Winter 2021

Lidar-Based Riparian Forest Assessment of the Nooksack River, Washington

Julia Tatum

Lidar-Based Riparian Forest Assessment of the Nooksack River, Washington

By

Julia Tatum

Accepted in Partial Completion
of the Requirements for the Degree
Master of Science

ADVISORY COMMITTEE

Dr. David Wallin, Chair

Dr. James Helfield

Michael Maudlin

Dr. Andrew Bunn

GRADUATE SCHOOL

David L. Patrick, Dean

Master's Thesis

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Julia Tatum

February 24, 2021

Lidar-Based Riparian Forest Assessment of the Nooksack River, Washington

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Julia Tatum
March 2021

Abstract

This paper addresses two applications of lidar remote sensing: an area-based watershed-scale analysis of forest structure used to prioritize riparian restoration projects for salmon, and an individual-tree-based analysis for tree species classification. Salmon conservation is extremely important in the Pacific Northwest, but restoration efforts have been hampered by insufficient data on riparian stand conditions. I used lidar to map riparian stand structure and composition along the Nooksack River, Washington, and developed a restoration priority model based on six factors: riparian stand conditions, shade potential, cause of riparian impairment, susceptibility to climate change, position in the watershed, and proximity to intact riparian forest. Nine reaches (out of 268 total) were identified as priority targets for riparian restoration. Four of these reaches were on the upper South Fork, two were in the lower South Fork, two were in the lower Middle Fork, and one was in the North Fork near Maple Falls.

At the individual tree level, I compared six different approaches using five different algorithms to sort a discrete-return lidar point cloud into segments representing individual trees. Using these segments, I built models to predict height, diameter, conifer/deciduous classification, and species. Using the best segmentation model, I was able to classify black cottonwood (*Populus trichocarpa)*, Douglas fir (*Pseudotsuga menziesii)*, and red alder (*Alnus rubra)* with user's accuracies up to 89%, overall accuracies of 83-92%, and kappa values above 0.6. This was the first landscape-scale study attempting to classify tree species in natural landscapes in the Pacific Northwest.

Acknowledgements

Many thanks to my advisor, Dr. David Wallin, and my thesis committee members, Dr. James Helfield, Michael Maudlin, and Dr. Andy Bunn, for their support and feedback throughout this project. Working with a study area of this size presents special challenges, and I would like to thank the Whatcom Land Trust, Washington DNR, Seattle City Light, Sierra Pacific Timber Industries, USFS, Whatcom County Parks, Maberry Packing LLC, the Nooksack Tribe, and other landowners for allowing me access to their land to collect ground-truth data. Special thanks are due to Chris Hankey of the DNR for his help with fieldwork logistics. Collecting ground-truth data at this scale would not have been possible without the help of Nicole VandePutte, whose assistance with both fieldwork and data entry was invaluable. I would also like to thank Maria Tatum for her assistance with GIS and python related issues, and I thank the WWU Geology Department and the Nooksack Tribe for their loan of field equipment. This work was funded by the Nooksack Indian Tribe.

Table of Contents

List of Figures

List of Tables

1.0 - Introduction

1.1. Salmon Recovery Planning in the Pacific Northwest

Salmon have great economic, environmental, cultural, and spiritual importance in the Pacific Northwest. Pacific Northwest salmon stocks are declining, so conservation is a high priority in the region. Declines are partly due to reduced quality and availability of freshwater habitat (Gregory and Bisson 1997; Paulsen and Fisher 2001), and climate change is expected to worsen current problems (Grah and Beaulieu 2013; Croizer 2016). Millions of dollars are spent on restoration efforts in the Pacific Northwest every year (Roni et al. 2002), but most projects are small in scale (typically addressing less than 10 percent of the watershed), and they are not always optimally targeted to support overall salmon population goals (Roni et al. 2018). As such, it is important to strategize restoration efforts across the landscape and to find ways to focus restoration efforts onto the areas where they will do the most good.

Restoration projects need to be designed to support overarching goals at the watershed level. Beechie et al. (2013a) proposed that an effective restoration plan would 1) determine what actions are necessary to restore habitat availability, quality, and diversity, 2) determine which habitats, if restored, will most affect overall salmon success, and 3) take into consideration how human infrastructure and land use may constrain restoration efforts. The relative priority of various habitat restoration actions (for example, reconnecting existing side-channel habitat or replanting riparian vegetation) is dependent on the scale of the proposed restoration project, the overall impairment status of the watershed, and the urgency of habitat needs (Roni et al. 2002). For example, if access to habitat is poor because of natural or human-made obstacles, then restoring vegetation will have little effect. In an area with acceptable connectivity, such as the

Nooksack River, Washington, restoring hydrologic, geologic, and riparian processes increases in importance (Roni et al. 2002).

In general, watershed-level habitat restoration has been limited by the difficulty of collecting detailed habitat information at that scale (Roni et al. 2018). Direct habitat assessment by field teams provides the best information, but it is prohibitively expensive over large areas (Beechie et al. 2013a). An increasingly popular approach is to use remotely sensed data to assess habitat conditions. Remotely sensed information must be calibrated with field data before it can be used, but it still greatly reduces the cost and effort required to document conditions for an entire watershed (Beechie et al. 2013a). Until recently, most projects used some form of satellite or airborne imagery (Beechie et al. 2013a), but technological and computing advances are providing other options. Light detection and ranging (lidar) is a rapidly expanding technology that has the potential to overcome many of the limitations of traditional remote sensing techniques. Lidar is particularly appropriate for assessing riparian vegetation because it is readily available for many Pacific Northwest streams (DNR 2018), and it can be used to measure 3D forest structure with high accuracy and quantifiable error (Moskal et al. 2015; Tompalski et al. 2017).

1.2. Influence of Riparian Forest on In-Channel Habitat

Riparian areas, the transition zone between terrestrial and aquatic habitat, critically influence in-stream habitat quality. Restoring riparian conditions is a major focus of salmon conservation projects (Naiman and Latterell 2005). Riparian forests contribute to in-stream habitat quality by providing shade, supplying wood debris, supplying leaf litter and terrestrial insects, stabilizing banks, retaining sediment, and filtering nutrients and chemical pollutants

(Gregory et al. 1991; Naiman and Latterell 2005). Mature and old growth riparian forests provide the greatest benefits, but the history of logging and land clearing in the Pacific Northwest has caused riparian patches of old growth or mature forest to be scattered and rare (Gregory and Bisson 1997; Brown and Maudlin 2007; Hyatt 2007). Because of this, restoration projects for salmon may involve managing existing riparian forests to encourage the development of mature forest characteristics in comparatively young stands that otherwise would not yet perform those roles (Roni et al. 2002).

Thermal regulation is a key function of riparian forests in the Pacific Northwest. Salmon live in cold water, generally between 3°C and 15°C depending on life stage and species (EPA 2003). Dense mature riparian forests that provide shade during the heat of the day are critical, especially in lowland areas (Naiman et al. 2000). Younger, shorter forests provide less protection. Riparian forests also produce large woody debris (root wads and logs with a diameter >10 cm) that contribute to log jam and pool formation. The effects of large woody debris depend on channel width relative to wood volume, and larger logs supplied by mature forests provide benefits to larger channels (Bisson et al. 1987; Fox and Bolton 2007). Under the right conditions, log jam pools can provide thermal refuges to salmon (Bisson et al. 1987; Beechie and Sibley 1997). Conifers are more influential than deciduous trees in terms of both shade and log jam formation because they grow taller and, once in the water, decompose more slowly (Bisson et al. 1987; Bilby and Ward 1989; Hyatt and Naiman 2001).

In addition to providing possible thermal refuges, log jams provide cover, help to maintain forested islands by reducing erosion, and contribute to flow rate variation across the landscape. Flow rate variation across the landscape allows for more spatially diverse deposition of coarse gravel and more widespread spawning habitat availability (Beechie et al. 2013b). Large

woody debris also contribute to side-channel formation, which provide flow-refuge areas for juvenile salmon. Low-flow areas where fish can shelter enable reduced metabolic rates for swimming and helps them conserve energy (Beechie and Sibley 1997; Beechie et al. 2013b).

Riparian vegetation stabilizes stream banks, reduces erosion, filters sediment, and contributes to channel stability (Naiman et al. 2000; Beechie et al. 2013a). Speaking in general terms, confined streams, in which bankfull width is restricted by natural rock formations or artificial structures, typically have low habitat potential because flow rates are too high. At the other extreme, unrestricted streams that become braided can be poor habitat because of elevated temperatures, shallow water depth, and reduced food availability caused by extreme channel instability and the corresponding lack of nearby vegetation. Riparian forests help keep stream ecosystems in the middle of these two extremes. Their roots slow erosion and create stability during moderate flow times without restricting the stream so much that flow increases and beneficial spawning gravel is lost (Beechie et al. 2013a). In addition to controlling in-channel coarse sediment, riparian vegetation also filters fine sediment present in terrestrial runoff. This helps maintain stretches of spawning-suitable coarse gravel by reducing inputs of fine sediment that, in excess, clog interstitial spaces in coarse gravel thereby restricting oxygen access to developing salmon embryos and can reduce food availability for juvenile salmon (Suttle et al. 2004; Kemp et al. 2011).

Riparian forests influence in-channel food web dynamics. Forest vegetation provides leaf litter which feeds invertebrates at the base of the food chain (Naiman and Latterell 2005). Riparian vegetation is also a source of terrestrial insects, which feed young fish directly (Nakano et al. 1999). Mature forests with vegetation overhanging the water tend to provide plentiful food

for juvenile salmon, whereas barren gravel stretches bordering unstable stream channels provide relatively little (Beechie et al. 2013b).

1.3. Salmon and the Nooksack Watershed

The Nooksack River in Washington State contains a diverse assembly of salmon, but much of the river needs to be restored in order to properly support them (WRIA 1 Salmon Recovery Board 2005). The river contains 25 salmon stocks, only three of which are considered healthy (WRIA 1 Salmon Recovery Board 2005). Several stocks in the Nooksack River are listed under the Endangered Species Act (ECOS 2020). Bull trout (*Salvelinus confluentus)*, Puget Sound steelhead (*Oncorhynchus mykiss),* and two stocks of Puget Sound chinook (*Oncorhynchus tshawytscha)* are listed, and Nooksack coho (*Oncorhynchus kisutch)* is considered a species of concern (WRIA 1 Salmon Recovery Board 2005). The South Fork early chinook sub-population is especially vulnerable, and the recovery of both the North Fork and South Fork early chinook is considered essential for the recovery of the Puget Sound chinook environmentally significant unit (WRIA 1 Salmon Recovery Board 2005). Unfortunately, salmon habitat quality in the Nooksack Watershed is impacted both by legacy effects of historic land use and current land use practices. The Salmon Recovery Plan for Water Resource Inventory Area 1 (i.e., WRIA 1, the administrative area that includes the Nooksack River and independent coastal drainages of Whatcom County) identified nine primary habitat factors that limit endangered salmon recovery in the Nooksack River. These nine factors are channel stability, sediment load, habitat diversity, key habitat quantity, obstructions, withdrawal structures, flow, temperature, and chemicals (WRIA 1 Salmon Recovery Board 2005). Most of these problems are closely tied to the removal and degradation of riparian forest and the loss of large woody debris.

The Nooksack River includes three main forks (North Fork, Middle Fork, and South Fork), each of which experience different local conditions and pose different challenges for salmon habitat restoration. The North Fork is the main fork. It is joined by the Middle Fork at river mile (RM) 40.5 and by the South Fork at RM 36.6 (WRIA 1 Salmon Recovery Board 2005). The North Fork is glacier-fed, so high water temperatures are usually not a problem except in the tributaries or at exceptionally low-flow times (Smith 2002). Much of the North Fork is too braided and unstable to make good incubation, spawning, and rearing habitat; historically the channels were likely more stable and vegetated islands were larger and supported older trees (WRIA 1 Salmon Recovery Board 2005; Hyatt 2007). These changes may have been partly due to changes in hydrology related to logging practices in the late 20th century and partly the result of decreased availability of large woody debris. In the lower North Fork, only 1.3 percent of the floodplain has trees large enough to contribute large woody debris comparable to historical sizes (Hyatt 2007). As a short-term fix for this problem, restoration groups have been building artificial log jams along the North Fork to encourage forested island and floodplain stability to increase side channel habitat and increase the river's salmon incubation success by diversifying habitat (NNR 2018).

The Middle Fork of the Nooksack is also glacier-fed, and the upper reaches are high quality habitat for salmon. However, the lower portion of the river is highly braided and suffers from high temperatures, high fine sediment load, and lack of key habitats (LNR 2011). This is driven by similar factors to those affecting the North Fork (WRIA 1 Salmon Recovery Board 2005). Until recently, access to the high-quality upper reaches was restricted by a diversion dam at RM 7.2 (WRIA 1 Salmon Recovery Board 2005). Removal of the dam has been considered the highest restoration priority for the North Fork/Middle Fork early chinook population in the

Nooksack Watershed. The dam removal process started in early 2020, and in-water work was completed in fall 2020, although the upland portions of the site will be under construction for several more months (City of Bellingham 2021).

The South Fork is listed as an impaired waterbody due to high water temperatures (Kennedy et al. 2020). Problems with temperature are related to land use practices as well as the natural hydrology of the river. The South Fork is fed by snowpack from Twin Sisters Mountain, so flow decreases substantially during the warm summer months (Soicher et al. 2006). The lower 16 miles of the river flow through an open valley, and water temperatures routinely exceed optimum levels for life stages present during the late summer and early fall (WRIA 1 Salmon Recovery Board 2005; Kennedy et al. 2020). Riparian vegetation is lacking along the lower reaches. Shade from riparian trees is still important in the lower reaches, but stream widths limit the magnitude of possible shade effects (Kennedy et al. 2020). The upper reaches, which are narrower, suffer from degraded riparian forests and related to legacy logging practices, although these areas are largely protected under the current rules (Brown and Maudlin 2007).

Lack of large woody debris is a key problem in both upper and lower reaches of the South Fork, so artificial log jams have been built to temporarily address this deficiency (NNR 2015). Similar to the other forks, logjams are very rare compared to historical levels, partly due to active removal during the 20th century and partly due to low replacement rates (Collins and Sheikh 2004a). Large woody debris recruitment is low because there are fewer mature trees relative to historical conditions and because of channelization and bank armoring. Bank armoring protects roads and structures but results in low habitat diversity and restricts large woody debris recruitment from channel movement (Maudlin et al. 2002; WRIA 1 Salmon Recovery Board 2005). Under current regulations riparian forests along the upper South Fork

should be on a trajectory for recovery, although without active intervention recovery may take 50-100 years (Brown and Maudlin 2007), but because of the different regulatory environment the lower South Fork is in need of active intervention to restore salmon habitat (Maudlin et al. 2002).

Restoration efforts are ongoing throughout the Nooksack, but the size of the watershed makes wholistic restoration challenging. It has been difficult to target sites to maximize the benefits of salmon habitat restoration, especially because the three forks have different problems and different factors restricting restoration. If restoration could be targeted more precisely to areas with high habitat potential, it would be possible to see greater benefits to salmon within a shorter timeframe and at lower cost (Roni et al. 2002). A key data gap is lack of information regarding the distribution and structure of riparian forest stands along the river (Puget Sound Partnership 2018). To address this need, my study aims to use lidar data to assess riparian forest structure and model some aspects of riparian function for the Nooksack watershed based on a combination of topography, stream channel morphology (including channel width), and riparian forest structure. This model and its component layers could then be used to identify opportunities for restoration efforts based on current conditions and estimated stand growth trajectories.

There have been a few studies of riparian vegetation along the Nooksack, but updated and more extensive information is needed. Studying a portion of the South Fork, Brown and Maudlin (2007) used high resolution airborne laser scanning to map potential for large woody debris recruitment. They found that less than 1% of their study area contains mature trees and that conifer-dominated stands, which contribute the majority of large woody debris, are spatially limited. Although the relative proportions of different tree species are mostly consistent with conditions in the 1880s (Collins and Sheikh 2004a), the natural scarcity of conifers means that land-use activities that have favored the removal of large conifers have greatly impacted large

woody debris recruitment (Brown and Maudlin 2007). Supporting these findings, Capuana (2013) mapped size classes and deciduous versus coniferous cover for a portion of the South Fork basin and concluded that the site is dominated by small, deciduous trees that are sub-optimal for logjam formation. In the lower North Fork, Hyatt (2007) used a spatially-limited airborne laser scanning dataset to quantify vegetation heights for trees in the floodplain and on portions of the valley walls. This paper concluded that vegetated area had been consistently declining and that riparian stands are much younger and smaller than they would have been historically.

In a more widespread image-based assessment, Coe (2001) investigated large woody debris recruitment potential and concluded that large forest stands and large woody debris recruitment potential are limited throughout the watershed. Building on this data, Hyatt et al. (2004) investigated shade and wood recruitment relative to channel size and prioritized areas for restoration based on the modeled pool-forming ability of large woody debris potentially recruited from adjacent stands. This study found that 74% of stands that were too small to provide pool-forming large woody debris are in agricultural land use areas (Hyatt et al. 2004). These prior analyses are extremely useful and have provided the background for many restoration projects in recent years. However, most of them do not provide a watershed-level view, and the studies reported in Coe (2001) and Hyatt et al. (2004) had some technical limitations due to the age and format of the available data.

