Silvilaser2012.com

Full Proceedings

first return

September 16–19th

SilviLaser

VANCOUVER *2012*

Vegetation classification in southern pine mixed hardwood forests using airborne scanning laser point data

Robert J. McGaughey[1], Stephen E. Reutebuch[2]

[1]USDA Forest Service, Pacific Northwest Research Station, bmcgaughey@fs.fed.us
[2]USDA Forest Service, Pacific Northwest Research Station, sreutebuch@fs.fed.us

Paper Number: SL2012-163

1. Introduction

Forests of the southeastern United States are dominated by a relatively small number of conifer species. However, many of these forests also have a hardwood component composed of a wide variety of species that are found in all canopy positions. The presence or absence of hardwood species and their position in the canopy often dictates management activities such as thinning or prescribed burning. In addition, the characteristics of the under- and mid-story layers, often dominated by hardwood species, are key factors when assessing suitable habitat for threatened and endangered species such as the Red Cockaded Woodpecker (*Picoides borealis*) (RCW), making information describing the hardwood component important to forest managers.

General classification of cover types using LIDAR data has been reported (Song et al. 2002, Brennan and Webster 2006) but most efforts focusing on the identification of individual species or species groups rely on some type of imagery to provide more complete spectral information for the study area. Brandtberg (2007) found that use of intensity data significantly improved LIDAR detection and classification of three leaf-off deciduous eastern species: oaks (*Quercus* spp.), red maple (*Acer rubrum* L.), and yellow poplar (*Liriodendron tulipifera* L.).

Our primary objective was to determine the proportion of hardwood species present in the canopy using only the LIDAR point data and derived products. However, the presence of several hardwood species that retain their foliage through the winter months complicated our analyses. We present two classification approaches. The first identifies areas containing hardwood and softwood (conifer) species (H/S) and the second identifies vegetation with foliage absent or present (FA/FP) at the time of the LIDAR data acquisition. The classification results were used to develop predictor variables for forest inventory models. The ability to incorporate the proportion of hardwood and softwood was important to the inventory as well as habitat assessments for the RCW.

1.1 Data

1.2 Study area

This study was conducted on the Savannah River Site (SRS). SRS is a National Environmental Research Park covering 80,267 ha (198,344 acres) located in the southeastern coastal area of the United States in west central South Carolina. In partnership with the Department of Energy (DOE), the USDA Forest Service's Savannah River Forest Station manages nearly 73,653 ha (182,000 acres) of commercial forest and more than 4,856 ha (12,000 acres) of non-forest land for a variety of natural resources.

Forests of the area are about 69% pine and 31% hardwood or mixed pine-hardwood. Dominant pine species include longleaf (*Pinus palustris* Mill.) and loblolly (*P. taeda* L.) pine and common hardwood species include various oaks, yellow poplar, blackgum (*Nyssa sylvatica* Marsh.), sweetgum (*Liquidambar styraciflua* L.), red maple, hickories (*Carya* spp.), and hollies (*Ilex* spp.). Bottomland hardwood forests are found along SRS streams and on the "islands" or "ridges" of the Savannah River swamp. In these forests, typical canopy species include water oak (*Q. nigra* L.), laurel oak (*Q. laurifolia* Michx.), sweetgum, elms (*Ulmus alata* Michx. and *U. Americana* L.), red maple, and yellow poplar. Swamp forests are common along the western boundary of the site, adjacent to the Savannah River. Baldcypress (*Taxodium distichum* L.) and water tupelo (*Nyssa aquatic* L.) are common in these forests. Within the forests of SRS, not all hardwood species are deciduous. SRS has a mixture of hardwood species that are evergreen, tardily deciduous, or that retain desiccated leaves through winter.

1.3 Laser data

LIDAR data were acquired for SRS in the spring of 2009 when deciduous trees were in leaf-off condition. Data were acquired using two Leica ALS50-II laser scanners (designated as sensor 46 and sensor 77 by the data provider) mounted in separate fixed-wing aircraft and operating during the same time period. The average overall pulse density was approximately 10 pulses/m^2. The total area covered by the acquisition was approximately 119,000 ha (294,055 acres). Acquisition specifications for the data are shown in Table 1.

Table 1. Flight parameters and scanning system settings.

