USING MULTISPECTRAL IMAGERY AND MULTI-RETURN LIDAR TO ESTIMATE TREE AND STAND ATTRIBUTES IN A SOUTHERN BOTTOMLAND HARDWOOD FOREST[1]

Curtis A. Collins, Research Associate
Robert C. Parker, Associate Professor
David L. Evans, Associate Professor
Department of Forestry
Mississippi State University
Mississippi State, MS 39762
ccollins@sorsc.cfr.msstate.edu
rparker@cfr.msstate.edu
dle@rs.cfr.msstate.edu

ABSTRACT

This study investigated the use of high-resolution multispectral imagery and small-footprint, multi-return LiDAR (Light Detection And Ranging) for tree and stand attribute estimation in a bottomland hardwood stand in East Mississippi. The multi-spectral images were acquired in June 2001 with a spatial resolution of 0.3 m (0.98 ft) in four bands (near-infrared, red, green, and blue). The LiDAR data were acquired in September 1999 using a five return, 15 kHz scanning sensor with a 20° field-of-view, one pulse per 3.45 square meters (37.17 sq. ft), and 0.6 m (~2 ft) approximate footprint diameter configuration. Field data were collected yielding height and crown radii values for 133 dominant/codominant trees along with trees per fifth acre (0.08 ha) plot values for 45 plots. Tree heights were extracted using a unique, non-overlapping crown polygon method while object-oriented analyses were performed to isolate crown extents, tree density, and species classification. Comparisons between field and remotely sensed metrics were performed through a variety of parametric and descriptive statistics. While height and classification results performed moderately well, the complexity of horizontal hardwood canopy structure hindered the automated tree recognition approach, which in turn led to poor crown extent and tree density estimation results.

INTRODUCTION

Light Detection And Ranging (LiDAR) is a relatively new tool in the remote sensing arena which many hope, along with more tested remote sensing technologies such as Multispectral (MS) imagery, will give natural resource managers an effective way to observe and note desired forest characteristics for a given area, particularly in the vertical domain. This vertical data usually culminates into a set of vegetation height values (i.e., returns or interpolated raster surfaces), which are computed from x, y, and z return data that is differenced, with respect to z, from a corresponding ground-level, base dataset (i.e., Digital Terrain Models (DTMs) or Digital Elevation Models (DEMs)). The use of this vertical LiDAR-derived and corresponding MS data is limited, however, when they are applied to hardwood stands at the tree-level. This is due to the complexity of hardwood stands and trees with regard to their vertical and horizontal structure (Yang, 1999).

Past projects using LiDAR data in forested conditions have dealt mostly with coniferous stands where uniformity in stand characteristics is more prevalent than in most uneven-aged, hardwood stands found throughout the southeastern portion of the United States. This degree of vertical and horizontal variability found in hardwood stands makes LiDAR utilization more complex and in need of further investigation. Similarly, many of these past works focused on the stand or landscape scales, as opposed to the individual tree-level. Previous studies also used either small-footprint (usually with smaller scaled areas) or large-footprint (usually with larger scaled areas) LiDAR systems (Means, 2000). Aside from the obvious differences in their names alluding to footprint size (the horizontal diameter of the light beam as it reaches the target), these systems also vary in the manner in which they record data.

Large-footprint systems usually record reflectance by digitizing, over small intervals, the amount of energy returned to the LiDAR sensor. These data form a more continuous stream per light pulse as compared to small-footprint systems, which record reflectance in the form of discrete returns. These returns are simple locations of points where some reflectance threshold was reached and recorded by the sensor yielding one or more (multi-return) returns per pulse (Wehr and Lohr, 1999). The decision of which system to employ is usually made depending on what data are desired for the mission at hand. Generally, large-footprint systems have been used in landscape scale studies, whereas small-footprint, multi-return systems are usually used in smaller areas of focus (Means, 2000). With these standards in mind, a small-footprint, multi-return system was used in this project.

Overall, this study focused on comparing integrated small-footprint, multi-return LiDAR and high resolution MS data acquired over a hardwood bottomland in eastern Mississippi to corresponding ground data gathered in order to quantify the individual tree attribute relationships between the two datasets. The tree attributes examined were species, tree density, crown size, and total tree height. The hope was to identify meaningful quantitative relationships, with the control variable being the field data, and use them in a manner where they may be statistically implemented to predict field conditions with a desired degree of confidence.

METHODS

Study Area
The study area in this project was in an uneven-aged hardwood stand in East-central Mississippi within the bounds of the Noxubee National Wildlife Refuge in Noxubee County. The site ranged from moderate to poorly drained conditions, indicative of those typically found in southern bottomlands. This area contained a variety of species from loblolly pine (*Pinus taeda* L.) and cherrybark oak (*Quercus pagoda* Raf.) in the moderately drained areas to overcup oak (*Q. lyrata* Walt.) and baldcypress (*Taxodium distichum* (L.) Rich.) in the poorly drained areas.

LiDAR Data
LiDAR data for the study were obtained in September 1999 by EarthData Technologies. The LiDAR system used included an Aeroscan sensor with a pulse wavelength of 1064 nm collected at a laser pulse rate of 15 kHz (Baltsavias, 1999), a scan rate of 10 Hz, a 20º field of view (10º maximum scan angle), and an average swath width of 698 m (2292 ft) at the desired aircraft altitude of approximately 2000 m (6600 ft) above target. This system could record up to five returns with the first being the closest to the sensor (usually tree crowns) and the last being the farthest (in many cases, the ground). The provider reported upon data delivery that this system could yield vertical and horizontal accuracies within 15 cm. The approximate mean pulse density was determined to be one pulse per 3.45 m^2 (37.17 ft^2) on the ground with a footprint diameter of approximately 0.61 m (2 ft).

Ground returns were differentiated from vegetation returns using proprietary software provided by the LiDAR contractor resulting in approximately 11% of the total LiDAR returns being assumed ground points. Ground and first return datasets were then surfaced with ArcMap 8.1 (ESRI, 2001) utilizing the power and neighbor optimization Inverse Distance Weighted (IDW) methodologies, in ArcMap's Geostatistical Analyst Extension. The first return and ground surfaces were then differenced to yield a Canopy Height Surface (CHS).

