Identifying Eastern Hemlock (*Tsuga canadensis*) Patches Using LiDAR, in New Brunswick Forests

by

Hunter Roberts

B.Sc. Environmental Biology, University of Guelph, 2012

A Report Submitted in Partial Fulfillment of
the Requirements for the Degree of

Masters of Forestry

in the Graduate Academic Unit of Forestry & Environmental Management

Supervisor(s): John A. Kershaw, Jr., PhD, FOREM

Examining Board: Charles Borque, PhD, FOREM, Chair
Dr. Mike Lavigne, PhD, FOREM (HRA)
Prof. Jasen Golding, MFE, FOREM

This report is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

January, 2017

ABSTRACT

Aerial light detection and ranging (LiDAR) has become an increasingly popular method of determining forest stand characteristics. The objective of this study was to create a predictive model to spatially identify small forest patches dominated by eastern hemlock (*Tsuga canadensis*), using leaf off LiDAR data. Specifically, 46 LiDAR attributes were evaluated for their capability to differentiate hemlock from 8 other patch types using randomForest regression analysis in New Brunswick, Canada. Two final classification models were created both achieving 92% hemlock classification accuracy, but had differing capabilities in classifying hemlock subclass types. Attributes used in the models were found to be strongly related to the canopy structure of hemlock-dominated patches. It is not believed that model accuracy could be improved using leaf on data; however new models could be designed using a similar method and larger sample of Acadian forest stand types to be more general in classification.

ACKNOWLEDGEMENTS

I would like to thank Dr. John Kershaw for accepting me as a graduate student and providing guidance, support and encouragement to complete my Masters at UNB. I would also like to thank Dr. Mike Lavigne and Prof. Jasen Golding for their support and input on my project. Funding for this project was partially provided by grants from the National Science and Engineering Research Council Discovery Grants program and the New Brunswick Innovation Foundation, Research Assistants Initiative.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTs iii

TABLE OF CONTENTS iv

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER 1. INTRODUCTION 1

1.1 Eastern Hemlock 1

1.2 Hemlock Woolly Adelgid 4

1.3 LiDAR 6

1.4 RandomForest 8

1.5 Objectives 10

CHAPTER 2. METHODS 11

2.1 Study Site 11

2.2 Field Data Collection 12

2.2.1. Systematic Inventory Data 12

2.2.2. Ground Reconnaissance Data 12

2.3 LiDAR Data 12

2.4 Data Analyses 13

2.4.1 Stand Types 13

2.4.2. LiDAR Attributes 14

2.4.3. RandomForest Modeling 14

2.4.4. Field Verifying New Hemlock Sites 15

CHAPTER 3. RESULTS 16

3.1 Model Development 16

3.2 Model Evaluation and Refinement 16

3.3 Model Application 23

CHAPTER 4. DISCUSSION 32

4.1 Model Implication 32

4.2 LiDAR Attributes 32

4.3 Leaf off LiDAR Data 36

4.4 Hemlock Subclass Prediction 38

4.5 Classification Accuracy 39

4.6 Method Analysis 40

CHAPTER 5. CONCLUSIONS 42
REFERENCES 43
APPENDIX A 47
APPENDIX B 49
APPENDIX C 50
Curriculum Vitae

LIST OF TABLES

TABLE 1. FOREST PATCH CLASSIFICATION TYPES 13
TABLE 2. FOREST PATCH CLASSIFICATION CONFUSION MATRIX FOR MODEL T09 17
TABLE 3. FOREST PATCH CLASSIFICATION CONFUSION MATRIX FOR MODEL T10 17
TABLE 4. FOREST PATCH CLASSIFICATION CONFUSION MATRIX FOR MODEL T15 18
TABLE 5. TOTAL CLASSIFIED HEMLOCK PATCHES AND ACCURACY RATINS FOR T MODELS 18
TABLE 6. NUMBER OF TRAINING PATCHES USED FOR CLASSIFICATION IN EACH MODEL 19
TABLE 7. ADDITIONAL FOREST PATCH CLASSIFICATION TYPES USED IN F MODELS 20
TABLE 8. FOREST PATCH CLASSIFICATION CONFUSION MATRIX FOR MODEL F1 21
TABLE 9. FOREST PATCH CLASSIFICATION CONFUSION MATRIX FOR MODEL F2 21
TABLE 10. TOTAL CLASSIFIED HEMLOCK PATCHES AND ACCURACY RATINGS FOR ALL MODELS . 22
TABLE 11. HEMLOCK PATCH CLASSIFICATION RESULTS FOR F1 AND F2 MODELS 23
TABLE 12. AVERAGE ATTRIBUTE VALUES OF THE ADJUSTED BASE TRAINING DATA FOR EACH CLASS TYPE 33

LIST OF FIGURES

FIGURE 1. GEOGRAPHIC RANGE MAP OF EASTERN HEMLOCK 1
FIGURE 2. PHOTOS OF UNDERSTORY STRUCTURE OF HEMLOCK STANDS 3
FIGURE 3. GOGRAPHIC RANGE MAP OF HEMLOCK WOOLLY ADELGID 5
FIGURE 4. LIDAR POINT CLOUD OF BLOCK OF NOONAN RESEARCH FOREST 7
FIGURE 5. AIR PHOTO INTERPRETED STAND MAP OF NOONAN RESEARCH FOREST 24
FIGURE 6. NOONAN RESEARCH FOREST MAP OVERLAID WITH MODEL T09 PREDICTED CLASS TYPE 25
FIGURE 7. NOONAN RESEARCH FOREST MAP OVERLAID WITH MODEL T10 PREDICTED CLASS TYPE 26
FIGURE 8. NOONAN RESEARCH FOREST MAP OVERLAID WITH MODEL T15 PREDICTED CLASS TYPE 27
FIGURE 9. NOONAN RESEARCH FOREST MAP OVERLAID WITH MODEL F2 PREDICTED CLASS TYPE 28
FIGURE 10. NOONAN RESEARCH FOREST MAP OVERLAID WITH MODEL F2 PREDICTED CLASS TYPE 29

Chapter 1. Introduction

1.1 Eastern Hemlock

Eastern Hemlock (*Tsuga canadensis*) is a coniferous tree species of the *Pinaceae* family native to eastern Canada and the United States. It is marginally slow growing, reaching up to 30m in height, and is very long lived, maturing at 200 to 300 years of age, and living for up to 800 years (Godman and Lancaster, 1990). Hemlock is predominantly found in cool and humid climates, on moist and well drained soils (Ashbel, 1960). In Canada, hemlock occurs in central and southeastern Ontario, the southernmost portion of Quebec and throughout the Maritime Provinces. In the United States, it occurs in northeastern Wisconsin and Michigan, throughout the New England states, in New York, New Jersey and Pennsylvania, south to the northern tip of Georgia and Alabama along the higher Appalachian Mountain ridges (Godman and Lancaster, 1990) (Figure 1). Remnant stands can be found in Indiana and Illinois in steep, narrow, north-facing ravines.

Figure 1. Geographic range map of eastern hemlock (*Tsuga canadensis*) (http://gec.cr.usgs.gov/data/little/tsugcana.pdf).

Hemlock is a dominant and minor species component in many different stand types throughout its range, regularly occurring in tolerant hardwood stands, conifer stands, mixedwood stands and pure stands. It is commonly found in association with sugar maple (*Acer saccharum*), yellow birch (*Betula alleghaniensis*), American

beech (Fagus *grandifolia*) and white pine (*Pinus strobus*) (Godman and Lancaster, 1990).

In stands in which it dominates, hemlock has been labelled a foundation species due to its prominent role in regulating ecosystem processes and the distinct habitats it creates for a unique assemblage of fauna (Ellison *et al.* 2005; Martin and Goebel, 2013). Hemlock stands are characterized by dense shade, with little or no vegetation in the understory and a moist, acidic duff layer (Jenkins *et al.* 1999) (Figure 2). This environment results in slow litter decomposition rates, reduced nitrogen cycling, nutrient poor soils (Jenkins *et al.* 1999), and ultimately reduced ecosystem productivity. Although this environment is of lower productivity, it creates a unique habitat for many different species, and creates heterogeneity in hardwood dominated landscapes (Martin and Goebel, 2013). The dense canopy provides important habitat for many wildlife species, in the form of thick over-story coverage for winter shelter (Ellison *et al.* 2005; Tingley *et al.* 2002), and has also been found to minimize seasonal fluctuations in both stream temperature and hydrological regimes (Snyder *et al.* 2002). These static seasonal conditions have been linked with unique benthic communities and a number of taxa which have been found to have close associations with hemlock dominated stands (Snyder *et al.* 2002). Hemlock regeneration also provides an important source of deer browse in harsh winter conditions (Mladenoff and Stearns, 1993).

Reproducing primarily through seed, hemlock seedlings regenerate mainly in shaded areas where moisture levels are constant; an environment produced by mature hemlock dominated stands. It is an extremely shade tolerant species and can persist in the forest understory for many years (Godman and Lancaster, 1990). Mladenoff and Stearns (1993) attributed successful hemlock regeneration to multiple contributing factors including climate conditions, disturbance patterns,

anthropogenic land use history, and life-history characteristics. Crediting reduced occurrence of *T. canadensis* regeneration largely to the reduction of the species within its natural range, thus a reduction in seed sources for reproduction.

Figure 2. Photo of understory structure of mature mixed softwood and hemlock stand (left) and young, hemlock dominated stand (right). Photos taken by Hunter Roberts, 2016.

Formerly used as a source of tannin in the leather industry, hemlock is currently used primarily as dimensional lumber and pulp (McWiliams and Schmidt, 2000). However, due to poor lumber quality, hemlock has much less market capability than other softwoods within its range; notably eastern white pine, balsam fir, and all eastern spruce species (Howard *et al.* 2000). Hemlock lumber has been described as having moderate strength, uneven texture, and to commonly crack and splinter when worked with tools (Alden, 1997). The largest deterrent in utilizing hemlock comes from cracking or splitting between annual rings known as ring shake, which renders the wood inadequate as structural lumber (Baumgras *et al.* 2000). In the New Brunswick forest industry, there is a limited market for hemlock products and trees are often left standing during

harvesting operations. However, due to its high abundance, hemlock is a more significant softwood lumber species in the New England states for both pulp and sawlogs (McWiliams and Schmidt, 2000).

Prior to European settlement, hemlock was a dominant species of the Acadian and Great Lakes St. Lawrence forest regions and along the eastern coast of North America. However, because of land clearing and logging practices used by European settlers over the last 300 years, hemlock populations have been significantly reduced in the northern and western portions of its range (Rogers, 1978). Notably, in northern Michigan and Wisconsin, hemlock is considered to occur over less than 0.5% of its pre-colonial settlement abundance (Mladenoff, 1995). More recently, hemlock populations have faced immense and expanding population declines in the southern portion of its range due to the spread of Hemlock Woolly Adelgid.

1.2 Hemlock Woolly Adelgid

Hemlock Woolly Adelgid (HWA) (*Adelges tsugae*) is a non-native forest pest currently decimating hemlock populations in eastern North America (Havill *et al.* 2014). Native to Japan and China, HWA was first discovered in North America in Richmond Virginia in 1951 and currently occupies nearly half of the native range of eastern hemlock in the U.S. (Canadian Forest Service, 2014) (Figure 3).

HWA feeds on two species of hemlock native to eastern North America, eastern hemlock (*Tsuga canadensis*) and Carolina hemlock (*T. caroliniana*), and both species exhibit high mortality rates after infestation (Orwig *et al.* 2002). Preferring to feed on new foliage, HWA diminishes the hosts' stored nutrients and reduces its ability to grow, causing needle and branch dieback (McClure, 1991). Repeated annual attacks eventually decrease tree vigor and causes mortality in 4-15 years, depending on the degree of infestation and other factors (Havill *et al.* 2014).