Flying height above ground (planned)	1432 m
Scan angle (flown)	±10°
Scan angle (delivered)*	±8°
Average scanning swath width (flown)	505 m
Swath overlap (flown)	62.5 percent
Scan frequency	58 Hz
Pulse rate	150 kHz
Beam divergence	0.22 mRad

*Returns from the outer 2 degrees of each scan were deleted prior to delivery. This reduced the scan angle, swath width, and overlap.

1.4 Field measurements

Tree measurements were collected on 194 plots in the spring of 2009. The plot protocol used a nested set of circular, fixed area plots to characterize all trees with diameter at breast height (DBH) larger than 2.54 cm (1 inch). The basic plots were 0.04 ha (0.1 acre) unless there were fewer than 8 dominant or co-dominant trees present on the plot. For these sparse stands, the plot size was increased to 0.081 ha (0.2 acre). On the basic plot the following measurements were taken for live and dead trees with DBH larger than 7.62 cm (3 inches): species, DBH, height, crown base height, and crown class. A smaller 0.004 ha (0.01 acre) plot, nested within the basic plot was used to collect detailed information for the smaller trees. On this plot the same measurements were taken for live and dead trees with DBH larger than 2.54 cm (1 inch) and less than or equal to 7.62 cm (3 inches). Smaller trees on the basic plot but outside the smaller 0.004 ha (0.01 acre) plot were tallied by species and size class (2.54 cm (1 inch) <= DBH < 5.08 cm (2 inch) and 5.08 cm (2 inch) <= DBH < 7.62 cm (3 inch)). For this study data were summarized by species, hardwood/conifer classes, live/dead, and foliage absent/present conditions.

A separate crew collected plot locations using dual-frequency, survey-grade GPS receivers (JAVAD Maxor. At least 600 positions were recorded for each plot center (10 minute occupation with 1-second epochs). Position data were post-processed using a continuously operating reference station (CORS) located close to the study site.

1.5 Photo measurements

Twenty five additional 0.04 ha (0.1 acre) plots were established using aerial photography acquired in leaf-off conditions. The 25 photo plots were located in swamp areas where the dominant species were baldcypress and water tupelo. Both species were without foliage at the time of the acquisition. Locations for these plots were digitized directly from the digital orthophotos in a GIS.

2. Methods

2.1 Intensity adjustment

Differences between the intensity values for the two sensors are visible in an image (7.62 m (25 feet) square pixel) produced using the first returns (Figure 1: left image). Intensity values for first returns were compared using 4,419 5- by 5-meter samples distributed evenly (10m spacing) along the area covered by the two sensors (red line in Figure 1). These samples were extracted from the overall point cloud and descriptive statistics were computed for the intensity values in each sample (Table 2). Overall, sensor 46 recorded higher values than sensor 77 over the same target area. In general, the difference in the intensity values recorded by the two sensors (Figure 2) is similar to the difference one would observe in photographs of the same area acquired using two different cameras or different exposure settings. Histogram matching is a relatively simple process that balances detector responses when dealing with images collected by different sensors or under different atmospheric conditions (Gonzalez and Woods 2007). A histogram matching procedure was implemented to adjust intensity values for first returns from sensor 46 relative to those from sensor 77.

