AN EVALUATION OF A 3D SAMPLING TECHNIQUE
AND LIDAR FOR THE DETERMINATION OF
UNDERSTORY VEGETATION DENSITY
LEVELS IN PINE PLANTATIONS

By
Matthew Thomas Clarkson

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Forestry
in the Department of Forestry

Mississippi State, Mississippi
May 2007
AN EVALUATION OF A 3D SAMPLING TECHNIQUE
AND LIDAR FOR THE DETERMINATION OF
UNDERSTORY VEGETATION DENSITY
LEVELS IN PINE PLANTATIONS

By
Matthew Thomas Clarkson

Approved:

____________________________  ______________________________
Robert C. Parker     David L. Evans
Associate Professor of Forestry     Professor of Forestry
(Major Professor)     (Committee Member)

____________________________  ______________________________
Jeanne C. Jones     Andrew J. Londo
Associate Professor of Wildlife and Fisheries     Associate Professor of Forestry
(Committee Member)     (Committee Member)

____________________________  ______________________________
James P. Shepard     George M. Hopper
Department Head and Graduate Coordinator     Dean of the College of Forest
Department of Forestry     Resources
A three dimensional sampling technique was used to compare field understory conditions in Southeastern Louisiana using a laser range finder at three height levels (0.5m, 1.0m, and 1.5m) to LiDAR generated understory conditions to determine if a relationship existed. A similar comparison was made between densitometer crown closure measurements and understory LiDAR vegetation counts. A comparison between overstory LiDAR counts and understory LiDAR counts was also performed.

LiDAR and understory counts exhibited a significant linear relationship but were poorly correlated at each sample level (Level-1 $R^2 = 0.34 – 0.38$, Level-2 $R^2 = 0.36 – 0.43$). The Level-3 LiDAR slope coefficient was non-significant. The crown closure versus understory linear model did not produce any significant results. The overstory LiDAR versus understory LiDAR model produced a moderate correlation ($R^2 = 0.5226$)
and was significant. The process of relating LiDAR points to understory conditions was not repeatable, even in the same geographic region.

Key words: LiDAR, understory, density, GIS, sampling techniques
ACKNOWLEDGEMENTS

I have had the pleasure of spending my last seven years in the company of some of the finest minds I have ever encountered. I would like to take this opportunity to thank all my professors that I have received tutelage from in my tenure here at Mississippi State University. I would especially like to thank Dr. Robert Parker and Dr. David Evans for taking a chance on a sub-par undergraduate student and giving me an opportunity to prove myself at a higher level of academia. I would also like to thank Dr. Andrew Londo and Dr. Jeanne Jones for serving on my committee and being there for guidance during this study.

I would also like to recognize Dr. Keith Belli for sitting through my often times convoluted questions and helping me to figure out what questions I really needed to be asking. Another group that deserves recognition is my second family, my friends and co-workers in the Spatial Information Technologies Laboratory, as well as my former roommate and friend Mr. Cliff Anderson. Finally, I would like to thank my family for their support (both emotional and financial) throughout my college career, and all the trials and tribulations associated with it.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... ii

TABLE OF CONTENTS .......................................................................................................... iii

LIST OF TABLES ..................................................................................................................... v

LIST OF FIGURES ................................................................................................................... vi

CHAPTER

I. INTRODUCTION ................................................................................................................... 1

  Overview ................................................................................................................................. 1
  Remote Sensing ...................................................................................................................... 2
  LiDAR ..................................................................................................................................... 2
  Forest Structure .................................................................................................................... 4
  Understory Sampling Techniques ...................................................................................... 5
  Study Objectives ................................................................................................................... 6

II. METHODS ......................................................................................................................... 8

  Data Collection .................................................................................................................... 8
  LiDAR Data Analysis ........................................................................................................... 10
  Data Issues ........................................................................................................................... 12
  Optimization of LiDAR Plot Size ....................................................................................... 13
  Data Modeling ..................................................................................................................... 15

III. RESULTS .......................................................................................................................... 20

  Understory Relative Density .............................................................................................. 20
  Crown Closure Comparison ............................................................................................... 21
  Overstory / Understory LiDAR Comparison ....................................................................... 22
IV. DISCUSSION ..................................................................................................................24
   Understory Relative Density ..........................................................................................24
   Crown Closure Comparison ..........................................................................................26
   Overstory / Understory LiDAR Comparison ..................................................................27

V. CONCLUSIONS AND RECOMMENDATIONS .........................................................29
   Understory Relative Density ..........................................................................................29
   Crown Closure Comparison ..........................................................................................30
   Overstory / Understory LiDAR Comparison ..................................................................31
   Overview .......................................................................................................................32

LITERATURE CITED .............................................................................................................34

APPENDIX

A. SCATTERPLOT CHARTS OF LEVEL 3 (1.5m) LIDAR RETURNS VS. VEGETATION INTERCEPTS FROM FIELD LASER ..........................................................................................36
2.1 Correlation of field to LiDAR point densities over different radii for field plots (column 1). Rho ($\rho$) is the correlation coefficient and the right column indicates if the correlation for each comparison was considered biased .....................................................14

3.1 Chart of all regression models used in understory density analysis. UW = un-weighted and W = weighted .................................................................21

3.2 Chi Squared test where the $H_0$ failed to be rejected. Therefore the distributions are not the same and the model did not work for the validation dataset .................................................................23
LIST OF FIGURES

1.1 Three-dimensional scatterplot of raw X, Y, and Z LiDAR points depicting overstory, midstory, and understory vegetation, as well as ground level .................................................................4

2.1 Example of the overlap between flight lines of LiDAR on the actual flight path traversed during data capture .................................................................13