1.4. Background of Lidar as a Forest Inventory Tool

Lidar is an active remote sensing technology that generates highly accurate 3D images of surfaces and objects. Active remote sensing systems emit their own illumination rather than

relying on energy emitted by the sun. A comparatively recent technique, lidar was first applied to forestry in the 1990s in Scandinavian countries (Hyyppä et al. 2008). Data quality has increased markedly in the past three decades (Mazza and Gatzolis 2018). With these advances, lidar now can measure a wide range of forest attributes that are relevant to conservation decisions and difficult to assess by other means (Wulder et al. 2012; Zlinszky et al. 2015).

One of the most widely used types of lidar is small-footprint discrete-pulse airborne laser scanning, commonly abbreviated as ALS (Hyyppä et al. 2008; Wulder et al. 2012). These systems can be used to collect detailed measurements of entire forest stands far more quickly and at a lower cost than an equivalent undertaking by field crews. High-quality ALS data commissioned by state, local and tribal agencies are publicly available for many stream systems throughout the Pacific Northwest (DNR 2018). Airborne lidar systems use an instrument mounted in an airplane to scan the ground with laser pulses of near-infrared light. They record the amount of time it takes the pulse to return, the precise location of the pulse, and the intensity with which it returns (McGaughey 2018). Pulse intensity is the energy with which the pulse returns and is dependent on the reflectivity of the surface it encountered, as well as other factors. When the instrument is flown over a forested landscape, some of the pulses bounce back from the top of the canopy, some penetrate the canopy but are returned by understory structures, and some penetrate all the way to the ground before returning. The result is a georeferenced point cloud that is detailed enough for individual trees to be clearly visible. Their individual heights can be measured with comparable accuracy to ground-based methods (Hyyppä et al. 2008; Gatziolis et al. 2010; McGaughey 2018), and the vertical distribution of points can be used to quantify the vertical distribution of biomass. Vertical distribution of biomass is difficult to quantify by any other method, including field surveys, so this is one of the greatest strengths of

lidar (Zlinszky et al. 2015). These capabilities combine to make ALS lidar a feasible alternative to manual field surveys for riparian habitat assessments (Moskal and Cooke 2015).

Measuring forest structure and topography with ALS lidar is well established in the forest sciences (Hyyppä et al. 2008; McGaughey 2018). However, differences in forest structure and topography mean that methods developed in Scandinavia are not directly applicable to the forests of the Pacific Northwest, which are typically more structurally complex and situated on more rugged terrain (Mazza and Gatziolis 2013). In response to these challenges, Pacific Northwest scientists have created a list of general specifications for optimal lidar acquisition intended to remove serious biases from standard forest inventory calculations (Gatziolis and Andersen 2008). These specifications have been adopted by the Oregon Department of Geology and Mineral Industries and form the "best practices" standard for lidar work in the region (Mazza and Gatziolis 2013). Within this framework, vertical accuracy of vegetation height measurements over flat ground is typically ± 15.25 cm, and vertical accuracy in steep forested areas is typically ±30-60 cm, both of which are comparable to error in metrics collected by ground crews (McGaughey 2018). The accuracy of these measures is greatly influenced by the density of the vegetation, the density of the laser pulses, and the quality of the filtering algorithms that are used on the raw data (Vauhkonen et al. 2012).

1.5. Characterizing Riparian Structure and Function with Lidar

Lidar is frequently used to help assess riparian conditions. Detailed topographic information derived from lidar can be useful for assessing channel characteristics such as gradient, width, and sinuosity, and lidar can predict many vegetative attributes with high accuracy including canopy cover, canopy height, canopy structure, and shading (Tompalski et al.

2017). Lidar also can be used to predict large woody debris availability (Richardson and Moskal 2016), although this is often derived from density estimates, for which accuracy depends on stand structure. Density estimates tend to be relatively accurate for tall stands with large trees, but stand density tends to be underestimated in shorter stands < 20 m tall (Richardson and Moskal 2011).

Shade modeling from lidar is still in its infancy, but it has recently received a fair amount of interest, and a number of different approaches have been studied. Loicq et al. (2019) found that a lidar-based method outperformed older, less computationally intensive methods for modeling stream temperatures along a river in France. For their model, they calculated the amount of direct and diffuse solar radiation for each water pixel in their study area using a digital surface model derived from lidar. They assumed a consistent transmissivity across their study area (transmissivity is the fraction of solar radiation that passes through the canopy) that was derived from previous research on deciduous trees in the same general region. Because their study area was relatively flat, they didn't need to account for topographic shading.

In the Pacific Northwest, Tompalski et al. (2017) used a voxel-based approach from raw point cloud data to model shade. Their approach was relatively simple: categorizing pixels as shaded or not shaded from direct solar radiation on an hourly basis that was summed up to give an estimate of total hours of shade throughout the day. They did not consider diffuse radiation or transmissivity, although these could be incorporated into a voxel-based approach as was done in a more recent study (Richardson et al. 2019). Richardson et al. (2019) tested the accuracy of two lidar-based solar insolation models, one voxel-based and one raster-based, along heavily forested streams in Oregon. They accounted for transmissivity by calculating a light penetration index based on the ratio of ground first return points relative to the total number of lidar first return

points. With this ratio, they used the Beer-Lambert law (Richardson et al. 2009) to estimate light extinction rates for the canopy. In order to account for solar angle, transmissivity was calculated with a slight offset relative to the ground it was shading. The raster-based approach was easily applied across their entire study area using tools in ArcGIS, and the researchers concluded that both methods were accurate enough not to need field-based calibration.

1.6. Using Remote Sensing for Salmon Habitat Assessments

Laser scanning has great potential as a tool to support salmon habitat assessments (Moskal et al. 2017), but it only recently started being widely used. Researchers have used remote sensing to map habitat suitability for fish in the Pacific Northwest, but few of them focused exclusively on lidar methods. As an exception, Tompalski et al. (2017) used lidar to derive a suite of characteristics for riparian forests in northern Vancouver Island, British Columbia, and their study demonstrates the utility of lidar for this purpose. However, such wide scale lidar-based studies are still relatively rare. It is far more common for researchers to classify land cover type using some form of aerial photography, as was the case for Fullerton et al. (2006) in their study of salmonid habitat availability in the interior Columbia River Basin. Other researchers have used lidar to map select components of an overall salmon habitat suitability model. For example, Richardson and Moskal (2016) used airborne lidar data collected during leaf-off (winter) conditions to quantify existing large woody debris in stream channels and to estimate the large woody debris recruitment potential of riparian forest stands. Their measurements of existing large woody debris were excellent, but their measurements of forest stands were only moderately accurate, which they attributed to having used a leaf-off rather than leaf-on dataset.

Mollot and Bilby (2008) used remote sensing techniques to create a salmonid habitat suitability model for the Cedar River Watershed, Washington. They combined stream channel factors, gradient and confinement, with riparian forest metrics that related to tree size and whether the stand was coniferous, deciduous, or mixed. All of their riparian forest metrics were derived from MASTER (Modis/Aster) 5m resolution airborne-collected hyperspectral imagery; their only use of lidar methods was for determining 3D terrain information. This was practical for them because they had access to already-processed imagery data that covered their study site, but few watershed managers have such resources on hand.

Capuana (2013) used ALS to study riparian landcover and riparian forest structure along a portion of the Lower South Fork of the Nooksack River. Capuana (2013) compared ALS and high-resolution satellite imagery (Worldview-2) methods and found that, at least in that region, high-resolution imagery provided no significant advantage over ALS for forest classification. The ALS metrics were height, canopy cover, and rumple (which is the ratio of 3D canopy surface model area to ground model area, and which can be used as a proxy for structural complexity). Although its inclusion in future studies was recommended, Capuana's study did not extend to distinguishing coniferous from deciduous stands using lidar data alone. Also, the spatial scope of the study was limited by the coverage of the high-resolution imagery that could be purchased.

Meixer and Bain (2010) used satellite-derived land cover data to assess stream channel and riparian restoration priority at the reach scale along a river in Ontario, Canada. Their restoration prioritization model considered steam channel condition and riparian condition (calculated from percent forest cover, patch density, and convexity), land ownership, slope, position in the sub-watershed, and adjacency to high-quality habitat. With these inputs, their final

model had 81% agreement with restoration recommendations based on field surveys and stakeholder input. In another satellite-based analysis, Burnett et al. (2007) modeled species-specific intrinsic potential for high-quality salmon habitat in coastal Oregon under current and future conditions based on stream flow, valley constraint, stream gradient, ownership, land use, and land cover type. An even earlier attempt at a GIS-based salmon habitat assessment calculated habitat suitability solely from slope and seral stage (Lunetta et al. 1997). Seral stage was defined in terms of percent cover, relative proportions of deciduous vs coniferous trees, and diameter. These studies contributed useful information to policymakers, but they were necessarily limited in their accuracy because of their reliance on two-dimensional satellite images.

1.7. Individual Tree Detection and Species Identification

Results of lidar analyses may represent an average over the landscape or specific values for individual trees or individual tree clusters. When individual trees are the focus, researchers can extract tree counts, crown area, canopy closure, canopy gaps, and estimate volume and biomass (Hyyppä et al. 2008). With the inclusion of intensity data, tree species (Holmgren and Persson 2004; Vaughn et al. 2012, Vauhkonen et al. 2014; Eitel et al. 2016) and dead wood (Meng et al. 2018) can be classified, although methods are by necessity highly specialized to local conditions and accuracy varies greatly. Many of these attributes are useful when predicting how the forest will affect nearby streams. It is important to note that objects identified as individual trees more often represent “tree approximate objects” consisting of a large tree canopy and several smaller subordinate ones (North et al. 2017; Jeronimo et al. 2018).

There are lots of methods for extracting individual trees, and success depends on forest structure, with different algorithms performing differently under different forest types. In general terms, segmentation algorithms used to identify individual trees either rely on the morphological shape of the canopy surface, or they look for patterns in the spatial configuration of the points in the point cloud (Chen et al. 2020). Tree stem locations, which may be used as seed points for segmentation algorithms, are generally set either as the highest point in a local area or as the region with the highest return density, which is assumed to correspond to the center of the tree (Mongus and Žalik 2015). Chen et al. (2020), Dai et al. (2018), Pirotti et al. (2017), and Vauhkonen et al. (2012) provide detailed comparisons of multiple segmentation approaches and can be referred to for additional background information which, in the interests of brevity, is not included here.

Individual tree species identification from lidar has not been studied as widely in the Pacific Northwest as it has in Europe, and the structural complexity of Pacific Northwest forests introduces an additional level of difficulty. The capacity of lidar to estimate Pacific Northwest tree species was classified as "low" by Tompalski et al. (2017), and Moskal et al. (2017) stated that, as of that time, there were "no known large scale, high accuracy methods for identifying species from lidar". Moskal et al. (2017) identified accurate tree location detection, sensor illumination angles, target tree species, and stand density as features limiting Pacific Northwest species identification accuracy. These authors were considering the application of species identification to riparian areas, which are notoriously composed of dense stands and heterogonous structures; they also only considered one segmentation algorithm when generating their own model. However, under more tightly controlled conditions and superior segmentation methods, other researchers in the Pacific Northwest have reported more encouraging results.

Promising tree-identification rates have been reported by researchers working in the University of Washington Arboretum. Vaughn et al. (2012) compared discrete point data to another type of lidar, waveform data, and reported that the discrete point data allowed them to distinguish five local tree species (two conifers and three deciduous trees) with 79.2 percent accuracy overall, increasing to 97.8 percent when comparing pairs of species (especially deciduous to conifer). A second group of researchers, Kim et al. (2011), used stepwise cluster analysis and a comparison of leaf-on and leaf-off datasets to classify individual trees to genera. They only attempted to classify trees whose crowns did not overlap other trees excessively, and they found that the leaf-on dataset considered by itself produced higher classification accuracies than either the leaf-off dataset or the leaf-on/leaf-off datasets combined. The best classification accuracy from structural variables was 74.9%. The study area used in both these studies was not a natural stand and was both small and topographically homogeneous, so their methods are not directly applicable to non-homogeneous field sites such as along the Nooksack River, and application to such sites would be expected to decrease accuracy. Even so, these studies indicate that some potential does exist for individual tree species classification in the Pacific Northwest, and that further research could be worth pursuing.

1.8. Research Objectives

There were two separate research goals addressed in this thesis. The first was to use area-based lidar to create a reach-scale restoration prioritization model that would support salmon habitat recovery in the Nooksack River watershed. I approached this by generating both a riparian stand conditions assessment and a model of current and prospective shade. These were then combined with morphological and land-use characteristics to give a multi-faceted view of the restoration actions needed for each reach. The second of my research objectives was to

explore methods of individual tree species identification from lidar. I approached this by comparing multiple segmentation methods and quantifying how each method influenced the accuracy of species identification models when applied across the heterogeneous riparian landscapes along the Nooksack River.

2.0 – Riparian Restoration Priority Analysis

2.1. METHODS

2.1.1. Study Area and Lidar Data

The Nooksack River Watershed is located in northwestern Washington State in Water Resources Inventory Area 1. The river is fed by the western slopes of the North Cascades. Upper reaches are mostly confined by bedrock, whereas middle and lowland reaches flow through a mixture of glacial sediments and alluvium at a lower gradient. The three forks of the river join together near Deming, Washington, and then flow into Bellingham Bay in Puget Sound, Washington (Capuana 2013). Multi-year lidar coverage is available for some portions of the river (DNR 2018).

The lidar dataset used for this project was collected in summer 2016 under leaf-on conditions. The overall dataset was contracted by the United States Geological Survey through Quantum Spatial LLC of Corvallis, Oregon, and the subset of data used in this study was subcontracted to Eagle Mapping Ltd of Port Coquitlam, British Columbia. The all-returns point cloud, a digital surface model, and a digital terrain model were made publicly available through the Washington State Department of Natural Resources Lidar Portal (DNR 2018). This dataset was intended for topographic and geophysical analysis and was of high-resolution QL1 data type (Table 1). The intensity values in the point cloud were normalized using proprietary software by the vendor (Lowe et al. 2017). At the time of my study, this was the most up-to-date lidar data available for this region.

My study area consisted of the riparian zones of most of the three forks and a portion of the upper main stem (Figure 1). The up- and downriver boundaries of the study area were

determined by the maximum extent of where there were both records of salmon presence (WDFW 2019) and recent lidar coverage (DNR 2018). The boundary for the study area was defined by a 100 m buffer on the river's recent migration zone, here defined as the area that had either been submerged for 50% or more of the time from 1933-2002 (as determined from aerial photographs, see Collins and Sheikh 2004b) or that was designated active channel in the National Hydrography Dataset published by the USGS (2018). This definition was intended to facilitate remote mapping, minimize accidental inclusion of upland forests, and maximize inclusion of true riparian stands. A 100 m buffer is wider than state- and federally mandated buffers and was intended to account for short-term channel migration.

Table 1: Lidar specifications for USGS 2016 lidar dataset (Lowe et al. 2017).

