outFile.open(OutputFileName);

//OUTPUT COLUMN HEADINGS TO OUTPUT FILE
outFile <<          "Data set" << "    " << "# Segments" << "     " << "# Total Elements" <<
                    "          " << "M1" << "    " << "WSV1" << "          " << "BSV1" << "          " << "RATIOBW1" << endl;

// DATA READ
strcpy(InputFileNameTemp,InputFileName);          //Copies root input filename to temp filename;
                                                  //Temp is used to reset input filename to read in next file + # + .txt
int RunSizer = 0;
int NumCounter = 0;

// For-loop to read in data for each file, and calculate variances, "outer" For-loop
for (int NumRuns = 1; NumRuns <= NumFiles; NumRuns++) //Repeat # files times

{
        Rows = 0;                                 //Counter to run through array rows

        if (NumRuns < 10)                         //If Numruns<10; only regular inputfilename needed
        {
                itoa(NumRuns, CharNumRuns, 100);                                    //Convert NumRuns to CharNumRuns, a character
                inFile.open(strcat(strcat(InputFileName, CharNumRuns), ".txt"));    //Open the concatenated filename
        }
        else if (NumRuns >= 10 && NumRuns < 100)          //If Numruns > 10 and < 100, need to handle in 10-increments
        {
                NumCounter = 0;                           //Counter that holds the 10 decimals count

                RunSizer = NumRuns;                       //Counter that holds the 1th count
```

```
        while (RunSizer >= 10 && RunSizer < 100) //Runs until Runsizer < 10; then concats Numcounter + Runsizer
        {
                NumCounter++;

                RunSizer = RunSizer - 10;
        }

        itoa(NumCounter, CharNumRuns2, 100);
        itoa(RunSizer, CharNumRuns, 100);
        strcat(InputFileName, CharNumRuns2);
        strcat(InputFileName, CharNumRuns);
        strcat(InputFileName, ".txt");
        inFile.open(InputFileName);
}

else if (NumRuns >= 100 && NumRuns < 200)          //Same procedure, except a 100th counter, 10th counter and final
                                                   //Runsizer or 1th counter is used
{
        NumCounter = 0;
        RunSizer = NumRuns - 100;

        if (RunSizer < 10)                         //For values 100-109
        {
                itoa(1, CharNumRuns3, 100);
                itoa(0, CharNumRuns2, 100);
                itoa(RunSizer, CharNumRuns, 100);

                strcat(InputFileName, CharNumRuns3);
                strcat(InputFileName, CharNumRuns2);
                strcat(InputFileName, CharNumRuns);
                strcat(InputFileName, ".txt");

                inFile.open(InputFileName);
        }

        if (RunSizer >=10)                         //For values 110-199
        {
                while (RunSizer >= 10 && RunSizer < 100)
                {
                        NumCounter++;
```

```
                                        RunSizer = RunSizer - 10;
                                }
                                itoa(1, CharNumRuns3, 100);
                                itoa(NumCounter, CharNumRuns2, 100);
                                itoa(RunSizer, CharNumRuns, 100);

                                strcat(InputFileName, CharNumRuns3);
                                strcat(InputFileName, CharNumRuns2);
                                strcat(InputFileName, CharNumRuns);
                                strcat(InputFileName, ".txt");

                                inFile.open(InputFileName);
                        }
                }

else if (NumRuns >= 200 && NumRuns < 300)               //Same procedure, except a 100th counter, 10th counter and final
                                                        //Runsizer or 1th counter is used
                {
                        NumCounter = 0;

                        RunSizer = NumRuns - 200;

                        if (RunSizer < 10)              //For values 100-109
                        {
                                itoa(2, CharNumRuns3, 100);
                                itoa(0, CharNumRuns2, 100);
                                itoa(RunSizer, CharNumRuns, 100);

                                strcat(InputFileName, CharNumRuns3);
                                strcat(InputFileName, CharNumRuns2);
                                strcat(InputFileName, CharNumRuns);
                                strcat(InputFileName, ".txt");

                                inFile.open(InputFileName);
                        }

                        if (RunSizer >=10)                      //For values 110-199
                        {
                                while (RunSizer >= 10 && RunSizer < 100)
                                {
```

```
                    NumCounter++;
                    RunSizer = RunSizer - 10;
                }

                itoa(2, CharNumRuns3, 100);
                itoa(NumCounter, CharNumRuns2, 100);
                itoa(RunSizer, CharNumRuns, 100);

                strcat(InputFileName, CharNumRuns3);
                strcat(InputFileName, CharNumRuns2);
                strcat(InputFileName, CharNumRuns);
                strcat(InputFileName, ".txt");

                inFile.open(InputFileName);