2.2 Example of fishnet overlay and the same sample with every thirtieth cell selected to reduce data size .................................................................18

2.3 Example of masking the age polygons by the flight-line dimensions to capture stands of interest ........................................................................19

3.1 The relationship between percent crown coverage and LiDAR counts in loblolly pine plantations in Southeastern Louisiana ..........................................22

4.1 Illustration of the densitometer sampling technique where each point represents a sample location for crown closure and the circle represents the total area sampled from the LiDAR ........................................................................................................27
CHAPTER I
INTRODUCTION

Overview

Forest understory density is important to land managers for monitoring wildlife habitat, forest health, forest competition, and wildfire concerns; however, fieldwork to classify understory density is costly and time prohibitive. A remotely sensed method for estimating understory density could enhance evaluation of wildlife habitat, forest competition and fire prediction models while decreasing costs for estimating density by other methods. Light Detection And Ranging (LiDAR) may provide the supplementary information with which to make meaningful estimations. This research seeks to investigate the relationship between forest overstory and understory densities using LiDAR interception counts. It also examines the relationship between forest understory density in pinelands (specifically pine plantations) using field density observations and LiDAR understory density data. The primary objective of this research was to explore the relationship between understory vegetation densities and understory LiDAR count.
Remote Sensing

Aerial photography was one of the first remote sensing tools used, but aerial imagery has advanced dramatically in the last few decades. There are many sensors available for remote sensing, ranging from satellites that can acquire high-resolution imagery to Radio Detection And Ranging (RADAR) and more are becoming available every year (Donoghue 2000).

Although much work has been done to quantify crowns and merchantable volumes using satellite imagery, the methods for determining understory characteristics are labor intensive and time consuming (Riaño et al. 2003). Several articles have been published dealing with remote sensing as a tool to help locate fires and reduce impacts on surrounding areas (Lavrov et al. 2003, Utkin et. al. 2002, Sunar and Özkan 2001, Roy and Prévost 1993). However, very few have dealt with remote sensing techniques for fire prediction by attempting to quantify understory fuel loading levels.

LiDAR

In the late 1960’s, scientists began experimenting with a new tool known as Light Detection And Ranging (LiDAR) for describing bathymetry in oceanographic mapping (Nelson et al. 1988). There are two types of LiDAR, continuous-wave and discrete-return. Currently, there are very few commercial continuous-wave lasers. The majority of LiDAR systems used for natural resource management are discrete-return systems (Lim et al. 2003).
LiDAR is obtained from an aircraft equipped with a Geographic Positioning System (GPS) and an Inertial Navigation System (INS), that permits the identification of the absolute location and altitude of the aircraft (Lim et al. 2003). The laser sends out a pulse and when the pulse intercepts sufficient mass (biomass or other) that it exceeds a specified threshold (i.e. some level higher than an ambient noise condition), a return is recorded. The time from initiation of the pulse to the time that a return is recorded are differenced and recorded by the receiving unit. During post processing, this time is converted into a distance and translated into a coordinate in terms of X, Y, and Z based on the location of the laser at the time of the pulse (Baltsavias 1999). Many systems in use by natural resource professionals today record up to five returns per pulse. Millions of returns over any given area can produce a cloud of points that can represent trees or stands in three dimensions from the ground through the understory and midstory, to the crown (Means et al. 2000).

A relatively recent development in LiDAR data is the inclusion of an intensity (I) value. Since a laser is a near infrared device, and the strength of the return is used to threshold values for inclusion as returns, newer systems also record this signal strength as a separate variable. Although research is still on-going in terms of the accuracy and functionality of including I-values for forestry and natural resource applications, there have been some studies that have had positive results in tree species classification using this intensity value (Douglas 2004). Because LiDAR has a three dimensional aspect, as well as the inclusion of a pseudo-spectral characteristic (intensity), LiDAR could be a useful tool in estimating forest structure parameters.
Upon viewing LiDAR data, it becomes clear that even with raw X, Y, and Z coordinate groupings, a discernable pattern can be recognized in the data. Data groupings clearly represent the patterns of overstory, midstory, and understory vegetation as well as ground level (Figure 1.1).

![Figure 1.1: Three-dimensional scatterplot of raw X, Y, and Z LiDAR points depicting overstory, midstory, and understory vegetation, as well as ground level.](image)

**Forest Structure**

Overstory is the upper most portion of the forest structure and includes dominant, co-dominant, and intermediate trees. Overstory has a strong effect on forest understory because of its interception of light and moisture. As forest crown closure increases it has a strong influence on the decrease of understory density. However, the dense leaves of hardwoods usually result in more complete crown closure than do the needles of pines.
The sparse nature of pine needles allows more light to reach the ground thus allowing the development of higher density understories (Oosting 1956). Light is not the only factor in understory density, competition levels are also important, as shown by studies in which plots in the understory were trenched to remove root competition from the overstory trees. However, with sufficient light, root competition can be overcome. In a similar fashion, lack of light can be overcome by sufficient ability to compete with overstory trees (Oosting 1956). Using this information it can be interpreted that once crown closure occurs, understory density levels will be reduced due to competition and reduced light levels.

Understory characterization is important for a variety of reasons. It can help natural resource managers predict potential fire hazards, interpret understory structure, and decide management polices that influence wildlife habitat (Riaño et al. 2003, Nagendra 2001). The understory is a source of fuel loading, which is a key component of predicting the probability and severity of a fire outbreak. The higher the biomass levels, the more intense the fire may become (Riaño et al. 2003). Understory density is a function of these biomass levels; the higher the biomass level, the greater the density.

