2020

REMOTE SENSING APPROACHES TO PREDICT FOREST CHARACTERISTICS IN NORTHWEST MONTANA

Ryan P. Rock

REMOTE SENSING APPROACHES TO PREDICT FOREST CHARACTERISTICS IN NORTHWEST MONTANA

By

Ryan P. Rock

B.S., Virginia Polytechnic and State University, 2012

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science in Geography

The University of Montana

Missoula, MT

December 2019

Approved by:

Scott Whittenburg, Dean
Graduate School

Dr. Anna Klene, Chair
Department of Geography

Mr. Kevin McManigal,
Department of Forest Management

Dr. David Affleck
Department of Forest Management

Rock, Ryan, M.S., December 2019 Geography

REMOTE SENSING APPROACHES TO PREDICT FOREST CHARACTERISTICS IN NORTHWEST MONTANA

Chairperson: Dr. Anna E. Klene

Remote sensing can be utilized by land management organizations to save money and time. Mapping vegetation using either aerial photographs or satellite imagery and the applications for forest management are of particular interest to the Montana Department of Natural Resources. In 2018, the organization began a pilot program to test the incorporation of raster analysis of remotely sensed data into their inventory program and had limited success. This analysis identified two areas of improvement: the selection method of inventory plots and the imagery used for classification and metrics. This study found that selecting inventory plots using a generalized random tessellation stratified (GRTS) sampling design in spectral space would likely improve the representation of the population as the sample distributions for mean and standard deviation in all spectral bands were more concentrated about the population means. Analysis using Sentinel-2 based predictors produced results that were comparable to predictions built using predictors derived from high resolution National Agricultural Inventory Program (NAIP) imagery. The increased spectral/radiometric/temporal resolution of Sentinel-2 imagery may have compensated for its lower spatial resolution.

ACKNOWLEDGMENTS

Thanks to all who have helped me along at various steps of this project. Anna Klene for her consistent input and willingness to adapt to a slightly unconventionally thesis project. Kevin McManigal for all of the cartographic knowledge he imparted upon me and for helping me keep things in perspective during hectic times. Dave Affleck for his advice on all things statistics and for helping me narrow the scope of the initial project. DNRC staff members, especially Wayne Lyngholm, for working with me through the painful early attempts and taking the time to answer many data/inventory method questions. John Hogland for his raster analysis guidance and availability to answer questions and troubleshoot. To Lauren Sharp for providing incredible support outside of school which allowed me to focus on working and for allowing me to vent my frustrations often. Finally, to my family for offering words of encouragement when I needed them throughout this long process.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The Montana Department of Natural Resource Conservation's (DNRC) Forest Management Bureau (FMB) is responsible for maximizing long-term revenue on over 780,000 acres (316,000 ha.) of forested trust lands (Montana DNRC, 2018). To do this, reliable measurements of forest characteristics are required. Three metrics (stand-level forest type, size class, and stocking level) are used to determine an estimation of annual sustainable yield, which the agency is required to have an independent 3rd party calculate at least every 10 years (MCA 77-5-222). For the 2015 calculation, the Forest Vegetation Simulator (FVS) Forest Growth Model (Dixon, 2002) was used to estimate annual sustained yield (Stander et al., 2015). The sustainable yield calculation (SYC) provides an estimate of annual harvestable timber volume in million board feet (MMbf), allowing managers to maintain healthy and diverse forests across the state of Montana. Data collection for the SYC is time-consuming and expensive. Remote sensing may provide a viable alternative to solely field-based estimates of forest metrics used in the SYC as well as other forest planning objectives.

1.1. Background

In 2018, the DNRC started a pilot project to test the use of raster analysis as a supplement to their field inventory program for collecting data necessary for the SYC. The organization hoped to develop a raster analysis workflow for these three forest characteristics, and then to eventually expand the program to predict all of the variables in their stand-level inventory (SLI) dataset. The area of interest for the pilot project was the Lazy Swift Acquisition (LSA), a chunk of land near Olney, MT that DNRC had recently purchased from Weyerhaeuser. Inventory plot data collected during the summer of 2018 were used to build models using NAIP 2015 imagery to predict the desired forest characteristics across the LSA. Results from the initial models were not matching what inventory crews reported on the ground, so the DNRC decided to generate new models in late

2018. I was working as an intern at the time, interested in developing a thesis around the DNRC's raster analysis project, and began working on building new models. After repeated attempts to produce more reliable results, it became clear that something other than the structure of the models was impacting the accuracy of the outcomes. Because of the lack of success of these models, DNRC decided to census the LSA at a stand level. In the summer of 2019, crews completed a stand-level inventory in which they walked every stand in the study area to collect the same variables collected at the plots in 2018. This thesis seeks to refine certain points where alternative decisions may have improved the remotely sensed analysis, using the new stand-level data to confirm and aid decision making.

1.2. Objectives

The DNRC is seeking to become more cost efficient in both calculating sustained yield and maintaining their state-wide timber inventory. This analysis sought to develop a remote-sensing based method for achieving this goal. There are a number of factors that determine the success of predicting land-cover metrics including the spatial, temporal, spectral, and radiometric resolution of imagery used, the predictor variables and algorithm selected, as well as the collection methods and preparation of training data. While many studies focus on the predictors and algorithms, this study will examine two decisions that may affect the accuracy and usefulness of the mapped forest characteristics:

1) the method of selecting inventory plot locations, and

2) the type of imagery used to build the models.

Exploring and comparing these factors will aid the DNRC as they incorporate more image analysis into their management program.

1.3. Study Area

This project focuses on the Lazy Swift Acquisition (LSA) in northwest Montana (Figure 1). The 13,000 acre (5,260 ha) area, which the DNRC acquired from Weyerhaeuser in 2018, lies in the western foothills of the Whitefish Range and south of Stryker Ridge. The LSA is dominated by a mix of conifers, primarily subalpine and grand fir (*Abies lasiocarpa* and *Abies grandis*), Engelmann spruce (*Picea engelmannii*), and western larch (*Lasix occidentalis*). Much of the area contains a multi-storied canopy with thick understory although there are ephemeral wetlands and wet meadows as well.

There are two main drainages, Lazy Creek and Swift Creek. Swift Creek remains a single reach throughout the whole LSA, while Lazy Creek is divided into three forks which combine to form the main fork just outside the area's southern boundary. Both creeks flow into the northern end of Whitefish Lake.

Elevations within the LSA range from 965 m to 1600 m (3160 – 5250 ft), lower than most of the Stillwater State Forest. The study area is within the northwestern Montana climate division, which experiences a maritime climate similar to the interior Pacific Northwest (Whitlock et al., 2017). The nearest Snotel site is located 40 km north along Grave Creek, at an elevation of 1400 m (4593 ft). Since 2017, the site reported a mean annual precipitation of 112 cm (44 in.) and a mean annual temperature of 4.0°C. Precipitation here is lower than the 1980 - 2010 mean of 135 cm (14 in.) (Natural Resources Conservation Service, 2016) and higher than the northwestern Montana mean of 82.2 cm (32.3 in.) (Whitlock et al., 2017). Mean annual temperature is cooler than the northwestern average of 4.8°C (Whitlock et al., 2017).

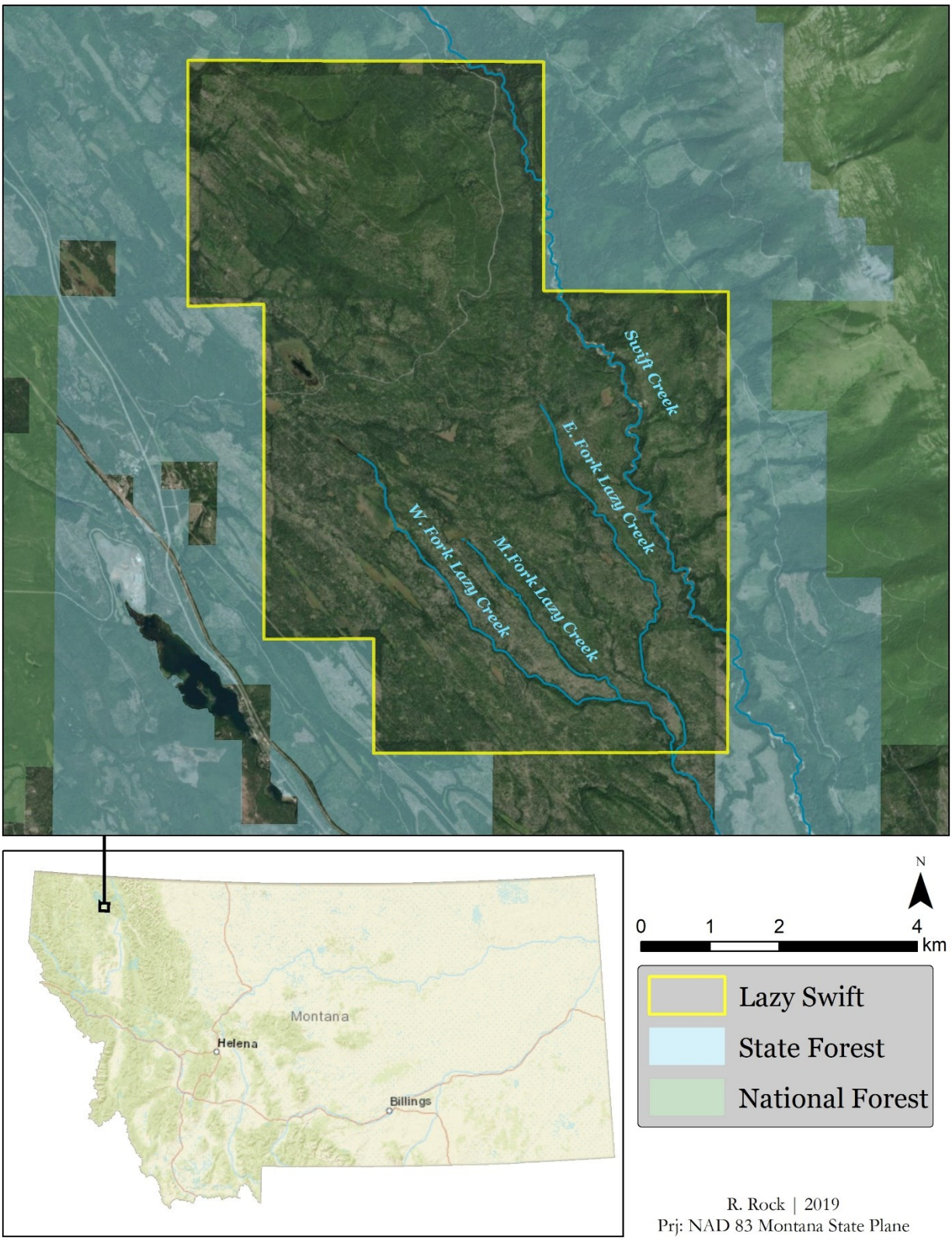

Figure 1: Map of the Lazy Swift Acquisition (LSA). Boundary (yellow line) is overlaid on NAIP 2015 image.

1.4. Sustained Yield Calculation

The algorithm to determine Sustainable Yield uses three input variables: stand size class (SSC), forest type (FTYPE), and stocking (STK; Table 1). The definition of SSC is dependent on the density and diameter at breast height (DBH) of measured trees (Table 2), while the definition of STK is unique to each size class (Table 3). When determining SSC for a stand, inventory crews use a hierarchical approach based on the density of trees in the stand. Crown density the percent canopy coverage in a stand. If trees larger than 9 in. (23 cm) DBH compose more than 10% of the crown density in the stand, the stand is recorded as a sawtimber stand. If this requirement is not met, the surveyor would move down to the next largest size class and apply the same test. Forest type (FTYPE) is done in a similar manner but without a density requirement. The species that is observed to dominate the stand is recorded as the stand's FTYPE.

Table 1: Table of stand size class, forest type, and stocking levels and the possible values of each within the study area. DNRC maps seven additional classes of forest type which were not found to occur on the LSA.

Stand size class (code)	Forest type (abbreviation)	Stocking
Nonstocked (6)	Subalpine fir (AF)	Nonstocked
Seedling/Sapling (7)	Aspen (AS)	Low
Pole timber (8)	Cottonwood (CO)	Adequate
Saw timber (9)	Douglas fir (DF)	
	D. fir/W. larch (DL)	
	Engelmann spruce (ES)	
	Grand fir (GF)	
	Lodgepole pine (LP)	
	Mixed Conifer (MC)	
	W. Red Cedar (RC)	
	W. larch (WL)	
	W. white pine (WP)	
	Nonforested (NF)	

Table 2: Table of stand size class, which depends on both the size and density of trees. Nonstocked stands are considered to be deforested or sparsely vegetated.

Stand Size Class (SSC)	DBH (in.)	Crown Density Requirement
Sawtimber (SSC = 9)	> 9	> 10% crown density
Poletimber (SSC = 8)	5 – 9	> 10% crown density if sawtimber density is < 10%
Seedling/Sapling (SSC = 7)	Less than 5	> 10% crown density if poletimber density < 10% and sawtimber density < 10%
Nonstocked (SSC = 6)	NA	

Table 3: Table of stocking criteria for each size class. Seedling/sapling stocking is based on trees per acre, while pole and sawtimber stocking is derived from the basal area per acre (ft^2/ac.).

	Seedling/Sapling (trees/ac)	Poletimber (ft^2/ac)	Sawtimber (ft^2/ac)
Nonstocked	0 – 50	0 - 4	0 - 10
Low	51 - 100	5 - 9	11 - 39
Adequate	101+	10+	40+

While the DNRC eventually hopes to predict a wider range of forest characteristics using remotely sensed data, this analysis will focus on only the three that are needed for the SYC for two reasons. First, these are perhaps the most useful in differentiating one stand from another in terms of the market value. With only these three metrics, one can determine the utility of a timber stand with some degree of certainty. Secondly, many of the other parameters of interest for the DNRC are quite similar to these. For instance, they are also interested in dominant species at each canopy level which is essentially a tiered FTYPE, and habitat type which is a forest type that includes both upper and lower level species. Therefore, if the SYC parameters can be predicted, the same methodology could possibly be adapted to include the larger list of parameters of interest.

2. INVENTORY PLOT SELECTION

2.1. Introduction

Fixed-area plots are commonly used to sample elements of a population that are distributed spatially over a landscape. Sampling locations are often randomly selected points that serve as the center of a circular plot (Gregoire & Valentine, 2007). To select inventory plot locations in 2018, the DNRC used a modified simple random sample with replacement method, assessing that sample with the Kolmogorov-Smirnov test (K-S test) to compare, in spectral space, the sample distribution to that of a larger sample being used to represent the full population.

This method was outlined by John Hogland and executed by forestry interns in 2018 (K. Shank, pers. comm., Feb. 2019). To begin, two simple random samples with replacement (SRSwR) were drawn within the study area (n = 120 and N = 5,000). Sample 1 represented potential plot locations, while the second sample (N=5,000) was meant to oversample the study area to capture all variation, allowing it to serve as a proxy for the population. Focal mean, focal standard deviation, and horizontal contrast from gray level co-occurrence matrices (GLCM) were derived from NAIP imagery and extracted for each site. The final sample (n = 103; Figure 2) was selected by comparing the distribution of various subsets of Sample 1 to find the one that best matched the distribution of Sample 2 (the proxy population) using the K-S Test p-value (K. Shank, pers. comm., Feb. 2019).

Determining how representative a sample is relative to the population of interest is highly dependent upon the user's definition of representative (Gregoire & Valentine, 2007). The DNRC's method depends on the Kolmolgorov-Smirnoff test to assess the level of agreement between sample and population distributions. This test compares the cumulative frequency distribution of the population with the cumulative step-function of a random sample by determining the maximum distance between distributions (Massey, 1951). The d-statistic

represents the maximum absolute difference between the two distributions and can be converted into a p-value based upon the sample size to determine if the two samples were drawn from the same population. If the p-value is below the desired significance level then you would reject the null hypothesis that the two samples were drawn from the same population (Massey, 1951).

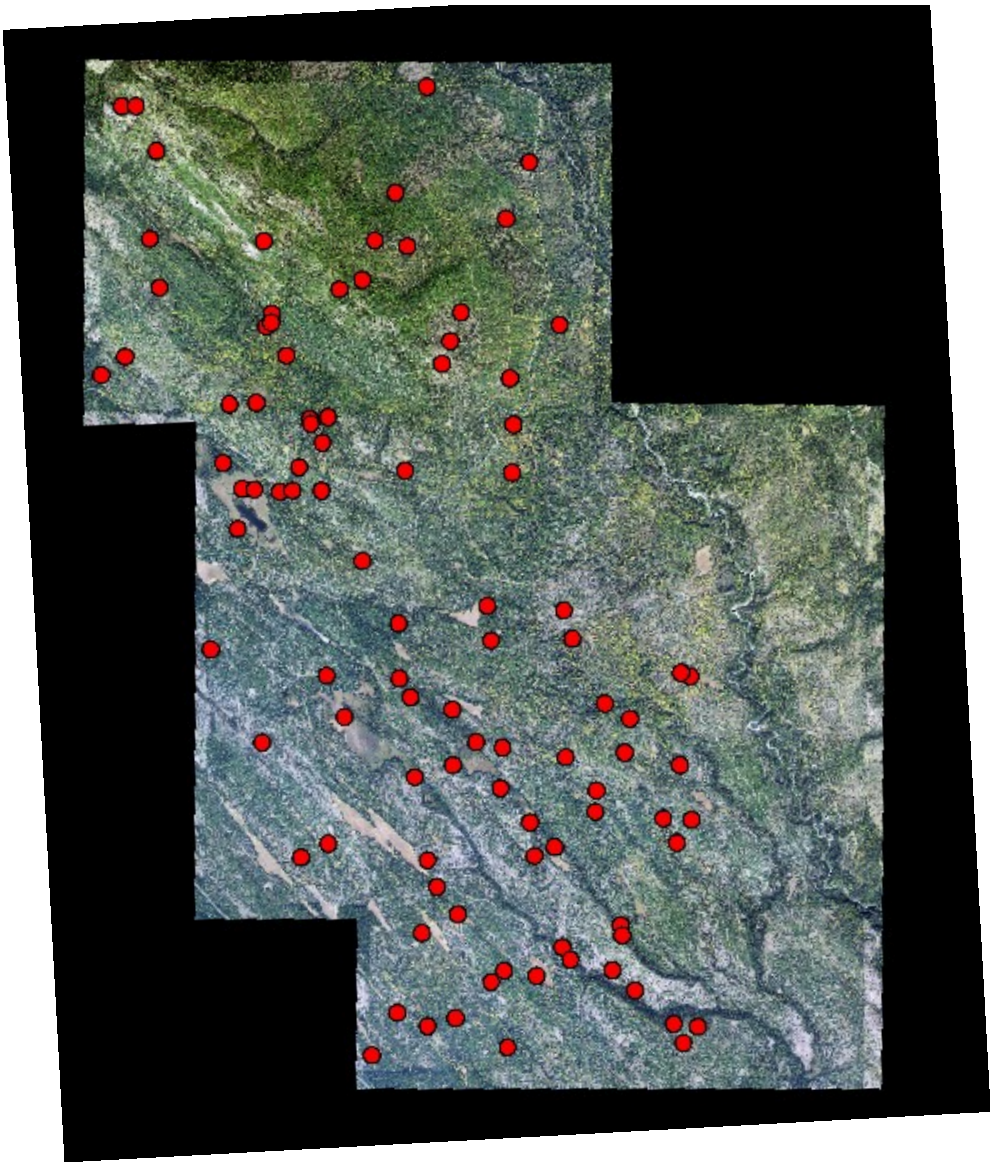

Figure 2: Map of locations selected by the DNRC using the SRSwR/K-S Test selection method and visited in 2018. The plots appear biased toward the western side of the LSA and spatially clustered.