Data Source	**Specifications**
Vendor	Eagle Mapping
Vertical accuracy (cm)	20.7 cm (vegetated); 8.1 cm (bare earth)
Maximum horizontal error	5 cm
Maximum returns	Unlimited
Sensor and laser system	Riegl - LMS Q1560
Nominal pulse spacing	0.35 m
Laser pulse footprint diameter	33.8 – 40 cm
Resolution density	Average 8 pulses/m^2
Central wavelength	1064 nm
Field of view	60°
Swath width	1560 – 1900 m
Swath overlap	60%
Survey altitude	1350 – 1600 m
Pulse mode	Multi Pulse in air
Intensity	8 bit, scaled to 16 bit

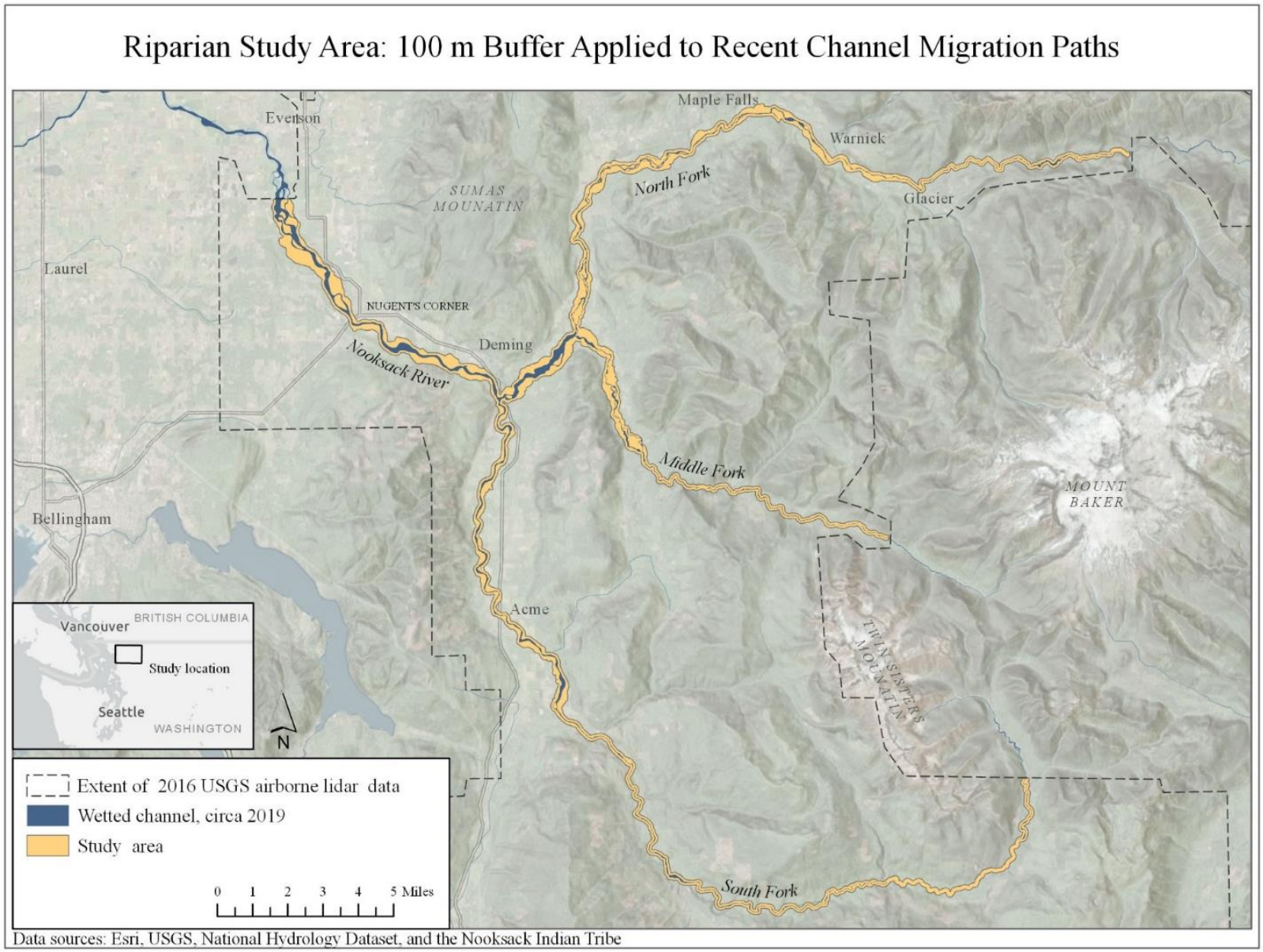

Figure 1: Map of the study area along the Nooksack River, Washington.

2.1.2. Lidar Data Pre-Processing

The lidar point cloud was processed in Fusion 3.8, a non-commercial lidar processing package developed by the US Forest Service (McGaughey 2018). Fusion is a widely used, highly adaptable program that operates primarily at the script level and allows for better direct data quality assessment and visualization than most equivalent commercial packages (Gatziolis and Andersen 2008). The vendor-provided point cloud was normalized and filtered for outliers, then converted into a series of rasters at 1-meter and 30-meter resolution using the Gridmetrics tool embedded in Fusion's AreaProcessor (McGaughey 2018). For inputs intended for use in forest models, height and intensity metrics were calculated from first returns, and returns below 3 meters were excluded. Over 30 different metrics were generated as potential model inputs, but many were correlated with each other so could not be used together in the same model. These metrics represented measures of height, variability in height, intensity, variability in intensity, canopy relief, canopy cover, and canopy structure (Table 2).

Table 2: Lidar metrics considered during model building. All metrics were generated with Fusion's Gridmetrics tool in AreaProcessor. Except when otherwise specified, they represent first returns.

Gridmetrics output	**Description**
Elev_mean	Mean return height
Elev_sd	Standard deviation return height
Elev_var	Variance of return height
Elev_P10 – P99	Mean return height of 10^{th} – 90^{th}, 95^{th}, and 99^{th} percentiles
Relief ratio	Canopy relief ratio ((mean - min) / (max – min))
Int_mean	Mean return intensity
Int_sd	Standard deviation intensity
Int_var	Variance of intensity
Int_P10 – P99	Mean intensity of 10^{th} – 90^{th}, 95^{th}, and 99^{th} percentiles
Percent cover	(All returns above 3m) / (total first returns) * 100
1^{st} returns above 3m	Percentage of 1st returns above the height cutoff
All returns above 3m	Percentage of all returns above the height cutoff
1^{st} returns above mean	Percentage of first returns above the mean height
All returns above mean	Percentage of all returns above the mean height

2.1.3. Ground-Truth Data Collection

Potential plot locations for ground truth data collection were chosen using a proportional random sampling design. The objective was to collect data from at least 100 plots in the riparian zone at varying distances from the active channel. To select potential locations, I adapted methods described in the literature and used principal components analysis (PCA), k-means clustering, and GIS to determine how the primary factors influencing variability in the lidar data were spatially distributed (Frazer et al. 2011; Wulder et al. 2012; Moskal et al. 2017). With 15-meter radius circular plots that were equivalent to the planned ground-truth plot dimensions, I used ArcGIS to randomly sample 5900 non-water locations across the landscape. Random samples were clustered based on significant principal components calculated in R (R Core Team 2020) for nine lidar metrics: maximum return height, mean return height, standard deviation of

return height, 80th percentile of return height, canopy relief ratio, mean intensity, mode intensity, standard deviation of intensity, and percent cover. All these metrics were of all-returns over 3 meters elevation and were therefore slightly different from the first-return metrics that were later used in the riparian forest structure models. The characteristics of each cluster were used to define a series of unique structural groups within the 2016 lidar data. Each structural group was mapped across the landscape, and then proportional random sampling was used to randomly place a corresponding number of plots within each structure type (Cochran 1977).

Potential plots were then evaluated using GIS for logistical constraints such as land ownership and accessibility. Most sites were also visited in person prior to sampling because the majority of randomly selected sites turned out to be inaccessible or unsafe. When this occurred, a nearby location in the same forest structural group was substituted. In the end, 37% of the plots were randomly placed, and the remaining 63% of all plot locations were chosen from locations near to the original randomized point ahead of time in the lab or (rarely) on the fly in the field.

All field data were collected between July 1 and September 12, 2019. Plots were circular with a 15 m radius (area = 707 m^2 or 0.07 ha). This plot size was chosen to balance minimizing edge effects with logistical practicality, based on work by Frazer et al. (2011), who found that plots $\geq$ 707 m^2 produced substantially more accurate models than smaller size plots. I recorded height (measured with a Nikon Forestry Pro II hypsometer), diameter (measured by hand with a DBH tape), coordinates relative to plot center (measured using a Leica TC600 Total Station and Trimble GeoXH Geoexplorer 6000) and various structural characteristics for all stems $\geq$ 10 cm in diameter and taller than 3 m (Table 3). I also recorded a brief description of the plot vegetation, gave a subjective assessment of the relative canopy closure, and noted terrain features such as steep slopes, marshy areas, or small streams running through the plot. Using the

coordinates obtained using the total station, I sketched a rough tree-stem map for each plot. This helped me match height measurements to the correct trees and sped up data collection by making it easier to determine which trees were within the plot area.

Table 3: Ground truth attributes measured in the Nooksack River riparian zone, summer 2019. In the context of this study, a stem was defined as a discrete trunk ≥ 10 cm in diameter at breast height, 1.37 m. A single tree could have multiple stems if it forked below breast height.

Attribute	**Description**
Plot location	GPS coordinates of plot center
Stem ID	Unique ID for the stem
Tree ID	Unique ID for the tree
Height	Stem height
DBH	Stem diameter measured at breast height (1.37 m)
Species	Tree species
Dominance Y/N	Whether or not the stem canopy would be visible from above
Snag Y/N	Whether the stem was alive or dead
Broken Y/N	Whether the top of the stem was broken off
Forked Y/N	Whether the stem was forked above breast height
Leaning Y/N	Whether the stem canopy was significantly offset from the base
Stem location	XY location of the stem relative to plot center

Because not all plot locations were truly random, I used k-means cluster analyses in R (R Core Team 2020) to confirm that randomly selected plots were not different from non-randomly selected plots (Appendix A). I also checked for spatial autocorrelation because it can, if present, affect which analyses are appropriate to use on a dataset. I used R to create directional variograms and calculate Moran's I to assess potential spatial autocorrelation between plot-level variables (Pebesma 2004; Gräler et al. 2016; R Core Team 2020).

2.1.4. Ground Truth Data Pre-Processing – Interpolating Missing Height Data

Tree height was challenging to measure due to the prevalence of thick underbrush, unstable ground, and closed canopies in the study area. As a result, most plots contained several trees where either it was impractical to measure the height, or I could not definitively match canopy heights to the correct stems. When this occurred, I would note the heights of trees in the immediate area (inside or sometimes outside of the plot) that appeared representative for that species, dominance, and size class. Later, this supplemental height information was used to estimate values to fill in missing data prior to calculating overall plot metrics.

I created a series of height-diameter curves from the existing data by regressing tree diameter against tree height (Appendix B) then using these species-specific relationships to predict heights from stem diameter. When supplemental height data were available for a given plot, they were used to refine the estimates given by the height-diameter curves. A value would be randomly selected from within the range of heights typical for that diameter and type of tree, then averaged with the predicted value given by the height/diameter curve. This weighted the final prediction towards site-specific conditions. This method was selected for its efficiency because each missing value had to be individually corrected. The resulting height data fell under the following four categories: 1) height was directly measured in the field, 2) height was estimated using height-diameter curves and was refined by field observations of typical tree heights within the plot, 3) height was estimated using height-diameter curves and was refined by field observations of tree heights from outside of the plot, and 4) height was estimated using height-diameter curves only. The processing method used for each height observation was documented and used to calculate the mode height method for each plot. To see if the estimated heights had introduced a detectable bias, I plotted the mean ground truth height data against

various lidar height metrics and looked for plot-level patterns or trends corresponding to the frequency and method of estimated height data in the plot.

2.1.5. Modeling Riparian Forest Structure with Area-Based Methods

Lidar values from the 30-meter resolution raster were regressed against ground truth plot averages to create models of height, diameter, stand density, and stand composition. Although ground truth composition data were available as a percentage, these data had a strongly binary distribution so were modeled using logistic regression as "mostly coniferous" or "mostly deciduous". Statistical analysis and model development were conducted in R statistical analysis and graphics software version 3.6.3 (R Core Team 2020), and model skill was determined with leave-one-out cross-validation. Choice of the best model was based on Akaike information criterion (AIC) and cross-validated root mean squared error (RMSE).

To give a rough estimate of the amount of time it would take for restoration efforts to take effect, I estimated stand growth rates at each of the 104 field plots by comparing normalized canopy surface models from previous lidar acquisitions. Full lidar coverage of the study area was not available prior to 2016 (DNR 2018), but there are datasets that cover portions of the river from 2013 (WSI 2013), 2009 (WSI 2009), and 2005 (Terrapoint 2005a; Terrapoint 2005b). These datasets have varying resolution and accuracy, so they were only analyzed at the 30-meter scale. All 104 plots had data from both 2016 and 2013, but only seven plots had data from all four years. There were 299 total observations from the 104 plots across the four acquisitions. To determine how growth rates varied by height class, the data was sorted into three size classes and the average size-class-specific rate of change from each observation to the next was calculated.

2.1.6. Riparian Condition Index

I calculated a riparian condition index from the percent canopy cover, coniferous vs deciduous classification, and modeled average diameter at breast height. This raw riparian condition index was calculated at the 30-meter scale (Figure 2). For the inputs later used for restoration planning, the river was divided into 0.5 km long reaches and overall riparian condition was assessed at that scale. I determined overall riparian condition classification for each reach using the full width of the study area but excluding pixels classified as water. At the reach scale, reaches were sorted into four categories based on the relative proportion of "good" vs "bad" pixels (Table 4).

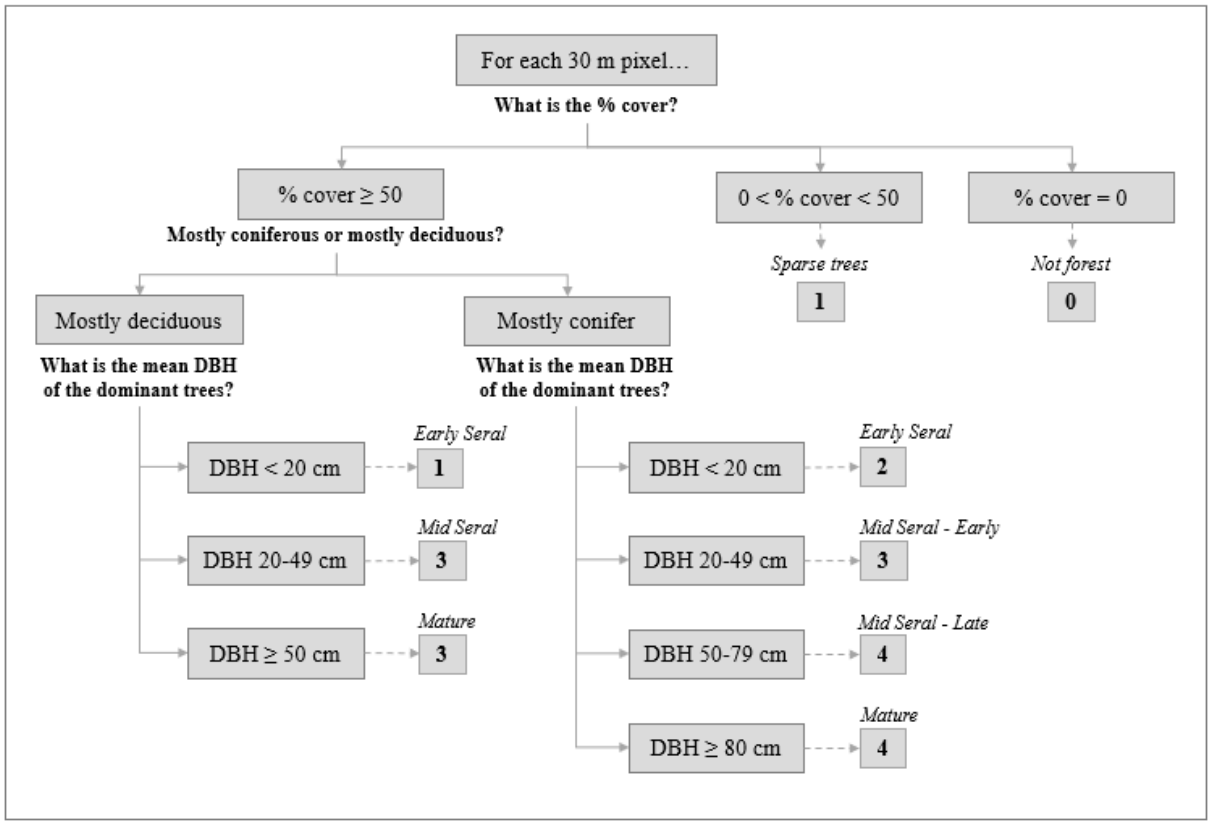

Figure 2: Decision process for assigning riparian condition at the 30-meter scale. Riparian Condition Index categories at the 30-meter scale were: 0 = Very poor, 1 = Poor, 2 = Fair, 3 = Good, 4 = Excellent. Cutoffs were derived from Oregon Land Cover Standard 2006; Beechie & Sibley 1997; Lunetta et al. 1997.

Table 4: Definition of riparian condition index categories at the reach scale (0.5 km).

Reach-Scale Category	**Definition:** % of non-water area that was classified as "good" or "excellent" at the 30-meter scale	**Dominated by:**
1 – Bad	0 – 29%	*Sparse forest, unforested areas, early seral forest*
2 – Poor	30 – 49%	
3 – Good	50 – 69%	*Mid seral and mature forest*
4 – Excellent	70 – 100%	

2.1.7. Shade Model

Within the reaches that had a reach-scale classification of "poor" or "bad", I modeled shade under a what-if scenario that projected what would happen if the quantity of riparian forest along the river banks was increased. I selected areas that were either sparsely vegetated or very short forests (up to 12m tall) and modeled them as if they were 25 meters tall and a moderately dense stand of mixed coniferous and deciduous composition. I selected sparsely vegetated areas by looking at the lidar data heights, supplemented by standard satellite imagery from Esri (Esri 2020) which helped me confirm vegetation was present in cases where the lidar data was unclear (usually due to large woody debris). I modeled the hypothetical stands by tracing the footprint in ArcGIS and then increasing the existing canopy surface model to 25 m tall. The result was hypothetical stands that were modeled as simplified blocks of the correct height and footprint. Because the amount of light that can penetrate a forest canopy depends on the leaf area index, it was necessary to estimate a reasonable leaf area index for the hypothetical new stands. Using previously developed regionally-specific leaf area index models, I used the proportion of first and ground lidar returns to calculate leaf area index for my entire study area (Richardson et al. 2009). Based on the range of leaf area index values in different stand types, I selected a leaf area

index value for the hypothetical stands that seemed consistent with existing conditions for similar stands in the region.