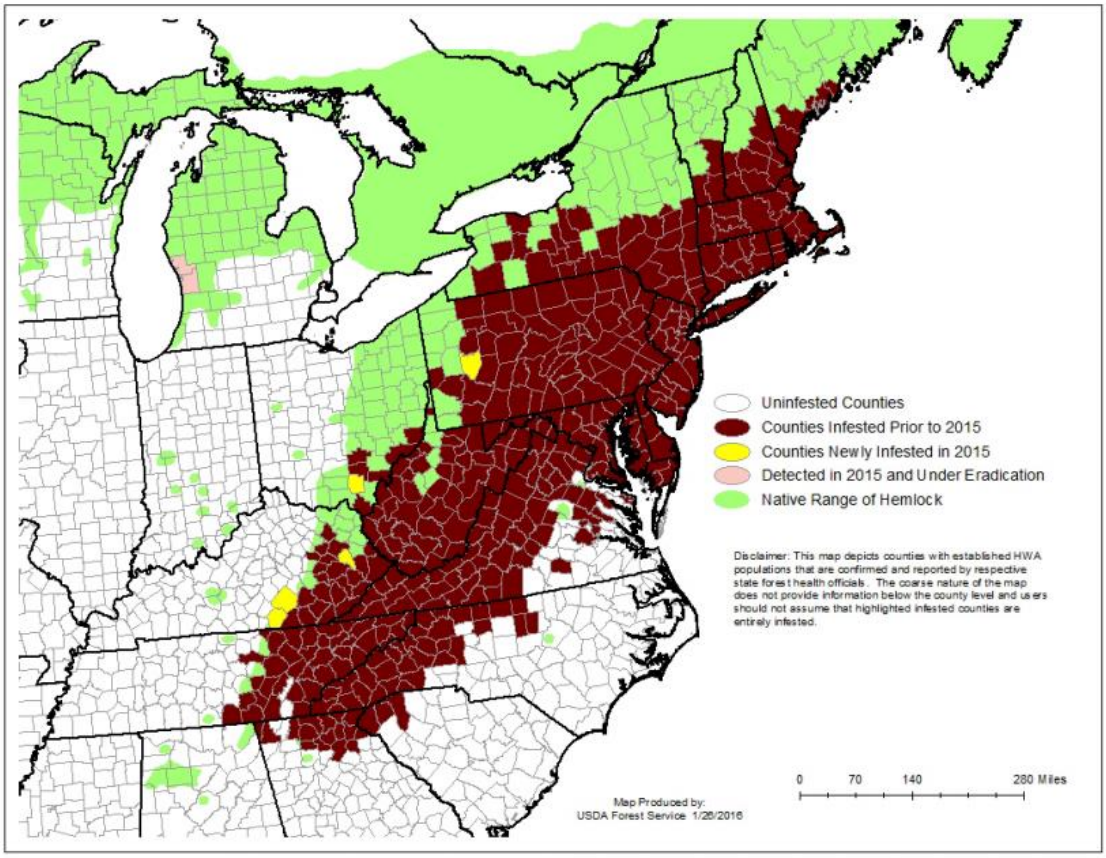

Figure 3. Geographical range of hemlock (*Tsuga canadensis*) overlaid with the 2015 range of Hemlock Woolly Adelgid (*Adelges tsugae*) (http://na.fs.fed.us/fhp/hwa/maps/2015_HWA_Infestation_Map_20160502.pdf).

HWA has been found to feed on trees of all size and age classes (Elliot and Vose, 2011). Mortality occurs first in hemlock saplings because they often are already stressed due to low light conditions (Preisser, 2011). In mature hemlock, infestation occurs in the upper to mid crowns, resulting in foliage loss and stunted growth (Canadian Forest Service, 2014). After repeated attacks, reduction of the characteristic thick canopy produced by hemlock allows light to penetrate to the forest floor, resulting in the establishment of deciduous tree species (Orwig and Foster, 1998). Hemlock stands have a poor ability to regenerate after attacks, leading to successive replacement of hemlock, and a major shift in species composition from coniferous to deciduous species. This leads to a loss of the unique conditions created by hemlock and to a more homogeneous forest landscape (Evans, 2002; Orwig and Foster, 1998; Sullivan and Ellison, 2006).

The extensive range and spread of HWA in North America has been attributed to its ability to reproduce asexually through parthenogenesis, meaning a single insect can start a colony (Fitzpatrick *et al.* 2012), as well as its ability to produce two generations per year, allowing HWA to colonize at a fast rate (Paradis *et al.* 2008). Currently HWA ranges from North Carolina and Tennessee, to the southern portions of Vermont, New Hampshire, Maine and New York, with two documented incidents in Southern Ontario (Canadian Forest Service, 2014) (Figure 3).

The northern range of HWA has been speculated to be limited by cold winter temperatures. However, although it has been found that there are significant reductions in HWA populations during prolonged winter temperatures below -15°C, mortality is not always 100 percent. This indicates that cold temperatures may only act to slow HWA's colonization in northern climates and not prevent it (Skinner, 2003). Additionally, warming winter temperatures due to climate change (Paradis *et al.* 2008), as well as the potential for HWA to adapt to colder winter temperatures (Havill *et al.* 2014) present a potential threat to hemlock populations previously thought protected by cold climate conditions.

Because of the devastating impact HWA has had on the hemlock population in the eastern United States, eastern Canada may become a refugium for hemlock as a species in the near future. Regardless, whether HWA is able to expand its populations into the northern range of hemlock or not, identification and mapping of hemlock dominated patches and stands across its native range may play an important role in both the preservation of hemlock as a species, as well as the unique ecosystem functions which the species provides.

1.3 LiDAR

An increasingly popular method of remote sensing used to estimate stand and forest characteristics is aerial Light Detection and Ranging (LiDAR). An aerial LiDAR scanner emits near-

infrared laser points downward from an aerial position (i.e. plane), and, by measuring the angle of emission, time to return, and aircraft movement, estimates locations of objects that reflect the laser beams. A LiDAR scanner is composed of a laser emission unit, a laser receiver, a global positioning system (GPS), and an inertial measurement system (IMU) to determine the position of the laser scanner in the air, and a computer for data processing and storage. The coordinate for each laser reflection is determined using a time measurement from when the laser point is emitted, reflected off an object's surface, and returned to the source. The time measurement is then calculated into a unit of distance from the scanner. The pulse arrangement depends on both scanner set up (i.e. pulse density and frequency) and flying parameters (i.e. scanner height and flight speed). The x, y, z location of the point is then determined and the combined points form a "point cloud" creating a 3D map of terrain and vegetation (Figure 4). This "biospatial data" provides the location of specific attributes rather than a generalized inventory (Reutebuch *et al.*, 2005). In forestry, typically the first point returns are interpreted to be the vegetation layer and the last returns are viewed as the ground layer (Wulder *et al.*, 2008). The elevations of the laser returns can be converted to heights of the vegetation by subtracting the elevation of the ground layer from the elevation of the LiDAR returns.

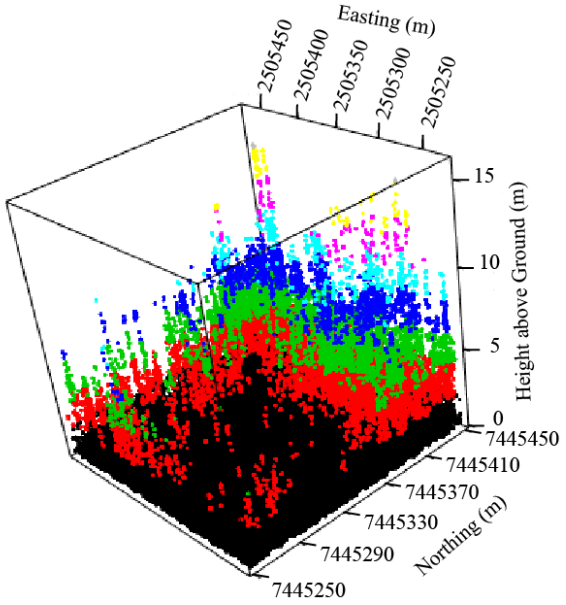

Figure 4. LiDAR point cloud for a 40,000 m^2 block of Noonan Research Forest.

Often in forest canopies, individual laser pulses become fragmented when a portion of a pulse is intercepted by an object, while the remaining portion continues to intercept one

or more objects below. When fragments of laser pulses reflect back to the scanner this generates what is termed “multiple returns.” To differentially target conifers, hardwoods, or ground terrain, LiDAR data can be obtained during leaf on or off seasons for hardwoods; leaf off producing fewer multiple returns and a larger amount of ground returns (Reutebuch *et al.*, 2005).

The strength of each return point, determined by the reflectivity of the surface from which the point is reflected, is known as the intensity value. In forestry applications, intensity value is strongly related to the reflective properties of the foliage, as well as foliage arrangement, and may be used to differentiate species (Kim, *et al.*, 2009).

Attributes of the LiDAR point cloud are extracted at desired geographic locations (usually associated with ground samples). For a point cloud of a given area, values for the mean, variance, and standard deviation can be extracted for each LiDAR attribute, such as intensity and point density, and can be averaged into different height quantiles within forest canopies to create a large number of LiDAR attributes. These extracted measures are then correlated with ground data using a variety of parametric and nonparametric statistical tools.

1.4 randomForest

One such tool that has gained considerable interest and application in analysis of the large data sets that result from LiDAR is the randomForest (RF) imputation algorithm designed by Breiman (2001). The RF procedure is used for multiple functions, predominantly classification and regression, and consists of an ensemble of machine learning methods in which multiple decision trees are produced and averaged to determine the best classification scheme. This method is extremely efficient in processing large datasets with many diverse variables and is resilient to outliers and noise. RF functions by building an assembly of decision trees from a training dataset. Each decision tree is built using a random independent subset, also known as a

'bootstrap' sample, of the training data to form a classification scheme. In traditional classification trees, nodes are split based on which factor among all the variables splits the data into two subsets that maximizes variance between the two subsets and minimizes the variance within each subset, and then each subset is further split until some criteria, either minimum variance or minimum set size, is reached (this process is call recursive partitioning). In random forest, each decision tree's nodes are split based on a random sample of the variables (Liaw and Wiener, 2002). This helps prevent overfitting of the training data. For classification, random forest then assumes the classification scheme that is the mode of all of decision trees. In this way, RF uses multiple weaker classifiers or learners and combines them to create a strong classifier/learner.

RF also has useful features for estimating a variable's importance and correlation. These values can be used for many purposes. In the case of large data sets with many variables, such as those generated by LiDAR, variable importance and correlation measures can be used to eliminate irrelevant variables which do not influence classification accuracy; decreasing dataset size and processing time.

A number of studies have examined classifying tree species type from LiDAR data. Often these studies combine LiDAR with other media, including multispectral data (Leckie *et al.,* 2003; Holmgren *et al.,* 2008) and aerial photographs (Heinzel *et al.*, 2008). As well, many studies have focused on the classification of individual trees, in which case LiDAR points for individual tree crowns also must be delineated (Heinzel *et al.*, 2008). This study adopts a new approach to tree species type classification using LiDAR data by classifying species types based on LiDAR characteristics from 10 m by 10m cells.

1.5 Objectives

The objective of this study was to create a predictive model to identify forest patches dominated by hemlock using LiDAR data. To achieve this, a ground reconnaissance was used to locate hemlock patches, within the matrix of the other common stand types within the Noonan Research Forest (NRF). Forest types were grouped into 9 types: softwood dominant; hardwood dominant; mixedwood; spruce dominant; hemlock; white pine; open water; bog areas; and non-forested areas. Training data were composed of 2M BAF plots collected using two different selection criteria: the first was a systematic inventory of 1480 plots collected on a permanent forest inventory grid at 100 m spacing; the second was targeted plots composed primarily of pure hemlock and mixtures of hemlock identified during the ground reconnaissance. Additional points in bogs, non-forested areas, and open water were selected from aerial photos. LiDAR data was extracted from a 10 m radius zone around the center of each point and a variety of LiDAR attributes extracted (Appendix A). The LiDAR data for these patch types were analyzed to identify important attributes for differentiating between the 9 stand types using randomForest regression analysis. Newly identified hemlock patches were then field verified to determine model accuracy.

Chapter 2. Methods

2.1 Study Site

The study site was the 1500 ha University of New Brunswick's Noonan Research Forest (NRF) in New Brunswick, Canada (45° 59' 12" N, 66° 25' 15" W). The site is in the Acadian Forest region of North America (Loo and Ives 2003), and is composed of a variety of softwood, hardwood, and mixedwood stand types with 271 stands total. Stands vary in size from 0.5 to 47 ha and range from pure species, single cohort stands to mixed species, multi-cohort stands. The site has been managed by the University of New Brunswick since 1985. Primary species include black and red spruce, balsam fir, large-toothed and trembling aspen, white and yellow birch, red and sugar maple and white ash (Appendix B).

Hemlock is a relatively minor species in NRF and is the dominant species in three different patch/stand types: 1) pure hemlock; 2) spruce hemlock mix and; 3) old mixed hemlock. Pure hemlock stands/patches are dominated by relatively even aged hemlock and occurs in a single stand, 3.5 ha in size, and in various smaller patches 0.05 to 0.25 ha in size. The spruce hemlock mix consists of patches of mature spruce over-story, with a hemlock dominated understory and a minor hardwood component typically consisting of white birch and red maple, occurring in patches 0.05 to 1 ha in size. The old mixed hemlock patches consist of large mature trees, dominated by hemlock with minor components of spruce, white pine and white cedar; ranging in size from 0.05 to 0.25 ha in size.