DNRC also identified "deferred areas" within the LSA which they were not interested in for timber management: roads, rivers, wetlands, lakes, and other non-forested areas. These areas were identified but not removed from the sampling or the simulated population and data were collected in the field at some of the plots that fell within these deferred areas.

2.2. Objective

The purpose of this chapter is two-fold. It will first compare two alternative methods of drawing samples from the population of interest using the Generalized Random Tessellation Stratified (GRTS) design. Then, it will suggest another method for assessing the degree to which a sample captures the desired population variability.

GRTS sampling appeared advantageous after examining the original plot locations selected by DNRC for two reasons. First, the original sample locations did not appear to be distributed across the entire study area (Figure 2), which because of the first law of geography, led to some skepticism as to whether the sample really captured the full range of variability in the population. Secondly, the GRTS algorithm can be utilized to operate in auxiliary space so that samples are evenly spread over a spectral space rather than a geographic one (Stevens & Olsen, 2004) ensuring sites would include the complete range of spectral reflectances.

2.3. Background

2.3.1. GRTS in Geographic Space

The Generalized Random Tessellation Stratified (GRTS) design was developed to monitor natural resources by Stevens and Olsen in 2004. The algorithm places a series of nested 2×2 grids over the study region and transforms the 2-dimensional space into a 1-dimensional list of cell addresses, where cell addresses have been locally randomized. Addresses in the list are then selected and mapped back to their actual locations so that potential plot locations are

randomized locally but also spatially distributed across the entire region of interest (Stevens & Olsen, 2004).

2.3.2. GRTS in Spectral (Auxiliary) Space

While the GRTS algorithm was designed to spatially distribute samples across a study region, it can be manipulated to sample any auxiliary domain that represents the study area. A useful application of this for vegetation, is sampling across spectral space, using an image of a study area. The commonly available implementations of the GRTS algorithm only allow for two axes however, so a multi-band image needs to be reduced to just two inputs. An efficient way to accomplish this is to perform a principal component analysis (PCA) on the image and retain only the first two components for use as inputs. In a PCA, successive components include progressively less of the total variation, and the first two components generally represent more than 90% of the total spectral variance in bands from a multi-spectral satellite such as Landsat (Lillesand, Kiefer, & Chipman, 2008).

2.4. Methods

2.4.1. Sampling

In order to assess the sample-to-sample variation from the different designs, a total of 3000 samples were drawn: 1000 using a simple random sample with replacement (SRSwR) to reproduce the DNRC's approach, 1000 using GRTS in geographic space, and 1000 using GRTS in spectral space. Each individual sample consisted of 100 sample cells/points representing plot locations. These plot locations were drawn in R (R Core Team, 2017) across the entirety of the study area. In order to save on processing time and avoid computer memory limitations, the NAIP 2015 image of the study area was resampled from 1 meter (3.28 ft) to 10 m (32.8 ft) pixels.

The easiest to implement was the simple random sample with replacement (SRSwR). Using the *raster* package (Hijmans, 2019) in R (R Core Team, 2017), a simple random sample of 100 cells was drawn from the image and the values of the four (red, green, blue, and near infrared) bands in the image were extracted at each chosen plot location. Mean, standard deviation, minimum and maximum values from the 100 cells (plot locations) were then calculated for that one sample.

The process for drawing samples using the GRTS design involved drawing samples in both geographic and in auxiliary (spectral) space. To select plots in geographic space, the GRTS algorithm from the *spsurvey* package (Kincaid, Olsen, & Weber, 2019) in R (R Core Team, 2017) was used. To implement this function, users must define a design object, indicating the strata, sample size, and selection method and provide a frame to evenly distribute plot locations across. The extent of the study area for the geographic GRTS design was defined by a shapefile matching the extent of the resampled NAIP 2015 image which matched the bounds of the LSA. For the auxiliary (spectral) space, the resampled NAIP image was transformed using a principal component analysis (Appendix A, Fig. A1 & Tab. A1) and samples were spread evenly across spectral space using principal components 1 and 2 as false x and y coordinates. This allowed the inputs to mimic the orthogonal nature of x and y coordinates. In both cases, plots were selected using an unstratified equal probability GRTS design with a sample size of 100, again using the *spsuvey* package (Kincaid, Olsen, & Weber, 2019) in R (R Core Team, 2017).

The one thousand samples of each design type were plotted and compared statistically to the full population of 10 m (32.8 ft) pixels.

2.4.2. Statistical Assessment

Each time a sample was drawn, summary statistics (mean, standard deviation, maximum, and minimum) in each of the four (blue, green, red and near-infrared) bands NAIP 2015 were calculated. In addition to spectral information, geographic information was also summarized for all of the

samples to examine the distribution of plots in geographic space. The mean easting and northing of all samples was reported as well as the mean size of the Voronoi polygons generated from the sample. These polygons were derived from Delaunay Triangles that connect plot locations such that no point is inside the circumcircle of any triangle. By connecting the centers of circumcircles, Voronoi polygons are created which represent the region of area that is closer to the plot it includes than any other. Voronoi polygons allow for a quantification of the amount of "area", each point represents in the study region (Evans & Jones, 1987).

2.5. Results

2.5.1. Spectral Statistics: Sample Distributions

Spectrally, the geographic GRTS and the SRSwR designs yielded similar results. The distribution of means (Figure 3) and standard deviations (Figure 4) in all spectral bands were almost identical. The spectral GRTS however, had means and standard deviations more tightly clustered around the population mean and standard deviation in all spectral bands (Table 4 and Table 5).

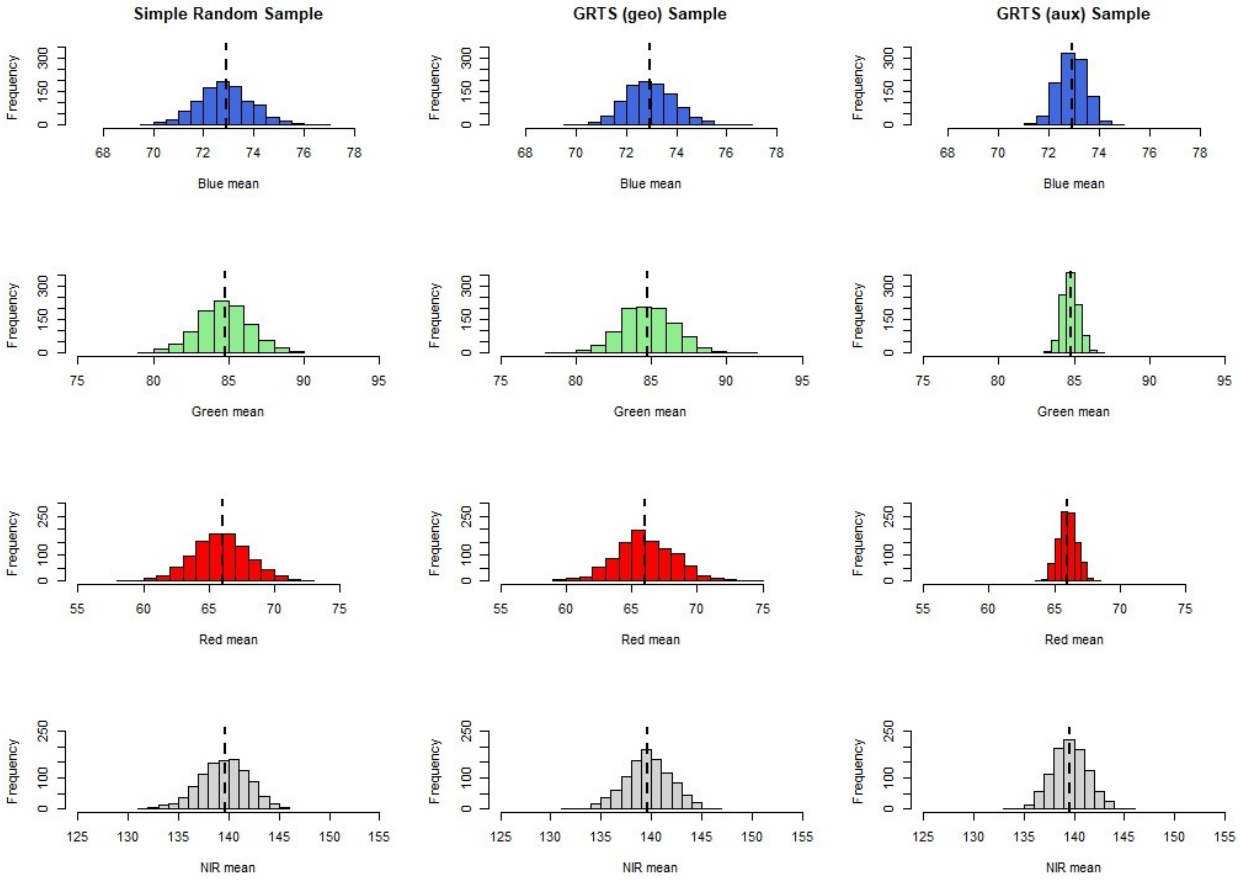

Figure 3: Histograms of the sample means within each spectral band. The rows show the blue, green red and near-infrared bands, respectively, from top to bottom. The dashed vertical line is the population mean.

Table 4: Table of the mean and standard error within each spectral band from each of the three sampling designs and the population. The values are in digital numbers and correspond to the histograms in Fig. 3.

	Population	SRSwR		GRTS Geo.		GRTS Aux.	
	Mean	Mean	SE	Mean	SE	Mean	SE
Blue Band	72.9	73.0	0.11	72.9	0.10	72.9	0.06
Green Band	84.8	84.7	0.17	84.8	0.18	84.8	0.05
Red Band	66.0	65.9	0.22	66.0	0.22	66.0	0.07
NIR Band	139.6	139.5	0.25	139.6	0.23	139.6	0.17

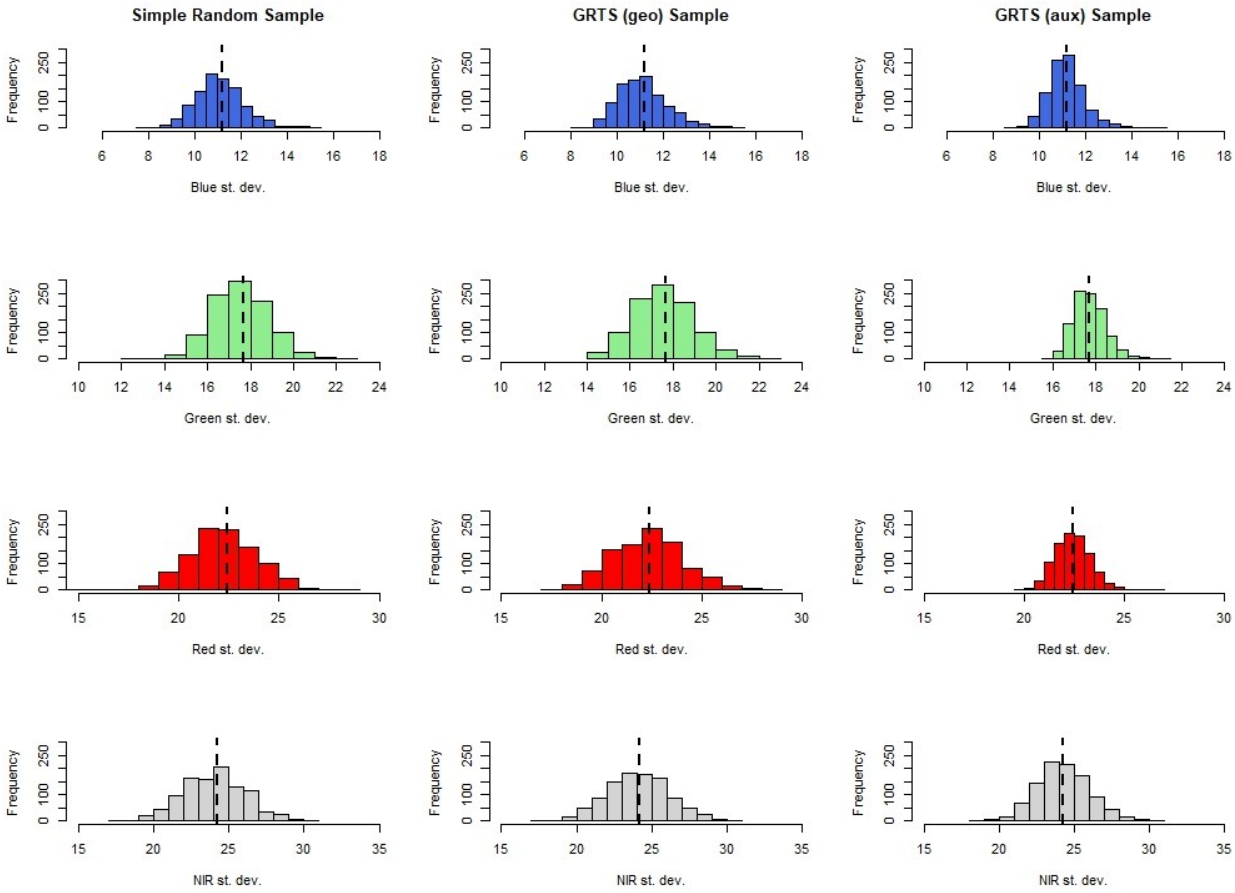

Figure 4: Histograms of the standard deviation of each sample in each spectral band. The population standard deviation is shown as a dashed vertical line. The rows represent values from the blue, green, red and near-infrared from top to bottom, respectively.

Table 5: Table showing the average standard deviation and variance of samples within each spectral band from each of the three designs and the population. The values are in digital numbers and correspond to the histograms in Fig. 4.

	Population	SRSwR		GRTS Geo.		GRTS Aux.	
	Std. Dev.	Avg. SD	Variance	Avg. SD	Variance	Avg. SD	Variance
Blue Band	11.2	11.1	1.1	11.1	1.1	11.2	0.6
Green Band	17.7	17.6	1.7	17.6	1.9	17.7	0.6
Red Band	22.4	22.2	2.9	22.3	3.2	22.4	0.8
NIR Band	24.2	24.1	4.2	24.1	4.3	24.3	3.1

With respect to the range of spectral values, all designs fell short of capturing the full range from each band in the population (Figure 5). All samples better approximated the minimum values than the maximum values, as all sample designs fell greatly short of the population maximum in all bands (Figure 5 and Table 6). Values from the raster image of the population can be found in Appendix A (Figure A2).

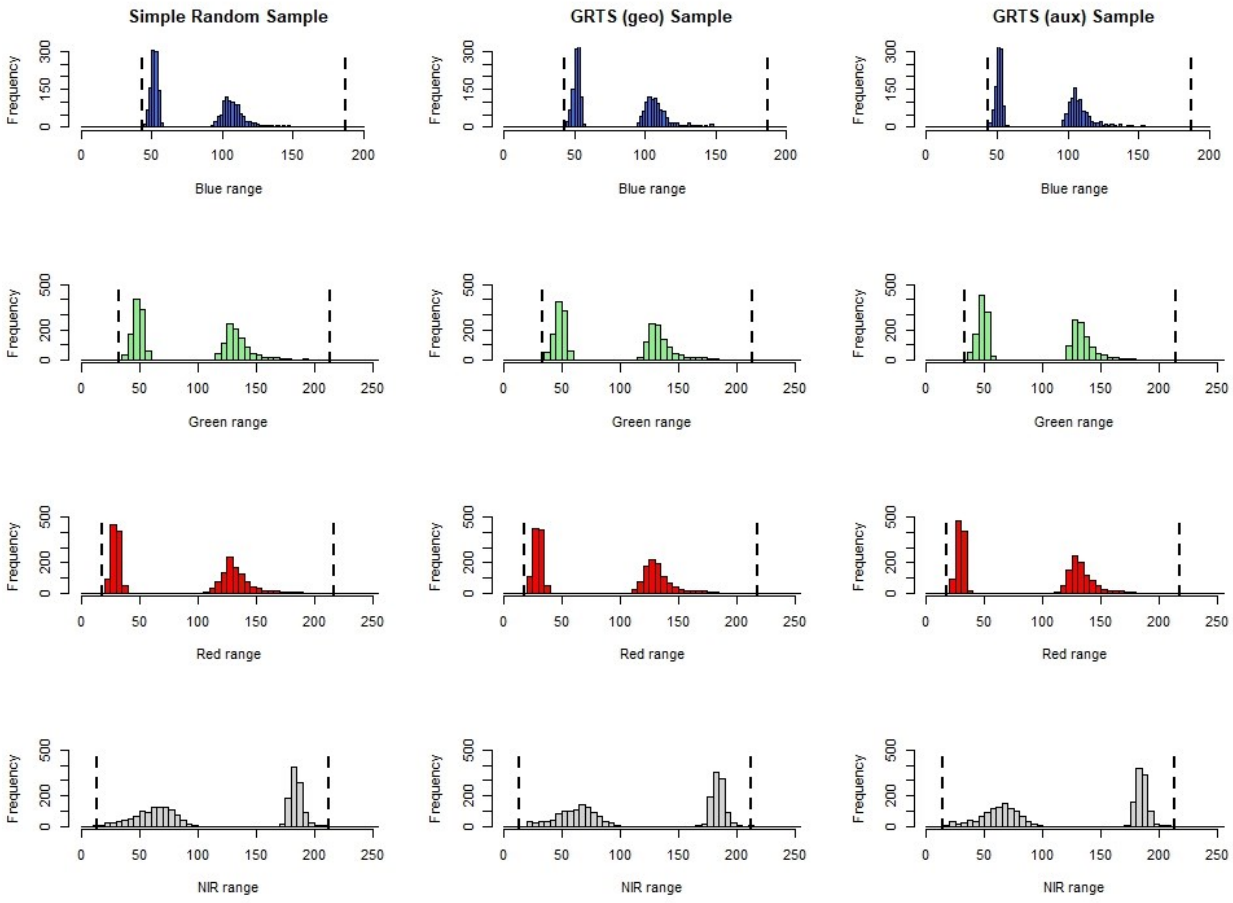

Figure 5: Histograms of the range of samples in each spectral band. These bimodal distributions show the distribution of sample minimum and maximum values in each band (blue, green, red, and near-infrared from top to bottom). The dashed lines display the population minimum and maximum for each band.

Table 6: Table showing the mean range values of samples from each of the three designs. These values are given in digital numbers and correspond to the bimodal histograms in Fig. 5.

	Population		SRSwR		GRTS Geo.		GRTS Aux.	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
Blue Band	43.4	187.0	51.6	109.8	51.4	109.6	51.3	109.8
Green Band	32.9	214.0	48.6	135.4	48.5	135.5	48.0	135.8
Red Band	17.6	218.0	29.6	133.0	29.6	132.7	29.2	133.5
NIR Band	13.9	213	61.9	184.2	61.4	184.3	61.8	184.6

2.5.2. Spatial Statistics: Sample Distributions

The geographic GRTS yielded samples with mean eastings most similar to the population mean (Figure 6). The mean easting and northing of all samples were within ~300 m (980 ft) of the population mean, or the center of the study area. Mean easting and northing values for the SRSwR and spectral GRTS, were as far as 1000 m (3280 ft) and 700 m (2300 ft), respectively, from the population mean (Figure 6), although still centered on the population mean.