I used the Area Solar Radiation tool from Esri's ArcGIS Spatial Analyst toolbox (Fu and Rich 2002) to estimate how much the hypothetical stands might reduce solar input on the water's surface. The Area Solar Radiation tool calculates solar insolation from an upward facing hemispherical viewshed created from the input surface layer, which is usually a digital terrain model. It accounts for time of year, estimated sky conditions, and the difference between direct and diffuse radiation, but it assumes the given surface is solid and is therefore most appropriate for modeling shade effects from topography or buildings (Fu and Rich 2002). I ran the solar insolation tool twice: once on the original (current conditions) canopy surface layer and then again on the modified one with the hypothetical 25 m tall stands. I ran the tool for the summer and early fall months, June through September, and based the sky parameters on the typical summer climate for Bellingham, WA under smoke-free conditions (NOAA 2020) (Table 5). Input canopy surface layers had a resolution of 1 meter. I divided the output by the total number of days to get an overall daily average for the summer in units of kilowatt-hours per square meter.

Table 5: Parameters used in shade model calculations. Transmittivity and diffuse proportion parameters were based on typical summer weather conditions for Bellingham, WA (NOAA 2020). Shade model was run using Esri's Area Solar Radiation tool (Fu and Rich 2002).

Parameter name	**Value**
Latitude	48.8
Sky size/Resolution	200
Date range	June 1 – September 30
Day interval	15
Hour interval	0.5
Calculation directions	40
Zenith directions	16
Azimuth directions	16
Diffuse model type	Uniform overcast sky
Diffuse proportion	0.2
Transmittivity	0.7

I calculated the difference in solar energy between the two shade model outputs. Any areas with values significantly above zero in the output layer represented zones where the hypothetical stands had reduced solar input. Using methods described in the literature, I calculated a light extinction coefficient based on Beer's Law and the estimated leaf area index for the hypothetical stands (Richardson et al. 2009). I used the light extinction coefficient to adjust the solar energy estimates to account for the transparency of the hypothetical stands (Loicq et al. 2018). I did not account for topographic shading, as all reaches under consideration were relatively flat. I determined the percentage of the total inundated area per reach that would experience increased shade under hypothetical stand conditions, and I also determined the proportion of pixels that would experience a reduction in solar input of > 1 kWh/m^2 per day. This was an arbitrary threshold that was based on the range and distribution of the data.

In addition to this analysis, I also calculated the solar energy input on the water's surface for the entire study area under current conditions. This was a deliverable that was requested by

the Nooksack Tribe, but it was not used as an input in the prioritization model. Details on how this layer was created are in Appendix C.

Figure 3: Annotated photograph illustrating areas considered suitable and unsuitable for hypothetical riparian plantings or silvicultural intervention as modeled in shade model.

2.1.8. Restoration prioritization matrix

To prioritize riparian enhancement projects at the reach scale, I developed a ruleset to assign priority scores and help indicate the relative usefulness of riparian plantings for each reach in the study area (Figure 4). I defined the edges of the reaches by dividing the study area into segments of about 0.5 km. The actual channel length varied slightly depending on the sinuosity of the river at that point. Reaches were prioritized based on the reach-level riparian condition index, which my rule set further modifies based on extent of human impacts (determined from freely available

recent satellite imagery (Esri 2020)), impaired waterbody listings (the South Fork is an impaired waterbody (Kennedy et al. 2020)), likely magnitude of possible shade increase, position of the reach in the watershed, and proximity to existing high-quality habitat (Tables 6-7). Shade potential was a major focus of this study, and I weighted it most heavily among the contributing factors. Shade effects were loosely classified as small, moderate, or strong based on the relative proportion of the water surface area that would experience increased shade if hypothetical stands were added. Strong shade effects meant that > 10 % of the stream area was shaded with a reduction of 1 kW/m^2 per day or more (taking into account the leaf area index of projected stands), moderate shade effects meant that 5 – 10% of the stream area was shaded to that degree, and small shade effects meant that < 5% of the stream area was shaded to that degree. When assigning final priority scores for the shade component, I took into account the summer wetted channel width of the channels or side channels that had received increased shade. Position of the reach in the watershed and whether or not the reach was identified as temperature limited (in this case, whether or not the reach was in the South Fork) were weighted the least heavily of the input factors.

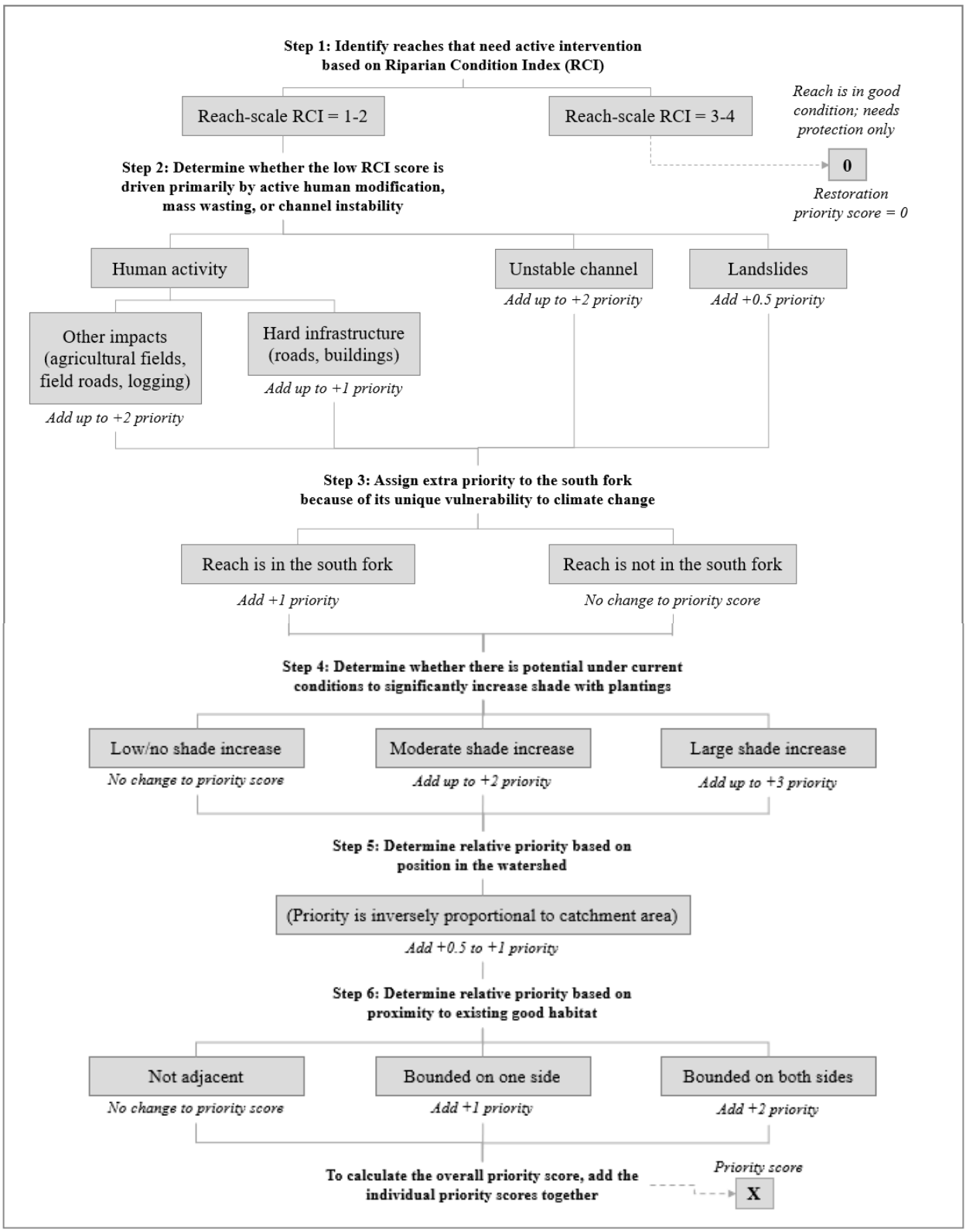

Figure 4: Conceptual flow used to prioritize reaches for the restoration prioritization matrix. See subsequent tables for rules governing each step.

Table 6: Rules for assigning priority scores for steps 1-3 from the restoration prioritization matrix (figure 4). At each stage the reach was assigned an action/priority score based on what condition it falls under. South fork priority was assigned based on findings by Kennedy et al. (2020).

Decision step	Condition of the reach	Action
Step 1: Reach level Riparian Condition Index (RCI)	< 50% of the non-water area was classified as "good" or "excellent" in the riparian condition index. RCI = 1 or 2	Active restoration is needed. Proceed to step 2.
	≥ 50% of the non-water area was classified as "good" or "excellent" in the riparian condition index. RCI = 3 or 4	Protection is more important than restoration. Active restoration priority score = 0
Step 2: Causes of low RCI	The primary cause for the low RCI ranking is a high proportion of bare alluvium, and there are multiple channels or braiding.	+ 2 to priority score
	The primary cause for the low RCI is a high proportion of bare alluvium, but there is only one channel.	+ 1.5 to priority score
	Poor RCI is due to wide areas of a mixture of alluvium and shrubby forest. The river is braided or multi-channel.	+ 1.5 to priority score
	Poor RCI is due to wide areas of alluvium and shrubby forest. The river has a single channel.	+ 1 to priority score
	Low RCI is primarily due to agricultural fields, logging activities, or other artificially cleared areas. These impacts are present within 50 meters of the edge of the water.	+ 2 to priority score
	Low RCI is primarily due to agricultural fields, logging activities, or other artificially cleared areas, and these impacts are present within 100 meters of the edge of the water.	+ 1.5 to priority score
	Semi-permanent human infrastructure (paved roads, railroads, buildings) are present within 50 meters of the edge of the water.	+ 1 to priority score
	Semi-permanent human infrastructure (paved roads, railroads, buildings) are present within 100 meters of the edge of the water.	+ 0.5 to priority score
	The primary reason for the low RCI ranking is recent landslides	+ 0.5 to priority score
Step 3: Assign priority bonus to temperature-vulnerable areas.	The reach is part of the South Fork	+ 1 to priority score
	The reach is not part of the South Fork	+ 0 to priority score

Table 7: Rules for assigning priority scores for steps 4-9 from the restoration prioritization matrix (figure 4). At each stage the reach was assigned an action/priority score based on its condition. Step 4 channel width cutoffs were the same used by Loicq et al. (2018) and were based on work by Teti (2006) and DeWalle (2008). Step 5 prioritization was based on Moore et al. (2005).

Decision step	Condition of the reach	Action
Step 4: Shade	Shade increases from modeled plantings are moderate and primarily affect channels > 30 meters wide.	+ 1 to priority score
	Shade increases from modeled plantings are substantial and primarily affect channels > 30 meters wide.	+ 1.5 to priority score
	Shade increases from modeled plantings are substantial and primarily affect channels 15 – 30 meters wide.	+ 2.5 to priority score
	Shade increases from modeled plantings are moderate and primarily affect channels 15 – 30 meters wide.	+ 2 to priority score
	Shade increases from modeled plantings are moderate or substantial and affect channels < 15 meters wide.	+ 3 to priority score
	Shade increases from modeled plantings are small and primarily concentrated on channels < 30 meters wide.	+ 0.5 to priority score
	Shade increases from modeled plantings are small and primarily concentrated on channels > 30 meters wide, or modeled plantings do not significantly increase shade.	+ 0 to priority score
Step 5: Position in watershed	Reach is located in the river headwaters. Catchment size is < 540 sq. kilometers	+ 1 to priority score
	Catchment size is 540 – 894 sq. kilometers	+ 0.75 to priority score
	Catchment size is 894 – 1248 sq. kilometers	+ 0.5 to priority score
	Reach location is among the lowest in the study area. Catchment size is > 1248 sq. kilometers	+ 0.25 to priority score
Step 6: Proximity to good habitat	The reach is bordered on both sides by reaches with a “good” or “excellent” RCI classification	+ 2 to priority score
	The reach is bordered on one side by a reach with a “good” or “excellent” RCI classification	+ 1 to priority score
	The reach is not bordered by reaches with a “good” or “excellent” RCI classification	+ 0 to priority score

2.2. RESULTS

2.2.1. PCA Results and Ground-Truth Data Pre-Processing

Based on the PCA analysis results, I determined that there were four forest structural groups that needed to be sampled in the ground truth data collection. An elbow plot indicated that six bins were optimum to describe all landcover types in the study area, but of those six, only four (Table 8) represented forest landcover categories. The other categories accounted for grass or low vegetation and exposed soil or gravel. The four forest structural groups were distinct enough to be readily recognizable in the field and seemed to accurately represent the range of forest types present in the study area. I found no evidence of spatial autocorrelation.

There were no detectable patterns related to whether tree height had been measured directly in the field or estimated in the lab. Plots with a high proportion of interpolated tree heights were no more likely to be outliers than plots where the majority of tree heights had been directly measured (Figure 5). Because of this, I concluded that all height data could be treated equivalently in later analyses.

Table 8: Unique structural groups defined by characteristics of clustered principal components and used for determining the number of field plots. Values given have ± standard deviation. Intensity values represent the relative amplitude of the return signal, scaled to 16-bit.

Group	Lidar characteristics	Field-observed characteristics
A	80^{th} percentile lidar height = 7 ± 7 m Mean return intensity = 20,806 ± 5768	Best thought of as a forest clearing, with 50/50 trees vs. low vegetation such as grass, trailing blackberry, or (in tall stands) knotweed. Natural sites included blow-down clearings and wetlands. Anthropogenic sites generally represented cleared areas where a few trees had been left.
B	80^{th} percentile lidar height = 13 ± 6 m Mean return intensity = 9937 ± 3283	Forest structure in this category was extremely variable, but in general seemed to represent early successional stages. Undergrowth was thick and tall, reaching to the base of the canopy.
C	80^{th} percentile lidar height = 29 ± 5 m Canopy relief ratio = 0.28 ± 0.12	An intermediate category. Moderate underbrush usually dominated by vine maple. Moderately large trees. Often, most of the stand was mature deciduous.
D	80^{th} percentile lidar height = 32 ± 8 m Canopy relief ratio = 0.47 ± 0.14	Large, widely spaced trees; underbrush generally short or absent; mostly coniferous.

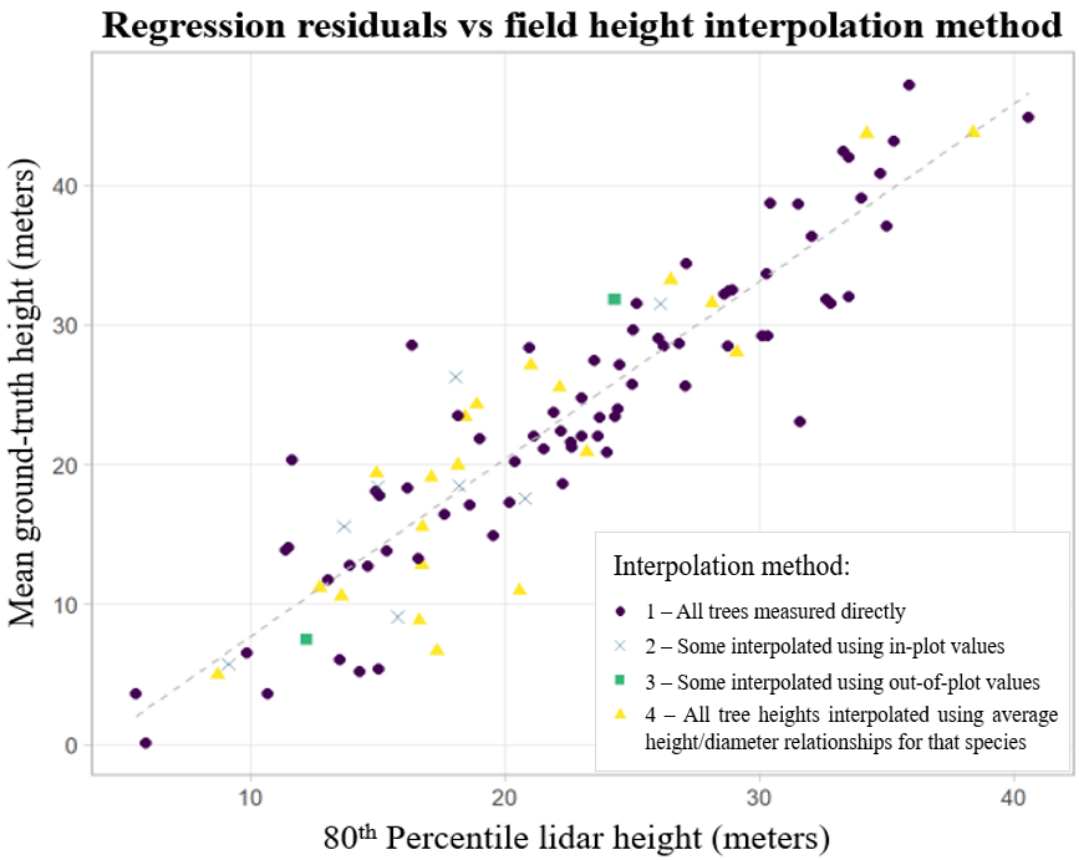

Figure 5: Mean height of dominant trees relative to 80th percentile lidar return height. Number of observations: 104 circular plots with 30-meter diameters. Points are symbolized according to the majority method used to generate the ground truth height data and demonstrate that interpolating height data did not affect model outcomes. The method used to generate the ground truth height data fell under the following four categories: 1) height was directly measured in the field, 2) height was estimated using height-diameter curves and was refined by field observations of typical tree heights within the plot, 3) height was estimated using height-diameter curves and was refined by field observations of tree heights from outside of the plot, and 4) height was estimated using height-diameter curves only.