**Understory Sampling Techniques**

Most techniques for measuring understory characteristics measure in only two dimensions (X and Y). However, LiDAR returns data in three dimensions (X, Y, and Z). Some sampling techniques currently used for understory characterization are point-center-quarter, random pairs, nearest neighbor, plot-less sample, quadrat sampling,
Nudd’s Board, line transect, and closest individual (Fidelibus and Mac Aller 1993, Beasom and Haucke 1975).

The method used in this study was a combination of the plot-less sample method and a variant of the point-center-quarter method. The plot-less sample method uses a number of points randomly or systematically distributed across a given area. Measurements are taken at those points for various metrics (Fidelibus and Mac Aller 1993). In the point-center-quarter method, the area around a point is divided into four quadrants, usually in the four cardinal directions and the distance to the nearest individual in each quadrant is measured (Beasom and Haucke 1975). Combining these methods and modifying them can quantify understory characteristics in three dimensions. By taking those characteristics that measure densities in the X and Y directions and adding in a height variable, the new measurement technique will account for not only the spread and density across a horizontal range, but also a vertical component.

**Study Objectives**

This project examined southern pine forests on industrial land in St. Tammany Parish, Louisiana. The research dealt primarily with the relationship between raw X, Y, and Z values from LiDAR and the three-dimensional field data. Secondary objectives were to test the relationship between crown closure and understory LiDAR counts as well as the relationship between overstory LiDAR points and the understory LiDAR points. The objectives of this study were:
1. Determine if understory field point counts are related to understory LiDAR point counts.
2. Determine if crown closure as measured from the field is inversely related to understory LiDAR point counts.
3. Determine if overstory LiDAR point counts are inversely related to understory LiDAR point counts as well as the impacts of stand age.

The hypotheses tested for the objectives were as follows:

1. There is no statistically significant relationship between forest understory field point counts in managed pine plantations and LiDAR point count data.
2. There is no inverse relationship between field-measured crown closure and the density of LiDAR returns.
3. There is no inverse relationship between the density of LiDAR crown returns and the density of understory returns.
CHAPTER II

METHODS

Data Collection

Field data were collected on Weyerhaeuser Company holdings in St. Tammany Parish, Louisiana and the LiDAR were acquired by Airborne1® in November 2003. Weyerhaeuser provided a Geographic Information System (GIS) coverage and aerial data of the study area. Flight line orientation and separation distances were determined by Weyerhaeuser to insure coverage of stands and age classes of interest. GPS tracking reports were used with the GIS to locate sample ground plots. There were 980 field inventory plots (20th acre) located for another project and 100 of those plot locations were selected for use in this study. The plots encompassed age classes ranging from 6-26 years consecutively and there were two additional stands age 28 and 42 respectively. Twenty plots were taken in each 5-year age interval for the stands 6-26 years. Ten plots each were planned for the 28-year-old and the 42-year-old stands. Unfortunately, the 42-year-old stand was found to have been cutover between the date of LiDAR collection and the time of the field data collection so the total number of plots for the study was reduced to 90. Field measurements were taken over a 3-week period from mid-December 2003 through the beginning of January 2004.
Each plot was located using a real-time differentially correcting GPS unit. Once the plot center was located, a GPS point was taken using a 20-point averaging method. If it was necessary to disturb the site in order to reach plot center, an offset was used. The points were converted into a GIS point coverage and buffered into a circular polygon with a radius of 32m. In the event that a plot was previously disturbed by industry crew members or other recent activity, the sample point was shifted one plot diameter distance (52.6 ft, 16.03m) from the initial point in a cardinal direction beginning with the east and shifting clockwise if the new location was not representative of the condition to be sampled. For example, if a sample point to the east was to fall outside of the stand or fall outside expected conditions, a new plot was established in one of the three remaining cardinal directions.

Field measurements were collected using a variant of the point-quarter method, thereby accounting for both horizontal and vertical spatial characteristics of each plot. A laser range finder was rotated in 10 degree intervals at 0.5m, 1.0m, and 1.5m above ground to determine distance from plot center to any hit encountered by vegetation. A measurement was taken every 10 degrees at 0.5, 1.0, and 1.5 m above the ground for a total of 36 hits per plot at each height level. Distance, direction, and vegetation type were tallied on field sheets. Field data were then processed with various plot diameters to create a pseudo-metric that would represent the relative vegetative density of understory within the stand. The processing was done iteratively by varying the diameter of a supposed plot diameter until scatterplots of the count showed no bias. The relative density pseudo-metric was assumed to represent the spatial distribution of possible hits.
Densitometer data representing overstory canopy density was also taken at each plot with a GRS Densitometer. The densitometer is a “T” shaped intersection of PVC pipe that has a mirror set at 45 degrees and two level bubbles. When the two levels are centered, then the densitometer is pointed upwards at a 90° angle to the ground.

Overstory observations were taken at 5 foot intervals for 25 feet in each cardinal direction to obtain an estimate of crown-closure. The center point of the densitometer is a dot that is either on vegetation or is not. Thus, the response for this field method is a simple yes or no and it follows the binomial distribution and the confidence intervals are known for a given sample size. This allows for a quick, accurate way to quantify the overstory on each plot (Stumpf 1993).

LiDAR Data Analysis

LiDAR were delivered in raw point format (ASCII X, Y, Z, and Intensity (I)), for extracted vegetation points (all returns except the last) and extracted ground points (last returns) from the November 2003 data collection flight. For this analysis only the raw and extracted ground points were used; intensity was not used. To separate out points by flight line from the raw LiDAR data a Visual Basic program was created especially for this project and was used to create and output the points in ArcInfo point coverage format. Once in the coverage format, each plot was intersected by the buffer polygon of the plot to add plot number and center X, Y coordinate.