The mean Voronoi polygon size for samples drawn using geographic GRTS was the largest of the three sampling designs, at 142.6 ac (57 ha) as opposed to 138.3 ac (56 ha) or 139.5 ac (56 ha) from the SRSwR and spectral GRTS, respectively (Figures 6 and 7, and Table 7). The Voronoi polygon sizes for samples from the geographic GRTS design were less variable compared to the other 2 designs (Figures 6 and 7, and Table 7) with most polygons containing a similar amount of area.

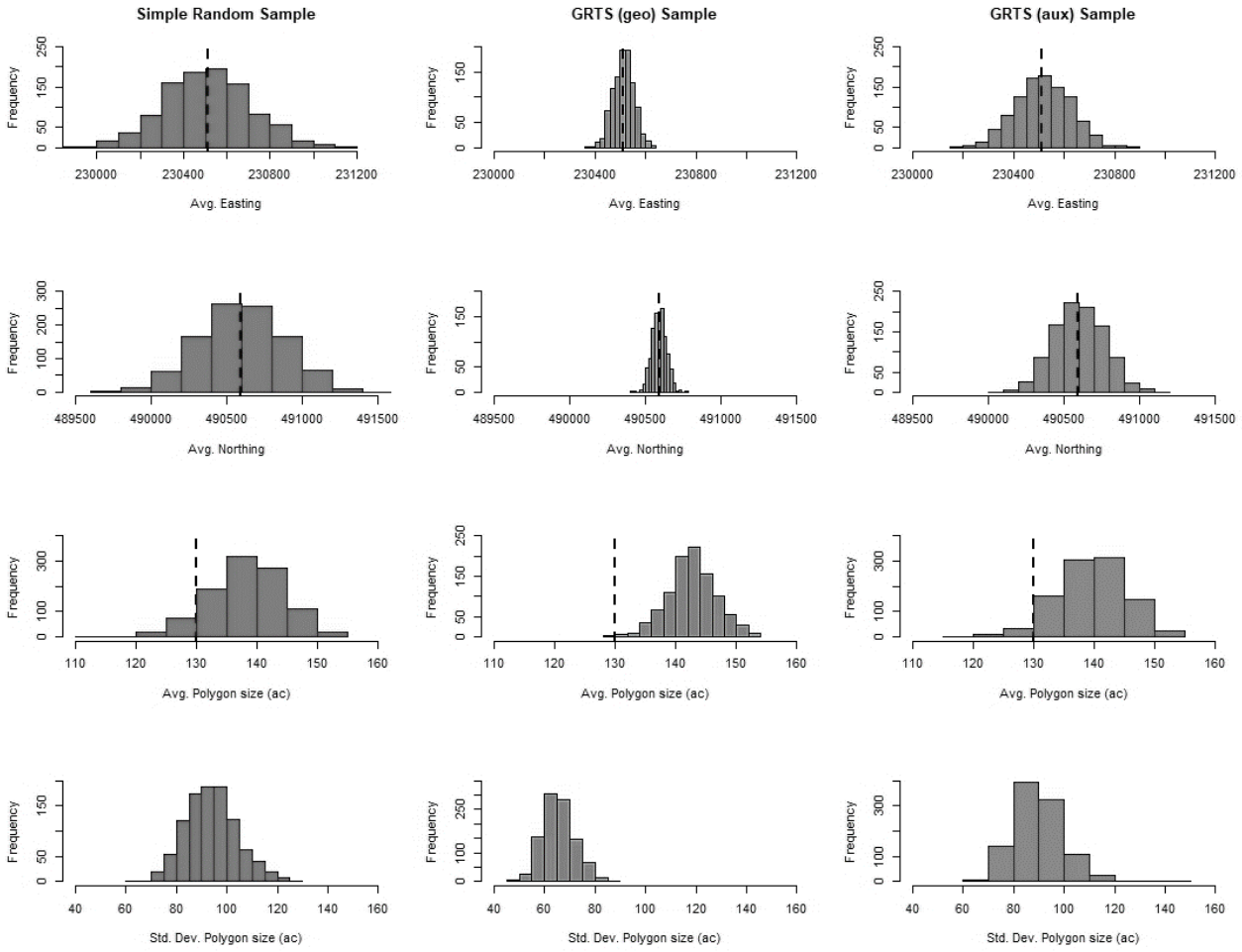

Figure 6: Histograms of the spatial statistics from each of the three sampling designs. The rows represent the mean easting, northing, Voronoi polygon size, and standard deviation of the polygons from top to bottom, respectively. The dashed lines in rows one and two represent the easting and northing of the center of the study area, while the dashed line in row three is the area of the study area divided by 100.

Table 7: Table showing the mean easting, northing, Voronoi polygon size, and standard deviation from each sampling method which are plotted in Figure 6. The rows in the table match the rows in the figure.

	Population	SRSwR	GRTS Geo.	GRTS Spectral
Mean Easting (m)	230513	230514	230512	230517
Mean Northing (m)	490589	490598	490589	490597
Mean Voronoi Polygon Size (ac)	NA	138	143	140
Mean Std. Dev. Voronoi Polygon (ac)	NA	94	66	90

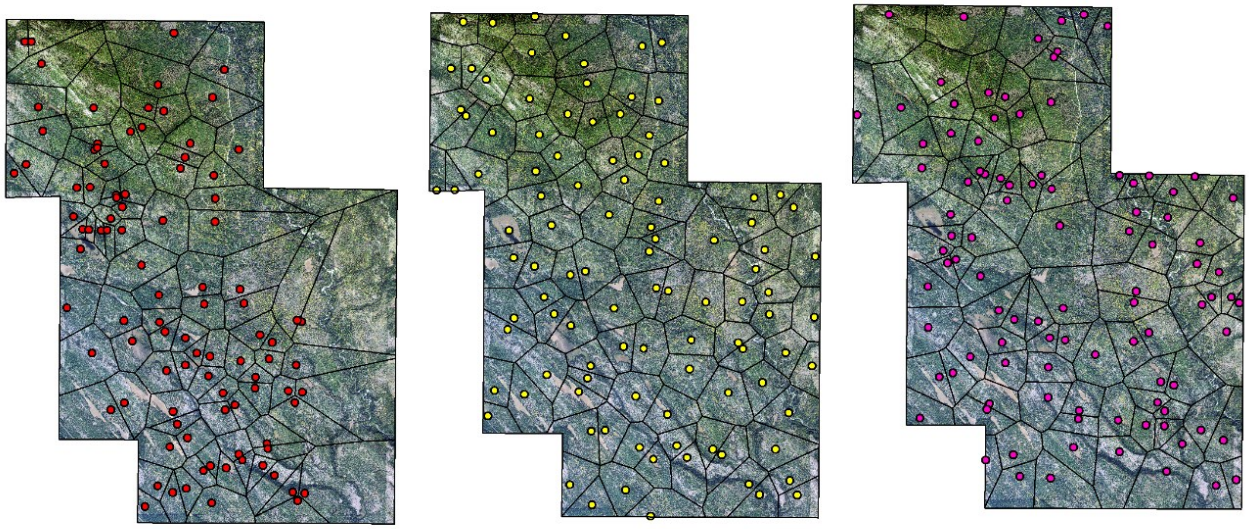

Figure 7: Maps of the Voronoi polygons generated from just one sample of each design (black lines around the prospective plot locations). The simple random sample map shows the plot locations the DNRC used for data collection in 2018.

2.6. Discussion

2.6.1. Spectral Distribution

All designs were capable of producing unbiased estimates of the population mean and standard deviation in each of the four spectral bands (Figures 3-4, Tables 4-5). That all designs fell short of capturing the full spectral range of values in the population (Figure 5 and Table 6) is not surprising given that each sample only contained 100 points. The full range of values in the population includes uncommon land-cover types that could only reliably be captured if the sample size was increased to impractical levels or those areas were specifically sampled through stratification. The distributions of spectral means for all bands were tightly clustered about the population means when samples were drawn using the GRTS in spectral space (Figure 3 and Table

4). This is a product of the inputs used. In this case, the NAIP image representing the population was transformed using a principal component analysis (PCA) which altered the population. The GRTS algorithm from the *spsurvey* package (Kincaid, Olsen, & Weber, 2019) only allowed two input axes, and the first two principal components contained 51% and 33% of the variability in the image (Appendix A, Fig. A1), respectively, essentially reducing the spectral information available to guide sample selection. While this may remove noise, it also may have removed useful information, which could be avoided if all four principal components were used as inputs. However, any sample drawn using the spectral GRTS methodology outlined here would have a higher chance of approximating the population than the other two designs.

Samples from any of the three designs explored here can produce an unbiased sample that approximates the spectral values in the population (Figures 3-5, Tables 4 - 6), but the spectral GRTS will produce samples that more precisely estimate spectral population parameters.

2.6.2. Spatial Distribution

According to Tobler's first law of geography, things that are closer together are more related than things that are farther apart (Tobler, 1970). When applied to selecting samples, if locations are randomly clustered in one area, it could homogenize the sample. Physical characteristics of the landscape (such as elevation, slope, soils, cover type, spectral reflectance, etc.) are likely to be more similar when sampling locations are closer together. In this respect, using a design like GRTS in geographic space may be advantageous.

The geographic GRTS effectively sampled the entirety of the study region and distributed the samples evenly (Figures 6-7; Table 7). Clusters of samples concentrated in any portion of the study area (as seen with the SRSwR) were eliminated. Additionally, each potential plot location more consistently represented approximately the same amount of area than the other methods

(Figures 6-7; Table 7). Visualizing plot locations using Voronoi polygons can show what portions of the study area may be oversampled and what areas may be under-sampled (Figure 7).

Spatial clustering was much more likely when samples were either random or balanced within spectral space. Spatially imbalanced samples can result in the exclusion of classes that cover less of the study area, like tree species with narrower habitat requirements or stands of old growth that may be isolated. Additionally, it may cause problems when sites are visited for data collection, for instance, if sample locations are highly clustered and the area is inaccessible for some reason (i.e. flooding, fire, land change, timber sale, etc.), then all of those sample locations will be more costly, or even impossible, to measure.

2.7. Conclusion

While all the sample designs examined might produce samples that adequately represent the population of interest, using a GRTS design in either geographic or spectral space offer advantages over the simple random sample with replacement (SRSwR).

If the spectral information in the population is of primary concern, using GRTS in spectral space ensured a higher likelihood of representing this aspect. It will not however, necessarily provide a spatially distributed sample.

The GRTS in geographic space ensured a spatially distributed sample. This strategy would be especially advantageous in a highly topographically variable landscape, with correspondingly complex vegetation. In the original study by DNRC, plot locations were used to identify forest characteristics including species, forest type, and habitat type. The highly spatially clustered samples from the SRSwR certainly resulted in the oversampling of certain species/forest types/habitat types and the exclusion of others.

2.8. Future Work and Research

Whatever sample design is chosen, a sample can easily be compared to a population by using the sample statistics and parameters derived from the population of interest (Tables 4 - 7). This comparison of sample to population serves the same purpose as the K-S test and can result in samples which mimic the population spectral and spatial information. The K-S test may not be the best method for assessing samples. The original sample selected by DNRC passed the K-S test but resulted in a sample that was deemed to not represent the population of interest (Wayne Lyngholm, pers. comm., Dec. 2018). The sample was so unrepresentative of the population that it was ultimately discarded and crews were sent out the following summer to census the entire study area. Additionally, the K-S test as it was implemented by DNRC in 2018 only incorporated spectral information. This immediately makes the assumption that the variables that will be predicted from the data can be differentiated using only spectral information when in reality, this may not be the case.

A GRTS design implemented in geographic or spectral space was shown to offer inherent advantages over the SRSwR method in this analysis. Another option would be to use topographic and climactic variables to stratify the study area, and ensure that plot locations are well-distributed in this regard as well. Stratification can be incorporated into the GRTS methodology easily and would likely reduce or eliminate the exclusion of less common variable combinations, as the spatial distribution of vegetation is a product of topography, soils, and climate variation. Additionally, although the most common implementation of the GRTS algorithm only allows for two input axes, there is no reason that the algorithm could not be implemented in more than two dimensions. This would allow for the possibility of incorporating both spectral and spatial information (or even additional variables) in tandem to guide sample selection.

Finally, the deferred areas were not removed from the sample or the population in this analysis. In the original process undertaken by the DNRC, deferred areas were identified but not removed. Upon closer inspection, the deferred areas layer only partially masked out these undesired land-cover types, and they were sometimes included in portions of the analysis. Creating a more accurate deferred area layer to use as a mask prior to drawing any samples would increase the likelihood that a sample represents the managed forest landscape well.

3. IMAGERY SOURCE COMPARISON

3.1. Introduction

The coupling of machine learning algorithms and remotely sensed data has surged due in part to the availability of high-quality data and an increase in computing power. Land cover and vegetation mapping using this combination have been the focus of a large number of studies over the past decade (Hogland et al., 2018; Jones et al., 2018; Zhu & Liu, 2014).

Many remote sensing studies in the United States make use of high-resolution, 4-band imagery from the National Agricultural Imagery Program (NAIP) e.g. Hogland and Anderson (2015), Davies et al. (2010), and Hayes et al. (2014). NAIP acquires 1 m or 60 cm (3.28 ft or 1.97 ft) resolution imagery during the growing season every other year across the continental Unites States (USDA Farm Service Agency, 2019).

Other than NAIP, one other source of imagery was explored in this analysis: Sentinel-2. The European Space Agency's Copernicus Sentinel-2 mission is comprised of two polar orbiting satellites with a revisit period of roughly 5 days. The constellation has been collecting data since 2015 in 13 spectral bands at spatial resolution varying from 10 – 60 m (32.8 – 196.9 ft). This imagery has been free to access since launch and multiple studies have used Sentinel imagery to map vegetation (e.g. Liu et al. (2018), Immitzer et al. (2016), and Mura et al. (2017)).

Random forest is a supervised ensemble classifier that has increased in popularity dramatically due to its ability to efficiently and effectively produce results, while requiring users to set few model parameters (Belgiu & Dragut, 2016). The algorithm works by creating a forest of decision trees using randomly selected variables to partition the data. Results of these trees are then evaluated against out-of-bag data (data withheld from model training) to determine which split variables result in the highest classification accuracy of the target dependent variable (Breiman, 2001).

Google Earth Engine (GEE) is a cloud-based raster analysis platform that leverages Google's servers to permit computationally expensive operations to be completed from any computer with an internet connection. The public data catalog is extensive (currently it includes the entire Landsat and Sentinel archives and many other geospatial datasets) and it is continuously updated with nearly 6000 scenes per day from active missions (Gorelick, et al., 2017). The incorporation of machine-learning classifiers into the GEE environment has permitted the completion of regional and global classifications that would be otherwise impossible (Robinson, 2017; Moreno-Martinez, et al., 2018; Jones, et al., 2018).

The purpose of this chapter is to compare the accuracy of RF classifications of forest characteristics using different sources of imagery available at no cost to the DNRC. NAIP imagery offers a higher spatial resolution, while Sentinel-2 provides spatially coarser data across a broader range of the radiometric spectrum at more frequent intervals.

3.2. DNRC Pilot Project

The 2018 Raster Analysis Pilot project completed by the DNRC only used NAIP imagery from 2015 to create predictor variables with which to construct classifications.

3.2.1. Pilot Project Imagery and Predictors

In Montana, four-band NAIP imagery was collected at a spatial resolution of ~1 m (3.28 ft) biennially from 2009 – 2015 and at a spatial resolution of 60 cm (1.97 ft) in 2017. This imagery is free to access and curated by the Montana State Library in local datums and projections.

For the pilot project, 13 predictor surfaces were generated from the NAIP imagery corresponding to the mean (4) and standard deviation (4) in each band, the gray-level co-occurrence matrix horizontal contrast in each band (4), and a normalized difference vegetation index (NDVI; 1). The focal mean of each surface was computed at the scale of the plot.

3.2.2. Pilot Project Field Data Collection

During this pilot project, predictor variables were extracted at 98 sites that were visited by inventory technicians in the summer of 2018. Plot centers were selected at random and each plot represented a 36 × 36 m (118 × 118 ft) square about the chosen center. These square plots were then divided into four circular subplots each with a 9 m (29.5 ft) radius. Each subplot contained a nested subplot with the same center but a radius of only 2 m (6.6 ft; Figure 8).

Data was collected at multiple scales at each sampling location. Stand size class (SSC) and forest type (FTYPE) determinations were made while walking through the larger plot area (36 × 36 m; 118 x 118 ft). Species and diameter at breast height (DBH) were recorded for all trees within each of the four subplots. To reduce collection efforts, smaller trees (< 9 in. DBH; 23 cm DBH) were measured within 2 m (6.6 ft) of the subplot center while larger trees (≥ 9 in. DBH; 23 cm DBH) were measured within 9 m (29.5 ft) of the subplot center.

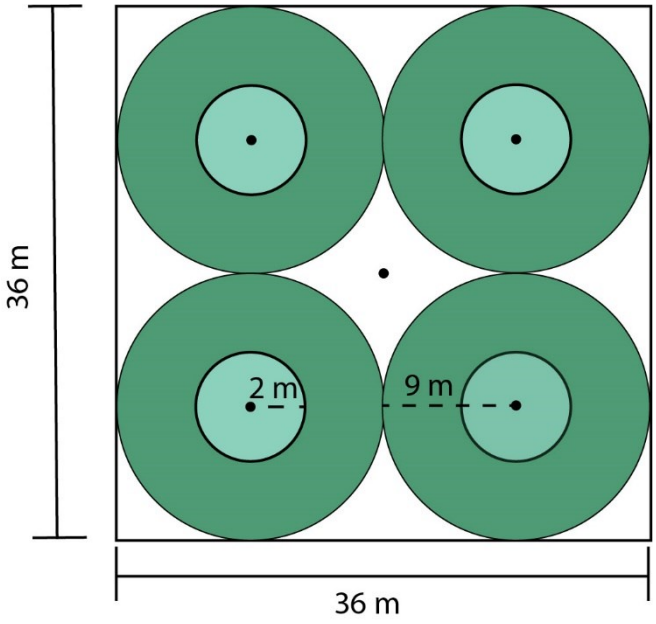

Figure 8: Diagram of the sub-plots within each 36 by 36 meter (118 x 118 ft) plot at which some variables were measured (diagram not to scale).

3.2.3. Classification

Field data collected at subplots were aggregated to the plot level. Plot centers were used to extract mean values across the plot for each predictor and these values were combined with field data. This new dataset was used to train random forest models to predict the desired forest characteristics which included SSC, FTYPE, and BAA, along with others of interest to the DNRC. Initial model results did not match what crews had reported and DNRC began discussing how to improve the predictive power of their models. Exploring different sources of imagery from which to derive predictor surfaces was determined to be the logical place to start.

3.3. Methods

3.3.1. Field Data Re-processing in 2019

There had been inconsistencies in the recorded values for SSC collected in 2018, primarily when a plot was located in an area that was not forested. Sometimes a value of 0 for SSC was recorded if a plot was inaccessible due to standing water or the plot was located in a river, but sometimes such a plot would be given a value of 6 (Nonstocked). An FTYPE value of "NF" was recorded at some of the plots that fell in these undesired areas.

Because this inconsistency would cause problems for classification, plot locations were examined using aerial imagery. An SSC of 0 was used to denote wet meadows (common in the LSA), while an SSC of 6 was used to denote areas of sparse vegetation. After this process was complete only a few data points existed in each of those classes, so supplemental points were digitized by the author using the imagery in wetland and nonstocked areas, given the appropriate SSC value (0 or 6), an FTYPE value of "NF" (Figure 9), and values of zero were recorded for TPA and BAA. Although class 6 could potentially have a small number of trees, it was assumed that trees made up less than 10% of the total area, based upon visual analysis.