Multispectral Data
Multispectral imagery was acquired for the study area by GeoVantage, Inc. in June 2001. This dataset was taken using SONY XCD-SX900 digital cameras. The four bands acquired were chosen using spectro-radiometer readings from another project in the Department of Foresry because they seemed sensitive to differentiating several hardwood species[2]. The portion of the electromagnetic spectrum covered by this dataset encompassed the 488-498 nm (blue), 546-556 nm (green), 668-678 nm (red), and 941-959 nm (NIR) wavelength intervals with an approximate spatial resolution of 0.3 m (0.98 ft). This configuration required 45 images to cover the study area.

A differential illumination problem was apparent in the delivered MS imagery (i.e., vignetting and bidirectional reflectance (BDR)). A method of empirical normalization was thus developed and applied [3], resulting in more homogeneous frame-to-frame spectral values among the MS images. The imagery was then orthorectified to the

---

[2] Selected by Dr. David L. Evans, Department of Forestry, Mississippi State University. Based on Knight (2003).
[3] John McCombs, Curtis Collins, and David Evans, 2003, *Normalization of differential illumination problems in high-resolution aerial imagery through response surface modeling*, paper presented at the 2003 ASPRS National Convention, Anchorage, AK. 6 May 2003.

first-return LiDAR surface and mosaicked using ERDAS Imagine. The contribution of each image to the final mosaic was defined by creating a set of cutlines which utilized Thiessen polygons centered at individual image principal points. These cutlines excluded image edges that were represented by fewer pixels in the normalization procedure. The cutlines were also modified so that misalignments and shadow discrepancies between images did not cause problems by splitting field-sampled crowns and plots into multiple image contribution areas.

Field Sampling

Field data was collected using circular one-fifth acre (0.08 ha) plots located at three-chain intervals (198 ft ~ 60.35 m) along lines spaced by ten-chains (594 ft ~ 181.05 m). The lines were oriented perpendicular to the approximate centerline of a LiDAR swath from the study area. This scheme resulted in a nine-plot width on each of ten lines. This spacing yielded 90 total plots in the study area with half being reserved as alternates, and half designated as "takes", or initial target plots.

The "take" plots were located on every other line and initially considered for sampling. The alternate lines permitted the replacement of a take plot falling in an undesirable area (e.g., a clearing, body of water, etc.) with an alternate plot located in either longitudinal direction (east or west) at the same approximate distance from the LiDAR swath center. The goal was to collect a minimum of 45 forested sample plots.

A real-time, differentially correcting Trimble AgGPS132 receiver linked to a Juniper Pro2000 field computer was used to navigate to plot center locations. Using the OmniSTAR real-time satellite service, this DGPS (Differential Global Positioning System) unit permitted the location of plot centers at a reported accuracy of "submeter" (<3.28 ft) in the field.

Traditional forest measurements were taken consisting of: Diameter at Breast Height (DBH) (on trees ≥ 4 in. or 10.16 cm) to the nearest 0.1 in. (0.254 cm), total height to the nearest 0.1 ft (3.048 cm), and merchantable height (to an estimated 10 in. (25.4 cm) top diameter and ≥ 10 ft (3.048 m) to the nearest 0.1 ft). Species and crown classification (dominant, codominant, intermediate, or suppressed) were also recorded. DBH was measured with a diameter tape and distances were measured with either a "logger's" tape or a Haglof Vertex III hypsometer. The Haglof Vertex III (in its hypsometer capacity) was also used to measure tree heights.

Each plot had to contain at least two crown data trees, which were trees that were classed as dominant or codominant and were the farthest from plot center in their respective Cartesian quadrant. If this situation did not exist for a particular plot an alternate plot was taken longitudinally, ten chains away. An Atlanta Optics laser rangefinder (mounted on a monopod with a built-in level) was used to obtain the distance and azimuth of each tree's location. Crown radii for these sample trees were measured in eight directions (north, northeast, east, southeast, south, southwest, west, and northwest). Radii directions were obtained with a hand compass and radii lengths were found by using the DME mode on the Vertex III while visual alignment between the DME and crown edges was made using a seven-foot range pole. This procedure yielded an overall measurement of 133 crowns.

Coordinate data such as plot, tree, and crown extent locations were computed in a spreadsheet as Universal Transverse Mercator (UTM) values and converted into vector datasets using ARC/INFO 8.0.1. The resulting Arc coverages were then built as either point features (e.g. tree and plot centers); polygon features, where no overlap was allowed (e.g. plot extents); or polygon region features, where feature overlap was allowed (e.g. crown extents). Other sample measurements (e.g. merchantable height, crown heights, DBH, species, plot number, and tree number) were saved to the individual spatial features as attributes.

Remotely Sensed Metrics

LiDAR Height Extraction. The 133 field-measured tree crowns and CHS datasets were used in conjunction with a zonal maximum operation to extract LiDAR-determined tree heights. The vector layer used in this process was constructed by using either the shortest distance from one of the eight vertices of an individual polygon or the closest vertex from a neighboring polygon, if an overlap existed, to the crown feature's centroid. This procedure created a set of Minimum Crown Extent Circles (MCEC), which intended to avoid height extraction from overlapping crown regions (Figure 1). The zonal attributes function in Imagine (ERDAS, 2001) was next used with the MCECs as the zonal dataset on the CHS to attribute LiDAR heights to sampled tree crown features.