2.2 Field Data Collection

2.2.1 Systematic Inventory Data

During the summer of 2012, 1410 sample points were established in the NRF located at the intersections of a permanent 100m x 100m sample grid. All points were measured using a 2 m^2/ha basal area factor (BAF; 2 m^2 ha^{-1} $tree^{-1}$ tallied) angle gauge, recording all trees with a DBH (diameter at breast height, 1.3 m above ground) greater than 6cm. Species and diameter (nearest 0.1cm) were recorded for all "in" trees. Plots were initially classified based on species into 8 different forest types (Table 1).

2.2.2 Ground Reconnaissance Data

In the fall of 2014, 50 hemlock dominated clumps were identified by walking the NRF in parallel transects. Hemlock clumps were classified using a 2M BAF angle gauge, recording all trees with a DBH greater than 6cm. The center of each clump was marked using a GPS waypoint.

2.3 LiDAR Data

A LiDAR scan was obtained on October 21 and 22, 2011 using an airborne Riegl LMS Q680i scanner mounted on a plane, at a density of 48 returns per m^2, in leaf off conditions. The mean flying altitude above sea level was approximately 724m. The sensor generated the pulse repetition frequency of 150 KHz, and the laser wavelength was 1550 nm with a scan angle of <28.5° from the nadir. The mean pulse density was 3 pulses m^2 per swath (this was flown at 50% overlap, providing a final density of 6 pulses m^2) with a footprint size of 0.35 m^2 and the sensor collected up to 8 returns per pulse. For this analysis, the full waveform LiDAR data were converted to discrete returns with an overall mean point density of 6.2 ± 1.3 points m^2 (± standard deviation).

2.4 Data Analyses

2.4.1 Stand Types

Using the systematic inventory and ground reconnaissance data, forest patches surrounding the center of each BAF sample point were classified as shown in Table 1. Patches were classified into 5 forest types and 3 non-forest types (Table 1). Hemlock patch type was defined as patches with a Hemlock composition of 50% or greater; these plots were identified during the 2014 ground reconnaissance. Spruce, Softwood, Hardwood and Mixedwood patch types were classified based on species composition using the 2012 inventory data and the classification scheme listed in Table 1. Areas with 5 trees per plot or less were classified as Non-Forest, and Water and Wetland areas were visually determined through interpretation of aerial photos of the study site.

Table 1. Forest patch classification, based on species composition.

Classification	Composition	Species[1]
Hemlock	Hemlock > 50%	eastern hemlock
Spruce	Spruce Species > 70%	black and red spruce
Softwood	Softwood Species > 70%	black spruce, red spruce, balsam fir, tamarack, eastern white cedar, white pine
Hardwood	Hardwood Species > 70%	American elm, alder, large-toothed aspen, gray birch, ironwood, mountain ash, paper birch, trembling aspen, red maple, sugar maple, stripped maple, white ash, yellow birch
Mixedwood	>5 trees/plot, < 70% of either hardwood or softwood species	Combination of softwood and hardwood species
Bog	bogs and grasslands with scattered trees	Delineated using map
Water	No trees	Delineated using map
Non-Forest	No or scattered overstory trees (dbh $\underline{\geq}$ 6 cm) or no BAF plot established	Roads, recent clearcuts, young precommercially thinned stands, and other area with few trees

[1]Appendix B contains common and Latin names for all species found in Noonan Research Forest.

2.4.2 LiDAR Attributes

The LiDAR data were read into the R statistical package from LAS-format files. GPS coordinates of field sample locations were used to extract 46 LiDAR attributes from 10 m radius plots centered on each patch location (Appendix A). A study by Hayashi *et al*. (2015) found that the 10 m LiDAR plot radius produced the best fit for both nonlinear mixed effects models and RF models for predicting above ground biomass on the NRF. Five canopy surface attributes were calculated using a canopy surface model: maximum canopy height; mean canopy height; and the 25th, 50th, and 75th height percentiles of the canopy surface model. Thirteen LiDAR density attributes were extracted: density of LiDAR returns per m^3 total, and densities below 10m and 5m height above ground; height of maximum point cloud density; and point density at the 25^{th} - 95^{th} quantiles in 10 percentile increments. Nine vertical distribution attributes were extracted: mean height of the point cloud and the 25^{th} - 95^{th} quantiles in 10 percentile increments. Eighteen intensity attributes were extracted: mean intensity value; and mean intensity value at the 25^{th} - 95^{th} quantiles in 10 percentile increments; standard deviation of the mean intensity value; and standard deviation of the mean intensity value at the 25^{th} - 95^{th} quantiles in 10 percentile increments. The Weibull shape and scale parameters were estimated from the point cloud distribution.

2.4.3 randomForest Modeling

Using the randomForest package (Liaw, 2015) in R (R Development Core Team, 2015), the randomForest classification and regression function was used to create a classification model using the 46 LiDAR attributes as predictor variables and the forest type (Table 1) of the ground plots as the dependent class variable. Using variable importance measures, variables with the least influence were systematically removed from the training data one variable at a time. Each time a variable was removed the classification function was re-run to ensure classification error

values of the Hemlock patch type remained true to the original values from the full attribute set. This process was repeated until a minimum subset of attributes with the highest influence but minimizing the original Hemlock patch type classification error was obtained. The final classification model was then used to predict the NRF study site based on a wall-to-wall 10m by 10 grid.

Due to the randomization of randomForest classification, variable importance measures fluctuated each time the classification was run. Although the importance ranking always displayed a similar trend, variables close in importance measure value were found to fluctuate in order. For this reason, a number of training or T models were evaluated using different numbers of variables and different variable combinations, based on the assessment of multiple randomForest test runs.

2.4.4 Field Verifying New Hemlock Sites

Model accuracy was verified by extracting the GPS coordinates for the center of each cell newly identified as Hemlock, and confirming the species composition in the field.

Chapter 3. Results

3.1 Model Development

A series of initial models were screened using the training data described above (the systematic inventory data and ground reconnaissance data). Three potential models were identified using the base training data: T09, T10, and T15. These models were then used to make a wall-to-wall classification of 10m by 10m cells across the NRF. All cells classified as "Hemlock" were field surveyed for verification and additional field data were gathered on any new sites that were correctly classified as "Hemlock". The additional data and careful assessment of the initial training models (T09, T10, and T15) were then used to improve the original base training data and create two final models (F1 and F2). Appendix A shows the variables included in each of the 5 models.

Models T09, T10, T15, F1, and F2 used 13, 14, 17, 16, and 15 variables, respectively. Several variables were shared among models; however, there were also a number of variables unique to each model. As well, despite common variables shared amongst models, all models were found to have different variable ranking orders. Important variables found in all 5 models included q50CHT, meanCHT, LdensityB05, q25INSmean and q35INSmean (see Appendix A for attribute definitions). There were also additional intensity measures at various quantiles found amongst the models, 3 of which were present in both Final models (q75INSmean, q65INSmean and q55INSmean).

3.2 Model Evaluation and Refinement

The confusion matrix, which measures the goodness-of-classification, for each of the training models had similar results (Tables 1, 2 and 3). All models achieved out of bag (OOB) estimated error rates of approximately 37%. In each model, this value is reflected in classification errors for

all class types except Hemlock; classification errors ranged from 30 to 80%, excluding Hemlock which had less than 4% error rates for all three models.

While the models appear to fit the Hemlock training data accurately, there is some confusion with Softwood and Spruce types. This confusion became more apparent when wall-to-wall classification was applied to the NRF (see below for more detail).

Table 2. Forest patch classification confusion matrix for randomForest model T09 using 13 predictor attributes, with out of bag estimate error rate of 36.7%.

Observed Type[1]	Predicted Type[1]								Classification Error
	BOG	HARD	HEM	MIXED	NONF	SOFT	SPRUCE	WATER	
BOG	6	0	0	0	13	1	0	0	0.7000
HARD	0	100	0	65	13	12	1	0	0.4764
HEM	0	0	49	0	0	0	1	0	0.0200
MIXED	0	38	0	313	16	63	20	0	0.3044
NONF	4	10	0	34	173	29	19	5	0.3686
SOFT	0	2	1	74	22	173	38	0	0.4419
SPRUCE	0	0	4	24	12	39	171	0	0.3160
WATER	0	0	0	0	16	0	0	8	0.6667

[1]Patch classification type based on tree density and species composition: BOG = bogs and grasslands with scattered trees; HARD = hardwood species composition > 70%; HEM= hemlock composition > 50%; MIXED = >5 trees/plot and < 70% of either hardwood or softwood species; NONF = none or scattered over-story trees (DBH $\geq$ 6 cm); SOFT = softwood species composition> 70%; SPRUCE = spruce species composition > 70%; WATER = open water with no trees.

Table 3. Forest patch classification confusion matrix for randomForest model T10 using 14 predictor attributes, with out of bag estimate error rate of 36.8%.

Observed Type[1]	Predicted Type[1]								Classification Error
	BOG	HARD	HEM	MIXED	NONF	SOFT	SPRUCE	WATER	
BOG	5	0	0	0	14	1	0	0	0.7500
HARD	0	96	0	71	13	10	1	0	0.4974
HEM	0	0	50	0	0	0	0	0	0.0000
MIXED	0	44	0	308	14	60	24	0	0.3156
NONF	3	9	0	31	175	31	20	5	0.3613
SOFT	0	3	0	80	22	173	32	0	0.4419
SPRUCE	0	0	4	21	14	34	177	0	0.2920
WATER	0	0	0	0	16	0	0	8	0.6667

[1]Patch classification type based on tree density and species composition: BOG = bogs and grasslands with scattered trees; HARD = hardwood species composition > 70%; HEM= hemlock composition > 50%; MIXED = >5 trees/plot and < 70% of either hardwood or softwood species; NONF = none or scattered over-story trees (DBH $\geq$ 6 cm); SOFT = softwood species composition> 70%; SPRUCE = spruce species composition > 70%; WATER = open water with no trees

Table 4. Forest patch classification confusion matrix for randomForest model T15 using 17 predictor attributes, with out of bag estimate error rate of 37.4%.

Observed Type[1]	Predicted Type[1]								Classification Error
	BOG	HARD	HEM	MIXED	NONF	SOFT	SPRUCE	WATER	
BOG	4	0	0	0	14	1	0	1	0.8000
HARD	0	97	0	70	14	9	1	0	0.4921
HEM	0	0	48	0	0	1	1	0	0.0400
MIXED	0	45	0	301	16	63	25	0	0.3311
NONF	3	13	0	37	162	31	21	7	0.4088
SOFT	0	1	0	75	23	181	30	0	0.4161
SPRUCE	0	0	4	26	8	31	180	1	0.2800
WATER	0	0	0	0	15	0	0	9	0.6250

[1]Patch classification type based on tree density and species composition: BOG = bogs and grasslands with scattered trees; HARD = hardwood species composition > 70%; HEM= hemlock composition > 50%; MIXED = >5 trees/plot and < 70% of either hardwood or softwood species; NONF = none or scattered over-story trees (DBH $\geq$ 6 cm); SOFT = softwood species composition> 70%; SPRUCE = spruce species composition > 70%; WATER = open water with no trees.

The 3 classification models were used to make wall-to-wall classifications of the NRF. Across the 3 models, 223 Hemlock patches were identified (Table 5). All Hemlock patches were field verified in November 2016; 50 patches were verified as Hemlock, 131 as Spruce, 41 as Softwood and 1 as Mixedwood. All 3 models had relatively low classification accuracy (31.7% and lower) for Hemlock when applied to the NRF in the predictive stage. Of the 50 Hemlock cells, 16 coincided with the locations of Hemlock patches used in the base training data and 34 were newly identified patches.

Table 5. Total classified Hemlock patches and accuracy rating for models T09, T10 and T15.

	T09	T10	T15	Total
Classified HEM Patches	142	173	192	223
Correct Hem Patches	45	44	39	50
Classification Accuracy (%)	31.7	25.4	20.3	22.4

The 34 new Hemlock patches were then combined into the training data and 9 of the original patches were removed because of initial reconnaissance errors, resulting in 75 Hemlock training patches. The 131 new Spruce patches misclassified as Hemlock, were correctly classified and added to the training data as well. The 41 Softwood patches misclassified as Hemlock were

subdivided into 25 Softwood patches and 16 patches into a new classification type called White Pine, and added to the training data. The White Pine type was defined as plots with a species composition greater than 70% white pine (Table 6). The new revised training data were then used to create 2 final models, F1 and F2.

Table 6. Number of training patches used for each classification type for 5 randomForest models.