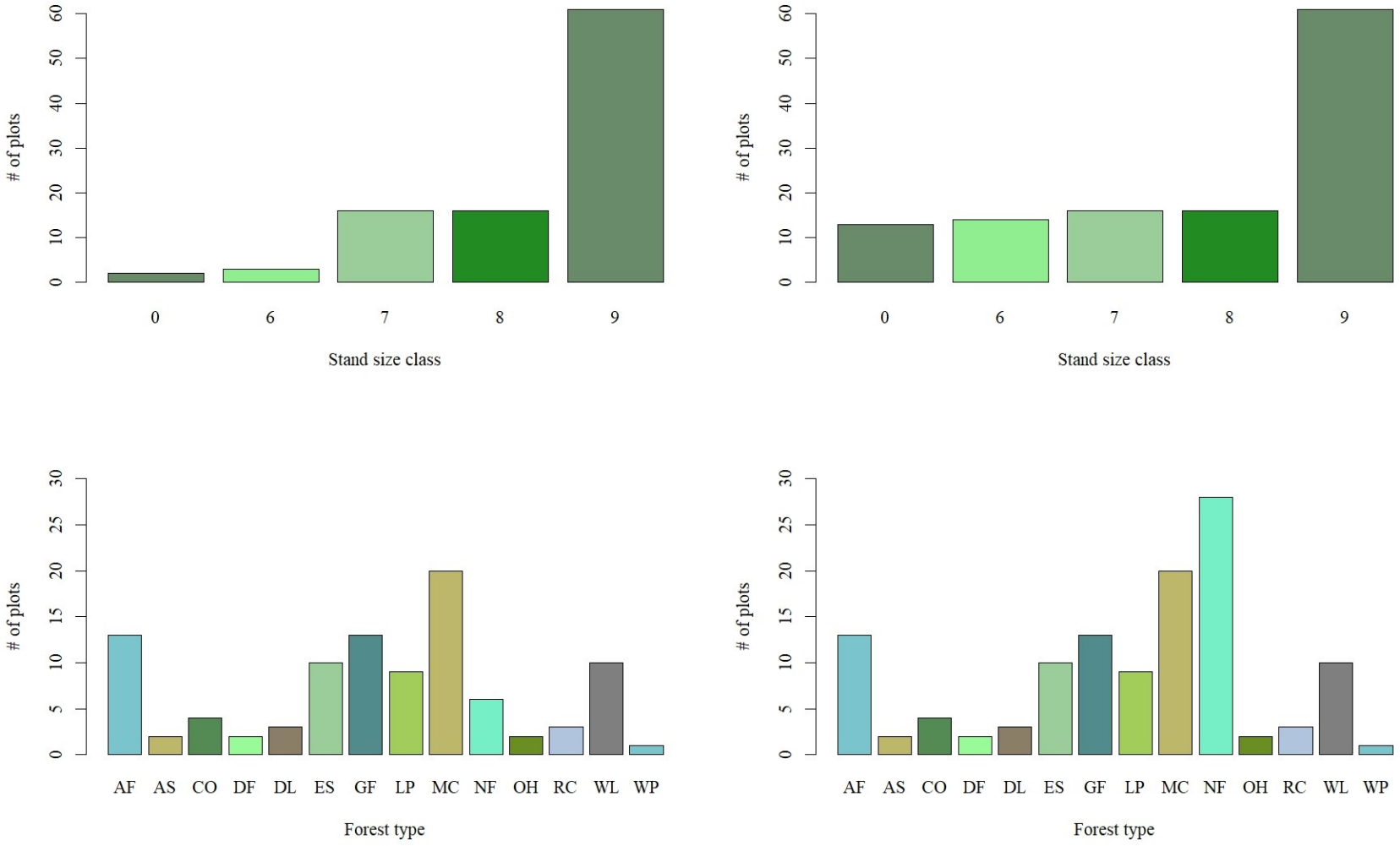

Figure 9: Pilot project plot data from 2018 (left column). Supplemental sites were added after the low vegetation classes were re-categorized (right column). Size class (top) and forest type (bottom) codes are in Table 1.

3.3.2. Imagery

Four sets of imagery were assembled for the study area using either NAIP or Sentinel-2 (Table 8) and used to derive predictor variables for comparison. NAIP for the area was collected in late September in both 2015 and 2017. Although NAIP 2017 was collected at a 60 cm (1.97 ft) spatial resolution, the product available through the GEE platform was scaled to 1 meter (3.28 ft) by Google for consistency across the collection.

Table 8: Table of the characteristics of the imagery and the bands utilized. Reducer refers to the way in which multiple scenes were combined into a single. The red-edge bands (RE 1 – 3) were resampled to 10 m (32.8 ft) from 20 m (65.6 ft) prior to analysis to match the other bands.

Predictor Image Source	**Spatial Resolution**	**Date(s)**	**Reducer**	**Bands Used**	**Bands**
NAIP 2015	1 m	Sept. 30, 2015	None	4	R, G, B, N
NAIP 2017	1 m	Sept. 28, 2017	None	4	R,G, B, N
Sentinel – 2 Oct.	10 m	Oct. 16, 2018	None	4	R, G, B, N
	20 m	Oct. 16, 2018	None	3	RE1, RE2, RE3
Sentinel – 2 Growing Season	10 m	May 12, 2018 – Sept. 6, 2018	Max	4	R, G, B, N
	20 m	May 12, 2018 – Sept. 6, 2018	Max	3	RE1, RE2, RE3

A single Sentinel-2 scene from 16 Oct. 2018 was selected as it was the closest (seasonally) cloud-free image to the timing of the NAIP. 2018 imagery was used because that was closer to the initial surveying of the plots. Additionally, a maximum growing season composite was created using 14 cloud-free images from 12 May 2018 to 6 Sept. 2018 with only the maximum value at each pixel from any date.

3.3.3. Response Variables

The 4 response variables were SSC, FTYPE, BAA, and TPA (Tables 1 - 3). Two of these were categorical and two numeric, so different types of decision trees were used to grow the random forest and different statistics were used to evaluate the success of the random forest classifications

(Table 9). Root mean squared error (RMSE) and mean absolute error (MAE) both report the differences in observed and actual values in the units of the variable of interest. The Kappa statistic compares the classification results to the results of a random assignment on a scale from 0 - 1, with 0 being no better than random chance. The Kappa calculation used here is a part of the Accuracy Assessment tool in the RMRS Raster Utility (Hogalnd and Anderson, 2010). Because the raw data for determining stocking were collected as numeric values (trees/ac or basal area/ac), TPA and BAA were used as response variables rather than the categorical variable of stocking level. This requires end-users to place the data into appropriate stocking categories, but allowed more thorough analysis of sources of confusion in the algorithm. Changing the data to categorical too early can reduce the amount of information within each pixel and lower the accuracy of resulting predictions.

Table 9: Table of characteristics of the four predicted variables as well as statistics used to evaluate the overall performance of the associated models, the individual contribution of input variables to the RF model, and validation or evaluation statistic.

Predicted variable	**Type**	**Number of classes**	**Decision tree type**	**Model performance statistic**	**Variable contribution statistic**	**Stand-level validation statistic**
SSC	Categorical	5	Classification	Classification accuracy	Decrease in accuracy	Classification accuracy; Kappa
FTYPE	Categorical	14	Classification	Classification accuracy	Decrease in accuracy	Classification accuracy; Kappa
BAA	Numeric	n/a	Regression	Variance explained; RMSE; MAE	Percent increase in MSE	MAE
TPA	Numeric	n/a	Regression	Variance explained; RMSE; MAE	Percent increase in MSE	MAE

3.3.4. Predictor/Explanatory Variables

Explanatory variables were derived from aerial imagery and were compiled into multi-band predictor images in Google Earth Engine (GEE; Gorelick et al., 2017). The process of generating predictors was the same for all sets of imagery described in Table 8, but the number of predictor

layers differed depending on whether the variables were derived from a NAIP or a Sentinel-2 image (Table 10). Sentinel-2 imagery offers 3 additional bands at the upper end of the visible portion of the spectrum, "red-edge" bands, which are situated between the red and near-infrared portions of the electromagnetic spectrum. NAIP required less processing as it only included a single image in each year and only 4 bands (red, green, blue and near-infrared) were available. Sentinel-2 imagery was first refined using a cloud mask function in GEE and only images that had a cloudy pixel percentage less than 1% were retained. The seven Sentinel-2 bands with data representing the visible and NIR portions of the spectrum (bands 2-8) were selected for each of the time slices in Table 8, and scaled to 10 m (32.8 ft) resolution if appropriate. Once the sets of imagery had been prepared, layers for the predictor stack were generated in the same way for all sets. First, the raw band values in each pixel were transformed into focal means using a 36 × 36 m (118 × 118 ft) moving window, which corresponded to the extent of the field plots. NDVI was calculated per pixel on the original data and then transformed into a focal mean using a window equal to the field plot (36 × 36 m; 118 ft × 118 ft). Finally, a texture variable was added to the stack of predictors by computing the gray-level co-occurrence matrix (GLCM) horizontal contrast of the NIR band, also using the same size focal neighborhood. All of these layers were then manipulated into a single multi-band predictor image for each imagery set (Table 10).

Table 10: Table of the characteristics of the predictor images used as explanatory variables in the random forest models.

Predictor Image Source	Spatial Resolution	Bands in predictor stack	Focal window	Total bands in stack
NAIP 2015	1 m	R, G, B, N NDVI GLCM (N)	36 × 36 m	6
NAIP 2017	1 m	R,G, B, N NDVI GLCM (N)	36 × 36 m	6
Sentinel – 2 Oct.	10 m	R, G, B, N RE 1, RE 2, RE 3 NDVI GLCM (N)	36 × 36 m	9
Sentinel – 2 Growing season	10 m	R, G, B, N RE 1, RE 2, RE 3 NDVI GLCM (N)	36 × 36 m	9

3.3.5. *Building Random Forests for Prediction*

Random forest models were generated utilizing both GEE (Gorelick et al., 2017) and R (R Core Team, 2017). To add spectral information to training data, the value of each predictor band was extracted at the plot center and exported as a table from GEE. These tables were brought into R and used to train random forest models to predict SSC (possible classes = 5), FTYPE (possible classes = 14), TPA, and BAA using the *randomForest* package in R (Liaw & Wiener, 2002). Although GEE has the functionality to run random forest classifications, the package in R-Studio provides users with more control over inputs and the ability to easily examine measures of variable importance. All models generated 500 trees with a training ratio of ~0.63 and a variable splitting factor approximately equal to the square root of the number of predictors ($\sqrt{p}$) for classification forests (SSC and FTYPE) and $p/3$ for regression forests (BAA and TPA; Table 11).

Table 11: Table of random forest characteristics for each of the 4 four predicted variables. The number of predictor variables is abbreviated "p" (see Table 10 for descriptions of the predictor images).

Predicted variable	Variable type	Decision tree type	Number of trees	Training ratio	Splitting factor
SSC	Discrete	Classification	500	0.6325	$\sqrt{p}$
FTYPE	Discrete	Classification	500	0.6325	$\sqrt{p}$
BAA	Continuous	Regression	500	0.6325	p/3
TPA	Continuous	Regression	500	0.6325	p/3

3.3.6. Prediction Accuracy and Variable Importance

The *randomForest* package (Liaw & Wiener, 2002) outputs separate metrics for evaluating model performance and variable importance. Overall prediction capability of the RF for each of the four predicted variables were evaluated using the out-of-bag statistics which differed depending on the type of the response variable (Table 9).

The importance of each input variable to the RF predictions for SCC and FTYPE was assessed using the mean decrease in classification accuracy and the percent increase in mean square error for BAA and TPA.

3.3.7. *Summarizing by Stand Level and Validation*

Stands were manually digitized by DNRC in 2019 over NAIP 2015 imagery and matching field data to that collected at the plots in 2018 was collected at all stands in 2019. This stand level inventory (SLI) data served as the validation data for this analysis. Pixel-wise predictions of all variables were made using GEE (Gorelick et al., 2017). Then pixels were summarized within each stand for the predicted size class, forest type, basal area per acre, and tree count per acre. SSC and FTYPE pixel values were summarized using the mode of pixels within a stand, while the BAA and TPA pixel values were summarized using the mean pixel value within each stand. Predicted values were then compared to the field data collected by DNRC in 2019 at the previously digitized stands and an accuracy assessment for each variable was completed using the RMRS Raster Utility

(Hogland & Anderson, 2010) add on for ArcMap (Environmental Systems Research Institute, Inc. (ESRI), 2016). Accuracy on the stand-level validation dataset was assessed using either the overall accuracy of classification for categorical variables or the mean absolute error (MAE) for continuous variables (Table 9). A Kappa statistic was also calculated for the categorical accuracy assessments (Hogland and Anderson, 2010).

3.4. Results

3.4.1. *SCC and FTYPE Results*

For each dependent variable, one model was created from each stack of predictor images. For SSC and FTYPE predictions were the most likely class at each training site. Accuracy rates were surprisingly low from all algorithms.

Results for the classification models were evaluated based on the out-of-bag classification accuracy (Table 12). Models generally showed consistency with NAIP 2017 models performing slightly better than all other models when predicting SSC and Sentinel-2 from October models performing slightly better than others when predicting FTYPE. Both Sentinel-2 models outperformed NAIP models when FTYPE was the target dependent variable.

Table 12: Table of classification accuracy for the SSC and FTYPE prediction on the withheld data.

Parameter	Imagery	OOB Classification Accuracy
SSC	NAIP 2015	62.5 %
	NAIP 2017	66.7 %
	Sentinel: Oct. 2018	64.2 %
	Sentinel: May – Oct.2018	64.2 %
FTYPE	NAIP 2015	27.5 %
	NAIP 2017	27.5 %
	Sentinel: Oct. 2018	33.3 %
	Sentinel: May – Oct.2018	30.8 %

Random forests grown in R-Studio automatically output a confusion matrix to allow users to identify the classification error in each class. These were summarized and plotted in order to compare the classification accuracy utilizing the different imagery for SSC (Figure 10; Table 13) and FTYPE (Figure 11; Table 14). All models predicted Sawtimber plots most effectively while all models struggled to predict Poletimber plots. The NAIP 2017 model identified wetlands much more effectively than the other models, but did the most poorly on Poletimber.

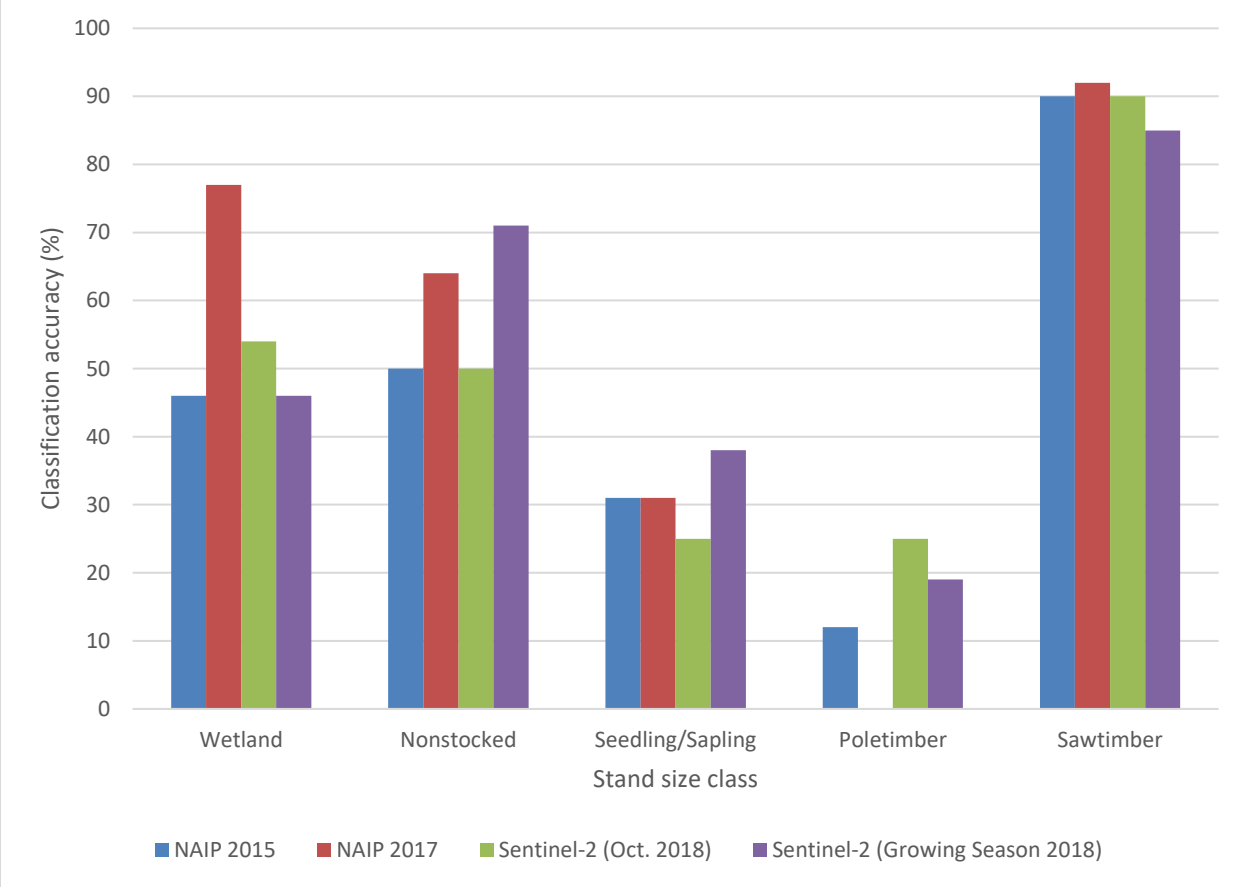

Figure 10: Graph of the classification accuracy on out-of-bag pixels within each size class based upon each imagery used (Table 13).

Table 13: Table of the classification accuracy on out-of-bag pixels within each size class based upon the imagery used (Figure 10).

	Wetland	Nonstocked	Seedling/ Sapling	Poletimber	Sawtimber
NAIP 2015	46	50	31	12	90
NAIP 2017	77	64	31	0	92
Sentinel-2 (Oct. 2018)	54	50	25	25	90
Sentinel-2 (Growing Season 2018)	46	71	38	19	85

With regard to forest type, all classifications predicted nonforested plots (wetland and nonstocked) with the highest degree of accuracy and struggled to predict most other forest types (Figure 11; Table 14). The model built using growing season Sentinel-2 imagery was the only model to identify any larch plots correctly, while the model built using October Sentinel-2 imagery identified Lodgepole pine plots and mixed conifer plots more accurately than all others.