Extracted ground points were transformed into a Triangular Irregular Network (TIN) and the ArcInfo command TINSPOT was used to add the Z-value at the center of
each plot to the coverage. Since the terrain at the scale of each plot had variations, sometimes as much as three to four feet, the LiDAR data had to be corrected to the same elevation plane as the field data. Therefore, the center points were expanded to create four corner points that encompassed all of the LiDAR points within each circular plot. These four points had equal Z-values and were converted into a TIN for each plot making a flat surface so that each data vegetation point in the plot had height measured from the same elevation. Each LiDAR plot was passed over the TIN with the TINSPOT command to add the Z-value for the plot center to every point. This was necessary because the height of the field data for the Z-value at plot center was used; not the height above ground where the data point occurred.

Once all the appropriate information was attached to each point, the table for each point cloud was exported into dBase table format. The distance of each point from plot center was determined using the Euclidian distance formula where x and y are UTM coordinates in meters (#1):

\[ Dist_E = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \]  

(#1)

Each plot was then broken into three separate tables, one for each of the height planes. Level 1 included points from the ground to 0.75m and its field counterpart was the 0.5m level. Level 2 included points from 0.75m to 1.25m and its counterpart in the field data was the 1.0m level. Level 3 encompassed all points that were 1.25m to 1.75m in elevation and its field counterpart was the 1.5m level. Each table contained a count of total number of LiDAR hits within various horizontal distances from 1m to 20m. The distances represented potential plot radii.
Data Issues

The LiDAR data used in this study contained errors relating to the sorting method that had been used to subset the full LiDAR dataset into individual flight lines. Data from cross-flight validation lines during data calibration, turn-arounds between lines, and geographic sorting method created double and sometimes triple coverage in the raw LiDAR data (Figure 2.1). Since the primary objective of this project was to examine the relationship between raw LiDAR counts and field data, the overlap caused serious data suitability issues. Many plots fell into portions of the flight lines that contained these data issues. All attempts to correct the duplication problem were unsuccessful, and all attempts to receive corrected data from the provider were not successful. Only 68 of 90 field plots were deemed useable for this study because they contained no duplicate data.
Optimization of LiDAR Plot Size

Due to the nature of the field plot and LiDAR plot data, any plot size can be assigned to a plot or height level within the plot. The linear correlation values between LiDAR and field data were used to determine the optimal plot size for developing the relationship between LiDAR data and field data in each stratum of the plots. Simple correlations were obtained between field plot sizes of 1m to 8m and LiDAR plot sizes of 1m to 15m (Table 2.1). For level 1 (0.5m), the optimal related LiDAR plot was a plot
with a 5m radius; 9m for level 2 (1.0m) and 10m for level 3 (1.5m). These were selected because they had the highest related correlation values to the field data after accounting for the artificial relationship at the selected field data plot sizes. To determine the field plot radii, a 10m standard LiDAR plot size was selected for all comparisons. The 1m field plot radius versus the 10m LiDAR plot radius showed a left side skewing in the scatterplot with a weak trend-line. The 2m field plot exhibited left side skewing in the scatterplot of and a stronger trend-line. The 3m field plot showed a fairly evenly distributed scatterplot with a stronger trend-line than any of the other plots exhibited. For the 4m field plot, a fairly evenly distributed scatterplot was observed, with a weaker trend-line than the 3m plot. The 5m and 6m plots were observed to have right side skewing with a weak trend-line. The 7m, 8m, and 9m, field plots all exhibited right side skewing of scatterplots with field sizes that led to an artificially inflated correlation value.

Table 2.1: Correlation of field to LiDAR point densities over different radii for field plots (column 1). Rho (\( \rho \)) is the correlation coefficient and the right column indicates if the correlation for each comparison was considered biased.

<table>
<thead>
<tr>
<th>Lvl. 3: Field vs LiDAR</th>
<th>Fig. Value (Appendix B)</th>
<th>( \rho )</th>
<th>Biased?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m v 10m</td>
<td>A</td>
<td>0.07464</td>
<td>Y</td>
</tr>
<tr>
<td>2m v 10m</td>
<td>B</td>
<td>0.20925</td>
<td>Y</td>
</tr>
<tr>
<td>3m v 10m</td>
<td>C</td>
<td><strong>0.20537</strong></td>
<td>N</td>
</tr>
<tr>
<td>4m v 10m</td>
<td>D</td>
<td>0.1806</td>
<td>N</td>
</tr>
<tr>
<td>5m v 10m</td>
<td>E</td>
<td>0.1642</td>
<td>Y</td>
</tr>
<tr>
<td>6m v 10m</td>
<td>F</td>
<td>0.19491</td>
<td>Y</td>
</tr>
<tr>
<td>7m v 10m</td>
<td>G</td>
<td>0.23841</td>
<td>Y</td>
</tr>
<tr>
<td>8m v 10m</td>
<td>H</td>
<td>0.23841</td>
<td>Y</td>
</tr>
<tr>
<td>9m v 10m</td>
<td>I</td>
<td>0.24176</td>
<td>Y</td>
</tr>
</tbody>
</table>
After analyzing each combination based on a scatterplot to determine correlation and bias, a 3m plot size was selected for the field data at all three levels. Although a 3m plot size for level 3 did not produce the highest correlation value, it was determined that the higher correlation values from the larger plots sizes were a result of artificial inflation due to the nature of the field sampling method. As the plot size increased beyond a 3m radius in level 3, more points fell at the maximum limit of the total points per plot (36). This created an artificial linear relationship in the data resulting in the inflated correlation values (Appendix A). Determination of optimal plot size was based on ocular selection of scatterplots that had even distribution and exhibited the best correlation values (Table 2.1).