Abbreviation	Full Type Name	# of Training Patches		
		T09,T10,T15	F1	F2
PHEM	Pure Hemlock	-	39	-
SPHEM	Spruce Hemlock Mix	-	13	-
OMHEM	Old Mixed Hemlock	-	23	-
HEM	Hemlock	50	-	75
SPRUCE	Spruce	250	381	381
SOFT	Softwood	310	335	335
HARD	Hardwood	191	191	191
MIXED	Mixedwood	450	450	450
BOG	Wetlands	20	20	20
WATER	Open Water	24	24	24
NONF	Non-Forest	274	274	274
WP	White Pine	-	16	16

In the F1 model, the Hemlock classification was subdivided into three new classification types based on species composition and size class of Hemlock patches found in the NRF: Pure Hemlock, Spruce Hemlock Mix and Old Mixed Hemlock (Table 7). In the F2 model, all 75 Hemlock plots were lumped together in a single class. Pure Hemlock (PHEM) classification was defined as stands with 90% or greater Hemlock composition. Spruce Hemlock (SPHEM) classification was defined as stands with 50% hemlock composition and at least 30% spruce composition. Old Mixed Hemlock (OMHEM) was defined as stands with 50% Hemlock composition and at least 30% softwood species composition, with 90% of trees being 30cm or greater DBH.

Table 7. Additional forest patch type classifications used in models F1 and F2.

Classification	Composition	Species
Pure Hemlock	Hemlock > 90%	eastern hemlock
Spruce Hemlock Mix	Hemlock > 50% Spruce Species > 30%	eastern hemlock, black and red spruce
Old Mixed Hemlock	Hemlock > 50% Softwood Species > 30%	eastern hemlock, black and red spruce, balsam fir, eastern white cedar, white pine
White Pine	White Pine > 70%	white pine

Models F1 and F2 were pared down to 16 and 15 predictor variables (Appendix C). The confusion matrix generated for F1 and F2 had similar OOB estimate error rates to those of the training models (36 and 37%, respectively). As well, the models produced similar classification errors in all class types excluding Hemlock (Tables 8 and 9). The final models had significantly higher classification error rates for Hemlock than that of the training models. In the F1 model PHEM, SPHEM and OMHEM had high classification error rates of 23, 23 and 26%. Most of this classification error for all three class types was caused by the misclassification of spruce types. The F2 model had a Hemlock classification error of 16%, with most misclassified cells were spruce as well.

Table 8. Forest patch classification confusion matrix for randomForest model F1 using 16 predictor attributes, with out of bag estimate error rate of 37.2%.

Observed Type[1]	Predicted Type[1]											Classification Error
	BOG	HARD	MIXED	NONF	OMHEM	PHEM	SOFT	SPHEM	SPRUCE	WATER	WP	
BOG	5	0	0	12	0	0	1	0	0	2	0	0.7500
HARD	0	90	75	14	0	0	11	0	1	0	0	0.5288
MIXED	0	47	296	16	0	0	62	1	28	0	0	0.3422
NONF	3	14	30	165	0	0	32	0	21	9	0	0.3978
OMHEM	0	0	0	0	17	0	0	1	5	0	0	0.2609
PHEM	0	0	0	0	1	30	1	0	7	0	0	0.2308
SOFT	0	6	78	24	0	0	174	1	50	0	2	0.4806
SPHEM	0	0	0	0	0	1	0	10	2	0	0	0.2308
SPRUCE	0	0	27	11	0	2	36	0	305	1	1	0.2037
WATER	1	0	0	18	0	0	0	0	0	5	0	0.7917
WP	0	0	0	0	0	0	1	0	1	0	14	0.1250

[1]Patch classification type based on tree density and species composition: BOG – bogs and grasslands with scattered trees; HARD – hardwood species composition > 70%; MIXED – >5 trees/plot and < 70% of either hardwood or softwood species; NONF – none or scattered over-story trees (DBH $\geq$ 6 cm); OMHEM – Hemlock composition > 50% and softwood species composition > 30%; with 90% of trees > 30cm DBH.; PHEM – Hemlock composition > 90%; SOFT – softwood species composition > 70%; SPHEM – Hemlock > 50% spruce species > 30%; SPRUCE – spruce species composition > 70%; WATER – open water with no trees; WP – White Pine composition > 70%.

Table 9. Forest patch classification confusion matrix for randomForest model F2 using 15 predictor attributes, with out of bag estimate error rate of 36.1%.

Observed Type[1]	Predicted Type[1]									Classification Error
	BOG	HARD	HEM	MIXED	NONF	SOFT	SPRUCE	WATER	WP	
BOG	5	0	0	0	12	1	0	2	0	0.7500
HARD	0	93	0	72	13	12	1	0	0	0.5131
HEM	0	0	63	2	0	0	10	0	0	0.1600
MIXED	3	43	1	302	15	60	29	0	0	0.3289
NONF	0	13	0	30	167	30	24	7	0	0.3905
SOFT	0	3	2	75	22	176	55	0	2	0.4746
SPRUCE	1	0	2	25	12	40	303	0	1	0.2089
WATER	0	0	0	0	16	0	0	7	0	0.7083
WP	0	0	0	0	0	2	1	0	13	0.1875

[1]Patch classification type based on tree density and species composition: BOG = bogs and grasslands with scattered trees; HARD = hardwood species composition > 70%; HEM= hemlock composition > 50%; MIXED = >5 trees/plot and < 70% of either hardwood or softwood species; NONF = none or scattered over-story trees (DBH $\geq$ 6 cm); SOFT = softwood species composition > 70%; SPRUCE = spruce species composition > 70%; WATER = open water with no trees; WP = White Pine composition > 70%.

When used to build the wall-to-wall classification of the NRF study site, both models individually classified 70 Hemlock patches, with much overlap between identified patches for a combined total of 76 Hemlock patches. Of the 76 identified Hemlock patches, all 50 Hemlock patches previously identified by the T models were re-identified. The remaining 26 Hemlock patches were field verified in November 2016; of the 26 patches, 19 were Hemlock, 6 were Spruce and 1 patch was Softwood. Both F models identified a significantly larger number of correct Hemlock patches, at a much higher accuracy rate than the previous T models; both achieving accuracy ratings of 91.4% (Table 10) despite their training error rates being significantly greater than the T models (compare Tables 8 and 9 with Tables 2, 3, and 4).

Table 10. Total classified Hemlock patches and accuracy rating for all 5 models.

	T09	T10	T15	F1	F2
Identified HEM Patches	142	173	192	70	70
Correct Hem Patches	45	44	39	64	64
% Correct Hem Patches	31.7	25.4	20.3	91.4	91.4

Although both F models had the same prediction accuracy, they proved to have differing abilities to identify the 3 Hemlock subclasses. F1 was able to identify a larger number of PHEM patches in both the training patches and new patches, and F2 identified a larger number of new SPHEM patches (Table 11). However, although both models identified equal or greater numbers of patches in both the PHEM and SPHEM class types in comparison to the training data, both models predicted significantly less OMHEM patches than what was included in the training data.

Table 11. Hemlock patch classification results for F1 and F2 models. Comparing total base training patches used to build models to the number of training and new patches identified by each model under 3 Hemlock subclasses; OMHEM, PHEM and SPHEM.

	Base Training Patches	F1			F2		
		Training	New	Total	Training	New	Total
OMHEM	23	3	0	3	4	0	4
PHEM	39	34	10	44	32	7	39
SPHEM	13	12	5	17	12	9	21
Total	75	49	15	64	48	16	64

3.3 Model Application

A leaf on photo-based stand type map for the NRF was used as the base map for examining the overall classification accuracy of the models developed here (Figure 5). For each of the five models, the predicted class type of each 10 m by 10 m cell was overlaid on the map using a small dot in the center of each cell, shaded using the same color scheme as used in Figure 5 for the air photo types (Figures 6 - 10). Because of the rarity, Hemlock dots were magnified in size to increase visibility. It should be noted that, although a cell prediction may not agree with the air photo type, the model-based classification may not necessarily be wrong. The 10 m by 10 m cell sized used in the study is a much higher resolution that that which is used in photo-based stand interpretation. Because of this, the cell prediction may have identified species variations within stand polygons that were ignored by photo interpreters due to their minimum stand polygon requirements.

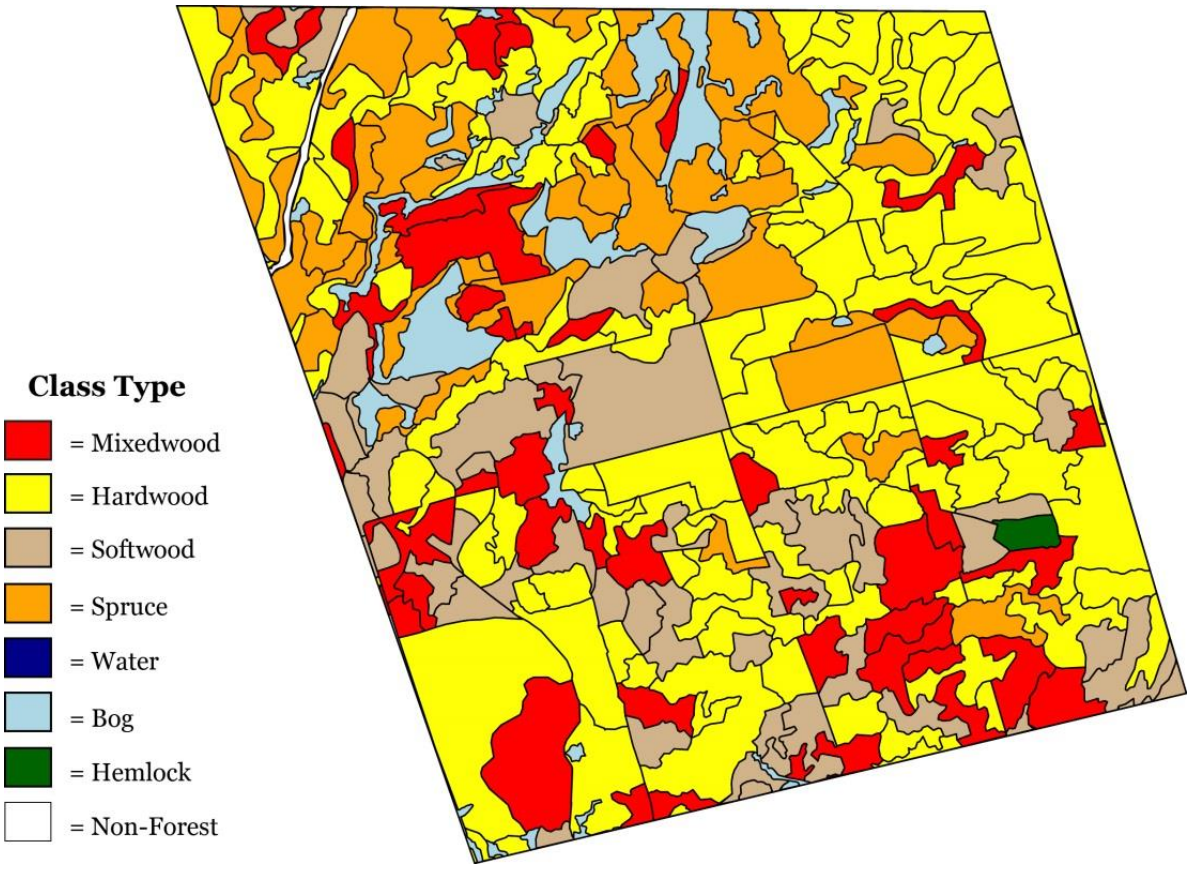

Figure 5. Air photo interpreted stand map of Noonan Research Forest.