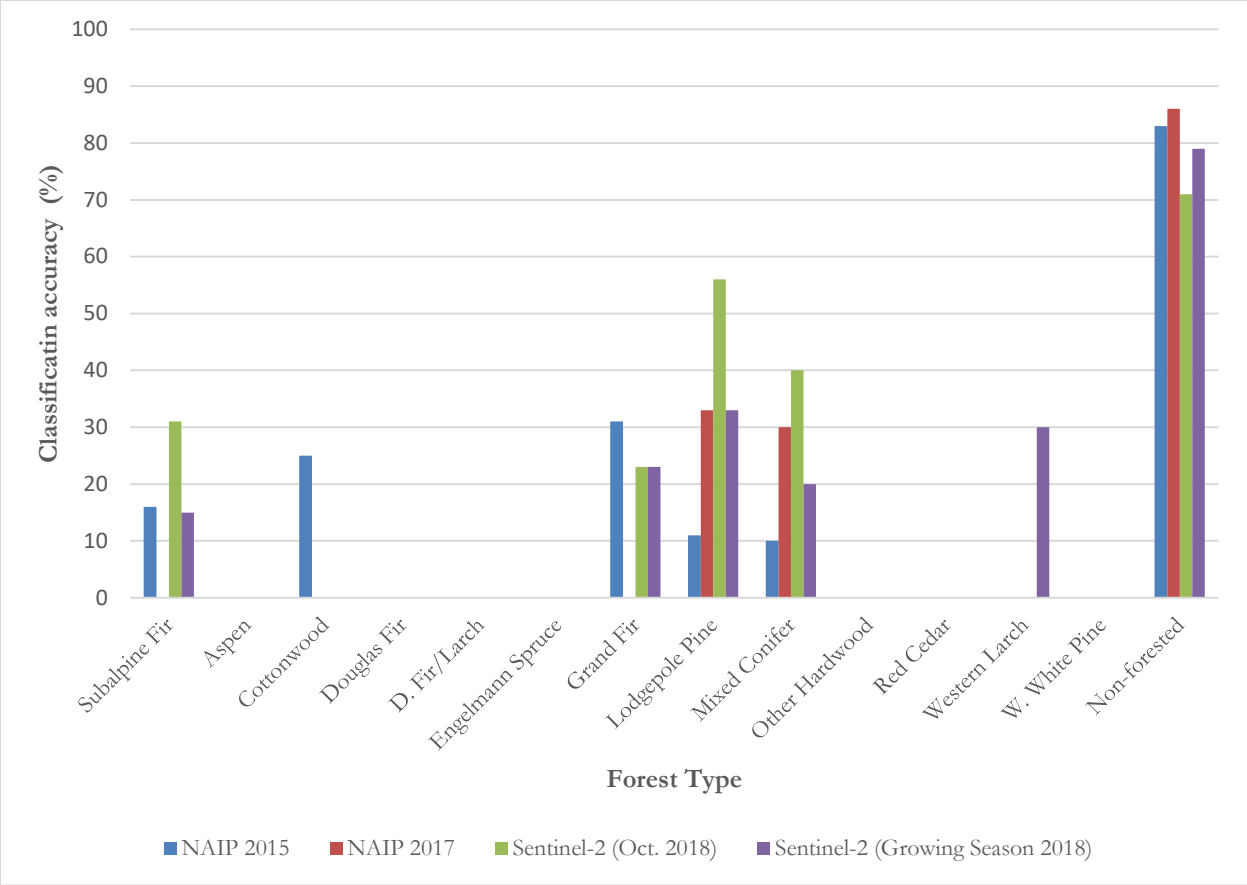

Figure 11: Graph of the classification accuracy on out-of-bag pixels within each forest type class based upon each imagery used (Table 14).

Table 14: Table of the classification accuracy on out-of-bag pixels within each size class based upon the imagery used (Figure 10).

Imagery	AF	AS	CO	DF	DL	ES	GF	LP	MC	OH	RC	WL	WP	NF
NAIP 2015	16	0	25	0	0	0	31	11	10	0	0	0	0	83
NAIP 2017	0	0	0	0	0	0	0	33	30	0	0	0	0	86
Sentinel-2 (Oct. 2018)	31	0	0	0	0	0	23	56	40	0	0	0	0	71
Sentinel-2 (Growing Season 2018)	15	0	0	0	0	0	23	33	20	0	0	30	0	79

3.4.2. *BAA and TPA Results*

Observed values for BAA in the 2018 plots ranged from 0 – 316 ft^2/ac with a mean of 64 ft^2/ac (Figure 12). TPA observed values in the 2018 plots ranged from 0 – 5266 trees/ac with a mean of 1262 trees/ac (Figure 12). RF regression models were summarized using the amount of

variance explained by the model as well as the root-mean-square error (RMSE) and mean absolute error (MAE) to evaluate model fitness (Table 15).

Table 15: Table of BAA and TPA performance statistics: variance explained, and out-of-bag RMSE and MAE values. MAE is in ft^2/ac for BAA and trees/ac.

Parameter	**Imagery**	**Variance Explained**	**OOB RMSE**	**OOB MAE**
Total BAA	NAIP 2015	32.55 %	49.1	35.7
	NAIP 2017	39.19 %	46.6	34.0
	Sentinel: Oct. 2018	25.29 %	51.7	36.8
	Sentinel: May – Oct.2018	39.07 %	46.7	34.4
Total TPA	NAIP 2015	28.21 %	995	698
	NAIP 2017	33.36 %	959	695
	Sentinel: Oct. 2018	23.12 %	1025	732
	Sentinel: May – Oct.2018	21.62 %	1043	783

The regression forests grown using NAIP 2017 and Sentinel-2 across the growing season were almost identical in their predictions of total BAA. The Sentinel-2 October model was the weakest when predicting total BAA (Table 15).

TPA models were less successful overall in terms of variance explained. The model built using NAIP 2017 imagery was the most successful in predicting total TPA. Conversely, models built using growing season composites from Sentinel-2 were the weakest for total TPA (Table 15).

The error estimates in Table 15 correspond to the difference between observed (Figure 12) and predicted values with respect to the training data, which was collected at plots during the 2018 field season.

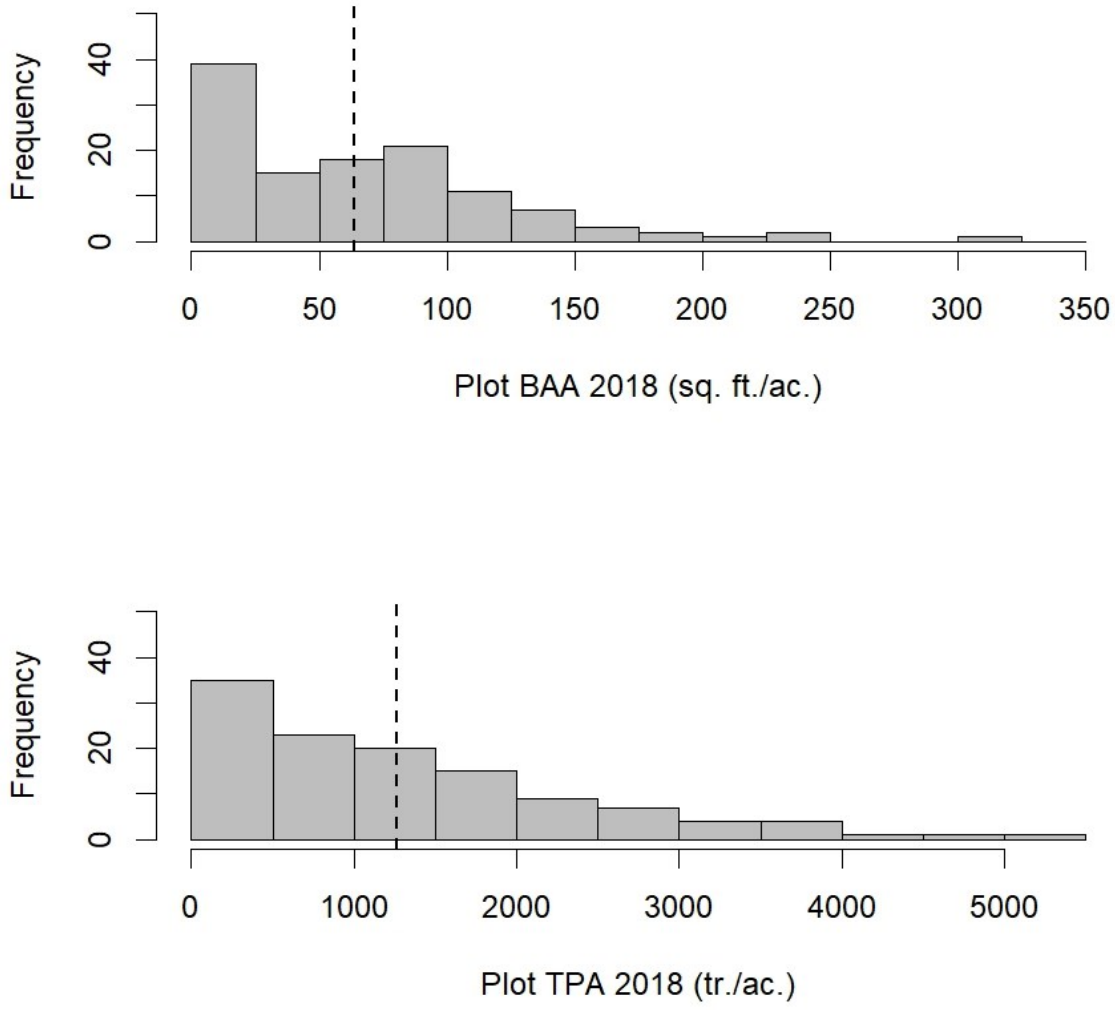

Figure 12: BAA and TPA distributions from the data collected at plots in 2018. The mean is shown as a dashed line.

3.4.3. *SCC and FTYPE Variable Importance*

Variable importance was also assessed for each classification and regression forest (Tables 16-19). When predicting SSC using NAIP models, NDVI was the most important variable while NIR and NIR contrast held the least predictive power (Figure 13, Table 16). In the model built using Sentinel across the growing season, the first red edge band emerged as the most important

with the NIR contrast being by far the least important. The October Sentinel model's blue band was the most important and the NIR band was the least important. The NIR contrast from October was nearly 5% more important than in any other models (Figure 13, Table 16).

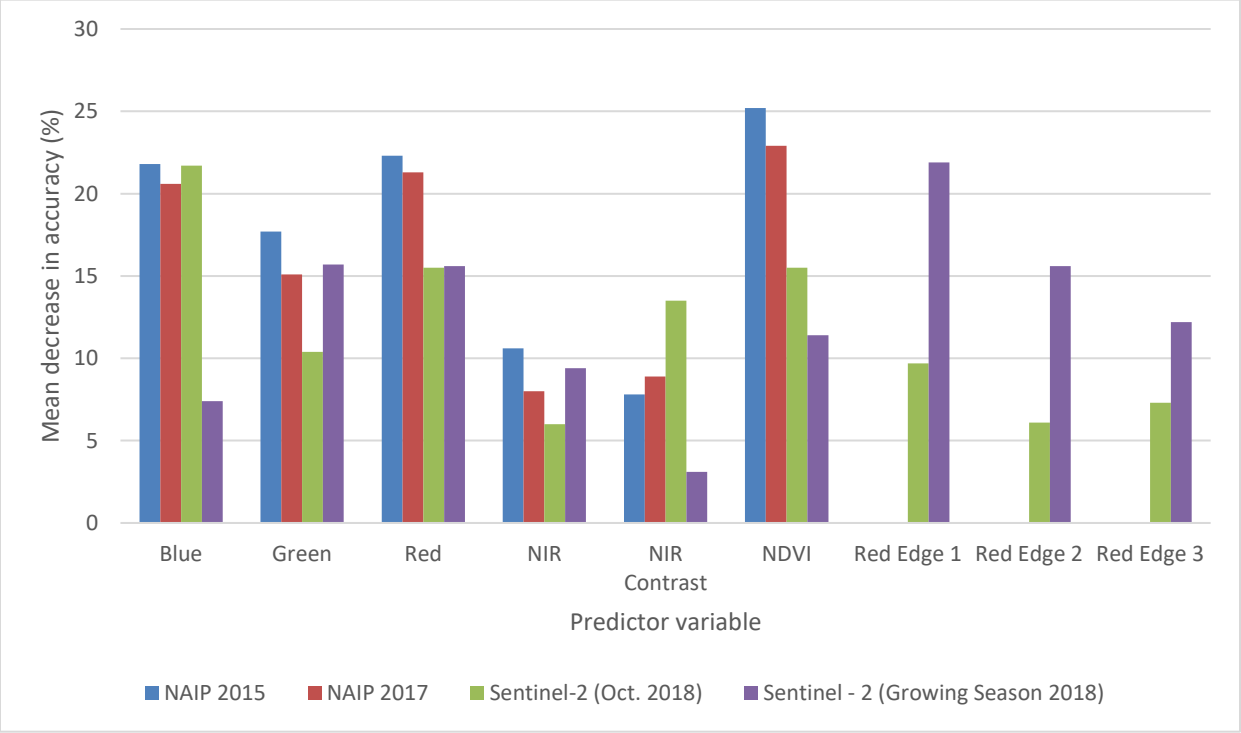

Figure 13: Bar chart of variable importance for predicting stand size class (SCC) in terms of the mean decrease in classification accuracy when the variable was removed.

Table 16: Table of variable importance for predicting stand size class (SCC) in terms of the mean decrease in classification accuracy when the variable was removed.

Imagery	Blue	Green	Red	NIR	NIR Contrast	NDVI	Red Edge 1	Red Edge 2	Red Edge 3
NAIP 2015	21.8	17.7	22.3	10.6	7.8	25.2	n/a	n/a	n/a
NAIP 2017	20.6	15.1	21.3	8.0	8.9	22.9	n/a	n/a	n/a
Sentinel – 2 Oct.	21.7	10.4	15.5	6.0	13.5	15.5	9.7	6.1	7.3
Sentinel – 2 Growing	7.4	15.7	15.6	9.4	3.1	11.4	21.9	15.6	12.2

The variable importance estimates for predicting forest type (Figure 14, Table 17) followed a similar trend to those from SSC (Figure 13). NDVI again was the most important variable in NAIP based models and Red edge 1 was the most important variable in the Sentinel growing season

model. Conversely, NDVI became one of the most important variables for the October Sentinel model and NIR for NAIP 2015 actually improved models when it was removed.

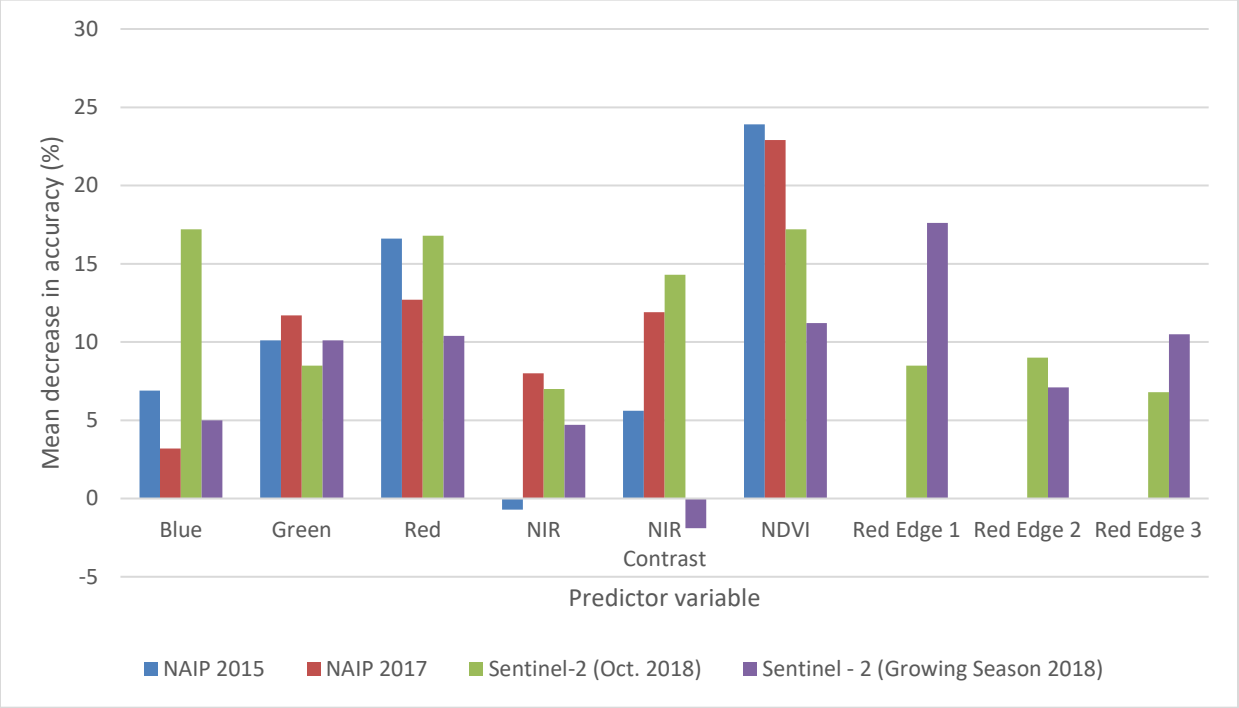

Figure 14: Bar chart of variable importance for predicting forest type (FTYPE) in terms of the mean decrease in classification accuracy when the variable was removed.

Table 17: Table of variable importance for predicting forest type (FTYPE) in terms of the mean decrease in classification accuracy when the variable was removed.

Imagery	Blue	Green	Red	NIR	NIR Contrast	NDVI	Red Edge 1	Red Edge 2	Red Edge 3
NAIP 2015	6.9	10.1	16.6	- 0.7	5.6	23.9	n/a	n/a	n/a
NAIP 2017	3.2	11.7	12.7	8.0	11.9	22.9	n/a	n/a	n/a
Sentinel – 2 Oct.	17.2	8.5	16.8	7.0	14.3	17.2	8.5	9.0	6.8
Sentinel – 2 Growing	5.0	10.1	10.4	4.7	- 1.9	11.2	17.6	7.1	10.5

3.4.4. *BAA and TPA Variable Importance*

In predicting BAA, the first red edge band across the growing season resulted in the largest increase in MSE when it was removed, ~20% increase in error (Figure 15, Table 18). The same was true for predicting TPA (Figure 16, Table 19). With the exception of October Sentinel-2 based

BAA, NIR contrast tended to produce the lowest increase in error when removed. NDVI was more important in predicting TPA than predicting BAA in all cases except models built using growing season Sentinel-2 imagery.

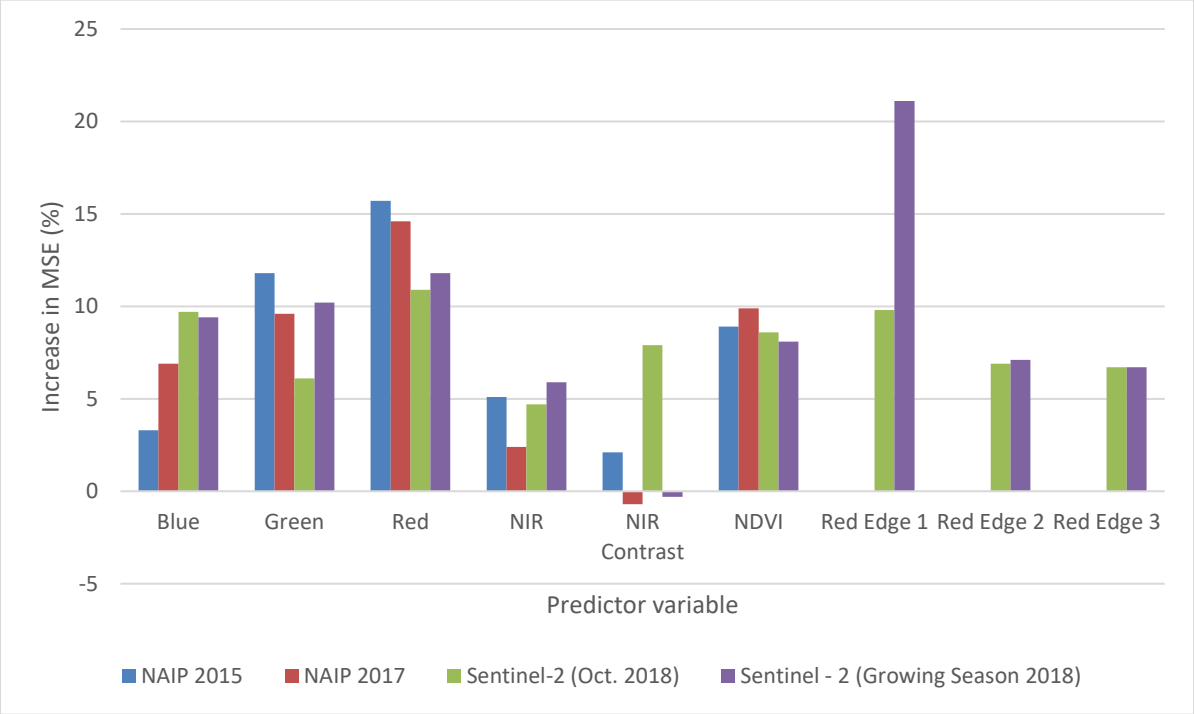

Figure 15: Bar chart of variable importance for predicting square feet of basal area per acre (BAA) in terms of the percent increase in mean square error (MSE) when the variable was removed.