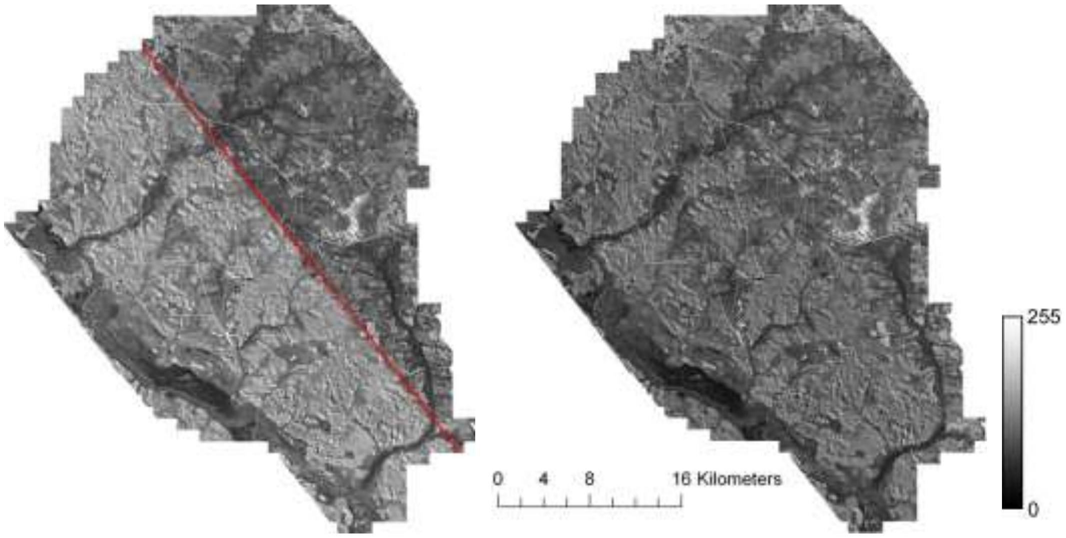

Figure 1. Intensity images created using original intensity values for first returns (left) and adjusted intensity values for first returns (right). The red line on the left image shows the center of the area covered by both LIDAR sensors. The NE portion was collected using sensor 77 and the SW portion using sensor 46.

2.2 Selection of training plots

For our supervised classification, we wanted to identify training plots that represented a “pure” condition. We selected a subset of field plots based on the mix of species present on each plot. We wanted plots with basal area composed of either all conifer or all hardwood species. From the 194 field plots, 19 conifer plots and 15 hardwood plots were identified.

Training plot selection was complicated by the presence of deciduous or tardily-deciduous hardwood species. When considering the presence or absence of foliage for trees on a plot, we were able to identify only three plots with 98% or more of the basal area in species without foliage. Initially, we wanted all trees on a plot to have the same condition but found that none of the 194 plots met the criterion. To augment the data for the “foliage absent” condition, we used the 25 photo plots in addition to the three field plots. Thirty six plots had 98% or more of the basal area in species with foliage.

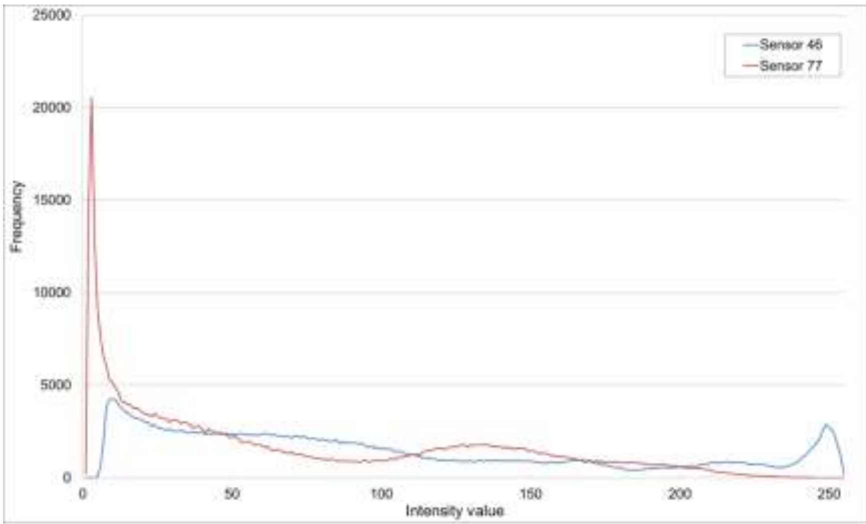

Figure 2. Distribution of first return intensity values for sensor 46 and sensor 77 for the portion of the acquisition covered by both sensors.

2.3 LIDAR-derived metrics

We extracted samples from the original point cloud and the first return data with adjusted intensity values using the location and size of each plot. We computed metrics for all of the plots using 2- by 2-m cells--resulting in 98 cells for the 0.04 ha (0.1 acre) plots and 197 cells for the 0.081 ha (0.2 acre) plots. For the H/C classification, there were a total of 4,189 cells with 2,071 cells representing pure hardwoods and 2,118 cell representing pure conifers. For the FA/FP classification, there were 6,660 cells with 2,831 representing vegetation without foliage and 3,829 representing vegetation with foliage.

Two types of metrics were computed for each 2- by 2-m cells; Intensity metrics using only first returns with adjusted intensity values within 2m of the canopy surface, and metrics computed using the height above ground for LIDAR returns above 2m. We also computed a pulse penetration metric based on first and last-of-many return surfaces but did not find it useful for classification.

To compute intensity metrics, a 0.5- by 0.5-meter resolution canopy surface model was used to isolate only first returns within 2m of the canopy surface. The returns in this sample were then compared to the LIDAR-derived ground model to eliminate all returns within 2m of the ground surface. The goal was to eliminate returns from understory vegetation, shrubs, and grasses that were not represented in field plot measurements. Finally descriptive metrics for the intensity values were computed using the remaining returns.