2.2.2. Riparian Forest Structure Results

Mean tree height was the cleanest metric to model from the lidar data. The model for predicting the mean height of dominant trees had the highest R^2 value, but the model for predicting the mean height of all trees was slightly more accurate on average (Table 9). In general, models were most effective at explaining overall variance when predicting attributes of

dominant trees. The classification of plots into conifer-dominated or deciduous-dominated had an overall accuracy of over 80%, and the user's accuracy for both classes was consistent with the overall accuracy (Table 10).

Table 9: Comparison of area-based models of forest structure. Number of observations: 104 plots. Plot area = 707 m^2. RMSE was cross-validated using leave-one-out cross-validation. All slopes were significant to at least $p < 0.01$. Best models were chosen based on RMSE and AIC. Pseudo-R^2 for GLM models was calculated as: 1 – (Residual deviance / Null deviance).

AREA-BASED ANALYSIS

	RMSE ± sd	*R^2*	*Model*	*Lidar Parameters*
Mean height *(dominant)*	2.50 ± 1.98 m	0.83	1st order linear model	80th percentile height
Mean height *(all trees)*	1.98 ± 1.79 m	0.69	1st order linear model	80th percentile height, Mean intensity
Maximum height	3.53 ± 2.68 m	0.81	1st order linear model	95th percentile height
Mean DBH *(dominant)*	7.90 ± 7.65 cm	0.65	1st order linear model	80th percentile height
Mean DBH *(all trees)*	6.53 ± 6.98 cm	0.46	1st order linear model	80th percentile height
Basal area (m^2/ha) *(all trees)*	0.89 ± 0.83 (transformed units)	0.74	1st order linear model (Square root transformed)	95th percentile height
Density (stems/plot) *(dominant)*	7.77 ± 7.2 stems	0.30	Generalized linear model Quasi-Poisson distribution	Variance in height, Mean intensity
Density (stems/plot) *(all trees)*	12.75 ± 10.1 stems	0.12	Generalized linear model Quasi-Poisson distribution	Variance in height, Percent cover

Table 10: Confusion matrix for area-based composition model for dominant trees. Classification (predominantly conifer or predominantly deciduous) was modeled using logistic regression on mean intensity and 95th percentile of lidar return height.

	Field control		
	Conifer	Deciduous	Row total
Classification			
Conifer	**22**	5	27
Deciduous	14	**63**	77
Column total	36	68	104
Producer's accuracy		**User's accuracy**	
Conifer = 22/36 = 61%		Conifer = 22/27 = 81%	
Deciduous = 63/68 = 93%		Deciduous = 63/77 = 82%	
Overall accuracy = (22+63)/104 = 82%			

My analysis of past lidar data indicated that growth rates varied by size class. Stands that were < 13 meters tall grew an average of 88 cm/year (± 63 cm standard deviation), stands 13 – 23 meters tall grew 76 cm/year ± 56 cm, and stands > 23 meters tall grew 42 cm/year ± 60 cm. In the ground truth data, the maximum mean tree height at the plot level was only 33.7 meters.

2.2.3. Riparian Condition Index and Shade Model

Stream shading was highly dependent on how closely the hypothetical new stand could be located to the water's edge. Most "shade" was on the order of 0.5 – 1 kWh/m^2 per day reduction in solar energy hitting the water's surface. More localized areas, generally not extending more than 6-12 meters from the base of the hypothetical new stand, had shade effects with an estimated magnitude of 1 – 2 kWh/m^2 per day reduction in solar energy (Figure 6, Figure 7). No areas experienced > 2 kWh/m^2 of per day reduction in solar energy. Values < 0.5 kWh/m^2 per day seemed to include noise, so I considered them unimportant. Predictably, north-facing

shadows (shielded from the south sun) were generally darkest. Steep topography contributed shade to the headwaters, but among the reaches identified as "poor" or "bad" condition by the riparian condition index, topography was generally not a major contributing factor: most existing and potential shade was provided by trees. Details of shade results per reach are presented in Appendix D.

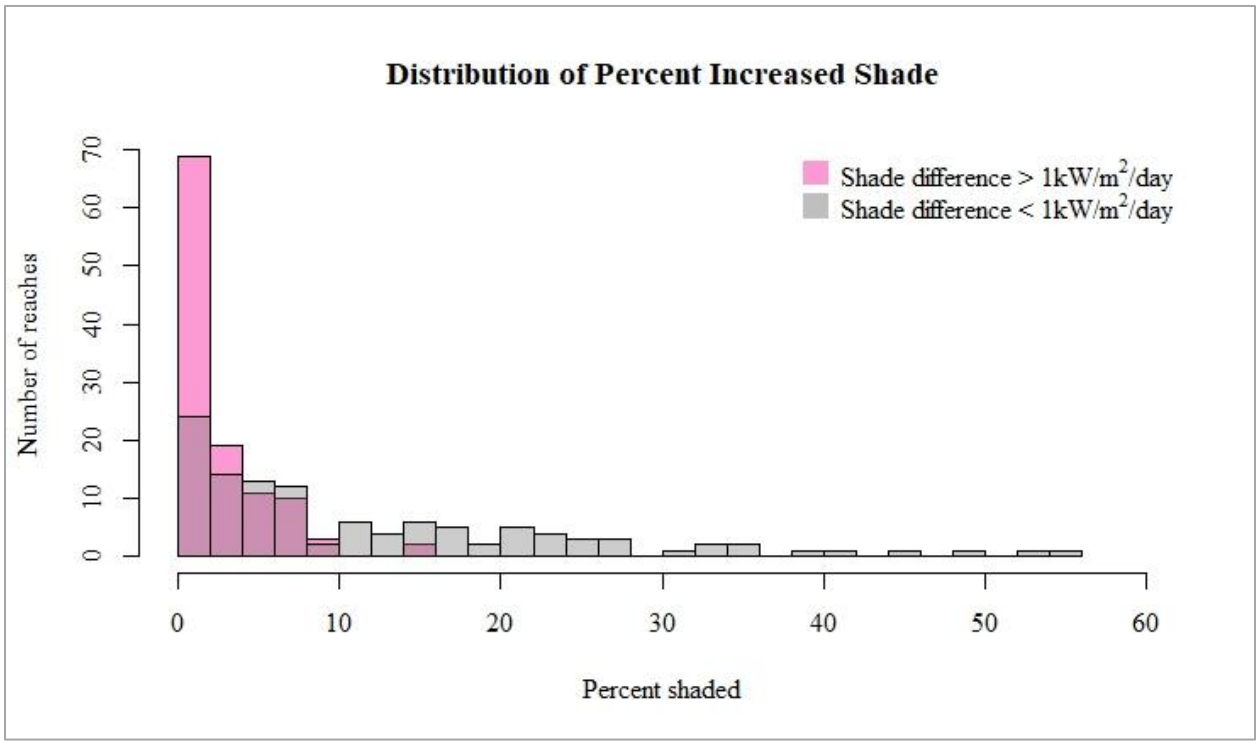

Figure 6: Histogram of reaches with different percent shade increases under hypothetical stand conditions. The number of reaches that had > 5% of the stream area shaded with ≥ 1kW/m^2/day less sunlight was comparatively small, and therefore more easily targeted through silvicultural intervention.

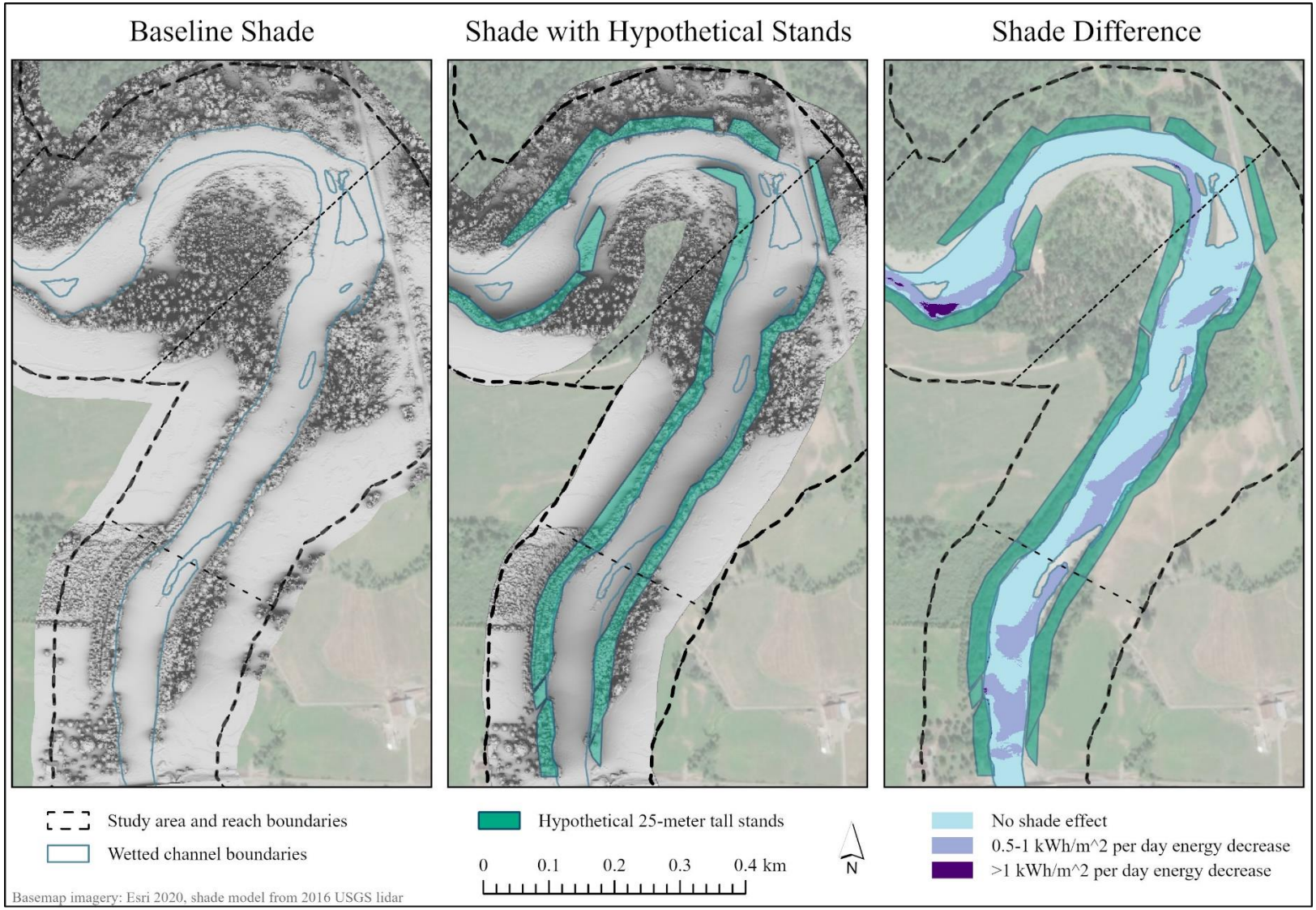

Figure 7: Example of shade model results from the lower South Fork

The outputs of the riparian condition index confirmed previously reported patterns, but in greater detail. For all three forks, the upper reaches were dominated by conifers and had more mature trees. The lower reaches and main stem were primarily dominated by deciduous trees, and braiding and large stretches of gravel and very young forest were much more common (Figure 8). Landslides and logging were the main contributing factors to areas classified as non-forest in the upper South Fork and upper Middle Fork; non-forest areas in the upper North Fork were mostly gravel bars. (Logged areas were present in the study area because the width of the study area was wider than legally set buffers.) Agriculture and gravel bars were the main contributing factors to areas classified as non-forest in the lower South Fork and mainstem reaches. Out of 265 total reaches, 114 were classified as having poor or bad riparian condition at the reach scale. Most poor-quality reaches were located in the lower parts of the watershed, but a few were located closer to the headwaters (Figure 9)

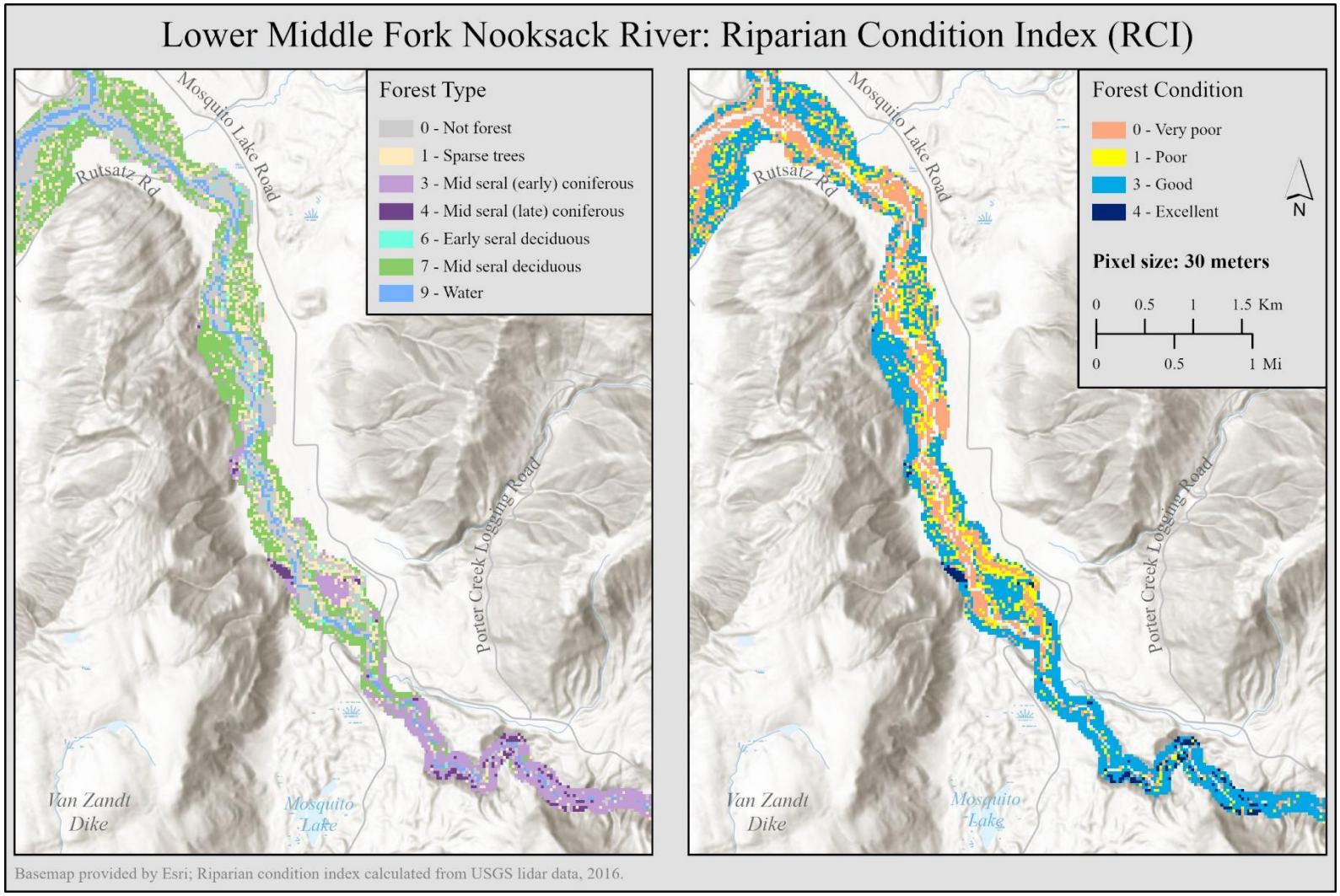

Figure 8: Riparian condition index output for a section of the lower Middle Fork. Pixels are 30 m