Table 18: Table of variable importance for predicting square feet of basal area per acre (BAA) in terms of the percent increase in mean square error (MSE) when the variable was removed.

Imagery	Blue	Green	Red	NIR	NIR Contrast	NDVI	Red Edge 1	Red Edge 2	Red Edge 3
NAIP 2015	3.3	11.8	15.7	5.1	2.1	8.9	n/a	n/a	n/a
NAIP 2017	6.9	9.6	14.6	2.4	- 0.7	9.9	n/a	n/a	n/a
Sentinel – 2 Oct.	9.7	6.1	10.9	4.7	7.9	8.6	9.8	6.9	6.7
Sentinel – 2 Growing	9.4	10.2	11.8	5.9	- 0.3	8.1	21.1	7.1	6.7

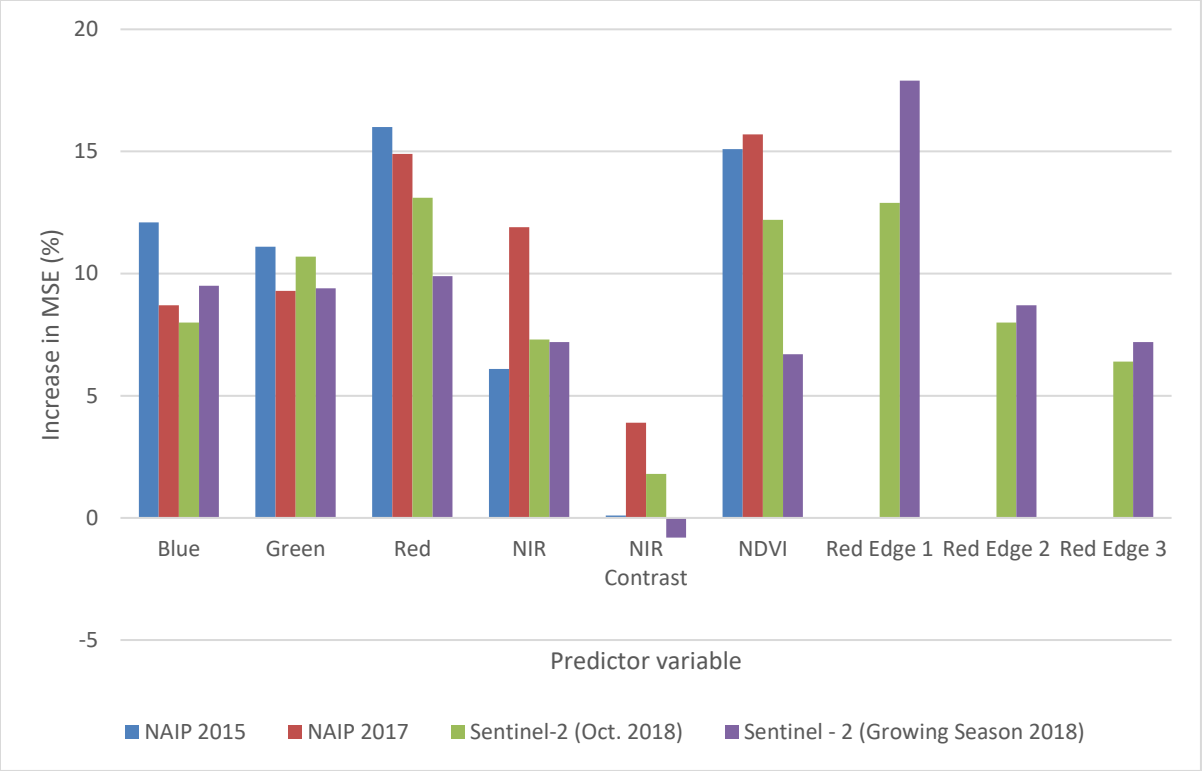

Figure 16: Bar chart of variable importance for predicting trees per acre (TPA) in terms of the percent increase in mean square error (MSE) when the variable was removed.

Table 19: Table of variable importance for predicting trees per acre (TPA) in terms of the percent increase in mean square error (MSE) when the variable was removed.

Imagery	Blue	Green	Red	NIR	NIR Contrast	NDVI	Red Edge 1	Red Edge 2	Red Edge 3
NAIP 2015	12.1	11.1	16.0	6.1	0.1	15.1	n/a	n/a	n/a
NAIP 2017	8.7	9.3	14.9	11.9	3.9	15.7	n/a	n/a	n/a
Sentinel – 2 Oct.	8.0	10.7	13.1	7.3	1.8	12.2	12.9	8.0	6.4
Sentinel – 2 Growing	9.5	9.4	9.9	7.2	- 0.8	6.7	17.9	8.7	7.2

3.4.5. *SCC and FTYPE Stand-Level Accuracy and Validation*

The accuracy of predictions on the 2019 stand-level validation data was low overall for all variables (Tables 20-21). For the categorical variables of SSC and FTYPE, the highest accuracy achieved was 28% and 24% respectively. Overall accuracy of FTYPE predictions on the validation data were even lower than the already low OOB accuracies, while the overall accuracy of SSC

predictions on the validation data were drastically lower (Table 20). The Kappa statistics show that the SCC were little better than chance, with FTYPE having higher but still low values. This may be due to sensitivity of the Kappa to the number of categories predicted rather than the quality of the classification's performance.

Table 20: Table of accuracies when predicted classifications were compared to the 2019 stand-level validation data. The out-of-bag classification accuracy shown here is identical to the column in Table 12.

	Imagery	**OOB Classification Accuracy**	**SLI Accuracy (validated)**	**Kappa**
SSC	NAIP 2015	62.50 %	28.34 %	0.03
	NAIP 2017	66.67 %	26.43 %	0.01
	Sentinel: Oct. 2018	64.17 %	27.07 %	0.02
	Sentinel: May – Oct.2018	64.17 %	26.75 %	0.02
FTYPE	NAIP 2015	27.50 %	21.34 %	0.09
	NAIP 2017	27.50 %	20.00 %	0.07
	Sentinel: Oct. 2018	33.33 %	24.20 %	0.11
	Sentinel: May – Oct.2018	30.83 %	19.75 %	0.07

Comparing the SSC predictions to the data used for validation reveals that the most successful model (and the others as well, not shown) significantly overestimated the amount of sawtimber in the study area while underestimating the seedling/sapling and poletimber size classes (Figure 17). For forest type, predictions from the most successful model appeared to capture more of the variability of forest types in the study area, despite the low classification accuracy (Figure 18).

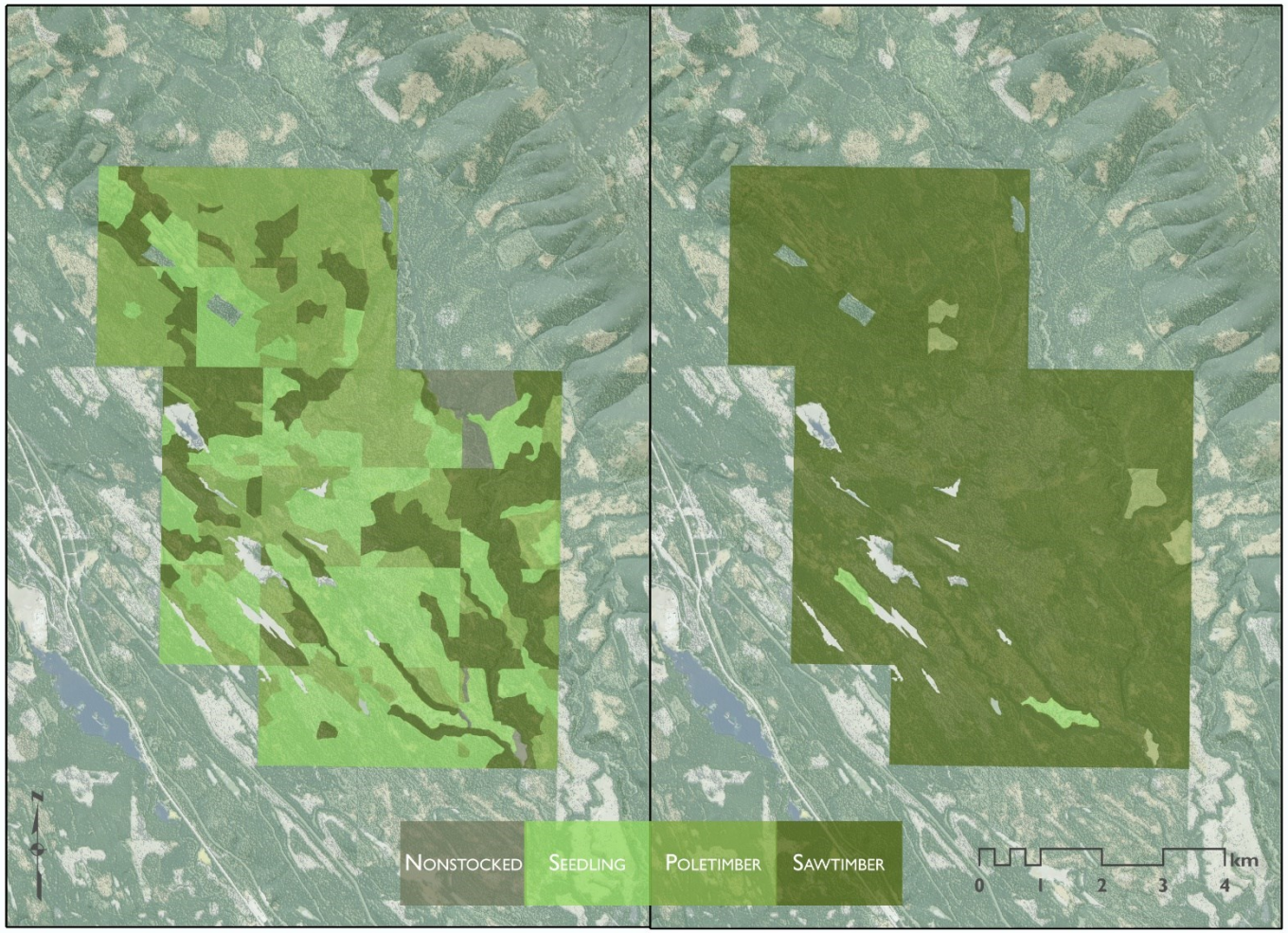

Figure 17: Map of the stand size class data used for validation (left) and the predicted stand size class (right) for the NAIP 2017 SSC model.

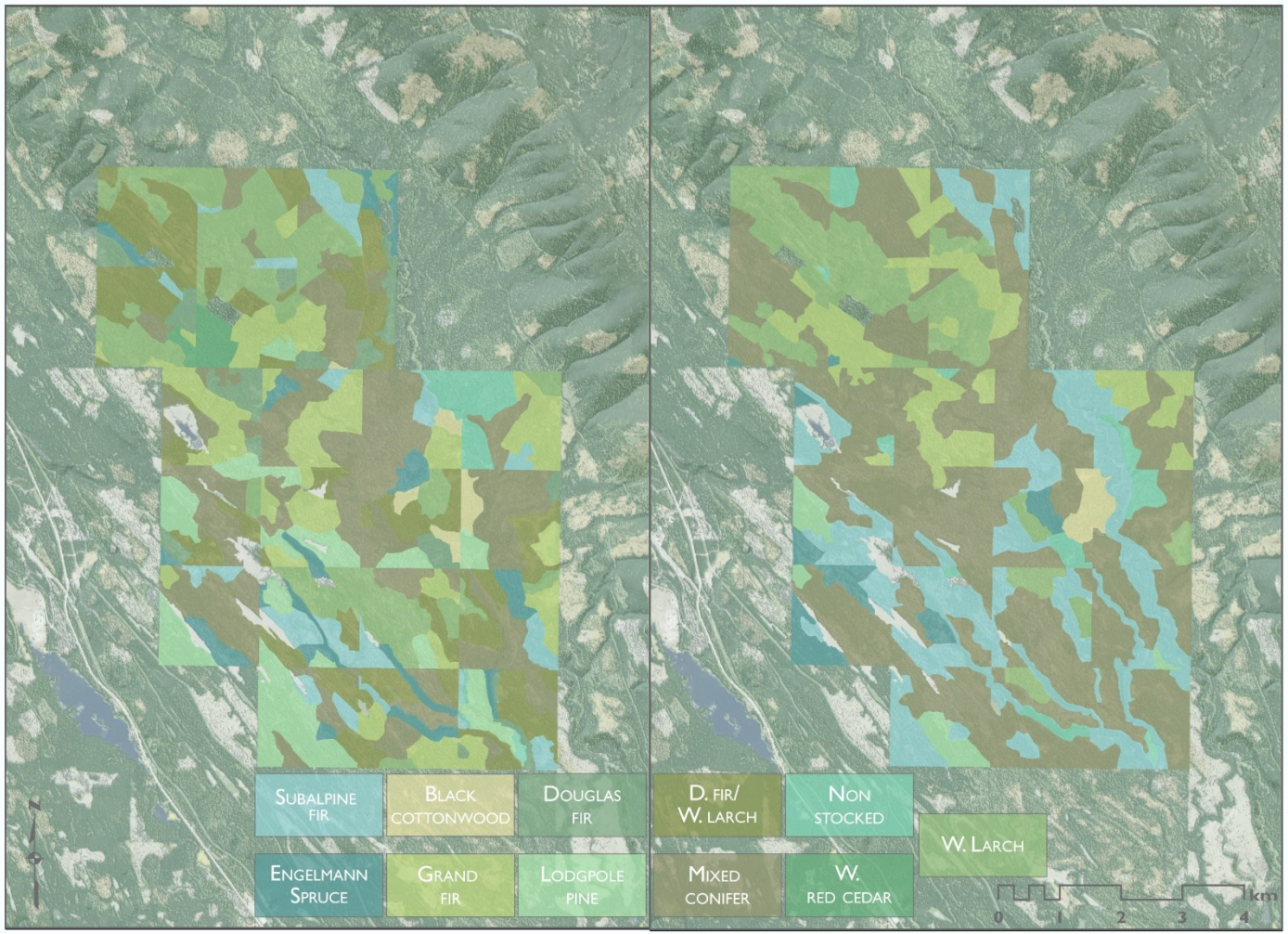

Figure 18: Map of the forest type data used for validation (left) and the predicted forest type (right) for the October Sentinel FTYPE model.

3.4.6. *BAA and TPA Stand-Level Accuracy and Validation*

Observed values for BAA collected at the stands in 2019 range from 0 – 220 ft^2/ac with a mean of 37 ft^2/ac and observed TPA values ranged from 0 – 5000 trees/ac with a mean of 558 trees/ac (Figure 19). MAE values for the stand-level predictions were substantially higher than the MAEs from the training data as would be expected. MAE values when predicting BAA and TPA on validation data were on average 37% and 48% larger than the training data, respectively (Table 21). The stand level MAE was calculated using the observed values collected in 2019 during the DNRC's stand level inventory (Figure 19).

Table 21: Table of the mean absolute error (MAE) comparing the Random Forest regression predictions to the validation data. The column out-of-bag MAE is identical to the one shown in Table 15. MAE is in ft^2/ac for BAA and trees/ac.

	Imagery	**OOB MAE (training)**	**Stand-level MAE (validation)**
BAA	NAIP 2015	35.7	49.2
	NAIP 2017	34.0	50.2
	Sentinel: Oct. 2018	36.8	49.2
	Sentinel: May – Oct.2018	34.4	44.2
TPA	NAIP 2015	698	1101
	NAIP 2017	695	1057
	Sentinel: Oct. 2018	732	1104
	Sentinel: May – Oct.2018	783	1027

Both BAA and TPA models tended to overestimate the observed values when predicted onto the stands used for validation (Figure 20-21). Basal area estimates mostly fell between 50-120 ft^2/ac, missing most of the stands which fell on the low or high end of the validation data range (Figure 20). Trees per acre estimates also missed stands at the ends of the spectrum, predicting a TPA value between 650-2500 for nearly all stands (Figure 21).

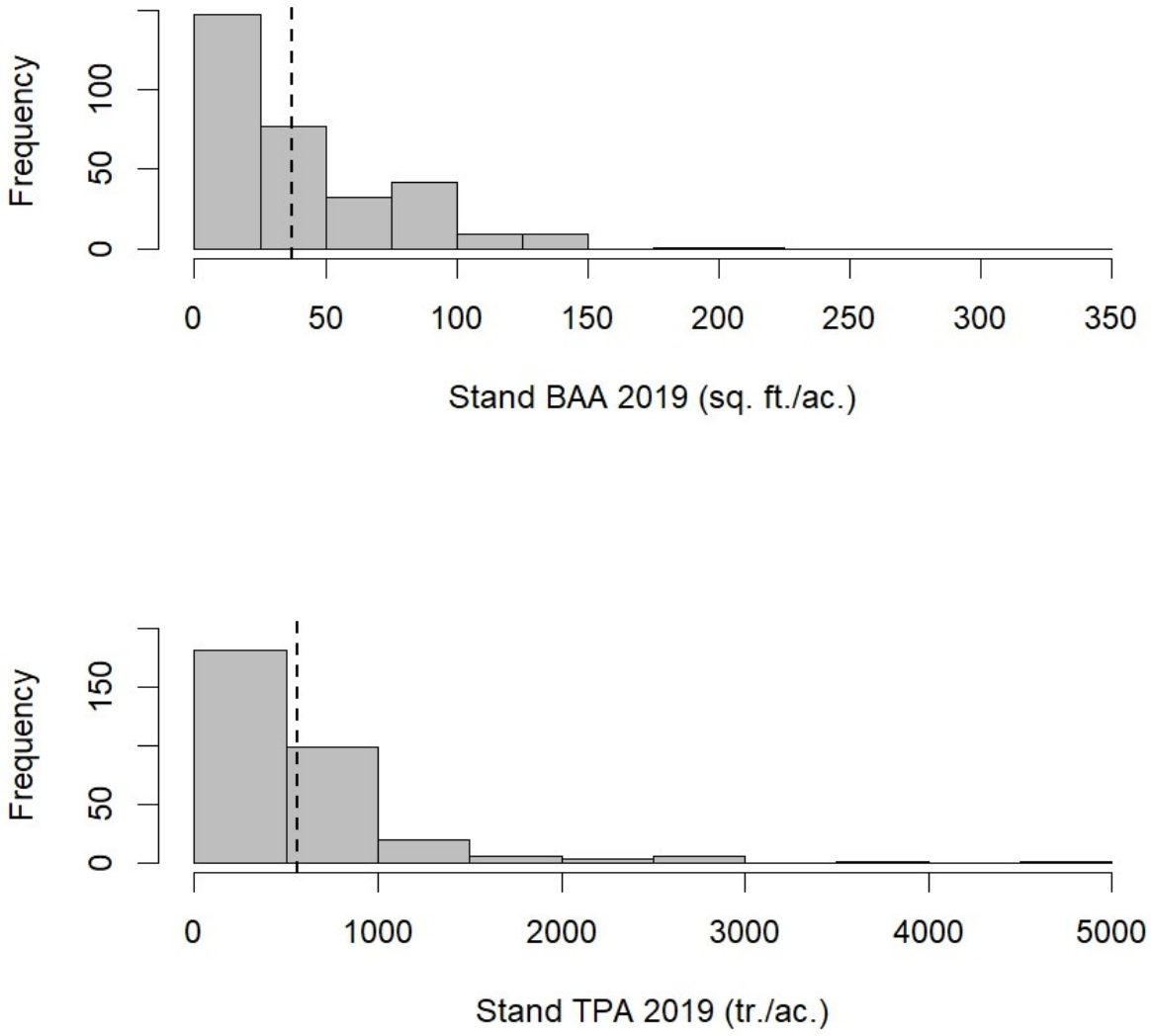

Figure 19: BAA (top) and TPA (bottom) distributions from the data collected at stands in 2019. The mean is shown as a dashed line.