Height metrics were computed using the plot samples from the original return data. Return heights were computed by subtracting the return elevation from an elevation interpolated from the bare ground surface models using the XY location of the return. Metrics were computed using all returns above 2m. In addition to the standard set of metrics output by the FUSION GridMetrics program (McGaughey 2012), we also computed relative height percentile values by dividing height percentile values by the 95th percentile value. The relative percentiles are not sensitive to the tree height on the plots and proved more useful in the classification process than the actual percentile values.

2.4 Classification

Metrics computed for the cells were used in the R statistical package (R Development Core Team 2010) along with the Rattle data mining GUI (Williams 2009) to conduct a supervised classification to identify conifer and hardwood vegetation and vegetation with foliage absent/present. We applied three classification methods: Random forest (Breiman 2001, Liaw and Wiener 2002), a simple decision tree approach (Breiman et al. 1984), and the adaptive boosting model (Friedman et al. 2000). The basic process used to build and apply the classification rules was:

- Conduct a Principle Components Analysis (PCA) to identify the components that explain the majority of the variation in the cell data,
- Evaluate correlations between components and LIDAR-derived metrics and select metrics most highly correlated with the components for use in building the classification rules,
- Conduct a classification using 70 percent of the cells to produce a set of 500 decision trees,
- Use the decision trees and data for the remaining 30 percent of the cells to evaluate the performance of the classifier,
- Apply the classification rules to cells for all plots at 2m cell resolution.
- Compute the proportion of each plot in each of the vegetation conditions at 20m cell resolution.

3. Results

3.1 Intensity adjustment

The adjusted intensity data were compared to the original data to evaluate the effectiveness of the intensity correction. An F-test indicated the original sample variances were not equal ($p = 1.98 * 10^{-32}$) and the variance of adjusted values were equal ($p = 0.36$) at the 5 percent level. The results of a t-test indicated the mean values of the samples were not equal in both the original data and the adjusted data. To further correct values for sensor 46, we applied a simple bias correction using the difference between the mean values for all samples. Table 2 shows the summary statistics for the original and adjusted intensity values and the right image in Figure 1 shows an image generated using the adjusted intensity for first returns.

Table 2. Summary statistics for first returns after intensity adjustment in 4,419 5- by 5-meter samples.

	Intensity values			
	Sensor 77	Sensor 46	Sensor 46 (after histogram matching)	Sensor 46 (after bias correction)
Mean	68	97	70	68
Standard deviation	35	41	35	35
Minimum	5	15	4	2
Maximum	217	249	210	207

3.2 Classification

PCA led us to select four variables to build the H/C classification rules (Table 3) and four variables to build the FA/FP classification (Table 4). Overall, the four variables in each set were not highly correlated and they provided values that describe the return intensity and the shape of the return height distribution.

Classification tools used 70% of the data to build rules and the remaining 30% to evaluate classification error. For the H/C classification, the overall error for the random forest method was 1.8%. For the FA/FP classification, overall error was 3.7%. The adaptive boosting model produced errors of 5.5% for the H/C and 5.8% for the FA/FP classifications. The simple decision tree produced errors of 10.8% for the H/C and 8.7% for the FA/FP classifications.

Table 3. Variables used in classification rules to differentiate hardwood and conifer species.

Variable	Description
Int.Mean	Mean intensity value for first returns close to the canopy surface (± 2m from canopy surface and height above ground greater than 2m)
Elev.RP30	30^{th} percentile (height) for all returns above 2m divided by the 95^{th} percentile for all returns above 2m
Percentage.all.returns.above.mean	Proportion of all returns above the mean height
Int.L3	Third L-moment for adjusted intensity values for first returns close to the canopy surface

Table 4. Variables used in classification rules to differentiate vegetation with and without foliage.

Variable	Description
Int.Mean	Mean intensity value for first returns close to the canopy surface (± 2m from canopy surface and height above ground greater than 2m)
Elev.RP40	40^{th} percentile for all returns above 2m divided by the 95^{th} percentile for all returns above 2m
Percentage.first.returns.above.mean	Proportion of first returns above the mean
Int.L.CV	L-moment coefficient of variation for adjusted intensity values for first returns close to the canopy surface

The classification rules were applied to the 2- by 2-m grid cells for all plots and the results summarized to produce "hardwood fraction" (HF) and "foliage absent fraction" (FAF) metrics. We compared these proportions to those computed using the basal area measured on field plots. Figure 3 compares the summarized cell classifications for the H/S classification and Figure 4 compares those for the FA/FP classification. In Figure 3 plots are separated based on the proportion of the basal area for hardwood species that retain their foliage through the winter. The red markers represent plots where more than five percent of the basal area is from species that retain their foliage and the green markers represent plots with less than five percent.