Tree Crown Identification. Several methodologies were attempted, with little success, to isolate individual tree crowns in the remotely sensed datasets. These methods included: valley following approaches, similar to Gougeon (1995), and watershed analyses, similar to Hyypa and Inkenen (1999); edge detection or standard deviation models; crown template matching, similar to Quakenbush *et al*. (2001); and simple clustering techniques performed on the CHS and MS datasets. Finally, the object-oriented approach in eCognition 2.1 was chosen to perform the individual crown analyses for this project.
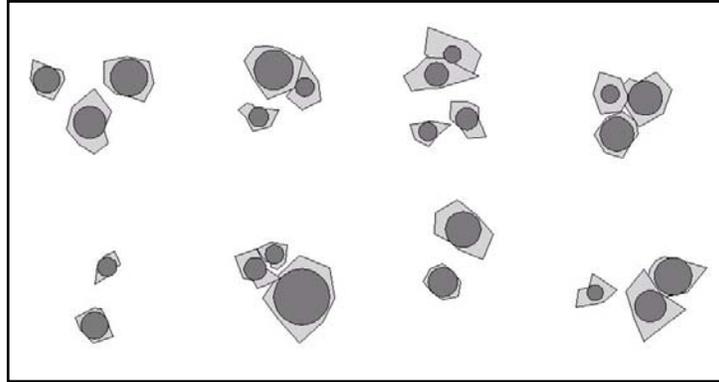
Figure 1. A view of a set of Minimum Crown Extent Circles (dark gray) placed inside their corresponding field-measured crown extents' boundaries (light gray).

In order to reduce processing time, object-oriented procedures were performed on a subset of the study area. Thus, a "partitioned-and-joined" subset was constructed by extracting data around each plot and manipulating the geographic coordinates of the subset so that they could be linked into a checkerboard-like pattern. A block design was then utilized with the block size being determined by the between-plot distance on a given column. This process yielded 45, three-chain (60.35 m) by three-chain blocks placed adjacent to each other.

The first step in the eCognition procedure involved the definition and construction of image segments or objects. These segmentation results were processed using a "bottoms up region-merging technique starting with one-pixel objects (Baatz *et al*., 2001)." In this project, however, the multi-resolutional segmentation capability of eCognition was not utilized in the initial phases. It was decided instead to find optimal segmentations at a single-level before using multiple levels of segmentations due to the increased complexity caused by these procedures. To initiate the single-level segmentation, the input raster layers were loaded into the system with corresponding image weights, homogeneity criteria, a scale parameter, and a segmentation mode.

The MS and CHS datasets were used independently in the initial segmentation stage with combined trials being performed in later stages. The independent MS step dictated that weights be set equal for each of the four bands from the two MS datasets (NIR = 1.0, red = 1.0, green = 1.0, and blue = 1.0), respectively, with the CHS dataset being ignored (weight = 0.0). Since the CHS data was a single layer it was weighted individually with the MS weights being set to zero in its independent testing. The combined weighting process for MS and the CHS datasets used five weighting combinations divided into odd-numbered proportions, each of which summed to one. The MS weights for all four layers were the same and based on the MS data's allotted proportion for the trial in question.

The next step was to incorporate the homogeneity criteria into the segmentation process. First, color versus shape criteria were assigned to form the spectral (which, in the case of the CHS dataset, was vertically spatial) differentiations essential to this type of object-oriented analysis. This protocol involved assigning proportional weights (summing to one) to spectral values as opposed to object shape. Next, the shape weight was partitioned between smoothness and compactness. Large smoothness weights tended to yield more circular segments while large compactness weights tended to yield rectangular or square segments. Color versus shape and smoothness versus compactness weights were examined by changing each pair of criteria at one-tenth (10%) intervals. This process yielded 100 different segmentations with any protocol that held scale and layer weights constant.

The segmentation mode and scale parameter were examined last. The segmentation mode allowed for a normal (eight-neighbor kernel) or diagonal (four-neighbor kernel) option. The normal mode was selected over the diagonal mode because the diagonal option was designed for extracting linear objects (Baatz *et al*., 2001), which was not the case among the sampled tree crowns. The scale parameter was optimized through a heuristic approach. This method's results varied with different layer weights and homogeneity criteria, showing little consistency throughout the attempted trials. When combined with the other criteria and weights, a large number of segmentation outputs were possible, leading to the importance of a visual decision–making system.

Visually Examining Identified Tree Crowns. A large number of segmentation routines were generated by the previous section's procedures. For this reason, the image segmentation process had to be narrowed by a protocol that was relatively fast and more substantial than purely aesthetic comparisons before any quantitative tests could be made. The result was the development of a visual interpretation process conducted on all candidate segmentations before they were allowed to be quantitatively compared.

The interpretation process followed a stepwise manner, analyzing one segmentation with one color-shape weight and one smoothness-compactness weight combination in each step. This segmentation was found by running several routines with the scale parameter being adjusted after each routine by comparing the segmentation output to 50 randomly selected field-obtained crown polygons (from the original 133). When the segmentation with the best match to the test crowns was displayed, the optimal scale was determined and a visual count (VC), enumerating the number of segments in the output that appeared to fit the 50 field measured test crowns, was recorded. The VC values were recorded for each segmentation that was within the test interval with the exception of some that produced grossly poor object-to-tree polygon fits, visually.

Following this testing, a minimum of five segmentation sets from each layer/weight combination were designated for quantitative comparison. If a tie existed when choosing the minimum five, all segmentations with the same rank of the fifth best segmentation were allowed to be quantitatively compared. These segmentations were then compared with another five chosen across all tested layer/weight combinations for ancillary testing. With all segmentation parameters being held constant, except scale, these five segmentations' parameters were processed over the subset dataset using two ancillary procedures. These processes applied both a Binary Forest/Non-forest Raster (BFNR), which was created by thresholding the CHS dataset by 15 meters, mask and two-resolution (mulit-resolution segmentation) methods. The masking protocol segmented areas that were identified as forest canopy (using a 15 m canopy definition and the CHS), focusing segmentation operations on these specific areas of interest.

The two-resolution method consisted of segmenting small crowns well at the first-level in the hopes that the large crowns would be correctly merged from the smaller segments comprising them, with the small segments that represented single crowns being maintained, in the second-level. After quantitatively comparing all of these subset segmentations, including the two-resolution and BFNR masked methods, a final segmentation criterion set was isolated and applied to the entire (non-subset) study area's datasets for final quantitative analysis.