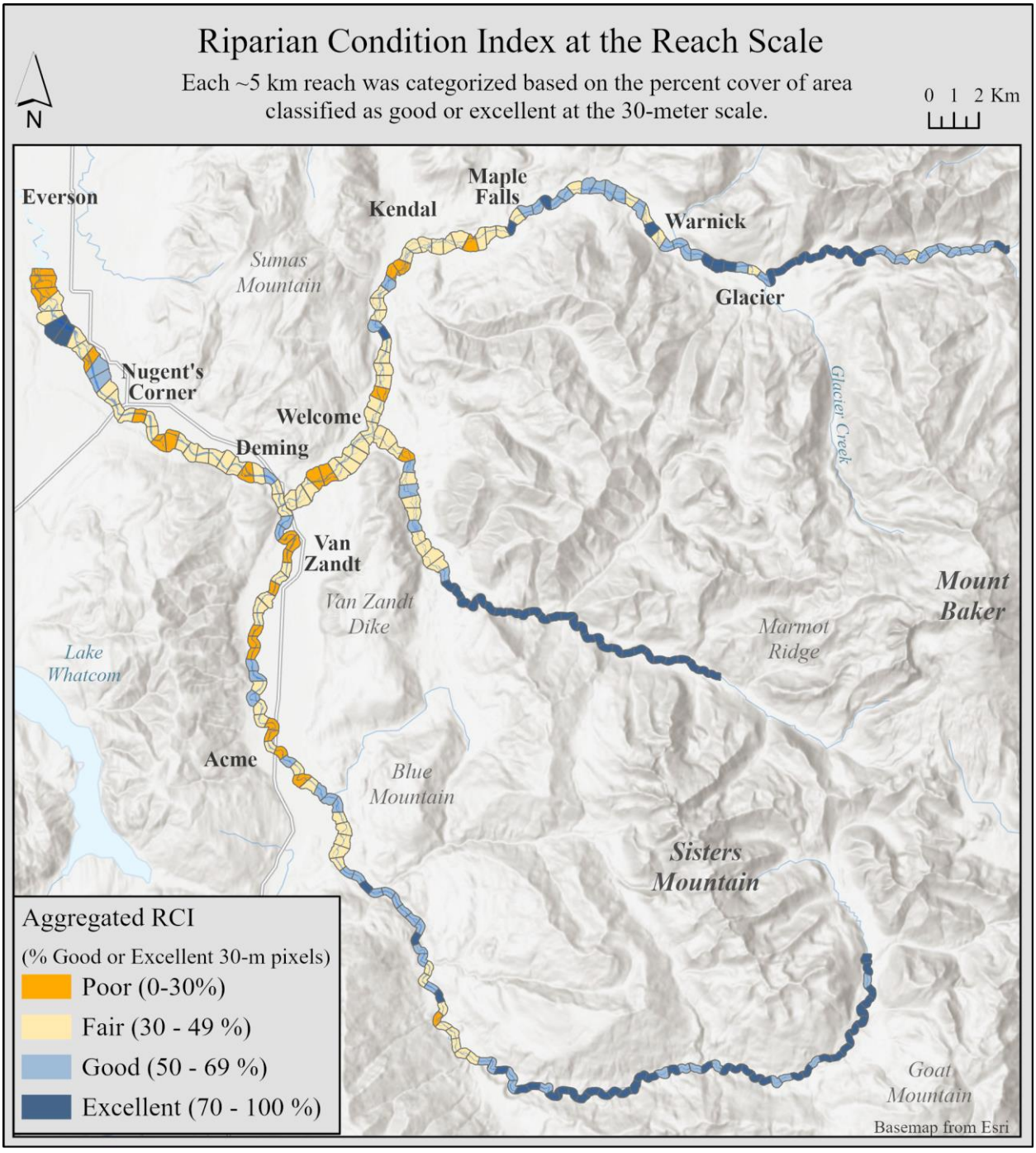

Figure 9: Riparian condition index at the reach scale

2.2.4. Restoration Priority Matrix

I gave the 154 reaches with good or excellent riparian condition very low priority scores for active restoration (priority score = 0) but considered them high priority for protection. The priority scores assigned to the 114 poor or bad reaches ranged from 1.25 to 6.5 (out of 11) with a median of 3.88 and mode of 4 (full results in Appendix D). There were nine reaches with a score of 6 or higher on the priority scale, making them the highest priority for intervention (Table 11). One of these was in the North Fork, two were in the lower Middle Fork, four were in the upper South Fork, and two were in the lower South Fork (Figure 10).

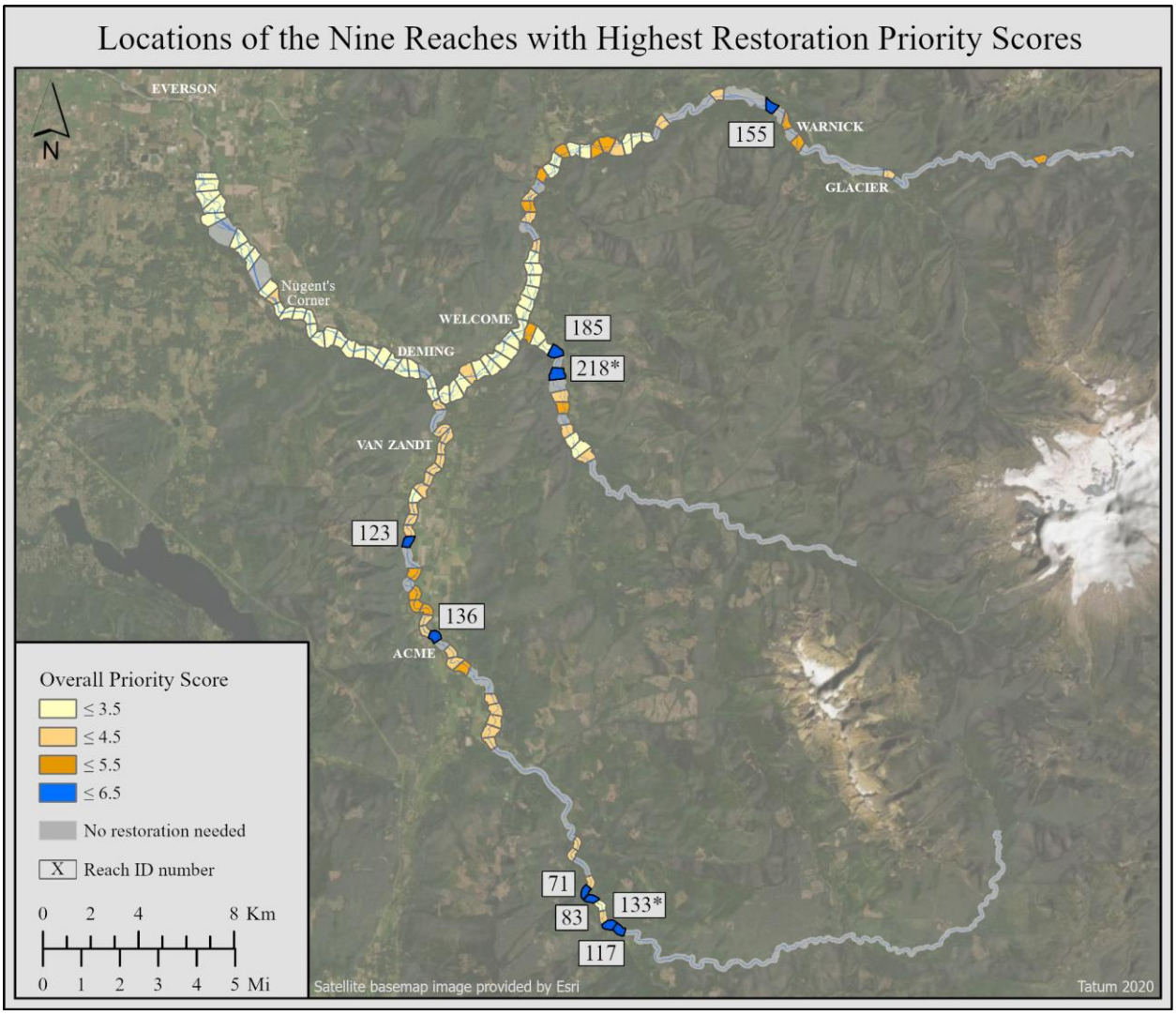

Figure 10: Map showing the locations of prioritized reaches. Grey boxed numbers show the reach ID numbers for the nine highest-priority reaches, asterisks indicate a priority score of 6.5.

The highest-priority reach in the North Fork (reach 155) was located at RM 53, between Maple Falls and Warnick. Its poor riparian condition score was primarily driven by a high proportion of gravel bars and low vegetation due to an unstable channel, but it also had Hwy 542 running through the northern half of the study area, and some of the land upslope of the highway had been logged (Tables 11-12). This reach contained a relatively high proportion of "good"

riparian forest (49%) and was adjacent to existing good riparian condition on both upstream and downstream sides.

The two top priority reaches on the Middle Fork (reach 185 and reach 218) were located roughly 2-3 kilometers upstream of the junction with the North Fork. Their low riparian condition score was primarily due to gravel bars and low vegetation, but they also had significant cleared areas associated with rural residential activity, primarily along the west bank. Although these cleared areas were included in the area to be hypothetically re-planted, in general the hypothetical new stands that resulted in the greatest increase in shade were either located on the other side of the river or on large existing islands and would not interfere with residential landscaping.

The four target reaches on the upper South Fork were all located in the same area, near RM 20. River miles are measured from the confluence with the main fork. The uppermost target reaches on the South Fork (reach 117 and reach 133) were adjacent to each other and comprised the two farthest-upstream poor-quality reaches on the South Fork. These reaches were downstream of Larson's Bridge and encompassed the junction with Plumbago Creek (RM 20). Several engineered logjams have recently been built here. The lower two reaches (reach 83 and reach 71) were located roughly a kilometer downstream. The poor riparian condition scores for all these reaches were due to gravel bars, both vegetated and unvegetated.

The two lowermost target reaches on the South Fork were located some distance from each other and under very different conditions from the uppermost four. The upstream one (reach 136) comprised the half-kilometer upstream of the Valley Highway bridge in Acme near RM 9. The reach immediately adjacent on the upstream side was classified as good quality habitat, but

most of the study area along the target reach was unforested and apparently used for agriculture. Despite the relatively large width of the channel (roughly 30 m wide), shade potential was found to be favorable because the river made a sharp bend to run east-west, allowing for hypothetical plantings on the south bank to cast long, deep shadows over half the length of the reach. The lowest South Fork target reach (reach 123) was located near Standard, south of the end of Hillside Rd (Near RM 4). Most of the study area along the reach was agricultural fields, and there was also a gravel bar with some low vegetation. Plantings close to the river on the east/south-east bank and gravel bar had the potential to create moderate levels of shade, although the river was over 30 meters wide at this point.

Table 11: Priority score breakdown for the nine highest-priority reaches in the watershed.

Reach ID	Overall score	Catchment area	Primary cause of low RCI	Modeled shade effects	Sides bordering good RCI	Reach location
71	6	210 km^2 priority score = 1	Bare alluvium priority score = 2	Moderate effects, channel 15-30 m priority score = 2	0 sides score = 0	South Fork score = 1
83	6	209 km^2 priority score = 1	Bare alluvium priority score = 2	Moderate effects, channel 15-30 m priority score = 2	0 sides score = 0	South Fork score = 1
117	6	187 km^2 priority score = 1	Bare alluvium priority score = 2	Moderate effects, channel > 30 m priority score = 1	1 side score = 1	South Fork score = 1
123	6	421 km^2 priority score = 1	Agriculture priority score = 2	Moderate effects, channel > 30 m priority score = 1	1 side score = 1	South Fork score = 1
133	6.5	205 km^2 priority score = 1	Bare alluvium priority score = 2	Moderate effects, channel > 30 m priority score = 1	0 sides score = 0	South Fork score = 1
136	6	396 km^2 priority score = 1	Agriculture priority score = 2	Moderate effects, channel > 30 m priority score = 1	1 side score = 1	South Fork score = 1
155	6	515 km^2 priority score = 1	Alluvium & shrubs priority score = 1.5	Strong effects, channel > 30 m priority score = 1.5	Both sides score = 2	North Fork score = 0
185	6	229 km^2 priority score = 1	Bare alluvium priority score = 2	Moderate effects, channel 15-30 m priority score = 2	1 side score = 1	Middle Fork score = 0
218	6.5	229 km^2 priority score = 1	Alluvium & shrubs priority score = 1.5	Moderate effects, channel 15-30 m priority score = 2	Both sides score = 2	Middle Fork score = 0

Table 12: Detailed conditions of the nine highest-priority reaches in the watershed.

Reach ID	Channel morphology	Roads or railroads?	Distance from water to nearest impacts	Primary factor driving low RCI	Secondary factor driving low RCI	Land use / land owner (primary)
71	Multiple channels	None	NA	Bare alluvium; channel instability	None	Commercial forestry
83	Multiple channels	None	NA	Bare alluvium; channel instability	None	Commercial forestry
117	Single channel	Unpaved roads	< 50 m	Bare alluvium; channel instability	None	Commercial forestry
123	Multiple channels	Unpaved roads	< 50 m	Agricultural fields	None	Commercial agriculture
133	Braided channel	None	NA	Bare alluvium; channel instability	None	Commercial forestry
136	Multiple channels	Paved roads, <50m away	< 50 m	Agricultural fields	Buildings; paved roads	Whatcom County
155	Multiple channels	Paved roads, >50m away	< 100 m	Mixture of exposed gravel and shrubby trees; channel instability	Logged areas (upslope of rd.)	WA Dept. of Nat. Resources
185	Multiple channels	Paved roads, >50m away	< 100 m	Bare alluvium; channel instability	Artificially cleared areas surrounding houses	Private owners (residential)
218	Braided channel	None	< 50 m	Mixture of exposed gravel and shrubby trees; channel instability	Artificially cleared areas surrounding houses	Private owners (residential)

2.3. DISCUSSION

Lidar-based analyses of forest structure have great potential to support land managers in watershed-scale management decisions (White et al. 2016; Moskal et al. 2017; Tompalski et al. 2017). Lack of information about riparian forest structure has been a problem when planning restoration projects for salmon conservation in the Nooksack watershed (Puget Sound Partnership 2018), and my study aimed to address this data gap. The lidar-based riparian forest models I created facilitate the mapping of forest structure across wider areas than would be possible from field-based sampling alone, and they are appropriate to use for planning at scales of 30-meters or smaller (less detailed) scales.

2.3.1. Context from Previous Research

My findings support previous studies that found that there were not enough large coniferous trees to support important riparian functions, including large woody debris recruitment and shade (Coe 2001; Hyatt et al. 2004; Brown and Maudlin 2007; Hyatt 2007; Capuana 2013). I found that conifer-dominated stands were overwhelmingly located in the upper portions of each fork; lower reaches and the mainstem were dominated by small cottonwoods (*Populus balsamifera*) and alders (*Alnus rubra*). My results also support previous assertions that the upper Middle Fork and upper South Fork are generally in good condition (Brown and Maudlinn 2007; LNR 2011). As observed in Collins and Sheikh (2004b), most of the North Fork and the lower Middle Fork are probably excessively braided compared to historical conditions. In addition, there are portions of the far upper North Fork (upstream of Glacier) that may be susceptible to channel instability if conditions deteriorate (Anderson and Konrad 2019). These areas should be monitored because their naturally unconfined morphology means that they will

be one of the first places to show problems from glacial retreat, sediment loads, and changing hydrology that are expected to occur with global warming (Anderson and Konrad 2019; Dickerson-Lange and Mitchell 2014).

My ability to predict individual metrics such as diameter, basal area, and height was lower than that reported for studies based in upland areas or on individual trees (Hyyppä et al. 2008; White et al. 2016), but they were consistent with other area-based research in riparian areas in the Pacific Northwest (Moskal et al. 2017). My height-related R-squared values were roughly 3-20% lower than those reported by Moskal et al. (2017), depending on method, but my cross-validated root mean squared error was on average about 0.5 m more accurate. My estimates of basal area explained over 70% of the variance in the data, which was comparable to models developed by other researchers and used fewer predictor variables (Moskal et al. 2017; Strunk et al. 2012). The accuracy of my density estimates from the area-based approach were poor (Table 8) because an area-based approach is not best suited to this application. Density estimates are better supported by an individual-tree approach (Richardson and Moskal 2011) which is covered in chapter 3.

There are multiple factors that can influence lidar-based model accuracies. Accuracy can be influenced by the filtering method used to remove outliers from the lidar point cloud (Vauhkonen et al. 2012), but this was not an issue in this case. I individually examined the point cloud for all 104 ground truth plots, and outliers were rare or nonexistent after filtering. A more likely source of error was the three-year time disjoint between the date of the lidar acquisition and ground truth data collection. Based on my growth trajectory estimates, an average tree would have grown approximately 1.26 – 2.64 meters during this time, depending on its starting height. This accounts somewhat for the discrepancy between the ground truth data and the lidar data,

especially if potential species-specific variations in growth rates are taken into account. Growth rates are almost certainly influenced strongly by species, but I was unable to account for this factor in my estimates. Another possible contributing factor to lower accuracy levels is error in ground truth measurements. Ground truth height measurements may have been less accurate than would be the case for upland stands, especially even-aged plantation stands where most assessments of lidar vs field-measured accuracy have been done (Hyyppä et al 2008; Strunk et al. 2012; White et al. 2016). The dense canopy, steep slopes, and unusually treacherous footing present throughout much of the study area made measuring tree height unusually challenging, but the magnitude of error arising from these factors is unquantifiable.

The biggest contributor to my lower model accuracies may have been the size and structural heterogeneity of the study area. Models developed for a small area are generally likely to have higher accuracies than those that attempt to describe a large structurally and compositionally complex forest comparable to the current study area (Strunk et al. 2012). In general, previous lidar studies of forest structure have been focused on areas that were more structurally homogeneous than my study area because they focused on commercially managed stands. A riparian forest study by Moskal et al. (2017) is an exception, with 130 plots spread over 57,000 acres of watershed and 530 miles of streams. My study was most similar to this study in scope, and my model accuracies were correspondingly similar, although I used fewer input parameters in my models.