**Data Modeling**

*Objective 1: Determine if understory field point counts are related to understory LiDAR point counts.*

A linear regression model was used to estimate the relationship between field point density and LiDAR return density. Other simple linear regression models were tested to determine if age and crown closure were significant variables of interest. After testing several regressions, it was determined that a simple linear regression was not an acceptable model for the data. A multiple linear model was constructed that incorporated the age variable as well as the LiDAR count variable to determine field counts. However, residuals from model #2 indicated that the model did not meet the assumption of non-

\[
Field = \beta_0 + \beta_1 (\text{LiDAR}) + \beta_2 (\text{Age}) \quad (#2)
\]
constant variance. A transformation was performed on the data and the refitted model’s (#3) residuals indicated that it met the assumption of non-constant variance.

\[ Field = \beta_0 + \beta_1 (\text{LiDAR}) + \beta_2 (1/\text{Age}) \]  

(#3)

A weighted and non-weighted regression was applied to each level of the data.

**Objective 2: Determine if crown closure as measured from the field is inversely related to understory LiDAR point count.**

In order to test the relationship of crown closure to LiDAR point count density, the densitometer data were used to determine field overstory conditions. Points along the transect that intersected overstory vegetation were assigned a value of one; those that did not intersect overstory were given a value of zero. Transect counts were totaled and divided by 25 (the total number of samples taken for each plot) to produce a crown closure percentage. The LiDAR data were then subset by the twenty-five foot radius to create a comparable region of interest. All LiDAR understory points that fell in that region were totaled and paired with the appropriate ground plot. A linear regression model was used to determine if the assumption of an inverse relationship between crown closure and ground point density was true.

**Objective 3: Determine if mid and overstory LiDAR point counts are inversely related to understory LiDAR point counts and the relationship of these counts to stand age.**

The third objective was tested by taking total mid and overstory counts (based on a midstory minimum of >10ft) and comparing these numbers with the total counts for
understory on the same area, as well as age class, ground elevation, slope, and aspect. There were 24 total flight lines to choose from, and at a posting density of 1.9 shots per square meter, the file sizes were quite large. It was necessary to choose a short flight line in order to reduce the file size for computation. Of the flight lines that were between 5,000 and 15,000 meters in length, line 18 intersected the most age classes so it was selected for the analysis. A ground Triangular Irregular Network (TIN) was created and ground heights were attached to each non-ground LiDAR point to give a height above ground using the ARCINFO command TINSPOT. Counts were computed by creating a fishnet lattice over the total area of line 18 in 20m² cells and selecting, every 30th cell for analysis (see Figure 2.2). This was achieved using the “select if” statement and performed on the Geodatabase’s tabular attributes using ArcGIS.

\[
\text{SelectIf: } ((\text{INTEGER(ID#/30)*30})-\text{ID#}) = 0 \quad (#4)
\]

Once the cells of interest were selected the points that fell within them were easily subset. The points were then divided, based on elevation, into either understory or overstory counts. A value of 1 was assigned to each point and using ESRI®’s Spatial Analyst toolset, the total counts of LiDAR points were summed across each cell for both understory and for mid and overstory. For each cell, age class was derived using the GIS provided by Weyerhaeuser. The coverage was queried for stands that fell within the line of interest, and was clipped to match the flight line (Figure 2.3). The clipped file was rasterized and multiplied by the lattice to attach age values to each cell. Stands outside Weyerhaeuser ownership were removed, as well as any non-pine stands and right-of-ways. A Digital Terrain Model was surfaced from the extracted ground points in the
LiDAR deliverables and was used to calculate ground elevation, slope, and aspect at a 5m resolution. The data were aggregated by plot and fitted to a multiple linear regression model. The elevation, slope, and aspect variables were found to be non-significant, so a refitted model was created (#5).

\[ \text{Understory} = \beta_0 + \beta_1(Overstory) + \beta_2(Age) \]  

(#5)

Every 15\textsuperscript{th} cell was used for validation (minus the cells used in the model).

Figure 2.2: Example of fishnet overlay and the same sample with every thirtieth cell selected to reduce data size.
Figure 2.3: Example of masking the age polygons by the flight-line dimensions to capture stands of interest.
CHAPTER III
RESULTS

Understory Relative Density

The understory density linear regression for optimal data model #3 at the first level (0.5m) was significant at the 0.05 alpha level with an $R^2$ of 0.337 and an adjusted $R^2$ of 0.309 and all slope coefficients were significant. The first level had a Root Mean Square Error (RMSE) of 7.60.

For level 2 (1.0m), the optimal regression model #3 was also significant and had an $R^2$ of 0.361 and an adjusted $R^2$ of 0.335. All betas were significant and the RMSE was 7.78. The third level regression model was significant, however, there was a large drop in the $R^2$ values. The $R^2$ value for the optimal third level (1.5m) model #3 was 0.124 and the adjusted $R^2$ was 0.089. A larger problem, however, was the lack of significance of the LiDAR beta coefficient. It did not pass the significance test with a t value of 0.228. The other betas were significant. The third level had an RMSE of 8.18.

The differences between the weighted regression versions of each level and their un-weighted regression counterparts followed the same pattern; the $R^2$ would increase, but the RMSE would more than double in each case (Table 3.1). All the regressions were significant for each level and all slope coefficients were significant with the exception of the level 3 LiDAR coefficient that, like its un-weighted regression counterpart, was non-
significant. Level 1 had a $R^2$ of 0.377 and an adjusted $R^2$ of 0.351, however the RMSE was 19.85. For level 2, the $R^2$ was 0.426 with an adjusted $R^2$ of 0.403, likewise it had a much higher RMSE at 17.23. Level 3 had an $R^2$ of 0.145 and an adjusted $R^2$ of 0.111.

The RMSE for level 3 was 18.40. Table 3.1 has a side-by-side comparison of all the levels, for both weighted and un-weighted regressions.