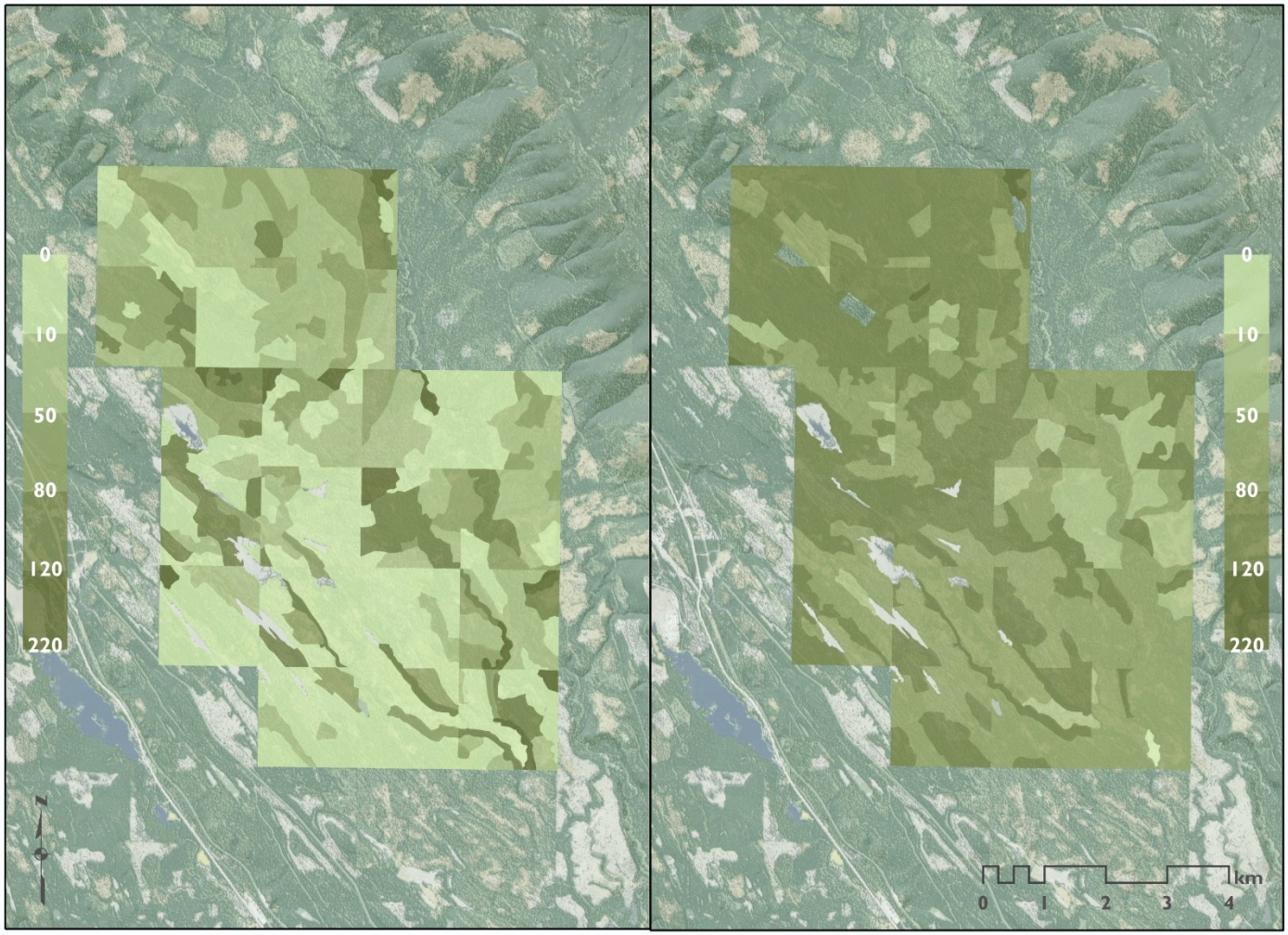

Figure 20: Map of the basal area data used for validation (left) from the 2019 stand-level survey and the predicted basal area (right) for the NAIP 2017 BAA model. BAA estimates are in units of ft^2/ac.

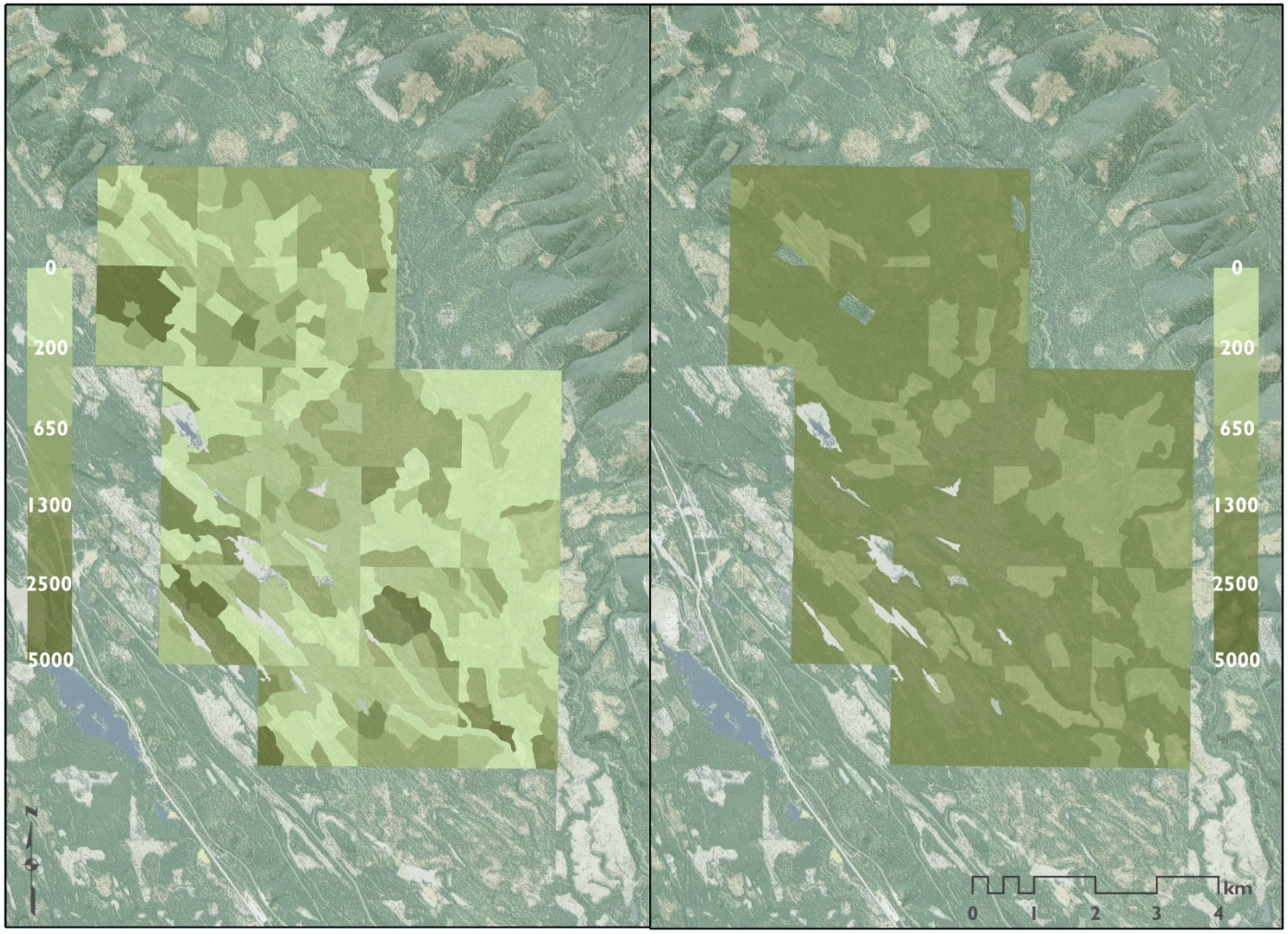

Figure 21: Map of the trees per acre data used for validation (left) and the predicted trees per acre (right) for the NAIP 2017 TPA model.

3.5. Discussion

3.5.1. *Overall Prediction Success*

Initially, in terms of predictive power, only models predicting stand size class approached OOB accuracy levels that indicated the models may prove useful. This seemed unusual given that forest type, the other categorical variable predicted, would appear to be better suited to classification using spectral information. Spectrally, conifer and deciduous trees should be easy to distinguish from each other, and due to the timing of the NAIP and October Sentinel-2 imagery, the larch trees were clearly visible to a human observer. Stand size class would appear to be more difficult to determine from spectral information alone, and would presumably benefit from some sort of height measure (such as a LiDaR point cloud) to be accurately classified at this detailed level. Additionally, the variability within the size classes as they were defined by DNRC was potentially large. For instance, the largest size class present only needed to account for 10% of the total area for the plot or stand to be designated as such. So a plot/stand comprised 10% sawtimber and 90% poletimber, and a plot/stand comprised of 10% sawtimber and 90% bare ground, would both be designated as sawtimber.

The lack of success of stand size class predictions likely depended on several things. First, the training data was imbalanced, with more than 60% falling into a single size class (sawtimber). This imbalance could be acceptable if it was an accurate depiction of the study area, but when the DNRC returned and censused the LSA in 2019, the stand-level data showed only 25% of the total area was sawtimber. The overrepresentation of sawtimber in the training data biased the algorithms towards sawtimber may have made them more likely to choose sawtimber as the SSC in cases where no strong relationship between explanatory and response variable existed, and the overabundance of data may have inflated estimates of classification accuracy. Second, the class designations for forest types contained overlap in the species within each, which likely caused confusion. For example,

while both Douglas fir (*Pseudotsuga menziesii)* and western larch (*Larix occidentalis*) were represented by their own classes, there was also a separate class representing a combination of the two species and a fourth class that represented a "mixed conifer" forest type. This redundancy and overlap between classes made the classification of forest type problematic as the spectral differences are decreased when multiple species are combined.

The incorporation of multiple sources of imagery uncovered some of the potential benefits that imagery with various spatial, temporal, spectral, and radiometric resolution may offer when building predictive models. Predictions results were generally similar in their relative ability to predict the variable of interest at the training plots, which were 36 × 36 m. Knowing that Sentinel-2 imagery is collected at 10 m (32.8 ft; B, G, R, NIR) and 20 m (65.6 ft) resolution (R-edge) compared with the 1 m (3.28 ft) resolution of NAIP, the increased temporal, spectral or radiometric resolution must be compensating for this lack of spatial detail. The Sentinel-based model results were similar in all cases except when predicting total BAA, in which case the model built using Sentinel-2 imagery throughout the growing season explained 14% more variance than all others. This similarity suggests that overall, the increased spectral or radiometric resolution contributed more to model success than the increased temporal resolution (higher revisit period of Sentinel) in this study.

3.5.2. Variable Importance

On a local level, the relative importance of each variable pointed toward increased temporal resolution as a factor in prediction success. The first narrow red edge band, located between the red and near-infrared bands, was the most important variable in all prediction models built using the growing season imagery and contributed to the overall success. In some cases, the removal of just this one band reduced accuracy by 20%. What makes this even more impressive is that all red edge bands are collected at a spatial resolution of 20 m (65.6 ft). The importance of the red edge band

even though it was much coarser than the other image bands only reinforces its utility in vegetation mapping projects and lends support to the continued use of Sentinel-2 imagery for the additional red edge bands/spectral resolution.

Another interesting finding was the lack of importance of the NIR and NIR contrast band. The contrast band was incorporated to add some measure of texture at each plot and was expected to be an important factor. Instead it was consistently one of the least important variables which may indicate that an error occurred when the image was generated in GEE or that the parameters used in building the layer should be adjusted. The contrast was computed for only the NIR band as it was expected that the NIR band itself would be an important factor in discerning differences in vegetation at field plots. Because the NIR band actually proved to be less important than expected, perhaps the horizontal contrast should be computed for a different band, or even across all bands in the future.

3.5.3. Accuracy Assessment: Moving from Plot-level to Stand-level

The addition of the full LSA stand-level data provided an interesting perspective on the original plot data used to train models. Upon closer examination of this data, discrepancy between plot and stand variables became apparent (Table 18). Discrepancy here does not necessarily indicate errors in data collection, but may indicate that stands are more heterogeneous than expected. The stand boundaries were drawn manually by a technician over 2015 NAIP imagery (Sec. 3.3.6). While stands may appear well-delineated to observers, much of the spectral heterogeneity in a stand may be undetectable to the human-eye. With this heterogeneity in mind, it is likely that some plots fell in areas which represented an anomaly compared to the rest of the stand.

Table 22: The agreement of size class (SSC) and forest type (FTYPE) between the plot data collected in 2018 and the stand data from 2019. The "Percent of agreement" is the percent of plots that reported the same value for the variable of interest as the stand that contained them.

Variable of interest	Percent of agreement
Stand Size Class (SSC)	39.8 %
Forest Type (FTYPE)	19.4 %

Segmentation could be used as a part of the stand delineation process in order to generate stands that are more homogenous in reflectance. If manual delineation of stand boundaries is necessary, computing some zonal statistics on the variability of spectral information within each stand would be useful and allow prediction based upon spectral un-mixing algorithms.

With all this in mind, the low accuracy of the stand-level predictions seem reasonable. The discrepancy between the out-of-bag results and the accuracy assessment results for the SSC models was striking. Prediction accuracies decreased from ~65% (training) to ~27% (validated) which further corroborates the concerns over in the training data utilized.

3.6. Conclusion

The incorporation of sources of imagery with increased spatial, spectral, or radiometric resolution is generally worthwhile in any classification. Sentinel-2 imagery offers additional spectral information beyond the standard visible and near-infrared NAIP product that generally differentiate vegetation more effectively. Sentinel-2's short revisit period also provides the opportunity to adjust the timing of the imagery and the ability to create seasonal composites using different reduction methods (i.e. mean, maximum, minimum, etc.). Sentinel-2 couples a high revisit period and increased radiometric resolution with a relatively higher spatial resolution when compared to other publicly available global datasets like Landsat.

None of these sensors must be used in isolation. The incorporation of bands or raster products derived from multiple sensors could all be used as explanatory variables in the same model.

For instance, the principal components derived from a NAIP image could be used as inputs alongside some combination (such as a mean) of the red-bands from Sentinel. These inputs could also be varied in their temporal resolution to better capture species-specific growth patterns and differences in the timing of green-up or leaf drop. Combining spectral information from multiple sensors allows users to leverage the advantages of each sensor without the trade-off associated with only using one.

Because models produced relatively consistent results despite the source of imagery, it can be assumed that each of the three image qualities that differed play some role in classification. The high spatial resolution that NAIP provides obviously helped to make up for its limited number of spectral bands, as well as the fact that it only captured a single slice in time which was taken quite late in the growing season in both 2015 and 2017. The added spectral/radiometric resolution of the Sentinel-2 imagery allowed predictions to perform at a similar level to the NAIP models even with coarser spatial resolution. Because there was little difference between the successes of Sentinel-2 models from October versus the models from the entire growing season, it would appear that temporal resolution had the least amount of influence on the predictive power of models. This does not mean that incorporating multiple time slices is not worthwhile however, as seen in the higher accuracy of Sentinel-2 growing season predictions against the validation data (Tables 16 & 17). The high level of importance of red-edge bands throughout the growing season speaks to the power of seasonality and advocates for the further exploration of red-edge bands for vegetation mapping.

3.7. Future Work and Research

In regard to the impact of imagery on the accuracy of the random forest predictions, this analysis did not reach definitive conclusions but showed some of the advantages that different

spatial, temporal, radiometric and spectral resolutions may offer. While neither NAIP nor Sentinel-2 imagery significantly outperformed the other, the topic deserves further investigation.

For this study, the bands used as predictors were intentionally limited to try to focus on specific differences between NAIP and Sentinel-2 and as such, this area that warrants further study. Transforming the visible and NIR bands using a principal component analysis (PCA) may eliminate some redundancy between these bands. Something similar could be done with the 3 red-edge bands from Sentinel-2. It may also be worthwhile to calculate a horizontal contrast for all bands in the images, rather than just the NIR band, and to also revisit the process of generating the GLCM in GEE to identify any error in the methodology.

This study also used a growing season maximum value as the only time-series composite. Future studies could explore using different reduction methods (such as mean or minimum) across other time slices such as spring/summer or leaf-on/leaf off.

Finally, additional indices and band combinations that exploit the additional spectral information in Sentinel-2 should be explored. NDVI can be computed using red edge bands rather than red for instance, and as Sentinel-2 becomes more commonly used in studies, other useful indices are likely to emerge.

With regard to the processing of field data, more time should be spent dealing with the nonforested data types. For instance, this analysis lumped all nonforested\non-stocked forest types into one category. Splitting these into more detailed cover classes may improve the model results by helping the classifier parse out differences in those classes.

4. FINAL RECOMMENDATIONS

The initial purpose of this study was to attempt to improve random forest predictions by incorporating additional imagery sources for the DNRC's project. This shifted to a deeper analysis of what was limiting the success of these predictions and eventually grew into an examination of the DNRC's work-flow from data-collection to stand-level predictions. The two aspects focused on in this thesis, the selection of inventory plots and the sources of imagery used in predictive models, have already been discussed in detail. This final chapter will summarize and make additional suggestions on how the DNRC could improve their efforts to incorporate remote sensing and raster analysis into their inventory program in the future.

In general, the largest area of potential improvement deals with the selection and collection of plot-level field data. First, the plot locations should be chosen carefully and a simple random sample, even one that passes the K-S test, is not ideal. A GRTS sample in either geographic or spectral space will inherently offer advantages over a simple random sample. Evaluating a sampling design to make sure that the sample satisfies spatial and spectral criteria would not take much additional time. Stratifying the study area by some physical attributes (such as elevation, aspect, or precipitation) to increase the amount of vegetative variation captured by the sample might be explored.