Tree Density Determination. The BFNR was also used to mask the final and subset quantitatively tested segmentation layers to remove non-canopy segments and attribute them to corresponding plot vector layers. Comparisons of these tree-object counts to field-measured, plot-level dominant/codominant tree counts were made to test whether or not the object-oriented method would yield viable tree density measurements. This testing was performed on all quantitatively examined segmentations mentioned above. Since this process hinged on the tree crown identification phase, tree density was not considered a goal for guiding segmentation. The assumption was that a procedure that accurately delineated individual tree crowns would also yield the best estimates of tree density.

Tree Species Classification. Species classification was performed on the non-subset, final segmentation using the CHS and MS datasets in the suite of classification functions in eCognition. These functions were either computed in eCognition using object samples (training objects) through nearest neighbor calculations or manual definition by the user, via membership functions, categorizing these methods as supervised classifiers. Seven target species were identified (baldcypress, cherrybark oak, loblolly pine, mockernut hickory (*Carya tomentosa* Poir. (Nutt.)), overcup oak, water oak (*Q. nigra* L.), and willow oak (*Q. phellos* L.)) for classification because they constituted 93.94% of the total merchantable volume sampled in the 45 field plots. Subsequent testing samples, for use in an accuracy assessment, were also required to analyze the classification's performance among these seven species.

The number of samples needed in the two sample types (training and testing) was difficult to determine because of the newness of object-oriented classification methods. Lillesand and Kiefer (2000) recommend a minimum training size of $n + 1$, where n is the number of bands, with a general accuracy testing suggestion of 50 samples per classification category in pixel-based supervised classification. Since individual objects in eCognition's classification scheme were treated similarly to pixels in pixel-based classification, these guidelines were accepted in the initial sample size determination. These standards required supplemental crown sampling, which was performed in late 2002 and involved recording dominant or codominant tree crown centers belonging to one of the target species with the DGPS. These locations were converted to polygon feature format using both the MS and CHS datasets to heads-up digitize the crowns where the individual DGPS points fell. For this reason, these new features were designated for testing because the previously-measured field crown extents were deemed more spatially certain, a characteristic of greater importance for training samples. These added crowns still did not meet the initial guidelines, but they were accepted for this project due to time constraints.

RESULTS

Field and LiDAR Tree Heights

A paired t-test was initially performed on the 133 corresponding field and LiDAR measured tree heights to

determine if these two datasets were significantly different. With a mean LiDAR height of 34.81 m (variance = 10.25) and a mean field-derived height of 33.95 m (variance = 9.87), the heights were different ($\alpha = 0.05$) with a t-statistic equal to 4.21 and a critical t-value of 1.96.

SLR (Simple Linear Regression) was then applied to the paired heights to determine if a linear relationship existed. This analysis used least squares fitting and was performed to detect the presence and significance of bias (β_0) and slope (β_1) as well as model fit and correlation (R^2) between the two height measures, all of which related to the predictability of field height from LiDAR height. The resulting model had a $R^2 = 0.5296$ and significant coefficients ($\alpha = 0.05$) (both β_0 and β_1), demonstrating the validity of the model as a predictive tool (Figure 2).
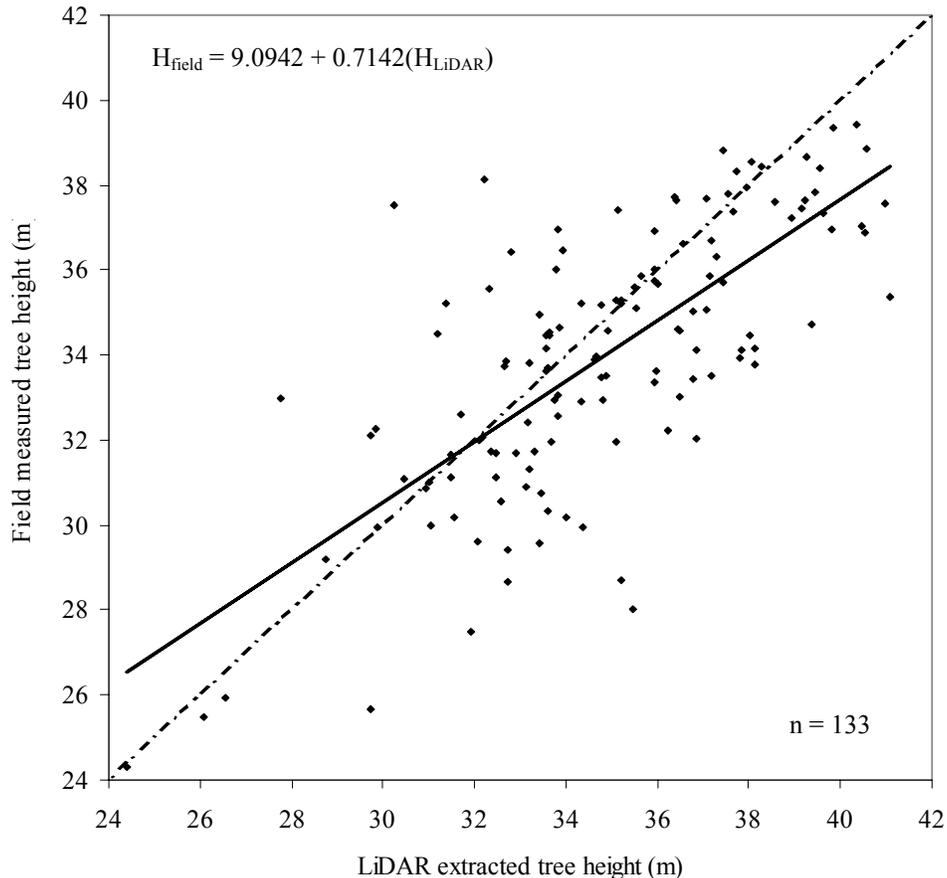

Figure 2. Simple Linear Regression (SLR) model (upper left), trend (solid line), and height data scatterplot between corresponding the 133 LiDAR and field-derived tree heights.