Previous research has generally emphasized channel morphological factors when modeling salmon habitat suitability and making recommendations for restoration priority, but I took a different route. Although some morphological characteristics (stream width and braiding) were indirectly included in my analysis, for the most part I avoided in-channel metrics in my

prioritization model, for two main reasons. First, in-channel metrics such as gradient are typically used to help predict salmon presence or absence for a given stream (Burnett et al. 2007; Mollet and Bilby 2008; Tompalski et al. 2017). This was irrelevant for my study because the limits of my study area were partially defined by the limits of recorded salmon presence: all reaches in my study area are known to support, or at least be passable, by salmon. Another reason for omitting in-channel metrics including gradient was that optimal habitat conditions usually are species-specific. In the Nooksack watershed, although 25 salmonid stocks are present, the stock of greatest concern is the South Fork early Chinook (WRIA 1 Salmon Recovery Board 2005). However, I wanted to keep my conclusions more general because the type and detail of my data was not appropriate for a rigorous species-specific fish study. Consequently, I built my prioritization model to make best use of the good-quality riparian condition data that I had, and I avoided including factors that were not supported by ground-truth data and would therefore have introduced a large extra measure of error and uncertainty.

My model complements, but does not replace, the recently released 2020 Salmon Recovery Funding Board grant restoration and protection strategy matrices put out by the WRIA 1 Watershed Management Board (SRFB 2020). These matrices are specific to early Chinook and provide reach-specific restoration action recommendations for a variety of restoration action strategies. In-stream habitat conditions and species-specific use and accessibility for a given reach are particularly important factors that should guide the decision-making process. The models I created focus on vegetation and shade potential in a more general sense and should be used together with this other information for a holistic view of local conditions. My models and GIS layer outputs are also intended to be able to be used as a jumping off point for new questions as restoration projects progress and the dynamic river system changes over time.

Suitability indices and prioritization models in general are coarse filters and need to be used at the scale for which they are designed (Burnett et al. 2007; Meixer and Bain 2010). My prioritization model was intended to function as a broad-scale strategic planning tool, and it was not intended to replace more detailed field studies which are important for determining what type of restoration action is needed in any given location. None of the results reported in this analysis should be considered at a more detailed scale than 30-meters, and most planning should happen at the reach scale.

There are several limitations and assumptions that should be considered when using my prioritization model and the shade model on which it depends. Shade was modeled as a "what if" scenario: what would be the outcome if nearby cleared areas and select gravel bars and islands were forested? How much of a difference would it make to river surface shade? As expected, the results were highly dependent on the site. I found that the amount of potential shade was dependent on the orientation and the location of the planting site relative to the river, and this was influenced by a variety of factors. In turn, the accuracy of shade predictions depends on how realistic the modeled plantings were.

Modeling potential planting sites required a series of judgement calls about what areas were stable enough to support potential plantings or other forest treatments. Note that, although my main focus was on planting new trees, there were several places where I outlined "plantings" on top of existing shrubby, sparse, or small trees in order to model the effects of thinning, interplanting, or other appropriate forest treatments. In real life, the actual edges of planted areas might be different than what was portrayed in my model due to movement of the river over time or inaccuracies in the lidar-based delineation of the river banks. Lidar is relatively effective at

outlining water features in open areas (Tompalski et al. 2017), but inaccuracies increase in areas where the vegetation overhands the bank.

It was also difficult to judge how transient (and therefore how worth shading) various side channels and branches were, since I only analyzed a snapshot of the river's location from 2016. In many reaches within my study area, channel position can move by many 10s of meters from one year to the next. In future analyses, it would be beneficial to look at multiple years of imagery or other time series data (such as those analyzed in Collins and Sheikh 2004b) to determine whether any given channel would be important and stable enough to be a focus for shade enhancement. Lidar in the near-infrared spectrum cannot assess water depth, so a different source of data (such as green-spectrum lidar) would be more appropriate for collecting this information, which is relevant to stream temperature modeling (Seixas et al. 2018). Assessing the relative permanence of individual side channels is particularly important because most of the main channel is too wide for shade to have a significant impact on water temperature (Seixas et al. 2018). Shading of smaller side channels is very important in these wide lower reaches.

I did not collect ground truth data on solar intensity, so I have no objective means of measuring how accurate the shade model was at predicting actual solar insolation at any given point. Such data would have been logistically challenging to collect due to the need to log insolation over at least one full day per location in order to compare ground-truth readings to outputs from the shade model. However, previous research in the Pacific Northwest has suggested that shade models derived from a lidar-based hemispherical viewpoint method (such as the one applied here) are relatively robust and do not need ground-truthing to produce usable outputs (Richardson et al. 2019). Including the light attenuation filter corrected shade levels to a more realistic level than has been reported by other shade modeling studies that modeled shade

under the simplifying all-or-nothing assumption that the tree canopy did not let through any light at all (Johnson and Wilby 2015; Tompalski et al. 2017; Sexais et al. 2018) and a similar approach to estimating light attenuation has been shown to be accurate both in the Pacific Northwest and in other regions (Bode et al. 2014; Loicq et al. 2019; Richardson et al. 2019).

Channel stability is the biggest factor influencing the applicability of my shade model projections. As the Salmon Recovery Funding Board matrices indicated, working to improve channel stability is a top priority for much of the watershed (SRFB 2020). Established vegetation can help stabilize a channel (Naiman and Latterell 2005), but before they become established, new plantings are at severe risk of being washed away in many locations, especially the lower North Fork, lower Middle Fork, and mainstem. In some locations I did not attempt to model shade because it appeared that the channel was much too unstable and, based on satellite imagery (Esri 2020), there were no potential planting sites that I thought had a chance of supporting new trees into maturity. For such locations, it may make more sense to situate any plantings some distance away from the channel (for long-term large woody debris recruitment and shade potential) and focus other restoration efforts on island-building measures such as artificial logjams. Plantings may also be more likely to be successful if they are planted in newly-sheltered areas such as behind artificial logjams.

2.3.2. Growth rates and time frame for restoration

The growth rates calculated from the multi-year lidar data provide some context for the necessary time frame for results from restoration actions, but they are subject to high variability. This variability is probably strongly tied to species-specific differences, but site specific differences almost certainly play a role. In addition to pronounced differences in stand density,

some sites were basically upland sites in their soil attributes, while others were wetlands with standing or flowing water even at the height of summer. Setting aside these caveats, overall growth rates suggest it would take approximately 33 years to grow an average 25-meter tall tree, the height of the trees modeled in the shade model. Growth rates could potentially be sped up by thinning and other active silvicultural treatments.

My study focuses on comparatively short-term effects (decades rather than centuries) because of concerns over near-future climate change stressors that will need to be addressed (Dickersen-Lange and Mitchell 2014; Kennedy et al. 2020). I also assumed that, in general, any planting that provides benefits in the short term will only continue to provide more and better benefits in the long term, although paying attention to species composition of restoration plantings will be important for determining the quality of long-term effects. To this end, quick-growing shade trees (spruce (*Picea sitchensis*) or cottonwood) should be interplanted with slower-growing shade-tolerant trees (such as western redcedar (*Thuja plicata*)) that have better rot resistance and provide long term benefits to large woody debris recruitment (Bilby and Ward 1989; Roni et al. 2002).

Because of uncertainties about river movements in future years, and because I was mostly focusing on relatively short-term rather than long-term effects, I did not emphasize large woody debris recruitment potential in this study. Large woody debris recruitment potential is challenging to model because it depends not only on size and type of trees, but also on the rate at which the channel shifts. If it shifts too fast, trees cannot grow to proper size before being eroded. If it shifts too slowly, recruitment is low because mature trees are not falling into the river. A truly rigorous large woody debris recruitment model would require a more extensive

time-series analysis than was carried out in this study and would be an interesting and useful topic for further research.

2.3.3. Pre-Processing and Ground-Truth Data Collection

Challenges measuring ground truth tree height are common (Slava et al. 2011), and many methods have been proposed to help correct for it (Vauhkonen et al. 2012; Mehtätalo et al. 2015) The method I used of estimating tree height from species-specific diameter curves is similar to methods described in Vauhkonen et al. (2012), but my approach had the additional refinement of using local out-of-plot trees to help refine in-plot predictions. This method, while not perfect, was comparatively computationally efficient, and does not appear to have significantly impacted height model accuracy.

The proportional random sampling strategy I used to collect ground truth data was not quite truly random, although no effects from this were detected. As described in section 2.1.2., safety and access concerns forced the repositioning of more than half of all field plots. Steep slopes were systemically under-sampled, as were modified forests in residential areas. Although almost all property owners that were approached granted me permission to sample on their land, I preferentially focused on large landowners rather than trying to get permission from many, many small residential properties. Riparian areas that were broken up into small parcels owned by private residents were not well-represented in my ground truth data. These stands are likely to be heavily modified and are more likely to contain exotic tree species, which might influence model accuracies. Even with these caveats, I am confident that my ground truth data was representative of the overarching condition of the watershed. The PCA approach to modeling forest structure from lidar metrics prior to sampling seemed very effective at identifying

predominant forest structures. By the end of my sampling period I was no longer encountering "new" forest structures in the field.

2.4. CONCLUSION

My study aimed to support salmon conservation efforts by mapping riparian conditions that had been identified as a key data gap in the Nooksack watershed (Puget Sound Partnership 2018). The results of my study confirmed the findings of previous localized smaller studies, but in greater detail and across a larger area. Model accuracies were comparable to similar studies on riparian forests in the Pacific Northwest, but the model accuracies were slightly lower than what is generally considered to be standard for lidar data. Both the size of my study area and the three-year lag between lidar data collection and field data collection may have contributed to this.

Using modeled forest stand attributes, potential for increased shade, and assessments of local human impacts based on free satellite data, I created a broad-scale planning tool to use for prioritizing restoration actions at the reach level. Modeled stand heights used to model potential for increased shade were 25 meters tall. Based on average stand growth trajectories in the recent past, such stands could be established on a timeframe of just over three decades. However, their effectiveness at providing riparian benefits would depend on their proximity to a wetted channel. Because the river is a very dynamic and changing place, and as future restoration projects gradually start to stabilize the channel, it will be important to iteratively re-visit watershed-level restoration prioritization and shade potential when planning new projects in the future.

3.0 – Individual Tree Species Analysis

3.1. OVERVIEW

The second of my research objectives was to explore methods of modeling tree species at the individual tree scale. Modeling accuracy for individual tree species is driven by two main factors: the accuracy of the segmentation model used to identify individual tree objects, and the accuracy of the species model itself, once lidar-derived canopies are matched with the correct ground truth data. Accurate individual-tree-level metrics have the potential to be useful for scientists and land managers alike, but such analyses have proved challenging in Pacific Northwest forests because of high structural heterogeneity (Jeronimo et al. 2018). Prior studies have noted that accurate tree segmentation seems to be the limiting step (White et al. 2016; Moskal et al. 2017). In response to this need, I investigated five different segmentation algorithms, comparing their accuracy both in terms of density and the spatial alignment with ground truth data. Building on the most promising of these segmentation approaches, I developed individual species models to classify several of the most common tree species in my study area.

3.2. METHODS

3.2.1. Study Area and Data Collection

See Chapter 2: sections 2.1.1. – 2.1.4. for methods related to study area, lidar data, and ground truth data collection and pre-processing.

3.2.2. Segmentation Methods

I compared six different approaches using five different algorithms for identifying individual tree canopies in the lidar data. Three of these approaches, (two versions of watershed segmentation and one version of multiresolution segmentation), were processed in Trimble's eCognition Developer 9.5 (Trimble 2020) on 1-meter resolution rasters which I generated using USFS Fusion's Area Processor (McGaughey 2018). The remaining three approaches were processed entirely in the R programming environment (R Core Team 2020) using the lidar processing package lidR (Roussel and Auty 2020). In the lidR package, these three methods are called the "Dalponte2016", "Silva2016", and "Li2012" algorithms, and they implement methods developed by Dalponte and Coomes (2016), Silva et al. (2016), and Li et al. (2012). For consistency with the lidR package, I will refer to these three methods by their lidR function names.

Multiresolution segmentation in eCognition is an iterative algorithm that segments a raster by grouping pixels together until each segment reaches a variance threshold. Separate parameters control the algorithm's sensitivity to spectral differences relative to the smoothness and compactness of the output segments. I used eCognition to run multiresolution segmentation on a 1-meter canopy surface raster representing the mean height of lidar first returns. After using a 3-meter height cutoff to remove non-forested areas from consideration, I processed all forested areas in my study area as a single unit under the same set of parameters. The scale, shape, and compactness parameters were set to 10, 0.1, and 0.5 respectively.

Watershed segmentation is a segmentation method that delineates image object boundaries from a canopy surface raster based on localized concavities and convexities in the data. The size and shape of output segments can be partially controlled by applying various smoothing filters to the input raster surface. I applied the eCognition watershed segmentation

algorithm to two different rasters: an unsmoothed 1-meter raster of the mean height of first returns, and a smoothed 1-meter raster that I created from the 80th percentile of lidar first return heights, smoothed over a 3x3 window.

The Dalponte2016 algorithm in the lidR R package is a region growing method developed by Dalponte and Coomes (2016). It requires a canopy surface model and a set of individual seed points representing tree tops. I generated the input canopy surface model by applying lidR's simple point to raster approach on the filtered, normalized point cloud (Roussel and Auty 2020). I filled null values and smoothed the 0.6-meter resolution surface across a 1-meter window using the R package "raster" (Hijams 2020). The parameters used in the Dalponte2016 segmentation were slightly modified from the defaults in order to make them more suitable for local conditions (Table 13).

I identified tree top locations in the lidar point cloud by using a local maximum filter to return the highest lidar return point in a given area (Roussel et al. 2020). The search window was defined using a multi-level function that allowed a different canopy radius for shorter trees than for large trees. I used a tree top detection function that assigned a circular search window diameter of 3 meters to trees < 2 meters tall, and a search window width of 5 meters to trees > 20 meters tall. Trees with heights between these cutoffs were assigned search window widths along a sliding scale. This balanced rates of accidental misclassification of tree branches as tree tops with accurate detection of smaller inter-canopy trees. Additionally, lidar returns located less than 5 meters off the ground were excluded from consideration as tree tops because visual examination indicated that too many of these belonged to tall shrubs rather than trees. To facilitate a visual check of the model performance, I plotted the tree top results on top of subsets of the filtered and normalized point cloud.

The Silva2016 algorithm in lidR is a nearest-neighbor segmentation algorithm developed by Silva et al. (2016). Like the Dalponte2016 algorithm, it uses a canopy surface model and tree top seed points. I used the same canopy surface and tree top points for both algorithms. This meant that both methods produced the same number of total segments per plot, though differences in segment size and shape could occur.

In contrast to the other two lidR algorithms, the Li2012 algorithm delineates trees directly from the point cloud without use of the pre-determined tree tops. The algorithm starts with the topmost lidar return point in the file, classifies all remaining points as either belonging to that tree or not based on a variable distance criteria, then removes the classified points and starts over again with the next highest unclassified point, continuing until all points in the dataset are classified (Li et al. 2012; Roussel and Auty 2020). After trial and error, I used parameters that were very close to the default parameters recommended by Li et al. (2012) (Table 13).

Table 13: Parameters used for Dalponte2016, Silva2016, and Li2012 algorithms in the lidR package.

lidR Parameter	Description
DALPONTE 2016	
th_tree = 3 m	Threshold below which a pixel cannot be a tree
th_seed = 0.45	Growing threshold 1 – default value. See Dalponte and Coomes 2016
th_cr = 0.55	Growing threshold 2 – default value. See Dalponte and Coomes 2016
max_cr = 15 m	Maximum allowable crown diameter
SILVA 2016	
max_cr_factor = 0.6	Maximum crown diameter, given as proportion of tree height
exclusion = 0.3	Minimum height threshold: Pixels below the tree height times this factor are removed
LI 2012	
dt1 = 1.5 m	Threshold number 1 – default value. See Li et al. 2012
dt2 = 2 m	Threshold number 2 – default value. See Li et al. 2012
R = 2 m	Search radius for local maxima.
Zu = 15 m	Height threshold determining whether threshold 1 or 2 is used.
hmin = 5 m	Minimum allowed height of a detected tree.
speed_up = 15 m	Maximum radius of a crown.

3.2.3. Accuracy Metrics used to compare Segmentation Methods

For all five methods, I assessed segmentation accuracy in terms of recall, precision, and F-score. Recall is the tree detection rate, also called producer's accuracy, precision is the correctness of detected trees, also called user's accuracy, and F-score is an overall accuracy measure that considers both false positives and false negatives but doesn't require quantifiable true negatives to calculate (unlike the overall accuracy metric calculated from a confusion matrix). I also assessed segmentation accuracy in terms of the correlation with ground-truth stem density, and I visually assessed how well the automatically generated segment shapes matched the shape and extent of the real tree canopy. Recall, precision, and F-score were calculated from the rates of true positives (TP), which represent correctly segmented trees, false positives (FP),

which were segments that did not represent a ground truth tree, and false negatives (FN), which were ground truth trees that were not assigned their own segment. Recall, precision, and F-score were calculated as follows (Li et al. 2012):

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Eq. 1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Eq. 2}$$

$$\text{F-score} = 2\,x\,\frac{recall\ x\ precision}{recall+precision} \quad \text{Eq. 3}$$

For the purposes of calculating segmentation accuracy metrics, I defined segments as "in" the plot if their centroid was located inside the plot. To avoid artificially inflated counts of true positives caused by overlapping segment boundaries (segments produced by lidR can overlap under certain conditions), the maximum number of true positives was not allowed to exceed the total number of dominant trees in a plot.