Table 3.1: Chart of all regression models used in understory density analysis. UW = un-weighted and W = weighted.

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Equation</th>
<th>Pr&gt; F</th>
<th>$R^2$</th>
<th>Adj. $R^2$</th>
<th>$B_0$ Pr&gt;t</th>
<th>$B_1$ Pr&gt;t</th>
<th>$B_2$ Pr&gt;t</th>
<th>RMSE</th>
<th>%Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lvl1, UW</td>
<td>$F = 13.03 + 0.40(L) + 104.65(1/A)$</td>
<td>&lt;0.001</td>
<td>0.34</td>
<td>0.31</td>
<td>0.000</td>
<td>0.001</td>
<td>0.002</td>
<td>7.600</td>
<td>21.1</td>
</tr>
<tr>
<td>Lvl2, UW</td>
<td>$F = 9.02 + 0.23(L) + 113.04(1/A)$</td>
<td>&lt;0.001</td>
<td>0.36</td>
<td>0.34</td>
<td>0.005</td>
<td>0.005</td>
<td>0.002</td>
<td>7.782</td>
<td>21.6</td>
</tr>
<tr>
<td>Lvl3, UW</td>
<td>$F = 8.37 + 0.1(L) + 75.26(1/A)$</td>
<td>0.039</td>
<td>0.12</td>
<td>0.09</td>
<td>0.019</td>
<td>0.228</td>
<td>0.037</td>
<td>8.179</td>
<td>22.7</td>
</tr>
<tr>
<td>Lvl1, W</td>
<td>$F = 13.03 + 0.40(L) + 104.65(1/A)$</td>
<td>&lt;0.001</td>
<td>0.38</td>
<td>0.35</td>
<td>0.022</td>
<td>0.001</td>
<td>0.008</td>
<td>19.853</td>
<td>55.1</td>
</tr>
<tr>
<td>Lvl2, W</td>
<td>$F = 9.02 + 0.23(L) + 113.04(1/A)$</td>
<td>&lt;0.001</td>
<td>0.43</td>
<td>0.40</td>
<td>0.156</td>
<td>0.029</td>
<td>0.000</td>
<td>17.229</td>
<td>47.9</td>
</tr>
<tr>
<td>Lvl3, W</td>
<td>$F = 8.37 + 0.1(L) + 75.26(1/A)$</td>
<td>0.021</td>
<td>0.15</td>
<td>0.11</td>
<td>0.038</td>
<td>0.258</td>
<td>0.021</td>
<td>18.400</td>
<td>51.1</td>
</tr>
</tbody>
</table>

Crown Closure Comparison

The result of a simple linear regression analysis for crown closure percentage versus understory LiDAR counts; indicates that little, if any, correlation existed between the two variables. The addition of age data to the model did not improve significance. A scatterplot of LiDAR counts versus the percent crown closure depicts the lack of trend to the data (Figure 3.1). The trend line shows almost no slope and the data points are bunched on the left hand side of the chart. There are a few extreme outliers that could not be attributed to known over sampling problems in the data. Postulation as to the cause of the poor relationship will be covered later.
Figure 3.1: The relationship between percent crown coverage and LiDAR counts in loblolly pine plantations in Southeastern Louisiana.

**Overstory/Understory LiDAR Comparison**

The variables age, aspect, and slope were found to be non-significant as they related to understory density but moderately related to overstory. LiDAR points in the understory were significantly related to the density of the overstory. The regression equation #6 had an $R^2$ of 0.5226, demonstrated the inverse relationship between overstory densities and understory LiDAR hits.

$$\text{Understory} = 881.99 - 0.618(\text{Overstory}) \quad (#6)$$
Using the regression model from the first test, the validation data set (every 15th cell except for those cells used in the original model) was tested. The expected value from each cell, based on the regression model, was paired with its observed value from the actual LiDAR data count. A Chi – Square test was used to compare the two distributions of the validation and the original data set. If the two datasets are from the same distribution, this will confirm the regression results. However, the result of the Chi-Squared test was to fail to reject the H₀ that the data follow different distributions (Table 3.2).

Table 3.2: Chi Squared test where the H₀ failed to be rejected. Therefore the distributions are not the same and the model did not work for the validation dataset.

<table>
<thead>
<tr>
<th>Mid-point of Range</th>
<th>Expected (E)</th>
<th>Observed (O)</th>
<th>(O-E)²/E</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>34.00</td>
<td>722.08</td>
<td>13925.00</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>78.15</td>
<td>547.10</td>
<td>2813.85</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>124.21</td>
<td>535.60</td>
<td>1362.59</td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>175.21</td>
<td>454.97</td>
<td>446.71</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>220.10</td>
<td>505.75</td>
<td>370.72</td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>272.44</td>
<td>478.44</td>
<td>155.76</td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>330.54</td>
<td>549.46</td>
<td>145.00</td>
<td></td>
</tr>
<tr>
<td>375</td>
<td>372.43</td>
<td>622.14</td>
<td>167.43</td>
<td></td>
</tr>
<tr>
<td>425</td>
<td>441.50</td>
<td>463.00</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>475</td>
<td>476.33</td>
<td>643.67</td>
<td>58.78</td>
<td></td>
</tr>
<tr>
<td>525</td>
<td>524.67</td>
<td>494.00</td>
<td>1.79</td>
<td></td>
</tr>
<tr>
<td>575</td>
<td>569.40</td>
<td>717.00</td>
<td>38.26</td>
<td></td>
</tr>
<tr>
<td>625</td>
<td>619.00</td>
<td>682.00</td>
<td>6.41</td>
<td></td>
</tr>
<tr>
<td>675</td>
<td>675.00</td>
<td>729.00</td>
<td>4.32</td>
<td></td>
</tr>
</tbody>
</table>