Field and Remotely Sensed Tree Crowns

The quantitative analysis of 83 (133 minus 50 visual guidance crowns) field-measured and segmented tree crown objects involved 55 quantitatively tested, 278 visually judged, and many heuristically tested segmentation routines. Overall this study examined 44 single-level subset segmentation routines to identify the best five routines for two-level and BFNR-modified trials to produce a final (non-subset) segmentation. This process yielded 16 segmentation routines for final quantitative analysis. Analysis results were examined using parametric and descriptive, non-parametric statistical values that were applied in the quantitative discrimination step.

Non-parametric, descriptive statistics, relating to the proportional matching area between corresponding segments and field-measured tree crowns, were first created. The three methods for determining these proportions were matched segment and crown overlap areas as a proportion of total segment area (commission errors), total crown area (omission errors), and summed mutually exclusive matched segment and crown areas in addition to matched overlap area (overall errors). These values were also calculated using two different methodologies: (1) where each tree was weighted the same, yielding weighted error indices; and (2) where total crown areas were used, across all trees, yielding total error indices.

The parametric methods employed to derive quantitative results used paired t-tests and SLR. Differences between matched field and segment radii were tested using these methods with t-test derived observed signficance and fit statistics from the SLR coefficients. Where multiple crowns were matched to the same segment, the largest crown was given the match and the other was not. Because only matched segments were tested using these methods, the number of unmatched crown features were also included in the resulting tables.

Non-parametric decision values for the 44 initial segmentations, displayed little support that the initially examined matched segmentation routines' objects and crowns resemble each other in size. Minimum and maximum (range), arithmetic mean, and median values among these routines indicated no trends of improvement among the various scale, layer, and criteria weight values (Table 1). These values also did not display proportional matching in any routine greater than 0.60 in the overall, weighted and 0.31 in the overall, total categories. The close proximity of the two measures of central tendency, mean and median, displayed changes of approximately five percent from the maximum overall assessment values in both weighted and total categories, indicating that there were no clear choices that could be made from these indices because the distribution of proportions was skewed toward the maximum end of the data range.

Table 1. Summary of the non-parametric crown area statistics (proportions) for the 44 initial quantitatively tested single-level subset segmentation routines.

Value	COM_{wt}[1]	OM_{wt}[1]	ALL_{wt}[1]	COM_{tot}[1]	OM_{tot}[1]	ALL_{tot}[1]
Minimum	0.2939	0.5927	0.4479	0.2292	0.5237	0.1956
Maximum	0.4356	0.8082	0.5946	0.3562	0.7404	0.3069
Mean	0.3750	0.7239	0.5494	0.3002	0.6397	0.2566
Median	0.3838	0.7252	0.5546	0.3099	0.6444	0.2603

[1] COM, OM, and ALL proportions represent commission, omission, and overall errors using both weighted ($_{wt}$) and total ($_{tot}$) methods.

Parametric decision values also displayed little support that the initially examined 44 matched segments and crowns resembled each other in size. Similar to the non-parametric tests, minimum and maximum (range), arithmetic mean, and median values among these routines indicated no trends of improvement among various scale, layer, and criteria weight values (Table 2). All paired t-test derived significance levels indicated that the crown and segment radii measures differed ($\alpha = 0.05$). The sole optimistic value was the maximum observed model significance (0.9930); however, this optimism was soon nullified with the R^2 indices, which did not exceed 0.1997.

Table 2. Summary of the parametric crown radii statistics for the 44 initial quantitatively tested single-level subset segmentation routines.

Value	T-test Observed Significance	SLR Analysis[1]			
		β_0	β_1	Obs. Significance	R^2
Minimum	< 0.0001	3.4804	0.0009	< 0.0001	0.0000
Maximum	0.0054	5.8157	0.2850	0.9930	0.1997
Mean	0.0002	4.6511	0.1430	0.2099	0.0489
Median	< 0.0001	4.6007	0.1498	0.0796	0.0419

[1] Using the model $R_{field} = \beta_0 + \beta_1(R_{object})$

The decision as to which of the five routines to choose for single-level BFNR-constrained and two-level subset trials was made using the overall, total values. The two-resolution and BFNR-modified routines were created and analyzed in a similar fashion as, and ranked with, the initial 44 routines. This process derived the same non-parametric and parametric indices to determine which routine would be applied to the non-subset study area. Again the overall, total values were used as the deciding factor and no trend of improvement along any criteria, weight, or scale lines was detected. This process led to the use of a two-level routine (Table 3) with its layer, color, and smoothness weights being applied to the entire study area.

Table 3. Descriptive, non-parametric crown area comparisons among the best single layer subset segmentation routines (first five layers), the two-level subset routines (second five layers), the single-level BFNR constrained subset routines (third five layers), and the final two-level non-subset segmentation (the last layer).

Layer Name	Weighted Comparisons						Totaled Comparisons					
Layer	COM_{wt}[1]	Rank[2]	OM_{wt}[1]	Rank[2]	ALL_{wt}[1]	Rank[2]	COM_{tot}[1]	Rank[2]	OM_{tot}[1]	Rank[2]	ALL_{tot}[1]	Rank[2]
L1_5971	0.3959	21	0.7884	2	0.5922	4	0.3439	8	0.7404	1	0.3069	6
L5_5756	0.3908	26	0.7687	5	0.5797	10	0.3367	11	0.7005	5	0.2943	10
L7_5772	0.4270	7	0.7622	8	0.5946	2	0.3562	6	0.6617	20	0.3013	8
L7_5870	0.4356	3	0.7247	28	0.5802	8	0.3425	9	0.6574	23	0.2906	11
L9_5545	0.3824	32	0.7209	31	0.5516	34	0.3420	10	0.6354	38	0.2859	12
L1_26_3871	0.3996	15	0.7556	10	0.5776	12	0.3542	7	0.7072	4	0.3089	4
L5_24_3556	0.4330	5	0.7664	6	0.5997	1	0.3343	13	0.7178	2	0.2955	9
L7_26_3770	0.3454	45	0.7547	13	0.5500	36	0.3087	30	0.7080	3	0.2738	22
L7_26_3872	0.4249	8	0.7203	32	0.5726	15	0.3599	5	0.6717	13	0.3061	7
L9_26_3445	0.4446	2	0.7328	23	0.5887	6	0.3668	3	0.6859	8	0.3141	2
L1_4071	0.4320	6	0.7470	17	0.5895	5	0.3612	4	0.6932	7	0.3114	3
L5_4156	0.3482	43	0.6966	42	0.5224	47	0.2599	50	0.6650	17	0.2298	47
L7_3672	0.4356	4	0.6830	47	0.5593	25	0.3669	2	0.6581	21	0.3081	5
L7_3970	0.4164	10	0.6986	40	0.5575	28	0.3086	31	0.6640	18	0.2669	27
L9_3545	0.3566	41	0.6365	53	0.4965	51	0.2775	42	0.6274	39	0.2382	43
L9_26_4445	0.5000	1	0.6312	54	0.5656	19	0.4258	1	0.5663	54	0.3211	1