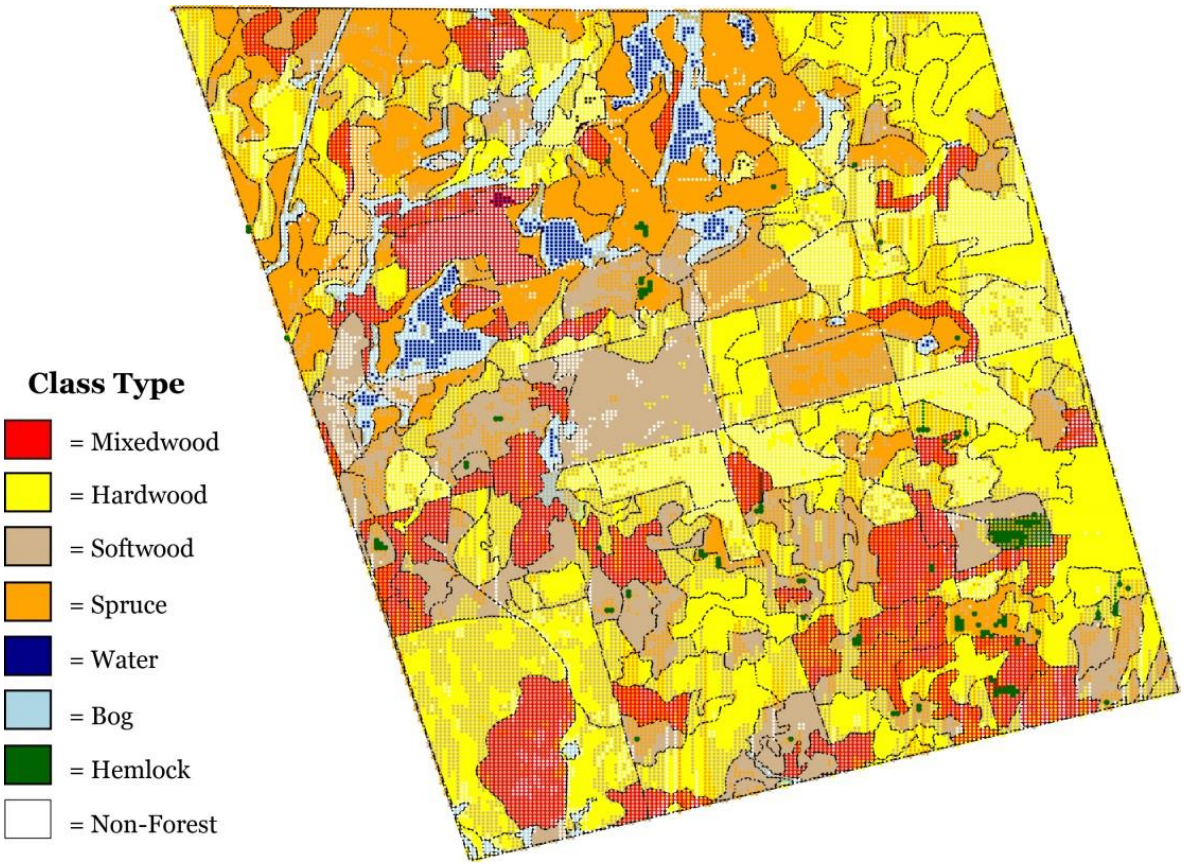

Figure 6. Air photo interpreted stand map of Noonan Research Forest overlaid with model T09 predicted class type of each 10 m by 10 m cell represented by a dot in the center of each cell. Each dot is shaded according to the predicted class type and Hemlock dots are magnified to x10 size to increase visibility.

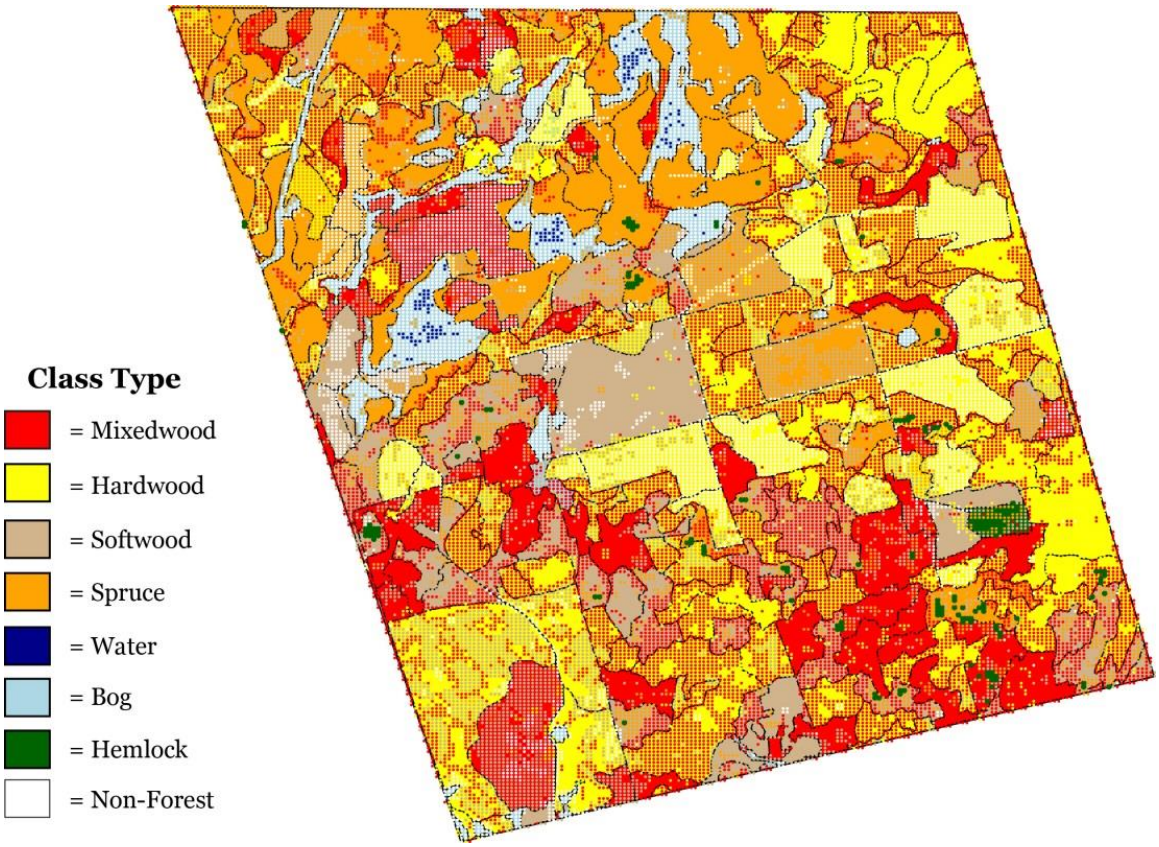

Figure 7. Air photo interpreted stand map of Noonan Research Forest overlaid with model T10 predicted class type of each 10 m by 10 m cell represented by a dot in the center of each cell. Each dot is shaded according to the predicted class type and Hemlock dots are magnified to x10 size to increase visibility.

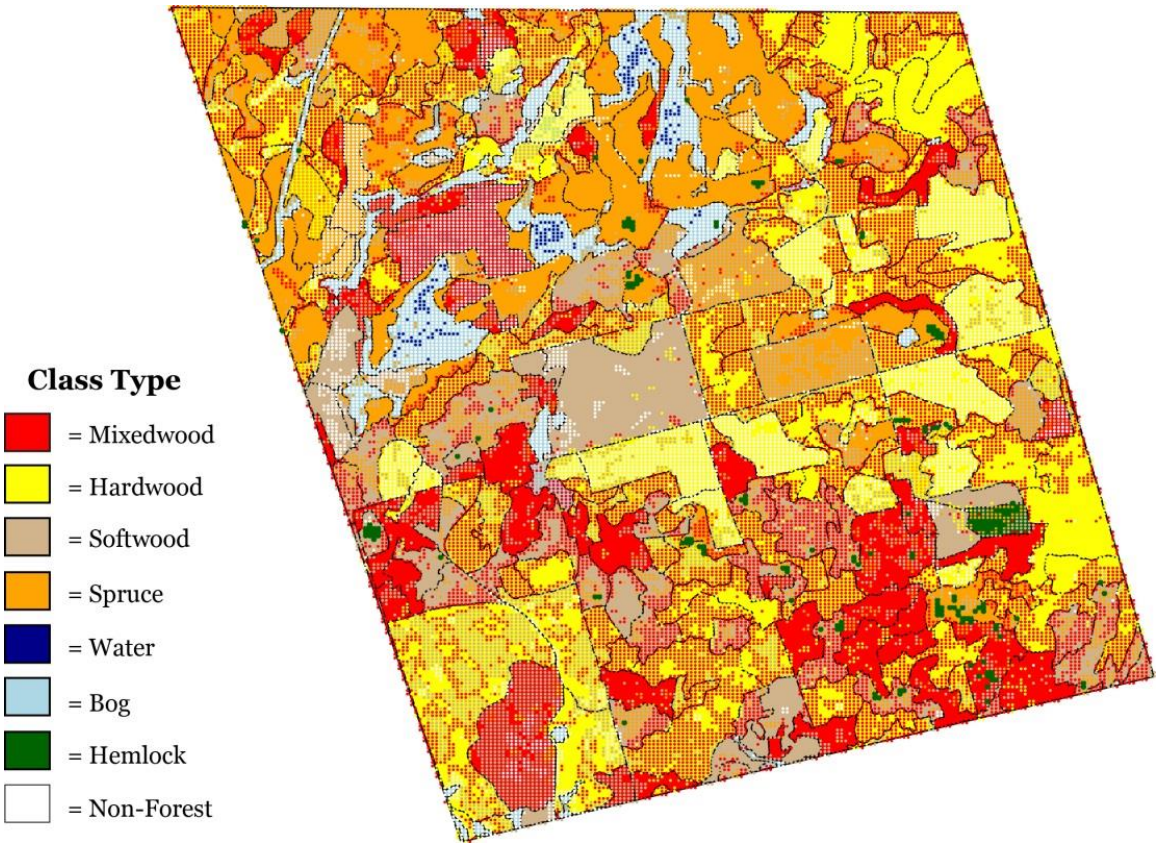

Figure 8. Air photo interpreted stand map of Noonan Research Forest overlaid with model T15 predicted class type of each 10 m by 10 m cell represented by a dot in the center of each cell. Each dot is shaded according to the predicted class type and Hemlock dots are magnified to x10 size to increase visibility.

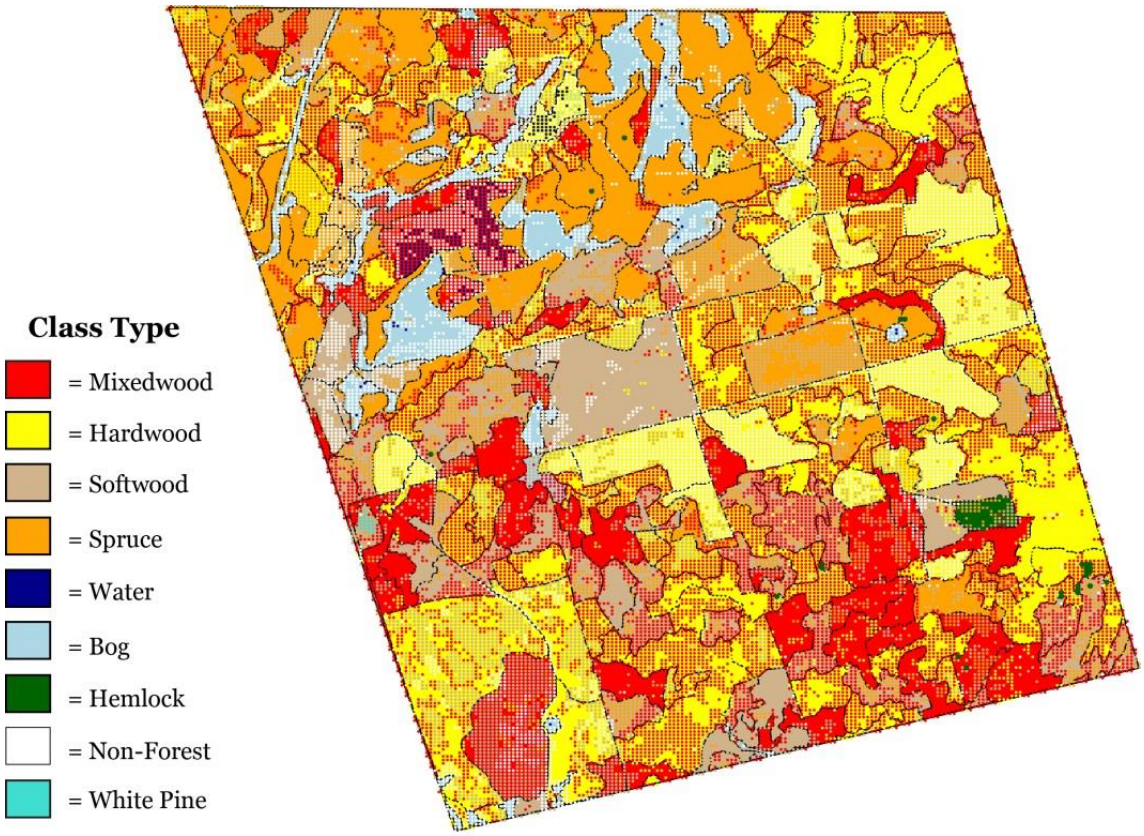

Figure 9. Air photo interpreted stand map of Noonan Research Forest overlaid with model F1 predicted class type of each 10 m by 10 m cell represented by a dot in the center of each cell. Each dot is shaded according to the predicted class type and Hemlock dots are magnified to x10 size to increase visibility.