19497.66  \( \chi^2 \) Test
CHAPTER IV
DISCUSSION

The results of this study show that the comparison of LiDAR and over/understory using the techniques describe earlier fails to detect the relationship between LiDAR and over/understory densities. While the 3D sampling structure in the field allowed for collection of data in a format similar to the data it was compared to, few discernable relationships were found. This study failed to detect the relationship between the number of LiDAR understory hits and the amount of crown closure. This research was based on the seemingly logical assumption that the higher the percentage of crown closure the fewer LiDAR pulses would reach the understory. While this study found a significant relationship between overstory LiDAR counts and understory LiDAR counts, the process does not appear to be repeatable, even across the same geographic area. Although plots with biased data from cross-flights were discarded, these data problems may have caused bias in all the LiDAR calculation in this study, and may have introduced an unknown effect on the results of the various tests in this study.

Understory Relative Density

Although levels 1 and 2 both had moderately acceptable $R^2$’s, the RMSE’s for each level translated into approximately 21% error. So for every plot, the total plot count
could have a prediction error of approximately a quarter of the total. The relationships between understory density and LiDAR hit count are significant, but they are not very strong. The coefficient results for model 3 indicate that $B_1$ (the LiDAR coefficient) was positive so the relationship between the LiDAR and understory counts is positive, but it has a very little influence when compared with the $B_2$ (the age coefficient). This points to age having a larger impact on the model than the LiDAR counts.

While the third level had some significant models, the LiDAR coefficient was non-significant in all tests of the model. One explanation is the spread of the understory as it goes up vertically. Based on observed conditions at the study site, grasses and forbs were at the lowest levels of the understory and they tended to be dense in nature. Even if there were not any grasses and forbs in the immediate vicinity of plot center, there was likely a patch very near by. In the mid-level understory, there were usually shrubs or small trees that filled the area. Since both these levels had high levels of biomass relatively close to plot center, a laser point would be recorded at a short distance and they would share similar distributions. The increased biomass also allowed for more potential hits for the LiDAR to record. In the upper portion of the understory (i.e. the third level), vegetation for both the field laser and the LiDAR system to intercept becomes much more sparse, thus accounting for the greater variability in the upper most level as well as for the lack of a significant relationship.
Crown Closure Comparison

This particular portion of the study did extremely poorly in terms of a functional model. It seems intuitive that a higher crown closure percentage would result in a lower number of understory LiDAR points. The higher levels of biomass should intercept the laser and reduce the number of understory hits. Likewise, lower levels of light being able to reach the ground should reduce overall understory biomass. Beyond the problems with data irregularities caused by the calibration, another explanation for the counter-intuitive result might be the plantation style planting rows of the study site. Since the densitometer data were taken in cardinal directions and some rows were also oriented in cardinal directions, this orientation may have caused an artificial inflation (or deflation, depending on starting location) of crown closure on some plots. In addition, the total LiDAR plot counts for the understory were taken in a full 25ft. radius around each plot center, instead of in a swath underneath the paths taken to estimate crown closure (Figure 4.1). This left a lot of area with no overstory crown closure information. Had the densitometer estimates been taken in eight directions instead of four, this might have accounted for more of the differences in the plots.
Figure 4.1: Illustration of the densitometer sampling technique where each point represents a sample location for crown closure and the circle represents the total area sampled from the LiDAR.

**Overstory/Understory LiDAR Comparison**

It was expected in the overstory versus understory LiDAR count test that most of the variables would be non-significant. Slope and aspect were not expected to have a strong relation to understory LiDAR count, and accordingly, regression analysis did show them non-significant. Age, however, was expected to have a significant relationship, but it did not. This may have been due to undetected influences of calibration data in the raw
data. Unlike the previous tests, there was no way to mitigate the influx of calibration points in this test due to the nature of the sampling scheme. This may also explain the lack of similarity between the test regression and the validation data set tests.

The understory and overstory relationship in the test regression was fairly strong for a natural systems regression, and was achieved without any transformations on the data. The negative overstory coefficient was logically expected; as LiDAR overstory counts increase, LiDAR understory counts decrease. So for most purposes, it would seem that this analysis succeeded and followed the hypothesized direction. Unfortunately, when the validation set was used, repeatability of the model became a problem. Some possible solutions to this issue might be to perform the study on a larger area with more age classes. The range of ages was very wide for this test, but there were not samples from all of the age classes available in the entire study area. Also, it may beneficial to perform the study again and compare the results with validation on a different geographic region that LiDAR has been flown on (assuming a similar posting density for the LiDAR).
CHAPTER V
CONCLUSIONS AND RECOMMENDATIONS

Understory Relative Density

This study has shown that, although there was statistical significance to the understory portion relative density of the study as well as a decent $R^2$, there are several problems with the results. One of the main problems was, even if a strong relationship was found, the resulting field count reveal very little. The counts themselves do not tell how much biomass was actually present. It did describe the distribution of that biomass, but a total amount could not be determined. If a better metric could be developed for the three-dimensional sampling technique, it could allow the investigator to more accurately quantify the actual vegetation density.

Another problematic characteristic of the results was the properties of a multiple linear regression. The solution to the model had the capacity to exceed the allowable counts for each plot. Since there were 36 maximum possible counts per plot, if one were to encounter a LiDAR plot with sufficiently high LiDAR point counts, it is conceivable that the field count would be more than 36 and any number of understory counts higher than 36 is undefined since the maximum number of points in the field was 36. While this could be overcome by defining these plots as extremely high density, it still seems to be problematic since there was no reference to compare of what a field count of 38 would
relate to. This also brings up the question of continuous versus discreet answers. In the selected regression model for this study, the model almost always returned a continuous answer. This has the same implications as the over-counting problem in that a field count of 24.3 is undefined in relation to reference field counts.