4. Discussion

A direct comparison of the H/S and FA/FP proportions is difficult since the plot data do not reflect the horizontal and vertical arrangement of trees and thus the proportion of a type visible from an aerial viewpoint may be different from the proportion measured on the ground. In reality, many of the hardwood species are shade tolerant and develop below a pine overstory. In many stands, the majority of the crowns associated with hardwood species would not be within 2m of the upper canopy surface and so returns from hardwood vegetation would not appear in the point-cloud data used to compute the intensity metrics.

For our data, the intensity values were not normalized to account for pulse strength, range, sensor gain, or atmospheric effects. While intensity normalization procedures have been described (Korpela 2008; Korpela et al. 2010), the process is not straight forward for the Leica ALS50-II sensors. This scanner provides automatic gain correction (AGC) to dynamically adjust the sensitivity of the detection circuitry to compensate for more or less reflective targets.

The gain settings and outgoing pulse strength were not available for use in a correction process so we adopted a more simplistic solution. The topography at SRS is generally mild with elevations ranging from 18 to 140 m. Flying height during the acquisition was consistent so the only range effects were due to the scan angle (±10 degrees as flown). Possible explanations for the different intensity values between the two sensors (Figure 2) include differences in laser power, detector sensitivity, manufacturer calibration, and hardware defects. Both instruments were configured identically and flown at the same height and during the same time period. Given the information available to us, we can offer no definitive explanation for the differences.

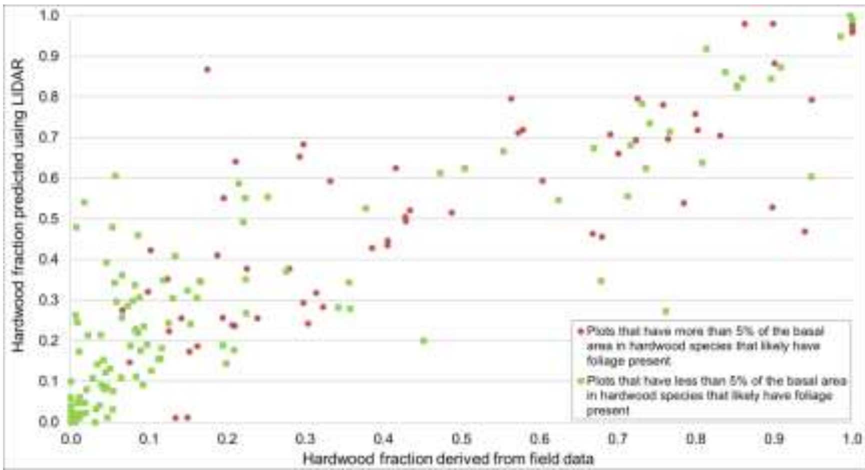

Figure 3. Scatterplot comparing hardwood/conifer fraction derived from field and LIDAR data.

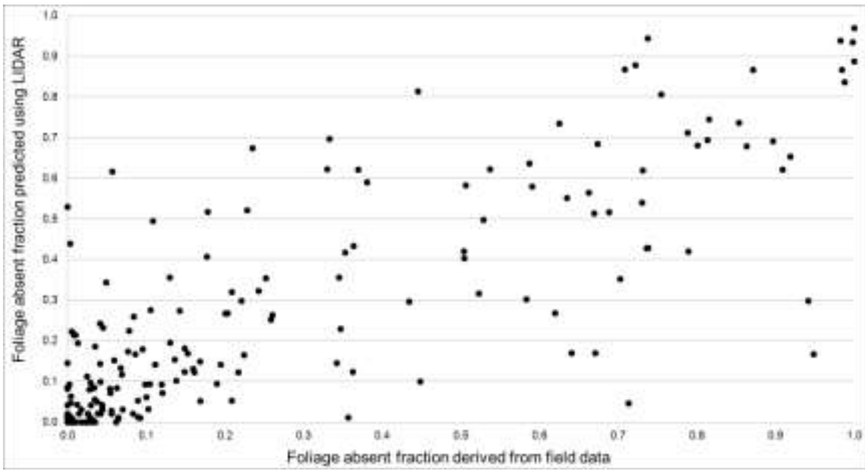

Figure 4. Scatterplot comparing foliage absent/present fraction derived from field and LIDAR data.