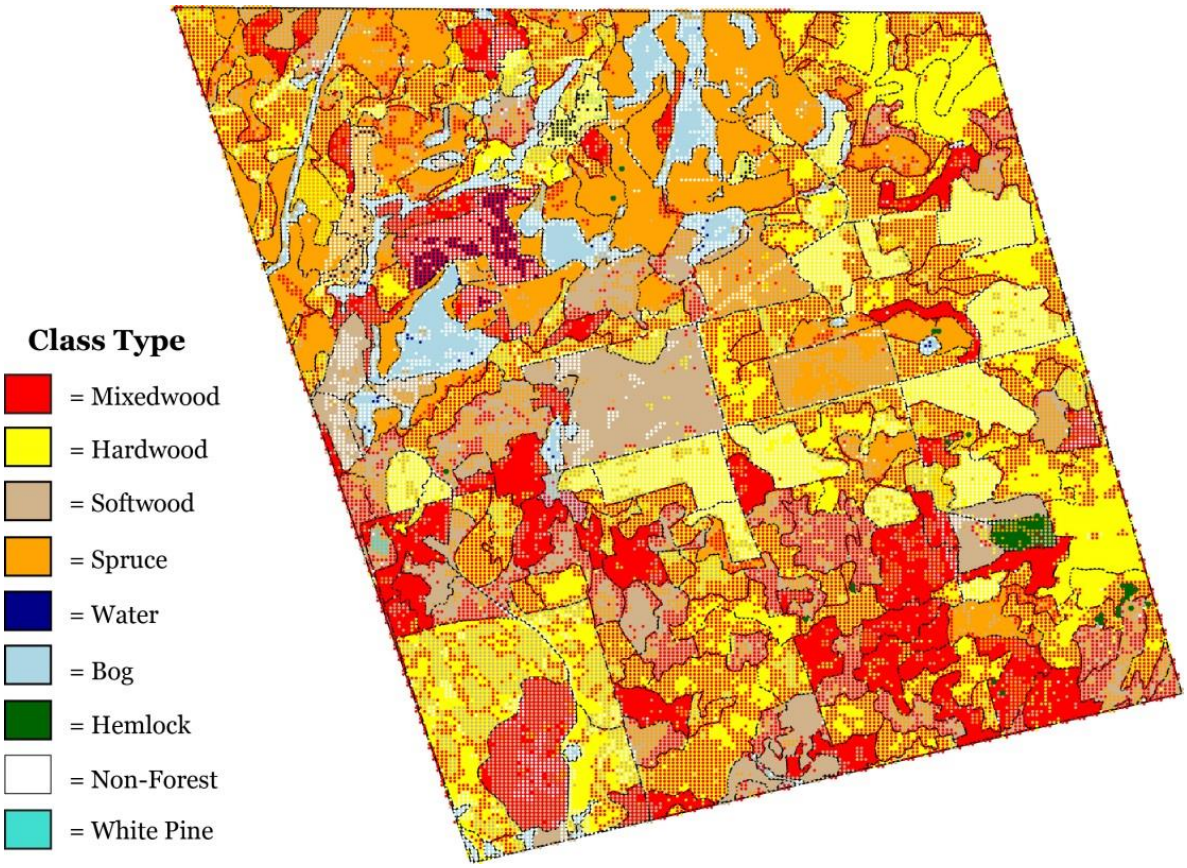

Figure 10. Air photo interpreted stand map of Noonan Research Forest overlaid with model F2 predicted class type of each 10 m by 10 m cell represented by a dot in the center of each cell. Each dot is shaded according to the predicted class type and Hemlock dots are magnified to x10 size to increase visibility.

From these maps the differences in stand classification for each model also can be compared. All Hemlock classified patches that disappear between the training models (Figure 6 – 8) and final models (Figure 9 and 10) were Spruce and Softwood patches misclassified as Hemlock. As well, it should be noted that many of these misclassified cells identified by the training models are surrounded by Spruce classified plots. For all class types, other than Hemlock, all models behaved very similar except for T09; T09 classified a significant number more Mixedwood and Water cells than the other models.

Although models T10, T15, F1 and F2 had similar overall classification, there are still many cell type discrepancies between model classification schemes. Common classification inconsistencies between models include: Spruce and Softwood; Hardwood and Mixedwood and; Hardwood and Softwood. However, Mixedwood tended to be more similar between photo and model classification.

An important note to make when interpreting Figures 6 to 10 is that the stand types for the air photo interpreted stand map were based on photographs taken during leaf on stand conditions. Conversely, the LiDAR data used in this study were acquired during leaf off stand conditions. As a result, many LiDAR returns that would have been intercepted by broadleaf canopies proceeded to lower elevations, intercepting conifer species in the mid-story; primarily balsam fir with a lesser spruce component. Accordingly, the classification of many Hardwood and Mixedwood cells may have been influenced by the mid-story conifer class type and the reflectance values associated with conifer species. This may account for the large amount of Mixedwood and Softwood cells classified within Hardwood stands, and Softwood classified in Mixedwood stands. Similarly, inconsistencies between model maps may be attributed to the specific LiDAR attributes used in

each model and the combined capacity of those attributes to be influenced by mid-story conifer species below over-story broadleaf species.

Chapter 4. Discussion

4.1 Model Implication

The F1 and F2 randomForest classification models created in this study represent a novel approach to identifying hemlock patches within the Acadian forest region. However, to determine their true capabilities, these models need to be tested on additional LiDAR data over the range of stand conditions in which hemlock naturally occurs. As Hayashi *et al.* (2015) point out, RF relies on the strength of training data, and may not transfer well when classifying other forests. It is likely these models are unique to the specific hemlock patch structures found within the matrix of stand conditions on the NRF from which the models were trained. Nonetheless, these models represent a steppingstone in the development and understanding of LiDAR-based stand type classification tools.

4.2 LiDAR Attributes

An extensive number of LiDAR attributes were evaluated and ultimately used in this study, and it is unlikely that there are any key LiDAR attributes that were overlooked or that could improve classification accuracy.

Both F models shared 13 of the same base LIDAR attributes and 5 of the same attributes occurred in the top 6 of the highest variable importance rankings for both models (Appendix C). The top 6 highest ranking attributes in the two F models include LdensityB05, q65INSmean, q55INSmean, q50CHT, q25LiDAR and Wshape.

LdensityB05, the density per m^3 of LiDAR returns below 5m height, was the most important attribute in the F2 model and played an important role, ranking in the top 6 attributes, for all three of the initial training models. The influence of this attribute on identifying Hemlock patch

types can be related to the dense canopy structure of hemlock. The dense canopy of hemlock significantly reduces the number of point returns that penetrate below 5m height in comparison to the more open canopy structures of the other class types. In addition, the dense canopy structure results in very little understory vegetation for the LiDAR beams to reflect from so that any beams that do penetrate the canopy end up reflecting from the ground and are thus classified as ground points prior to extraction of LiDAR attributes. As seen in Table 12, the average LdensityB05 for all Hemlock types is significant lower than that of all other class types. However this attribute, although it was still included in the F1 model, played a less significant role being ranked as the least important variable used in the model. It is likely that the other variables played a more important role in differentiating between the Hemlock subclasses used in the F1 model.

Table 12. Average attribute values of the adjusted base training data for each classification type.

LiDAR Attribute[1]	Classification Type[2]											
	HEM	OMHEM	PHEM	SPHEM	SPRUCE	SOFT	WP	BOG	HARD	MIXED	NONF	WATER
LdensityB05	0.6	0.3	0.8	0.5	1.9	2.6	1.3	2.9	3.0	2.7	3.1	2.0
q55INSmean	89.6	81.1	100.4	72.1	88.7	70.4	79.6	79.2	49.7	60.1	68.6	101.6
q65INSmean	90.4	82.3	101.2	72.2	91.1	71.1	80.5	77.2	49.2	59.9	69.0	97.7
q50CHT	13.1	15.4	11.0	14.9	11.6	9.8	23.3	1.7	13.7	12.8	5.4	0.8
q25LiDAR	8.0	8.8	7.1	9.1	6.0	4.8	13.5	0.8	4.5	4.9	2.3	0.5
Wshape	3.1	2.5	3.4	3.1	2.9	3.5	1.9	6.3	2.4	2.8	3.9	7.7

[1]LiDAR attribute descriptions can be found in Appendix A.

[2]Patch classification type based on tree density and species composition: BOG – bogs and grasslands with scattered trees; HARD – hardwood species composition > 70%; MIXED – >5 trees/plot and < 70% of either hardwood or softwood species; NONF – none or scattered over-story trees (DBH $\geq$ 6 cm); OMHEM – Hemlock composition > 50% and softwood species composition> 30%; with 90% of trees > 30cm DBH.; PHEM – Hemlock composition > 90%; SOFT – softwood species composition > 70%; SPHEM – Hemlock > 50% spruce species > 30%; SPRUCE – spruce species composition > 70%; WATER – open water with no trees; WP – White Pine composition > 70%.

Both q65INSmean and q55INSmean, the mean intensity values of the 65th and 55th percentiles of the point clouds, were ranked in the top 4 attributes used in both final models. As well, three other quantile intensity measures were used in the building of both models; q25INSmean,

q35INSmean and q75INSmean. The importance of these intensity attributes can also be linked back to the dense canopy structure of hemlock. Kim *et al.* (2009) found LiDAR intensity values to be weaker in more open canopies and with larger foliage spacing. Accordingly, Hemlock had much higher intensity values than all other class types excluding Spruce and Water, which can be attributed to the dense canopy structure of hemlock and spruce as well as the reflective capabilities of water. Within the Hemlock subclasses these intensity values can be found to be significantly lower in OMHEM and SPHEM and very high in PHEM. PHEM is likely higher in intensity value due to its 90% or greater hemlock component that creates consistent dense foliage coverage. The intensity values found in OMHEM and SPHEM are likely lower due to non-hemlock species found within these stands and canopy gaps. SPHEM often had a hardwood component, mainly red maple and white birch, which would break up the consistency of the spruce hemlock canopy. OMHEM was often found mixed with a variety of species including hardwoods but also located in wet areas in which there was not consistent tree cover. Given that the LiDAR scans used in this study were collected in leaf-off condition, the intensity values were very useful for differentiating hardwood from softwood stand types (Table 12).

The q50CHT, the 50th percentile of canopy surface height, was also found to be an important attribute, ranking in the top 5 of variable importance for all models. The average q50CHT for hemlock subclasses OMHEM and SPHEM was relatively high in comparison to all other class types excluding WP, which was significantly higher. This can be attributed to the large mature trees found in the OMHEM class type and the large mature spruce overtopping younger hemlock in the SPHEM class type; the relative size of these trees providing higher crowns and accordingly higher overall canopy surface heights. PHEM however, had a lower q50CHT value closer to that of Spruce; likely due to the relative size of the trees found in the PHEM class on the NRF. Although this stand type is mature, the trees are much smaller and younger than Hemlock found in OMHEM and the

mature spruce found in SPHEM. Similarly, the higher q25LiDAR values for WP can be attributed to the extremely large size of the white pine trees classified in the training data; accordingly having high crowns and higher overall canopy surface height. Hardwood and Mixedwood types had higher values than that of PHEM, Spruce and Softwood. This can likely be linked to the decurrent or branchy structure of NRF broadleaf species which intercept more LiDAR points in the upper canopy.

The q25LiDAR, the height of the 25th percentile of point clouds, was much higher in Hemlock and the 3 hemlock subclasses than all other class types excluding WP. Once again this trend can be related to the canopy structure of Hemlock; having a characteristic dense canopy and a bare understory. As a result, the height of the 25th percentile of point clouds is high in the tree canopy due to the bare understory structure of hemlock dominated stands in which there is little vegetation and regeneration to intercept LiDAR points before they become ground returns. Mature white pine stands create a similar bare understory structure, however as stated previously the white pine classified in the training data are exceedingly large in size and have very high crowns. This can account for the very large q25LiDAR value found in the WP classification.

Wshape, the Weibull shape parameter or Weibull slope, was calculated as a measure of the top down vertical height distribution of LiDAR points. A Wshape value of 1 indicates that all points are located near the top of the LiDAR profile. As the points become more evenly distributed on the vertical axis, the Wshape values increase to 3.6; a value of 3.6 represents a perfectly symmetrical point distribution. Values increasing beyond 3.6 indicate that a greater number of points are distributed towards the ground. The Wshape value can be directly related to the vertical distribution of tree foliage and branch structure. WP had a significantly lower Wshape attribute value than all other class types, indicating a very top heavy distribution of LiDAR points; accurately

describing the high emergent crown structure of the white pine used in the training data. OMHEM, Spruce, Hardwood and Mixedwood all had similar Wshape values, ranging from 2.39 to 2.87, indicating a more evenly, but still slightly top heavy distribution of LiDAR points. SPHEM, PHEM and Softwood had Wshape values were increasingly closer to a symmetrical shape distribution: 3.10, 3.36 and 3.5 respectively. In SPHEM this distribution of points can be associated with a two-story canopy structure with a spruce overstory and hemlock understory. In PHEM and Softwood plots this can likely be associated with the overall smaller tree sizes in these stands. The large Wshape values for Water and Bog class types is due to the lack of tree cover, the majority of the LiDAR returns being returned from the ground or lower vegetation.