Next, data collection methods at these locations needs to be re-examined. Currently, determinations for size class and forest type at the inventory plots are subjective rather than based on data collected. This can cause many issues with repeatability and consistency between different observers. It also requires observers to be extremely precise in the way they handle situations when multiple classes are seemingly equally represented at the plot. Additionally, reallocating classes to

avoid overlap, specifically in the case of forest type, would decrease confusion among predicted classes.

As DNRC personnel have already recognized, improving the accuracy of GPS units used to navigate to the plot locations would likely improve results. Using iPads with an accuracy of ~9 m (29.5 ft) is problematic since that could cause the subplots (9 m radius) to overlap within in the same main plot (36 × 36 m; 118 × 118 ft).

When generating stand boundaries, care should be taken to ensure that the stands are indeed homogeneous or nearly so. If stands need to be manually digitized, some measure of spectral and/or physical variability within the stands should be computed and evaluated.

Building predictive models should incorporate as much information as possible and the exploration of additional textural metrics may provide the most improvement in predictive power. Additional contrast bands, entropy, or standard deviation focal analyses are worth testing. When generating prediction surfaces, using a probabilistic outcome rather than a discrete one may be useful. Additionally, keeping data in numeric form (e.g. DBH) will likely result in more accurate predictions for size class, as lumping data into classes too soon will remove some information. Also, with TPA, it may be useful divide this into size classes. For instance, build separate models predicting TPA of trees less than or greater than a threshold or set of thresholds per stand.

Prior to generating predictions using RF or other algorithms, all undesired areas should be masked out. An accurate mask for water and roads would eliminate some spectral confusion.

Overall, the DNRC pilot project from which this thesis originated has largely accomplished its goal. Methods were tested, refined, and retested and valuable lessons learned from the outcomes. The organization is currently attempting to implement another raster-based project in another area

of the state and hopefully the results of this thesis will aid them as they continue to develop their program in the years to come.

WORKS CITED

Belgiu, M., & Dragut, L. (2016). Ramdom forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*.

Davies, K. W., Petersen, S. L., Johnson, D. D., Davis, D. B., Madsen, M. D., Zvirzin, D. L., & Bates, J. D. (2010). Estimating juniper cpver from NAIP Imagery and evaluating relationships between potential cover and environmental variables. *Rangeland Ecology and Management, 63*(6).

Dixon, G. E. (2002). *Essential FVS: A user's guide to the Forest Vegetation Simulator*. Fort Collins: USDA Forest Service, Forest Management Service Center.

Environmental Systems Research Institute, Inc. (ESRI). (2016). ArcMap 10.5. Redlands.

Evans, D. G., & Jones, S. M. (1987). Detecting Voronoi (area-of-influence) polygons. *Mathematical geology., 19*(6).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27.

Gregoire, T. G., & Valentine, H. T. (2007). *Sampling Strategies for Natural Resources and the Environment.* Boca Raton: Taylor and Francis.

Hayes, M. M., Miller, S. N., & Murphy, M. A. (2014). High-resolution landcover classifications using Random Forest. *Remote Sensing Letters, 5*(2).

Hijmans, R. J. (2019). raster: Geographic Data Analysis and Modeling. R package version 3.0-7. Retrieved from https://CRAN.R-project.org/package=raster

Hogland, J. S., & Anderson, N. M. (2015). Estimating FIA plot characteristics using NAIP Imagery, function modeling and the RMRS Raster Utility coding library. *Forest Inventory & Analysis (FIA) Symposium*. Portland: USDA Forest Service.

Hogland, J., & Anderson, N. (2010). RMRS Raster Utility.

Hogland, J., Anderson, N., St. Peter, J., Drake, J., & Medley, P. (2018). Mapping Forest Characteristics at Fine Resolution across Large Landscapes of the Southeastern United States Using NAIP Imagery and FIA Field Plot Data. *International Journal of Geo-Information, 7*(4).

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing, 8*(3).

Jones, M. O., Allred, B. W., Naugle, D. E., Maestas, J. D., Donnelly, P., Metz, L. J., Karl, J., Smith, R., Bestelmeyer, B., Boyd, C., Kirby, J.D., & McIver, J. D. (2018). Innovation in rangeland monitoring: annual, 30 m, plant functional type percent cover maps for U.S. rangelands, 1984 - 2017. *Ecosphere, 9*(9).

Kincaid, T., Olsen, A., & Weber, M. H. (2019). spsurvey: Spatial Survey Design and Analysis. R package version 4.1.0.

Liaw, A., & Wiener, M. (2002). Classifiaction and Regression by randomForest. *R News, 2*(3).

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2008). *Remote Sensing and Image Interpretation.* Hoboken: John Wiley & Sons, Inc.

Liu, Y., Gong, W., Hu, X., & Gong, J. (2018). Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sensing, 10*(6).

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association, 46*(253), 68-78.

Montana DNRC. (2018). *Trust Lands Management*. Retrieved December 5, 2018, from http://dnrc.mt.gov/divisions/trust

Moreno-Martinez, A., Camps-Valls, G., Kattge, J., Robinson, N., Riechstein, M., van Bodegom, P., Kramer, K. Corenelissen, J. H. C., Reich, P., Bahn, M., Niinemets, U., Penuelas, J., Craine, J. M., Cerabolini, B. E. L., Minden, V., Laughlin, D. C., Sack, L., Allred, B., Baraloto, C., Byun, C., Soudzilovskaia, N. A., & Running, S. W. (2018). A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sensing of Environment, 218*.

Mura, M., Bottalico, F., Giannetti, F., Bertani, R., Giannini, R., Mancini, M., Orlandini, S., Travaglini, D., & Chirici, G. (2017). Exploiting the capabilities of the Sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems. *International Journal of Applied Earth Observation and Geoinformation, 66*(2018).

Natural Resources Conservation Service. (2016, September). *National Water and Climate Center Report Generator 2.0*. Retrieved September 24, 2019, from https://wcc.sc.egov.usda.gov/reportGenerator/view/customSingleStationReport/daily/500:MT:SNTL%7Cid=%22%22%7Cname/-29,0/WTEQ::value,WTEQ::median_1981,WTEQ::pctOfMedian_1981,SNWD::value,PREC::value,PREC::average_1981,PREC::pctOfAverage_1981,TMAX::value,TMIN

Noi, P. T., & Kappas, M. (2018). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors, 18*(18).

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Robinson, N. P. (2017). Enhancing conservation with high resolution productivity datasets for the conterminous United States. *Graduate Student Theses, Dissertations & Professional Papers.*

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olma, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of nueral networks, random forest, regression trees and support vector machines. *Ore Geology Reviews, 71*.

Stander, H., Rasmussen, M., Long, B., Vickery, B., & Burton-Desrocher, J. (2015). *Montana Department of Natural Resources & Conservation, State Trust Lands Sustainable Yield Calculation.* Portland: Mason, Bruce, & Girard, Inc.

Stevens, D. L., & Olsen, A. R. (2004). Spatially Balanced Sampling of Natural Resources. *Journal of the American Statistical Association, 99*(465), 262-278.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*.

USDA Farm Service Agency. (2019, November 19). *NAIP Imagery*. Retrieved from USDA Farm Service Agency: https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/

Whitlock, C., Cross, W. F., Maxwell, B., Silverman, N., & Wade, A. A. (2017). *2017 Montana Climate Assessment.* Missoula: Montana Institute on Ecosystems.

Zhu, X., & Liu, D. (2014). Accurate mapping of forest types using dense seasonal Landsat time-series. *ISPRS Journal of Photogrammetry and Remote Sensing, 96*(2014).

APPENDIX A

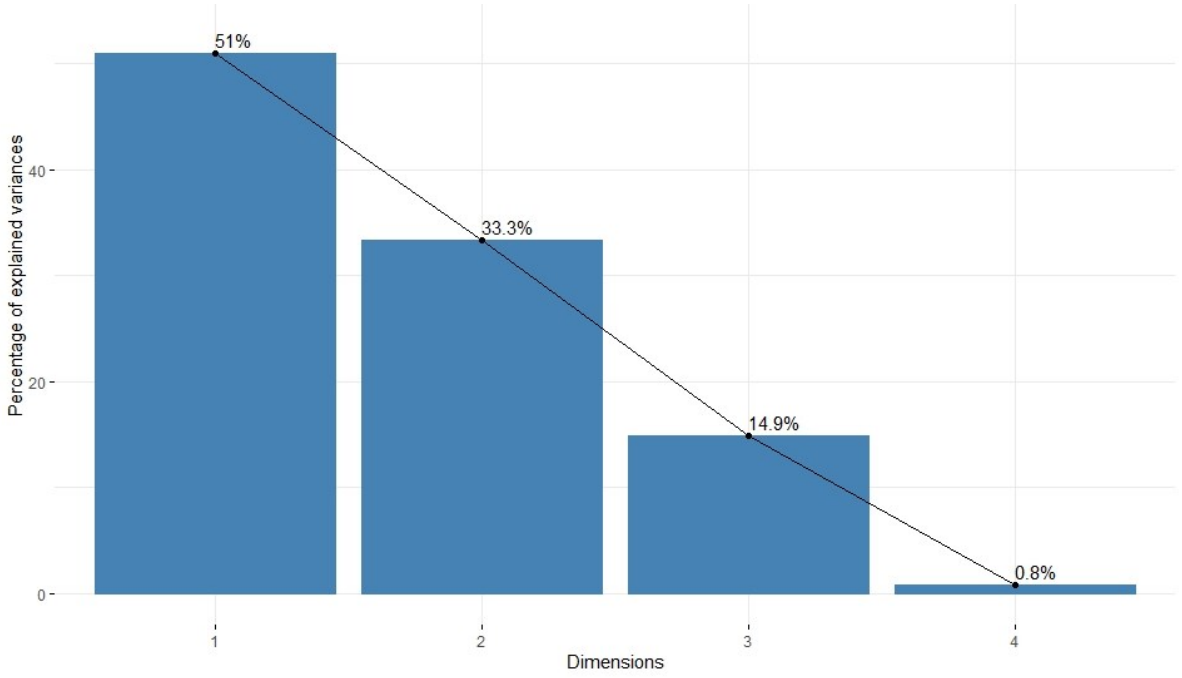

Figure A1: Results from the principal component analysis (PCA) performed on all bands of the resampled NAIP 2015 image.

Table A1: Results from the principal component analysis (PCA) performed on all bands of the resampled NAIP 2015 image

	Eigen Value	% Variance	Cumulative % Variance
PC 1	2.04	50.97	50.97
PC 2	1.33	33.30	84.27
PC 3	0.60	14.91	99.18
PC 4	0.03	0.82	100.00

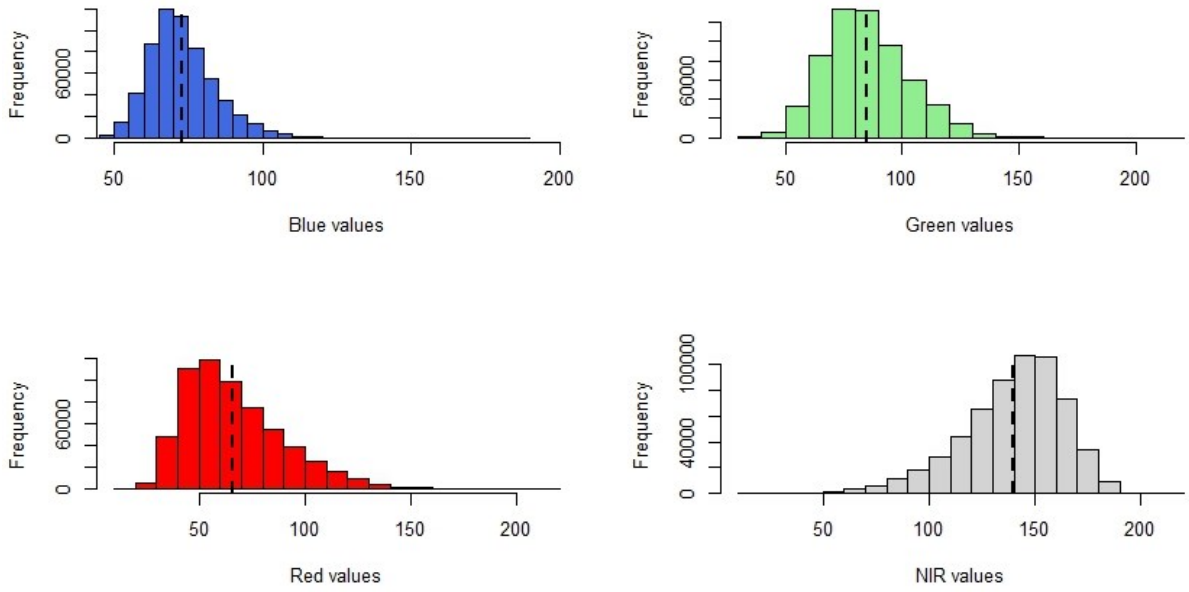

Figure A2: Histogram of the values in each spectral band of the NAIP 2015 image used to define the study area/population. The minimum and maximum from each are shown in Figure 5 for comparison to the samples.

Table A2: Classification error matrix for the NAIP 2015 random forest model predicting stand size class (SSC).

		Reference				
		Non-forest	Nonstocked	Seedling	Poletimber	Sawtimber
Mapped	Non-forest	6	3	0	1	3
	Nonstocked	2	7	2	0	3
	Seedling	0	1	5	2	8
	Poletimber	0	0	1	2	13
	Sawtimbber	1	0	2	3	55

Table A3: Classification error matrix for the NAIP 2017 random forest model predicting stand size class (SSC).

		Reference				
		Non-forest	Nonstocked	Seedling	Poletimber	Sawtimber
Mapped	Non-forest	10	1	0	0	2
	Nonstocked	3	9	1	1	0
	Seedling	0	2	5	2	7
	Poletimber	0	0	1	0	15
	Sawtimeber	0	0	2	3	56

Table A4: Classification error matrix for the October Sentinel - 2 random forest model predicting stand size class (SSC).

		Reference				
		Non-forest	Nonstocked	Seedling	Poletimber	Sawtimber
Mapped	Non-forest	7	3	0	0	3
	Nonstocked	3	7	1	2	1
	Seedling	0	1	4	1	10
	Poletimber	0	0	0	4	12
	Sawtimeber	0	2	1	3	55

Table A5: Classification error matrix for the growing season Sentinel - 2 random forest model predicting stand size class (SSC).

		Reference				
		Non-forest	Nonstocked	Seedling	Poletimber	Sawtimber
Mapped	Non-forest	6	1	2	1	3
	Nonstocked	3	10	0	0	1
	Seedling	1	1	6	1	7
	Poletimber	0	0	3	3	10
	Sawtimber	1	3	4	1	52

Table A6: Classification error matrix for the NAIP 2015 random forest model predicting forest type (FTYPE).

		Reference													
		AF	AS	CO	DF	DL	ES	GF	LP	MC	OH	RC	WL	WP	NF
Mapped	AF	2	0	0	0	0	3	1	1	3	0	0	3	0	0
	AS	0	0	0	0	0	0	1	0	0	0	0	1	0	0
	CO	1	0	1	0	0	0	0	0	1	0	0	1	0	0
	DF	0	0	0	0	0	0	0	0	1	0	0	0	0	1
	DL	1	0	0	0	0	0	1	0	1	0	0	0	0	0
	ES	4	0	0	0	0	0	2	0	3	0	0	0	0	1
	GF	1	0	0	0	0	1	4	1	4	0	1	2	0	0
	LP	1	0	0	0	0	0	1	1	4	0	1	0	0	1
	MC	5	0	0	0	0	2	3	3	2	0	1	2	0	2
	OH	0	0	0	0	0	0	1	0	1	0	0	0	0	0
	RC	0	0	0	0	0	0	2	0	1	0	0	0	0	0
	WL	1	1	0	0	0	0	0	2	5	0	0	0	0	1
	WP	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	NF	0	0	0	0	0	1	0	1	2	0	0	1	0	23

Table A7: Classification error matrix for the NAIP 2017 random forest model predicting forest type (FTYPE).

		Reference													
		AF	AS	CO	DF	DL	ES	GF	LP	MC	OH	RC	WL	WP	NF
Mapped	AF	0	0	0	0	0	2	1	0	6	0	0	2	0	2
	AS	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	CO	0	0	0	0	1	1	1	0	0	0	0	1	0	0
	DF	1	0	0	0	0	0	0	0	0	0	0	0	0	1
	DL	0	0	0	0	0	0	0	1	2	0	0	0	0	0
	ES	1	0	1	0	0	0	1	1	6	0	0	0	0	0
	GF	2	2	0	0	1	0	0	2	3	0	0	1	0	2
	LP	3	0	0	0	0	0	1	3	2	0	0	0	0	0
	MC	4	0	0	0	0	4	2	1	6	0	1	0	0	2
	OH	0	0	0	0	0	0	0	0	1	0	0	1	0	0
	RC	1	0	0	0	0	1	0	0	0	0	0	1	0	0
	WL	1	0	0	0	0	0	2	1	4	0	0	0	0	2
	WP	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	NF	1	0	0	0	0	1	0	0	0	0	0	2	0	24

Table A8: Classification error matrix for the October Sentinel – 2 random forest model predicting forest type (FTYPE).

		Reference													
		AF	AS	CO	DF	DL	ES	GF	LP	MC	OH	RC	WL	WP	NF
Mapped	AF	4	0	0	0	1	3	1	0	1	0	0	2	0	1
	AS	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	CO	1	0	0	0	0	1	0	0	0	0	0	0	0	2
	DF	1	0	0	0	0	0	0	0	0	0	0	0	0	1
	DL	1	0	0	0	0	0	0	0	1	0	0	1	0	0
	ES	1	0	0	0	0	0	1	1	3	1	0	0	0	3
	GF	1	1	0	0	0	0	3	3	3	0	0	1	0	1
	LP	0	0	0	0	0	0	2	5	2	0	0	0	0	0
	MC	1	0	0	0	0	1	4	1	8	0	1	1	0	3
	OH	1	0	0	0	0	0	0	0	0	0	0	1	0	0
	RC	1	0	0	0	0	0	0	0	2	0	0	0	0	0
	WL	3	0	0	0	0	0	1	0	2	0	0	0	0	4
	WP	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	NF	0	0	1	0	0	1	2	0	2	0	0	2	0	20

Table A9: Classification error matrix for the growing season Sentinel – 2 random forest model predicting forest type (FTYPE).

		Reference													
		AF	AS	CO	DF	DL	ES	GF	LP	MC	OH	RC	WL	WP	NF
Mapped	AF	2	0	0	0	0	2	1	0	5	0	0	1	0	2
	AS	0	0	0	0	0	0	1	0	1	0	0	0	0	0
	CO	0	0	0	0	1	0	1	1	0	0	0	1	0	0
	DF	1	0	0	0	0	0	1	0	0	0	0	0	0	0
	DL	0	0	0	0	0	0	0	1	1	1	0	0	0	0
	ES	1	0	0	0	0	0	2	1	1	1	1	1	0	2
	GF	0	0	1	0	0	1	3	2	1	0	0	2	0	3
	LP	0	0	0	0	1	1	1	3	2	0	0	0	0	1
	MC	2	1	0	0	0	3	2	2	4	0	1	0	0	5
	OH	0	0	0	0	0	0	0	1	0	0	0	0	1	0
	RC	0	0	0	0	0	2	0	0	1	0	0	0	0	0
	WL	0	0	1	0	0	0	2	0	2	0	0	3	0	2
	WP	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	NF	0	0	0	0	0	0	3	0	3	0	0	0	0	22