[1] COM, OM, and ALL proportions represent commission, omission, and overall errors using both weighted ($_{wt}$) and total ($_{tot}$) methods.
[2] Rankings compared to all 55 quantitatively tested segmentation routines
*Layer names indicate the layer and its weight used, scale parameter(s), as well as color and smoothness weights, respectively (i.e., L9_26_4445 used the CHS (or "L" for LiDAR) weighted nine times greater than the MS data with a first-level scale of 26, a second-level scale of 44, a color weight of 0.4 and a smoothness weight of 0.5).

The final segmentation layer ranked best with respect to the overall, total statistics, while ranking 19th with a difference of 3.4 % from the best ranked routine in the overall, weighted column among all 55 quantitatively tested routines. The parametric statistics from all quantitatively assessed routines were also calculated and ranked with the best five from the single-level subset, using the overall, total statistic, as well as the comparisons made from the two-level, BFNR-constrained, and final segmentations observed in Table 4. The number of non-matched crown features from the test set of 83 in the final routine ranked second among all 55 test segmentations with three. T-test and SLR probabilities both displayed significance ($\alpha = 0.05$) with a poor R^2 value (0.0043).

Field Remotely Sensed Tree Densities

Tree density comparisons were performed at the plot-level using the number of dominant/codominant trees sampled in the field compared to the number of canopy objects from the various quantitatively compared segmentations using paired differences' Root Mean Squared (RMS) and arithmetic mean averages. Density comparisons were not used as decision criteria. Assuming the segmentation level that matched tree crowns best also yielded the best density measures, tree densities served as a performance meter for how well eCognition identified individual tree crown extents and numbers.

In calculating and ranking, among all 55 routines, the mean and RMS values, no agreement was found between the crown identification stage and this step because none of the 16 trials examined in the final portion of that stage ranked better than 20 in mean and RMS comparisons. Overall the RMS values ranged from 1.56 to 4.06 with a

Table 4. Parametric and descriptive crown radii comparisons among the best single layer subset segmentation routines (first five layers), the two-level subset routines (second five layers), the single-level BFNR constrained subset routines (third five layers), and the final two-level non-subset segmentation (the last layer).

Layer	Mis-matches	Rank[1]	t-test Analyses		SLR Analyses[2]				
			Obs. Sig.	Rank[1]	β_0	β_1	Obs. Sig.	R^2	Rank[1]
L1_5971	5	9	< 0.0001	52	3.5722	0.2725	0.0005	0.1488	2
L5_5756	8	33	< 0.0001	30	3.4804	0.2850	0.0031	0.1135	4
L7_5772	9	37	0.0006	4	4.4005	0.1863	0.0821	0.0414	28
L7_5870	4	3	< 0.0001	27	5.5893	0.0276	0.7343	0.0015	51
L9_5545	6	16	< 0.0001	28	3.8703	0.2514	0.0085	0.0888	9
L1_26_3871	7	24	< 0.0001	17	3.8857	0.2382	0.0120	0.0823	11
L5_24_3556	4	3	< 0.0001	38	4.2461	0.1907	0.0092	0.0848	10
L7_26_3770	12	53	0.0004	5	4.5459	0.1466	0.1130	0.0360	29
L7_26_3872	4	3	< 0.0001	37	5.2373	0.0720	0.4376	0.0078	44
L9_26_3445	2	1	< 0.0001	44	5.2058	0.0710	0.4323	0.0078	45
L1_4071	4	3	< 0.0001	55	3.6651	0.2654	0.0028	0.1106	5
L5_4156	11	50	< 0.0001	54	5.0092	0.0920	0.2068	0.0227	40
L7_3672	7	24	< 0.0001	50	4.0069	0.2319	0.0183	0.0729	14
L7_3970	7	24	< 0.0001	48	4.9746	0.1027	0.1513	0.0276	35
L9_3545	12	53	< 0.0001	53	4.3556	0.1734	0.0425	0.0583	17
L9_26_4445	3	2	0.0014	2	5.3747	0.0682	0.5586	0.0043	48

[1] Rankings compared to all 55 quantitatively tested segmentation routines
[2] Using the model $R_{field} = \beta_0 + \beta_1(R_{object})$
*Layer names indicate the layer and its weight used, scale parameter(s), as well as color and smoothness weights, respectively (i.e., L9_26_4445 used the CHS (or "L" for LiDAR) weighted nine times greater than the MS data with a first-level scale of 26, a second-level scale of 44, a color weight of 0.4 and a smoothness weight of 0.5).

mean of 2.41 and a median of 2.31 while the arithmetic mean values ranged from 0.33 to 3.11 with a mean of 1.29 and a median of 1.16. This analysis also indicated that the final segmentation did not yield the best RMS value, which was preferred because it does not having a canceling effect like an arithmetic average. In fact the final, non-subset segmentation yielded a ranking of 51 out of 55.

Tree Species Classification

Image classification in this study intended to classify individual objects derived from the three segmentation routines of interest into one of seven species classes. These species were baldcypress (BAC), loblolly pine (LOP), cherrybark oak (CBO), overcup oak (OVO), mockernut hickory (MOH), water oak (WAO), and willow oak (WIO) with the "Non-forest" class being extracted on the basis of object majority to the non-forest portion of the BFNR thematic layer. After classifying the final segmentation routine, 131 test samples were selected for accuracy assessment and yielding an overall accuracy of 40.46% and a KHAT statistic of 29.19%, with an individual species breakdown given in Table 5.