4.3 Leaf off LiDAR Data

One particular method change that could lead to significantly different results would be the use of leaf on data as opposed to the leaf off data that were used in this study. However it is speculated that models built using leaf on LiDAR data would not improve classification from the current models built from leaf off data and could potentially have negative effects on attributes currently considered important within the models developed using leaf off LiDAR scans.

Leaf on data would primarily impact the LiDAR data gathered for broadleaf trees and not conifers; the only two classification types in the study that have significant broadleaf components are Hardwood and Mixedwood. Furthermore the models created in this study had no significant difficulty differentiating either of these class types from Hemlock. The two primary ways by which the interception of LiDAR points by the broad leaf foliage would adjust the current data is by altering point cloud arrangement and intensity values.

In leaf on conditions, LiDAR points that would otherwise proceed to intercept smaller trees and vegetation in the understory would now be intercepted by the hardwood foliage. This would

significantly alter point cloud arrangement from that of the leaf off data and affect a number of spatial attributes; specifically, LiDAR density measures, percentiles of canopy surface height, quantile measures of point clouds and the Weibull distribution parameters, particularly the shape parameter. The relative influence of LdensityB05, q50CHT, q25LiDAR and Wshape shown in Table 12, would likely be altered. The importance of these attributes in hemlock classification is attributed to the characteristic dense canopy and bare understory structure of hemlock dominated stands. These attributes easily distinguished Hemlock from that of Hardwood and Mixedwood class types, due to the reduction of broadleaf canopy returns in leaf off condition. In comparison to Hardwood and Mixedwood class types, this resulted in lower Ldensity measures for all hemlock classes, higher q50CHT values for OMHEM and SPHEM, and higher q25LiDAR values for all hemlock classes; all at significantly different values than the two broadleaf class types. Accordingly, the use of leaf on data would significantly change the values of these three attributes for the two broadleaf class types and potentially make them harder to differentiate from Hemlock values. In the case of Wshape however, increased LiDAR returns in the upper canopy would decrease Mixedwood and Hardwood values making them easier to differentiate from Hemlock class type.

A large amount of intensity attributes were used in the F1 and F2 models; 6 and 5 accordingly. Hemlock class types were found to have significantly higher average values for intensity attributes than that of Hardwood and Mixedwood class types for the leaf off data used; as seen with q65INSmean and q55INSmean. The low intensity values in the broadleaf class types is likely a result of the lack of foliage, causing all intensity values to be generated by ground returns and LiDAR points reflected off bark; Roberts *et al* (2004) found that bark has lower spectral reflectance than foliage. Using leaf on data it is likely that intensity parameters would have large increases in value; a number of studies have found broadleaf species to have higher mean intensity value than

that of conifers when using leaf on data (Kim *et al.* 2009; Song *et al.* 2002). This may reduce the accuracy of all models since each relied on a number of different intensity attributes to identify Hemlock class types; of which all Hemlock class types had higher average intensity attribute values, during leaf off conditions, than that of the two broadleaf classes.

Although it is not presumed that leaf on data would improve hemlock classification, it is however speculated it could improve the classification of Hardwood and Mixedwood stand types. As stated previously, the lack of broadleaf foliage in the leaf off data resulted in a large amount of Harwood and Mixedwood being classified according to mid-story conifer class types (Figures 6, 7, 8, 9 and 10). The NRF has a significant component of balsam fir in the understory of most Hardwood and mixedwood stands, given the leaf-off condition and the reflectance values associated with softwood species, many cells within these stand types were classified as Softwood or Spruce types. The use of leaf on LiDAR data could potentially increase the classification accuracy of Hardwood and Mixedwood patches, creating a more consistent relationship between LiDAR based classification and the aerial photo interpreted map (compare Figure 5 to Figures 6, 7, 8, 9 and 10).

4.4 Hemlock Subclass Prediction

When Hemlock was further divided into 3 subclasses, it was assumed that this would make the single Hemlock classification less general in its prediction and the model would misclassify fewer patches. Specifically, it was thought that the range of values for specific LiDAR attributes that was accepted for Hemlock classification may be quite broad, having to accommodate a range of Hemlock class types/patch conditions from OMHEM to SPHEM and PHEM. This broad classification range for the attributes may have produced classification errors where Spruce and Softwood patches were identified as Hemlock. However, this was not the result obtained here.

The single Hemlock class (F1) and Hemlock subclass (F2) models both proved to have the same prediction accuracy, though they had differing abilities to identify PHEM and SPHEM class types (Table 10). The differing capabilities of the 2 models to identify the two different Hemlock classes may be attributed to the specific attributes used in each model (Appendix C).

Both F models had difficulty identifying Old Mix Hemlock class patches; F1 and F2 models identifying only 3 and 4 patches from the 23 patches present in the training data (Table 11). A likely reason for this is that OMHEM structure was found to occur in relatively small patches and irregular shapes and consequently may have been too small to pick up if they were not centered within the 100 m^2 cells used in developing the wall-to-wall classification for the NRF.

4.5 Classification Accuracy

Many cells misclassified as Hemlock shared similar structural attributes as the Hemlock dominated cells, characterized by dense canopies with a bare understory. These plots included mature and uniform Spruce, as well as Softwood patches with large white cedar components. White cedar is not a major component of the NRF; however, a site with a higher abundance of cedar might significantly reduce model accuracy given the attribute structures used in this study. Common characteristics within the Spruce and Softwood plots identified as Hemlock such as stand density, as well as even-aged and uneven aged stand features, could be used to identify these cells and turn them into a new class type to facilitate differentiating these cells from Hemlock, as was done with the White Pine class type in the F models.

Classification error of Hemlock determined by the Confusion Matrix when building the models did not reflect overall classification accuracy when developing the wall-to-wall classifications of the NRF. Both F models had significantly higher Hemlock classification error in their Confusion Matrices (Table 8 and 9) than that of the T models (Tables 2, 3 and 4), yet they both had much

higher classification accuracies (Table 10). However, this is not a direct comparison due the use of different training data sets for the T and F models; Confusion Matrix classification error may be directly linked to the training data set used in each model.

4.6 Method Analysis

A large issue which may have significantly influenced results was the accuracy of the Garmin GPSmap 76CSx model used to record the center points of each Hemlock patch. The point cloud data may not have been positioned accurately over identified Hemlock patches, resulting in offset data collection (Reutebuch, *et al.* 2005). Consequently, the extracted LiDAR data would not be reflective of a Hemlock class type and reduce classification accuracy.

Another issue that may not have reduced the accuracy of the models but directly prevented identification of Hemlock patches within the NRF was the layout of the 10m by 10m cells. As stated previously, Hemlock patches can be relatively small in size, as with OMHEM, and missed if offset from the cell arrangement. This may account for the large number of OMHEM training patches that were not identified in the F models. As seen in Figures 6, 7, 8, 9 and 10, much of the Hemlock patch locations identified were located in close proximity to each other in the larger Hemlock stands where it was easier for cells to be centered on a number of patches.

A potential solution to the problem of small Hemlock patch sizes would be to use overlapping 10m by 10m cells that adjust by 1 meter increments for each prediction, this could potentially improve the opportunity for the model to detect and classifying the smaller Hemlock patches. Alternatively, smaller cell sizes could be used in an effort to capture smaller Hemlock patches; however Hayashi et al. (In Press) found little difference in volume estimates across a range of cell sizes of 2m by 2m – 100m by 100m.

Hemlock also had a relatively low number of training plots in comparison to the rest of the forest class types, except for White Pine. This was due to hemlock being a relatively minor component of the NRF, and often occurring in small patch sizes that were difficult to locate during the parallel transect sweeps used to gather the ground reconnaissance data. However, the method used to increase the number of Hemlock class patches in the base data, using the T models to locate additional Hemlock patches, and adding these patches to the base training data for the F models, may have overfit the model to the NRF Hemlock data. This could reduce the ability of the model to generalize to different forests. Although it can be noted that the use of both F models resulted in the identification of 19 new Hemlock patches.

A method that may have created great bias in the study was the use of species composition based on the number of trees, to define patch class type in the training data. This method may have poorly reflected which tree species dominated the canopy in each patch and consequently intercepted the most LiDAR points. Defining the class types based on species volume or species age class may have created a more accurate measure of species dominance. Additionally, the method of only recording trees with a DBH greater than 6cm may have created a bias in the inventory data against young stands.

Chapter 5. Conclusions

The results provided no clear choice between the two final models. Both models obtained the same accuracy values, although they both proved to have differing capabilities in terms of identifying Hemlock subclasses. As well, both models had significant difficulty identifying Old Mixed Hemlock class type; this issue may be corrected by predicting the study site in overlapping 10m by 10m cells that adjust by 1 meter for each prediction, or by predicting the study site using smaller patch cells. Additionally, The LiDAR attributes used in the final models were found to be heavily influenced by the characteristic dense canopy and bare understory structure created by hemlock dominated patches. Due to the particular LiDAR attributes found to be of highest variable importance in each of these models it is not believed that model accuracy would significantly improve using leaf on data, as opposed to leaf off data. For the continuation of this study it is recommended that the same methods be used, with correction for the OMHEM class type, to generate new general models using a larger sample of Acadian forest stand types. It may also be useful to compare and/or combine leaf on and leaf off data.

This study shows the potential for using LiDAR to identify rare forest types. It also highlights the need to understand the underlying forest structure and its influence on LiDAR returns. Much current LiDAR analyses focus too much on “kitchen sink” models with very little analysis of the underlying biology and its interaction with the physics of LiDAR pulses.

References

Alden, H.A. 1997. *Softwoods of North America*. U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisconsin.

Ashbel, H.F. 1960. *Silvical characteristics of eastern hemlock (*Tsuga canadensis*).* U.S. Department of Agriculture, Forest Service Northeastern Forest Experiment Station, Upper Darby, PA. Report No. NE-132.

Baumgras, J.E., Sendak, P.E. and Sonderman, D.L. 2000. Ring shake in eastern hemlock: frequency and relationship to tree attributes. In *Proceedings: Symposium on Sustainable Management of Hemlock Ecosystems in Eastern North America*. June 22-24, 1999, Durham New Hampshire. Edited by K.A. McManus, K.S. Shields and D.R. Souto. U.S. Department of Agriculture, Forest Service, Newton Square, PA. pp. 156-160.

Breiman, L. 2001. Random forest. Machine Learning 45, 5-32.

Canadian Forest Service. 2014. *Alien, invasive hemlock woolly adelgid found in Ontario*. Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie Ont. Report

Elliott, K.J. and Vose, J.M. 2011. The contribution of the Coweeta Hydrologic laboratory to developing an understanding of long-term changes in managed and unmanaged forests. *Forest Ecology and Management*. Vol. 261, pp. 900–910.

Ellison, A.M., Bank, M.S., Clinton, B.D., Colburn, E.A., Elliott, K., Ford, C.R., Foster, D.R., Kloeppel, B.D., Knoepp, J.D., Lovett, G.M., Mohan, J., Orwig, D.A., Rodenhouse, N.L., Sobc-zak, W.V., Stinson, K.A., Stone, J.K., Swan, C.M., Thompson, J., Von Holle, B., and Webster, J.R. 2005. Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment*. Vol. 3, No. 9, pp. 479–486.

Evans, R.A. 2002. An ecosystem unraveling? In *Proceedings, Hemlock woolly adelgid in the Eastern United States symposium*, February 5-7, 2002, East Brunswick, NJ. Edited by B., Onken and R., Lashomb. New Brunswick, NJ: Rutgers University. Pp. 23-33.

Fitzpatrick, C.F., Preisser, L.P., Porter, A., Elkinton, J., and Ellison A.M. 2012. Modeling range dynamics in heterogeneous landscapes: invasion of the hemlock woolly adelgid in eastern North America. *Ecological Application*s. Vol. 22, No. 2, pp. 472-486.

Godman, R.M. and Lancaster, K. 1990. *Tsuga canadensis* (L.) Carr. Eastern Hemlock. In *Silvics of North America, Vol. 1, Conifers*. Edited by R.M. Burns and B.H. Honkala. U.S. Department of Agriculture, Forest Service, Washington D.C. PP. 604-612.