        }
}

if (!inFile)
                {
                    cout << "CAN'T OPEN INPUT FILE " << NumRuns << endl << endl;
                }                               //End if


inFile.ignore(2000, '\n');                              // Ignore first, heading line
Rows = 0;
while (inFile)
        {
        // Loop to read in Segment_array
        for (int Columns = 0; Columns <= 3; Columns++)          // Reads in first six columns, up to first image mean value

                {
                inFile >> Segment_array[Rows][Columns];         // Input segment data in row
                inFile.ignore(10, ',');                         // Ignore comma between values in text file
                                                                // Ignore 10 in this case since 0.00 read in as 0 double and
                                                                // extra 0's cause problems
                }                                               // End For-loop
        inFile.ignore(2000, '\n');                              // Ignores rest of input line
        Rows++;                                                 // Increments Row counter for next loop run
        }                                                       //End while-loop
```

```
Rows = Rows - 1;                                        //To determine actual # of rows in array; over-counted by 1

// THIS SECTION OUTPUTS AN INFO MESSAGE
cout << "OUTPUT SECTION:" << endl << endl;
cout << "For file # " << NumRuns << " - Total number of segments read in: " << Rows << endl;
cout << endl;

// VARIANCE CALCULATION
Mi_total = 0.0;                                         //Initialize counter for total number of pixels

//STEP 1 - calculate OVERALL MEAN & WITHIN VARIANCE
for (int Rowscalc_within = 0; Rowscalc_within < Rows; Rowscalc_within++) //Runs through Segment_array
{
        if (Segment_array[Rowscalc_within][2] != 0.0)   //Only perform calculation when segment has actual values
                                                        //otherwise row with "illegal" values used in calculations
        {
                //Calculate # elements for segment, based on resolution
                Segment_elements = int((Segment_array[Rowscalc_within][3]/(Resolution * Resolution))+0.5);

                //Calculate sum of all pixels, not based on Segment_elements to avoid rounding errors
                Mi_total = Mi_total + (Segment_array[Rowscalc_within][3]/(Resolution * Resolution));

                //Overall Mean initial calculation
                Design_mean = Design_mean + (Segment_array[Rowscalc_within][1]* Segment_elements);

                //Initial calculation for WITHIN VARIANCE per row
                Design_w_var = Design_w_var + ((Segment_elements - 1) * (Segment_array[Rowscalc_within][2]*
                                        Segment_array[Rowscalc_within][2]));
        }
}

Mi_total = int(Mi_total);                               //Total pixels usmmed, revert to an integer for use in calcs

//Final OVERALL_MEAN & WITHIN VARIANCE calculation
Design_mean = Design_mean / (Mi_total);
Design_w_var = Design_w_var /(Mi_total - 1);

//STEP 2 - calculate BETWEEN VARIANCE
for (int Rowscalc_between = 0; Rowscalc_between < Rows; Rowscalc_between++)
```

```
		{
			if (Segment_array[Rowscalc_between][2] != 0.0)		//Again only use rows/segments that have values for images
			{
				Segment_elements = int((Segment_array[Rowscalc_between][3]/(Resolution * Resolution))+0.5);
				Design_b_var = Design_b_var + (Segment_elements * pow((Segment_array[Rowscalc_between][1] - Design_mean),2));
			}
		}

		//Final BETWEEN VARIANCE calculation
		Design_b_var = Design_b_var/(Mi_total - 1);

		//Calculate ratio for bewteen and within variance
		Ratio = (Design_b_var/Design_w_var);


		//OUTPUT SECTION
		outFile << InputFileName << "	" << Rows << "	" << Mi_total << "		";
		outFile		<< setprecision(12) << (floor(Design_mean*10000+0.5)/10000) << "	" <<
					(floor(Design_w_var*10000+0.5)/10000) << "		" <<
					(floor(Design_b_var*10000+0.5)/10000) << "		" <<
					(floor(Ratio*10000+0.5)/10000) << "		" << endl;

		//Reset all calculation array values to 0 before next iteration
		Design_mean = 0.0;
		Design_w_var = 0.0;
		Design_b_var = 0.0;
		Ratio = 0.0;

		inFile.close();		// Close input file
			strcpy(InputFileName, InputFileNameTemp);		//Reset InputFileName to root filename before next iteration
									//InputFileName = Root+#runs+.txt at this stage
									//InputFileNameTemp = root filename always

	}
	cout << "Thank you!  Please wait for your output file...." << endl;
	cout << endl;
	outFile.close();	// Close output file

}
```