5. Conclusions

The results from our classification, proportions of hardwood/softwood and canopy with/without foliage, were used as predictor variables when modeling forest inventory parameters for SRS. The proportion of canopy with/without foliage was found to be a significant predictor variable for most models that predicted values for softwoods and hardwoods. Having the proportion available for use in regression modeling significantly improved the final models.

In general, our approach of building the classification rules using high-resolution data (2- by 2-meter pixels) and then summarizing the results at lower-resolution (20- by 20-meter pixels) produced results useful to managers. When we compare the results at both resolutions to orthorectified aerial photos acquired in leaf-off conditions during the same year the LIDAR data were acquired, we see strong agreement (Figure 6). A similar approach would be useful for classifying live and dead trees in pure conifer types.

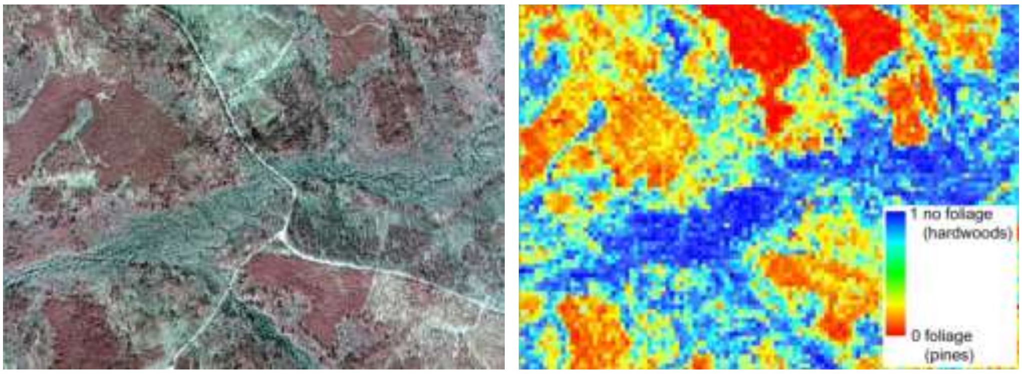

Figure 6. 2007 color infrared aerial photograph (left) and 2009 LIDAR proportion of canopy with/without foliage at 20m resolution (right).

Acknowledgements

Funding and support for this project was provided by the Department of Energy-Savannah River Operations Office through the U.S. Forest Service-Savannah River under Interagency Agreement DE-AI09-00SR22188. The authors would also like to thank the anonymous reviewer who made several suggestions that improved the paper.

References

Brandtberg, T. (2007). Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5): 325–340.

Breiman, L., Friedman. J.H., Olshen, R.A., Stone, C.J. 1984. Classification and regression trees. Belmont, CA: Wadsworth.

Breiman, L. 2001. Random forests. *Machine Learning*. 45(1):5-32.

Brennan, R. and Webster, T.L. (2006). Object-oriented land cover classification of lidar-derived surfaces. *Canadian Journal of Remote Sensing,* 32(2):162-172.

Friedman, J., Hastie, T., and Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*. 28(2): 337-374.

Gonzalez, Rafael C., Woods, Richard E. 2007. Digital Image Processing.Third Edition. Upper Saddle River, New Jersey: Prentice-Hall.

Korpela, I., 2008. Mapping of understory lichens with airborne discrete-return LiDAR data. *Remote Sensing of Environment*. 112(10):3891-3897.

Korpela, I., Orka, H.O., Hyyppa, J., Heikkinen, V., Tokola, T. 2010. Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*. 64(4):369-379.

Liaw, A.; Wiener, M. 2002. Classification and regression by random forest. *R News*. 2(3):18-22.

McGaughey, R.J. 2012. FUSION/LDV: Software for LIDAR data analysis and visualization. Version 3.10. May 2012. http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria. (http://www.R-project.org).

Song, J.-H., Han, S.-H., Yu, K., and Kim, Y.-I. (2002). Assessing the possibility of land-cover classification using LIDAR intensity data, *International Archives of Photogrammetry and Remote Sensing,* Graz, Austria, 2002, Vol. XXXIV, Part 3B, pp. 259-262.

Williams, G.J. 2009. Rattle: A data mining GUI for R. *The R Journal*. 1(2):45-55.