Hayashi, R., Kershaw, J.A. and Weiskittel, A. 2015. Evaluation of alternative methods for using LiDAR to predict aboveground biomass in mixed species and structurally complex forests in Northeastern North America. Math. Comput. For. Nat.-Res. Sci. Vol. 7, No. 2, pp.49-65.

Havill, N.P., Vieira, L.C., and Salom S.M. Biology and control of Hemlock Woolly Adelgid. 2014. U.S. Department of Agriculture. pp. 1-16.

Heinzel, J.N., Weinacker, H. and Koch B. "Full automatic detection of tree species based on delineated single tree crowns – a data fusion approach for airborne laser scanning data and aerial photograph" in *Proc. SilviLaser 2008*, Edinburgh, UK, pp. 76-85, 2008.

Howard, T., Sendak P., and Codrescu, C. 2000. Eastern Hemlock: A Market Perspective. In *Proceedings: Symposium on Sustainable Management of Hemlock Ecosystems in Eastern North America*. June 22-24, 1999, Durham New Hampshire. Edited by K.A. McManus, K.S. Shields and D.R. Souto. U.S. Department of Agriculture, Forest Service, Newton Square, PA. pp. 161-166.

Holmgren, J., Persson, A. and Soderman, U. 2008. Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing*. Vol. 29, No. 5, pp. 1537-1552.

Jenkins, J.C., Aber, J.D., and Canham, C.D. 1999. Hemlock woolly adelgid impacts on community structure and N cycling rates in eastern hemlock forests. *Canadian Journal of Forest Research*. Vol. 29, pp. 630-645.

Kim, S., McGaughey, R.J., Enrik, H. and Schreuder, G. 2009. Tree species differentiation using data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*. Vol. 113, pp. 1575-1586.

Liaw, A., Wiener, M. 2002. Classification and regression by randomForest. R News 2/3: 18-22.

Leckie, D., Gougen, F., Hill, D, Quinn, R., Armstrong, L. and Shreenan, R. 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Can. J.Remote Sensing*. Vol. 29, No.5, pp. 633-640.

Loo, J., Ives, N.2003. The Acadian forest: historical condition and human impacts. Forestry Chronicle 79, 462-474.

Martin, L.M. and Goebel, P.C. 2013. The foundation species influence of eastern hemlock (*Tsuga canadensis*) on biodiversity and ecosystem function on the Unglaciated Allegheny Plateau. *Forest Ecology and Management*. Vol. 289, pp. 143-152.

McClure, M.S. 1991. Nitrogen fertilization of hemlock increases susceptibility to hemlock woolly adelgid. *Journal of Arboriculture*. Vol. 17, pp. 227-229.

McWilliams, W.H. and Schmidt, T.L. 2000. Composition, Structure, and Sustainability of Hemlock Ecosystems in Eastern North America. In *Proceedings: Symposium on Sustainable Management of Hemlock Ecosystems in Eastern North America*. June 22-24, 1999, Durham New Hampshire. Edited by K.A. McManus, K.S. Shields and D.R. Souto. U.S. Department of Agriculture, Forest Service, Newton Square, PA. pp. 5-10.

Mladenoff, D.J. and Stearns, F. 1993. Eastern Hemlock Regeneration and Deer Browsing in the Northern Great Lakes Region: A Re-examination and Model Simulation. *Conservation Biology*. Vol. 7, No. 4, pp. 889-900.

Mladenoff, D.J. 1995. The role of eastern hemlock across scales in the northern lake states. In *Hemolock Ecology and Management, Proceedings of a Regional Conference on Ecology and Management of Eastern Hemlock*, Spetember 27-28, 1995, Iron Mountain, Michigan. Edited by G. Mroz and J. Martin. Department of Forestry School of Ntural Resources College of Agricultural and Life Sciences, University of Wisconsin, Madison. Pp. 29-42.

Orwig, D.A., Foster, D.R., and Mausel D.L. 2002. Landscape patterns of hemlock decline in New England due to the introduced hemlock woolly adelgid. *Journal of Biogeography*. Vol. 29, pp. 1475-1487.

Orwig, D.A. and Foster, D.R. 1998. Forest Response to the Introduced Hemlock Woolly Adelgid in Southern New England, USA. *Journal of the Torrey Botanical Society*. Vol. 125, No. 1, pp. 60-73.

Paradis, A., Elkinton, J., Hayhoe, K., and Buonaccorsi, J. 2008. Role of winter temperature and climate change on the survival and future range expansion of the hemlock woolly adelgid (*Adelges tsugae*) in eastern North America. Mitig. Adapt. Strat. Glob. Change. Vol. 13, pp. 541-554.

Preisser, E.L., Miller-Peirce, M.R., Vansant, J., and Orwig, D.A. 2011. Eastern hemlock (Tsuga canadensis) regeneration in the presence of hemlock woolly adelgid (Adelges tsugae) and elongate hemlock scale (Fiorinia externa). *Canadian Journal of Forest Research*. Vol. 41, 2433-2439.

R Development Core Team 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Reutebuch, S.E., Andersen, H. and McGaughey, R.J. 2005. Light Detection and Ranging (LiDAR): An Emerging Tool for Multiple Resource Inventory. Journal of Forestry. Vol. 103, No. 6, pp. 286-292.

Rogers, R.S. 1978. Forest dominated by hemlock (*Tsuga canadensis*): distribution as related to site and postsettlement history. *Canadian Journal of Botany*. Vol. 56, pp. 843-854.

Skinner, M., Parker, B.L., Gouli S., and Ashikaga, T. 2003. Regional Response of Hemlock Woolly Adelgid (Homoptera: *Adelgidae*) to Low Temperatures. *Environmental Entomology*. Vol. 32, No. 3, pp. 523-528.

Snyder, C.D., John, A.Y., Lemarie, D.P., and Smith, D.R. 2002. Influence of eastern hemlock (*Tsuga canadensis*) forests on aquatic invertebrate assemblages in headwater streams. *Canadin Journal of Fisheries and Aquatic Sciences*. Vol. 59, pp. 262-275.

Sullivan, K.A., and Ellison, A.M. 2006. The seed bank of hemlock forests: implications for forest regeneration following hemlock decline. *Journal of the Torrey Botanical Society*. Vol. 133, No. 3, pp. 393-402.

Tingley, M.W., Orwig, D.A., Field, R., and Motzkin G. 2002. Avian response to removal of a forest dominant: consequences of hemlock woolly adelgid infestations. *Journal of Biogeography*. Vol. 29, pp. 1505-1516.

Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T. and White, J.C. 2008. The role of LiDAR in sustainable forest management. The Forestry Chronicle. Vol. 84, No.6, pp. 807-826.

Appendix A. LiDAR metrics

No.	Abbreviation	Definition
1	maxCHT	Maximum canopy surface height
2	meanCHT	Mean canopy surface height
3	q25CHT	25th percentile of canopy surface height
4	q50CHT	50th percentile of canopy surface height
5	q75CHT	75th percentile of canopy surface height
6	Ldensity	Density of LiDAR returns number per m^3
7	LdensityB10	Density of LiDAR returns below 10 m^3
8	LdensityB05	Density of LiDAR returns below 5 m^3
9	HTmaxDens	Height of maximum point cloud density
10	q25Dens	Point density of the 25th percentile of point clouds
11	q35Dens	Point density of the 35th percentile of point clouds
12	q45Dens	Point density of the 45th percentile of point clouds
13	q55Dens	Point density of the 55th percentile of point clouds
14	q65Dens	Point density of the 65th percentile of point clouds
15	q75Dens	Point density of the 75th percentile of point clouds
16	q85Dens	Point density of the 85th percentile of point clouds
17	q95Dens	Point density of the 95th percentile of point clouds
18	meanLiDAR	Mean point cloud height
19	q25LiDAR	Height of 25th percentile of point clouds
20	q35LiDAR	Height of 35th percentile of point clouds
21	q45LiDAR	Height of 45th percentile of point clouds
22	q55LiDAR	Height of 55th percentile of point clouds
23	q65LiDAR	Height of 65th percentile of point clouds
24	q75LiDAR	Height of 75th percentile of point clouds
25	q85LiDAR	Height of 85th percentile of point clouds
26	q95LiDAR	Height of 95th percentile of point clouds
27	meanINS	Mean intensity value
28	q25INSmean	Mean intensity value of the 25th percentile of point clouds
29	q35INSmean	Mean intensity value of the 35th percentile of point clouds
30	q45INSmean	Mean intensity value of the 45th percentile of point clouds
31	q55INSmean	Mean intensity value of the 55th percentile of point clouds
32	q65INSmean	Mean intensity value of the 65th percentile of point clouds
33	q75INSmean	Mean intensity value of the 75th percentile of point clouds

34	q85INSmean	Mean intensity value of the 85th percentile of point clouds
35	q95INSmean	Mean intensity value of the 95th percentile of point clouds
36	stdevINS	Standard deviation of mean intensity value
37	q25INSstdev	Standard deviation of the mean intensity value of the 25th percentile of point clouds
38	q35INSstdev	Standard deviation of the mean intensity value of the 35th percentile of point clouds
39	q45INSstdev	Standard deviation of the mean intensity value of the 45th percentile of point clouds
40	q55INSstdev	Standard deviation of the mean intensity value of the 55th percentile of point clouds
41	q65INSstdev	Standard deviation of the mean intensity value of the 65th percentile of point clouds
42	q75INSstdev	Standard deviation of the mean intensity value of the 75th percentile of point clouds
43	q85INSstdev	Standard deviation of the mean intensity value of the 85th percentile of point clouds
44	q95INSstdev	Standard deviation of the mean intensity value of the 95th percentile of point clouds
45	Wscale	Weibull scale parameter
46	Wshape	Weibull shape parameter

Appendix B. Common and Latin names of tree species found in NRF.

No.	Common Name	Latin Name
1	balsam fir	*Abies balsamea* (L.) Carrière
2	stripped maple	*Acer pensylvanicum* L.
3	red maple	*Acer rubrum* L.
4	sugar maple	*Acer saccharum* Marshall
5	alder	*Alnus spp.* L.
6	yellow birch	*Betula alleghaniensis* Britton
7	paper birch	*Betula papyrifera* Marshall
8	gray birch	*Betula populifolia* Marshall
9	white ash	*Fraxinus Americana* L.
10	tamarack	*Larix laricina* (Du Roi) K. Koch
11	ironwood	*Ostrya virginiana* (Mill) K. Koch
12	black spruce	*Picea mariana* (Mill.) BSP
13	red spruce	*Picea rubens* Sarg.
14	jack pine	*Pinus banksiana* Lamb.
15	white pine	*Pinus strobus* L.
16	large-toothed aspen	*Populus grandidentata*
17	trembling aspen	*Populus tremuloides* Michx.
18	mountain ash	*Sorbus spp.*
19	eastern white cedar	*Thuja occidentalis* L.
20	eastern hemlock	*Tsuga Canadensis* (L.) Carrière
21	American elm	*Ulmus Americana* L.

Appendix C. LiDAR attributes used in each classification model.

No.	F1	F2	T09	T10	T15
1	Wshape	LdensityB05	q35INSmean	q35INSmean	q50CHT
2	q65INSmean	q55INSmean	q50CHT	q85INSmean	LdensityB05
3	q25LiDAR	Wshape	q85INSmean	meanCHT	meanCHT
4	q55INSmean	q65INSmean	LdensityB05	q50CHT	q55INSmean
5	q50CHT	q50CHT	q25INSmean	q25INSmean	q25INSstdev
6	q45INSmean	q25LiDAR	q25INSstdev	LdensityB05	q75CHT
7	meanCHT	q45LiDAR	meanCHT	q25INSstdev	q25LiDAR
8	q75CHT	q35LiDAR	stdevINS	q25CHT	q75INSmean
9	q35INSmean	q35INSmean	q25CHT	Wshape	Wshape
10	q35LiDAR	q75CHT	q85INSstdev	stdevINS	q45INSmean
11	q25CHT	q25CHT	q95LiDAR	q95LiDAR	stdevINS
12	q25INSmean	stdevINS	LdensityB10	LdensityB10	LdensityB10
13	q75INSmean	q25INSmean	q75Dens	q85INSstdev	q95LiDAR
14	LdensityB10	meanCHT		q75Dens	q25INSmean
15	LdensityB05	q75INSmean			q35INSmean
16	Ldensity				Wscale
17					q85INSmean

Appendix D.

CURRICULUM VITAE

Candidates Full Name: Hunter Roberts

Universities attended: University of Guelph, 2008-2012, B.Sc. Environmental Biology

University of New Brunswick, 2013-2017, Masters of Forestry