DISCUSSION AND CONCLUSIONS

The results from the LiDAR and field height comparisons demonstrated paired data sources whose differences were significant ($\alpha = 0.05$), but were modeled (SLR) with good fit and coefficients that were significant. The plotted data, however, showed the presence of four possible outliers (the two lowest and highest points in Figure 2). After refitting the SLR model the coefficients from the "with" and "without outliers" datasets changed as the

intercept decreased from 9.0942 to 7.6898, the slope increased from 0.7142 to 0.7543, and the R^2 value increased from 0.5296 to 0.6231.

Table 5. Final multi-level classified segmentation routine error matrix and accuracy assessment results at the species-level for the entire Noxubee NWR study area.

		Reference data[1]							
		BAC	LOP	CBO	OVO	MOH	WAO	WIO	Sum
Classified data[1]	BAC	4	2	1	2	0	0	0	9
	LOP	1	8	0	0	0	0	0	9
	CBO	0	1	8	1	3	7	1	21
	OVO	7	0	7	19	0	2	6	41
	MOH	1	6	0	0	5	3	1	16
	WAO	0	2	1	0	1	1	0	5
	WIO	0	1	9	6	3	2	8	29
	Non-forest	1	0	0	0	0	0	0	1
	Sum	14	20	26	28	12	15	16	131

Accuracies							
Producer	0.4444	0.8889	0.3810	0.4634	0.3125	0.2000	0.2759
User	0.2857	0.4000	0.3077	0.6786	0.4167	0.0667	0.5000
Overall	0.4046						
KHAT	0.2919						

[1] BAC = Baldcypress, LOP = Loblolly pine, CBO = Cherrybark oak, OVO = Overcup oak, MOH = Mockernut hickory, WAO = Water oak, and WIO = Willow oak.
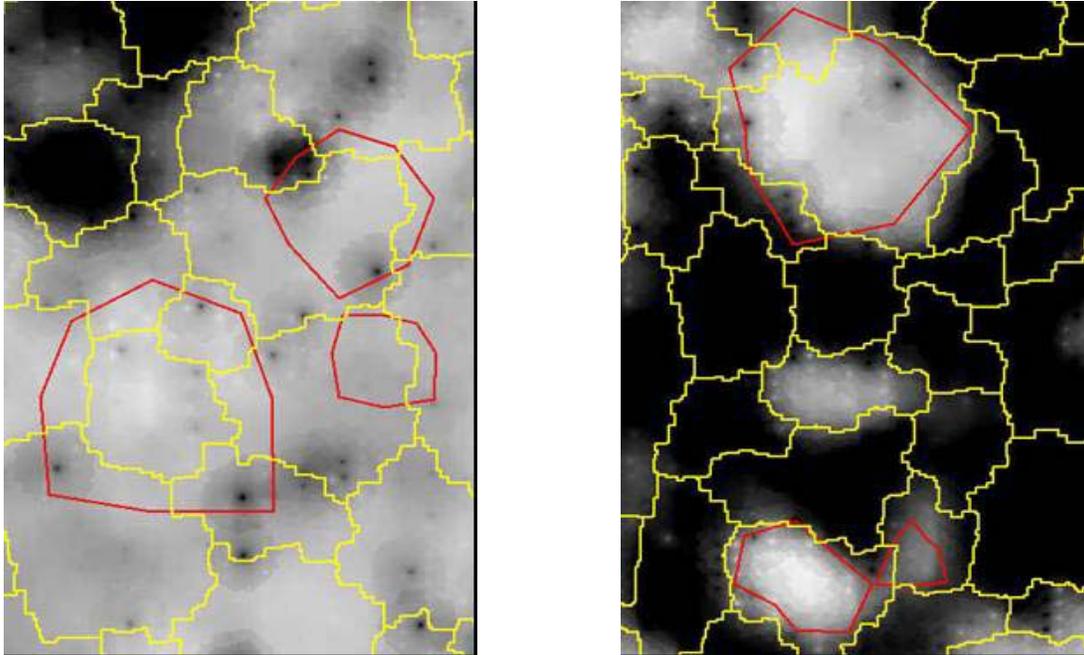
In similar studies, both Brandtberg *et al*. (2003) and Hyyppa and Inkinen (1999) noted uncertainties with regard to field height collection. Brandtberg and others noted, as in this project, an overestimation of shorter trees and an underestimation of taller trees using LiDAR height (compared to field height). This similarity was such that the SLR intercept (10.623), slope (0.6115), and R^2 (0.69) values were close to the ones derived in this study. This problem was probably the result of difficulties in identifying crown peaks, which is presumably due to the nearly flat surfaces of hardwood crowns as well as the uncertainties in viewing these crowns through multiple lower story canopy layers, which are present under most dominant hardwood canopies in the South. Even with these acknowledged limitations and discrepancies, the results of this testing were encouraging.

Although a multitude of crown-to-object size testing values were possible, six non-parametric (descriptive) and two parametric (inferential) tests were used to examine area and radii of paired crown and segments. These eight procedures were chosen to ensure confidence that an accurately predictive segmentation, if one existed, would be identified for generation of final results. The ideal segmentation, however, was never determined because none examined appeared to accurately identify field-measured tree crowns. A final segmentation was selected from the subset trials in order to perform species classification and to determine if overall crown size and tree density determination would improve when segmentation occurred over the entire study area.

The value selected for determining the optimal segmentation routine was the overall, total, non-parametric assessment value. This value was selected because it incorporated all field-measured tree crowns designated for testing, whether they matched a segmentation object or not. These values also allowed for direct influence of individual tree crowns based on matched field-measured and object feature sizes, giving more power to the larger tree crowns, which would be appropriate in most timber volume inventory analyses as larger crowned trees typically contribute more volume than smaller trees.