One solution to the problem of biomass definition would be to use an existing biomass metric to compare to the LiDAR data. One example, in particular, would be to use a variant of the Daubenmire method (Daubenmire 1959). Using three Daubenmire circles (a circle one square meter) set at the same heights as in this study and occularly determining percentage biomass at each level at plot center as well as in eight directions a few meters away from plot center, the investigator should be able to define the biomass levels of the plot much more accurately. One could then compare total LiDAR counts for the extent of the plot at each level and compare them to the biomass levels (which are continuous and could be defined for quantities higher than found in the study) to reflect a much more interpretable and useful result.

Crown Closure Comparison

This particular portion of the study did not produce any real useful results. The most likely way to solve the issue in this particular study would be to take fewer densitometer readings (as far as length of transect) and have more transects that are shorter. Using eight transects (one in each of the cardinal directions as well as in the sub-cardinal directions) that are only 15ft. long would cover more of the area that was also covered by the related LiDAR data to make a more comparable sample. Field data was
thought to be the main culprit in the differences since the overstory versus understory portion of this study showed that there was a fairly strong relationship when dealing only with the LiDAR data.

**Overstory/Understory LiDAR Comparison**

While age was an integral part of the field to LiDAR study, it was not as useful when comparing LiDAR to LiDAR models. This particular part of the study also showed that there was a strong correlation between the overstory LiDAR and understory LiDAR. This was important because it tends to illustrate the concept that as more overstory points occur in a given area, there are fewer understory points. The main problem that was encountered was the lack of repeatability for the model. This was believed to be primarily due to distortions stemming from the calibration cross-flight data.

One solution to the noise that is introduced by the cross-flight data would be to normalize the LiDAR counts across those areas that are affected, if not the entire dataset. A procedure for accomplishing this could be achieved on a plot by plot basis by summing the total LiDAR returns for a plot, dividing it by the total number of individual shots taken and dividing that result by the total area of the plot (#7). This number could be used

$$w = \left( \frac{\sum \text{LiDAR}(\text{return})}{\text{LiDAR}(\text{shots})} \right) / \text{Area}(\text{plot})$$  \hspace{1cm} (#7)

to weight the regression by allowing the investigator to compare normalized values.

Another factor that has serious implications on stand structure for this study is thinning. If a thinning occurs on a stand the immediate impact is the physical reduction in understory biomass. Also, in a case such as this, you might see lower levels of crown
closure as well as lower levels of understory biomass which would run opposite of what was hypothesized for objective 3. This could also have profound effects on the results of modeling for objectives 1 and 2. A few years after a thinning and before crown closure had reached the levels it would be at if no thinning had occurred, more light would reach the understory. This would likely cause an increase in understory densities. The number of years since a thinning occurred on any plot, as well as the type of thinning could be extremely useful data to include into a model to describe some of the unexplained variance that was found in the models tested for this study.

When choosing a flight line for future studies, it is recommended that a flight line with evenly distributed age classes (in terms of acreage) would be more desirable than those with those stands that have many slivers of different age classes. Doing this would allow more plots for each age class in areas with similar conditions rather than a few plots in many different age classes. Although the model did not find age to be a significant variable, the structure and therefore LiDAR count distribution will change over the life of a stand, so it would be imperative to sample these changes.

**Overview**

The conclusions to the hypotheses are as follows: 1) There was a statistically significant relationship between forest understory field point counts in managed pine plantations and LiDAR point count data. 2) Lower percentage of crown-closure did not result in higher relative density of understory derived from LiDAR. 3) A larger number of crown returns would result in a lower number of understory returns.
Some final thoughts on this study would have to include the possibility that those portions of this project that did not produce acceptable results may be viable with a LiDAR dataset of different specifications. A higher posting density would allow for more understory shots, as would a dataset that had more returns (i.e. a four or five return system as opposed to a two return system). One last concept that may have great implications in this particular study was the accuracy of LiDAR versus the accuracy of a field technician or researcher. The average accuracy in the vertical axis for current LiDAR systems is 15cm. It is possible that the dataset collected in the field and used as the reference set was less accurate than the dataset that was being defined. It is therefore postulated that a more accurate or robust field collection would most likely create a more comparable dataset for the development of statistical models.
LITERATURE CITED


APPENDIX A

SCATTERPLOT CHARTS OF LEVEL 3 (1.5m) LIDAR RETURNS VS.
VEGETATION INTERCEPTS FROM FIELD LASER
Lvl3 1m Field vs 10m Lidar

\[ y = 0.2534x + 24.478 \]

Vegetation Intercept with Field Laser

Lvl3 2m Field vs 10m Lidar

\[ y = 0.4127x + 21.286 \]

Vegetation Intercept with Field Laser
Lvl3 3m Field vs 10m LiDAR

\[ y = 0.3525x + 19.386 \]

Lvl3 4m Field vs 10m LiDAR

\[ y = 0.3513x + 17.211 \]
Lvl3 5m Field vs 10m LiDAR

\[ y = 0.3532x + 15.829 \]

Lvl3 6m Field vs 10m LiDAR

\[ y = 0.4801x + 11.314 \]
Lvl3 7m Field vs 10m LiDAR

\[ y = 0.6971x + 3.7269 \]

Vegetation Intercept with Field Laser

Lvl3 8m vs 10m LiDAR

\[ y = 0.8908x - 3.1576 \]

Vegetation Intercept with Field Laser
Lvl3 9m Field vs 10m LiDAR

$y = 0.9533x - 5.9275$