The final segmentation routine used a layer weight of 0.9 in favor of the CHS (0.1 to the GeoVantage or MS dataset), a scale of 26 in the first-level and 44 in the second-level, a color weight of 0.4 (shape weight = 0.6), and a smoothness weight of 0.5 (compactness weight = 0.5). The resulting overall, total test value of 0.3211 was deemed the best in this study (Table 3). In fact, parametric and non-parametric statistics for all 55 quantitatively compared segmentations displayed low non-parametric proportions, significant differences, and poor R^2 values. For this reason, the eCognition procedure for extracting image objects that optimally represented individual tree crowns was found to be inadequate.

The reason for this inability to delineate individual tree crowns is related to the "within crown" as opposed to the "between crown" vertical variance structure. This situation was attributed to the intermingled nature of hardwood tree crowns in high-density areas. When visually compared, this trend was observed in segmented areas where lower and higher crown density areas yielded different results (Figure 3). In these areas, trees located in low density areas, allowing lower height values in regions between individual tree crowns, were at times differentiated

Figure 3. Segmentation results (yellow outline) using the CHS layer (grayscale) weighted completely with field-measured crown extents overlayed (red) in high (left) and low (right) density crown areas.

well. In contrast, trees located in high density areas, where multiple tree crowns appeared continuous in the CHS, were rarely distinguishable in the various segmentation results. With this limitation noted, improper crown delineation using eCognition should decrease in stands where density is lower and individual crown forms become better defined.

Since crown size fitting served as the training phase of the segmentation process with tree counts serving entirely as a secondary test phase, tree count accuracy was judged, in addition to its independent accuracy, on the basis of whether or not optimal counts were achieved at the optimal crown size comparison routines. For this reason, plot-level tree counts were determined infeasible because of the misalignment between optimizing tree count and crown size routines.

The inadequacies of this object-oriented approach in deriving tree counts, or density measures, were consistent with the crown size limitations mentioned previously. With the entire optimization process for tree density extraction dependent on optimal crown size analyses, it stands to reason that the problems encountered in crown size trials were the major inhibitions in creating representative plot-level density values.

While the two softwoods, BAC and LOP, demonstrated good separability across the three classifications, the hardwoods, particularly WAO, appeared less separable. For this reason, species were joined to represent the baldcypress (BAC), southern yellow pine (LOP), red oak (CBO, WAO, and WIO), white oak (OVO), and hickory (MOH) merchantable classes. This combining of classes yielded improved overall accuracy (55.73%) and KHAT values (38.63%).

In order to determine the performance of this classification, a comparison of these results to two pixel-based classification studies performed in close proximity to this project's study area was made. These studies yielded overall accuracies of 65% (Casey, 1999) and 76% (Batten and Evans, 1998) across four species groups (hickory, oak, pine, and sweetgum). In reassessing the results from this project's classification with four classes so that they may be compared to these studies, hickory (MOH), oak (Oak), Southern yellow pine (LOP), and baldcypress (BAC) groups were identified. The overall accuracy from this analysis was 72.52%, surpassing the results achieved in one of these previous studies, and approaching the value achieved in the other.

As a whole, this study performed as expected with regard to the height and species classification portions. While the crown identification stage was preliminarily deemed problematic, it was surprising as to its actual level of difficulty. In any event, some obstacles that were encountered in this project were successfully mitigated, lending merit to this work as a benefit to the user and academic community. Continued work is stressed. If a piecewise research effort can be contributed from multiple sources, perhaps the problems addressed in this work may be overcome resulting in the practical application of these technologies in the natural resource management community.

ACKNOWLEDGEMENTS

REFERENCES

Baatz, M., U. Benz, S. Dehghani, M. Heymen, A. Holtje, P. Hofmann, I. Lingenfelder, M. Mimler, M. Sohlbach, M. Weber, and G. Willhauck (2001). *User Guide: eCognition 2.1*, Definiens Imaging, Munich, Germany, (pages numbered by section).

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

Batten, S., and D.L. Evans (1998). Forest stand assessments with high-resolution multispectral frame camera data. In Proceedings: *First International Conference on Geospatial Information in Agriculture and Forestry,* Lake Buena Vista, FL, 1-3 June 1998, ERIM International, Inc., Ann Arbor, MI: I-95–I-101.

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 Environmen,.* 85(2003): 290-303.

Casey, W. A. (1999). Forest species group classification using digital frame camera imagery, MS Thesis, Mississippi State University, Starkville, MS, 55 p.

Clark, A., D.R. Phillips, and D.J. Frederick (1985). Weight, volume, and physical properties of major hardwood species in the Gulf and Atlantic Coastal Plains, Research Paper SE-250, Southern Forest Experiment Station, United States Forest Service, Asheville, NC, 66 p.

ESRI (Environmental Systems Research Institute) (2001). *ArcMap 8.1 (PC), Online HelpFiles*, Environmental Systems Research Institute, Redland, CA.

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

Knight, T.C. (2003). Analysis of competing hardwoods in mid-rotation loblolly pine plantations using remote sensing technology, MS Thesis, Mississippi State University, Starkville, MS, 88 p.

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

Lillesand, T. M., and R. W. Kiefer (2000). *Remote Sensing and Image Interpretation.* John Wiley and Sons, New York, NY, 724 p.

Means, J. E. (2000). Comparison of large-footprint and small-footprint LiDAR systems: design, capabilities and uses. In Proceedings: *Second International Conference on Geospatial Information in Agriculture and Forestry,* Lake Buena Vista, FL, 10-12 January 2000, ERIM International, Inc., Ann Arbor, MI: I-185–192.

Quakenbush, L.J., P.F. Hopkins, and G.J. Kinn (2001). Using template correlation to identify individual trees in high resolution imagery. In Proceedings: *67th Annual Meeting of the American Society for Photogrammetry and Remote Sensing,* Washington, DC, 19-26 April 2002: CD-ROM.

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

Yang, X., J. J. Wichosky, and D. R. Miller (1999). Vertical overstory canopy architecture of temperate deciduous hardwood forests in the eastern United States. *Forest Science*, 45(3): 349-358.