3.2.4. Modeling Height, Diameter, Species, and Coniferous/Deciduous Classification

I calculated individual tree metrics including height, diameter, and coniferous or deciduous classification for segments produced by the smoothed watershed method and by the Dalponte2016 algorithm. Because of their size, segments produced by the smoothed watershed method were analyzed with a localized area-based approach. If there was more than one dominant ground truth tree in a segment derived from the smoothed watershed method, then the ground truth data for the trees in that segment were averaged and the average ground-truth value was regressed against the average lidar-based metric. In contrast, the segments produced by the Dalponte2016 algorithm were analyzed using an individual tree approach, in which the segment

lidar values were regressed against the individual ground truth values of the largest dominant tree.

Because it was important to ensure that all segments were associated with complete ground truth data, I used more stringent criteria for selecting in-plot segments for metrics generation than I had used for the segmentation accuracy metrics. Segments that I generated using the smoothed watershed method were paired with ground truth data and counted as inside the plot if they were completely within a 2-meter buffer around the plot and either the segment centroid was within the plot or the segment contained at least one dominant ground truth stem.

Segments that I generated with the Dalponte2016 algorithm were paired with ground truth data based on combined location and height criteria. By default, I paired tree tops with stems that were in the same segment, but in some cases (due to the known presence of leaning trees in the ground truth data) some tree tops were matched with stems that were located just outside of the segmented crown boundary but were otherwise an excellent fit based on height criteria. Potential ground truth matches were restricted to dominant trees that were taller than 5 meters in order to match the height cutoff applied to the lidR tree top detection function. Taller ground truth stems were prioritized over shorter ones from the same segment. Of the segments that could not be matched to a ground truth stem, if any part extended beyond a 2-meter buffer of the plot, they were classified as outside the plot. If the segment was completely contained by the 2-meter plot buffer but it appeared very likely that the stem associated with that canopy would have been outside the plot (and therefore not recorded by field crews), that segment was also classified as outside the plot. If a segment was completely contained by the plot, then it was classified as being in the plot regardless of whether it could be matched with a ground truth stem or not.

The models were built using either linear mixed models or generalized linear mixed models (Bates et al. 2015; Pinheiro et al. 2020) depending on the distribution and the variance structure of the data. The mixed models were run with random intercepts, with the plot location set as the random effect. Potential input predictor variables were mean first return height, 70^{th}, 80^{th}, 90^{th} and 95^{th} percentile of first return height, variance of first return height, standard deviation of first return height, mean first return intensity, standard deviation of first return intensity, and percent cover (the proportion of first returns above three meters height). For the Dalponte2016 segments, I also considered the height of the modeled treetop as an input variable. Predictor variables that were correlated (using the cutoff of $r \geq 0.3$) were not combined in the same model. I chose best models based on AIC and cross-validated root mean squared error.

To estimate the amount of error produced by segments that did not correspond to ground truth data (false positives), I assessed the average prediction accuracy when the segment-based results were aggregated to a 30-meter scale. This 30-meter scale corresponded to the area-based methods discussed in chapter 2. I checked the performance of the model at this scale by using the segment-based models to predict outcomes for segments across the entire landscape and then using ArcGIS to determine modeled plot-level averages from the segments that were within the ground-truth plot boundaries. I compared the modeled plot-level metrics to the actual ground-truth plot-level metrics and calculated the average root mean squared error using leave-one-out cross-validation.

I also developed species models at the individual tree level to predict the three most common tree species. Unlike in the other models discussed above, the largest tree in the segment was used as the ground truth species for both the smoothed watershed method and the Dalponte2016 method. Segments were classified separately for each species using a logistic

generalized linear mixed model (Bates et al. 2015) with random intercepts and the ground truth plot as the random effect. In addition to the input variables discussed above, I included metrics of canopy slope and canopy roughness as predictor variables. The canopy slope metric was generated by applying the Slope tool from Esri's Spatial Analyst toolbox to the 1-meter resolution mean first return height raster and then averaging across each segment. I estimated localized canopy roughness by taking the standard deviation of the slope layer. The best model for each species was considered to be the one that had the highest classification accuracy while remaining as parsimonious as possible.

3.3. RESULTS

3.3.1 Segmentation model accuracy

Each segmentation method had its strengths and weaknesses, but overall the Dalponte2016 algorithm was the most effective in the Nooksack watershed. The Dalponte2016 and Silva2016 algorithms are closely related and were the top performers in all categories, with almost identical recall, precision, and F-score (Table 14), but they differed in subtle but important ways that the accuracy metrics did not capture. Although all three of the lidR segmentation methods can produce overlapping segments, the Silva2016 segments were less likely to overlap each other than the Dalponte2016 segments were (Figure 11). Non-overlapping segments are desirable because they make geoprocessing tasks, such as matching ground-truth stems to segments, much easier. However, on visual examination of the data it was also apparent that the Silva2016 algorithm tended to be slightly larger and often over-extended into non-canopy area (such as surrounding clearings, or inter-canopy gaps) (Figure 12). On the basis of

this shape-based inaccuracy, I decided that Dalponte2016 was the more accurate of the two overall as applied in my study area.

Among the remaining algorithms, the smoothed watershed segmentation performed the best of the eCognition methods. It had the highest precision and F-score (Table 14) and produced segments that were usually realistic, although it under-segmented even-aged stands (Figure 13) which resulted in a very poor correlation with stand density (Table 15). Of the eCognition methods, the unsmoothed watershed segmentation had the highest correlation with stand density, but it had the lowest precision of any method (Table 14; Table 15). In terms of segment shape, the unsmoothed watershed algorithm did reasonably well when applied to stands with a low overall variability in return height, but it was too sensitive when applied to larger, more structurally complex stands, tending to incorrectly segment individual tree branches (Figure 13).

The multiresolution segmentation performed best on midrange moderate-rumple canopies, especially in lower-density stands. It did not do well with either very tall, heterogenous stands or short even-aged stands (Figure 13). The multiresolution segmentation algorithm was poorly suited to capturing the range of conditions in the study area. In a localized area surrounding a single plot, it could generally be tuned to produce acceptable results, but when those same parameters were applied to a different area, accuracy plummeted. In contrast, the Li2012 algorithm did a reasonably good job of adapting to different canopy sizes across the landscape. However, it produced segments that overlapped hugely, sometimes as many as three or more segments layered at a time, and this was a serious problem. 36% of all Li segments overlapped, in contrast to 10% of Dalponte2016 and only 5% of the Silva2016 segments, and the degree of overlap was much higher for Li2012. The degree of overlap made it difficult or impossible to assign ground-truth points to some segments, and it contributed to low precision

(Table 14) and unrealistic segment shapes (for example, an elongated segment on the left side of the plot in Figure 8). However, as a side effect it did mean that the Li2012 algorithm had a better correlation with ground-truth tree density than most other methods (Table 15). In general, I observed that the Li2012 algorithm performed better with high-rumple conifer stands than with amorphous dense deciduous stands.

Table 14: Comparison of accuracy metrics for all segmentation methods. Recall, precision, and F-score were calculated separately for each ground truth plot (n = 97), and then averaged to give overall scores for the study area. Means are given ± standard deviation. Metrics shown are for dominant stems only. See Roussel and Auty (2020) and Trimble (2020) for parameters.

Algorithm	Recall	Precision	F-score	Parameters
Multiresolution	0.51 ± 0.17	0.42 ± 0.19	0.41 ± 0.11	scale = 10 shape = 0.1 compactness = 0.5 Input raster: canopy height model from mean height of first returns
Watershed (smoothed)	0.43 ± 0.15	0.78 ± 0.17	0.52 ± 0.13	Input surface: raster (1 m resolution, smoothed over a 3x3 window) of 80th percentile height returns
Watershed (unsmoothed)	0.61 ± 0.16	0.40 ± 0.18	0.44 ± 0.13	Input surface: canopy height model from mean height of first returns
Dalponte2016	0.50 ± 0.17	0.80 ± 0.20	0.58 ± 0.14	th_tree = 3 m th_seed = 0.45 th_cr = 0.55 max_cr = 15 m
Silva2016	0.51 ± 0.21	0.80 ± 0.19	0.59 ± 0.13	max_cr_factor = 0.6 exclusion = 0.3
Li2012	0.54 ± 0.18	0.58 ± 0.20	0.52 ± 0.13	dt1 = 1.5 m dt2 = 2 m R = 2 m Zu = 15 m hmin = 5 m speed_up = 15 m

Table 15: Correlations between stem counts (dominant- or all trees) and segment counts per plot.

Method	**r (dominant)**	**r (all stems)**	**Mode, median, and maximum dominant stems per segment**
Multiresolution	-0.27	-0.00	Mode: 0 stems (64%) Median: 0 stems Max.: 16 stems (0.04%)
Watershed (smoothed)	-0.07	-0.15	Mode: 1 stem (29%) Median: 2 stems Max.: 20 stems (0.12%)
Watershed (unsmoothed)	0.35	0.14	Mode: 0 stems (65%) Median: 0 stems Max.: 10 stems (0.03%)
Dalponte2016	0.61	0.34	Mode: 1 stem (34%) Median: 1 stem Max.: 10 stems (0.07%)
Silva2016	0.61	0.34	Mode: 1 stem (37%) Median: 1 stem Max.: 8 stems (0.07%)
Li2012	0.50	0.23	Mode: 0 stems (42%) Median: 1 stem Max.: 24 stems (0.05%)

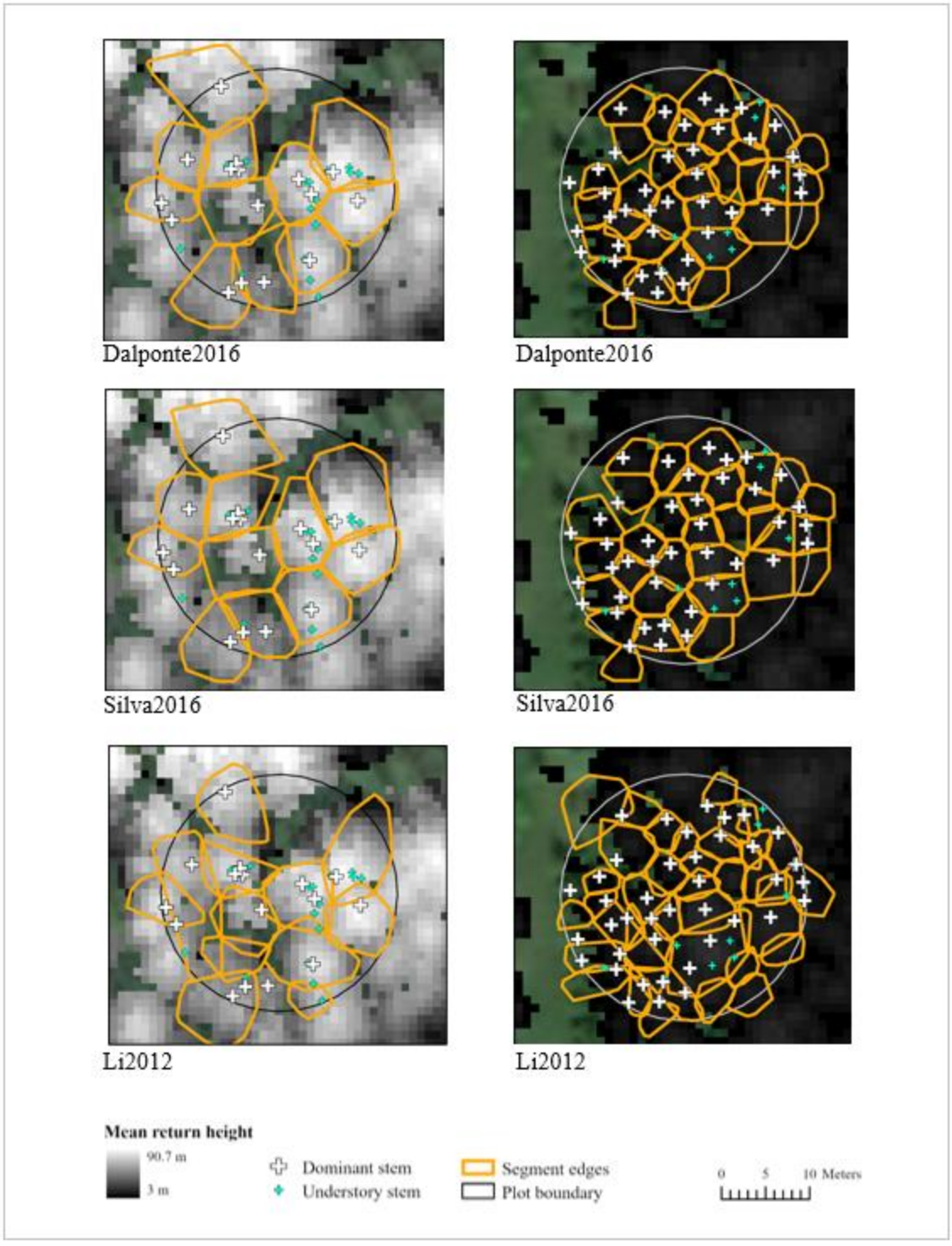

Figure 11: Examples of lidR segmentation outputs for a plot with large coniferous trees in the upper North Fork (left panels) and a plot with densely spaced small conifers on the lower main fork (right panels). Note that segments produced using the lidR packages can sometimes overlap each other.

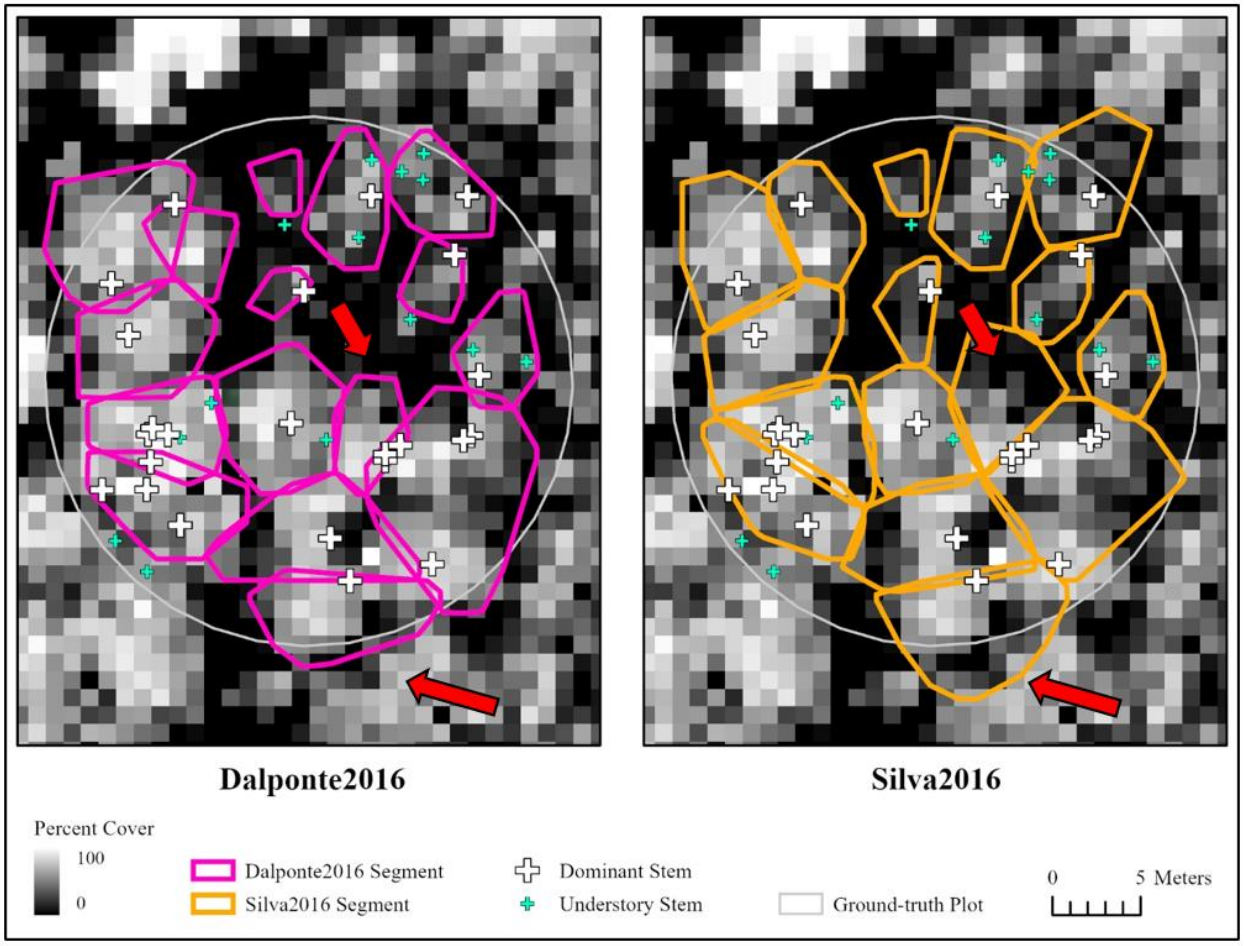

Figure 12: Comparison of the Dalponte2016 and Silva2016 segmentation outputs in plot 68. The Silva2016 algorithm tended to over-grow into canopy gap regions (see arrows).

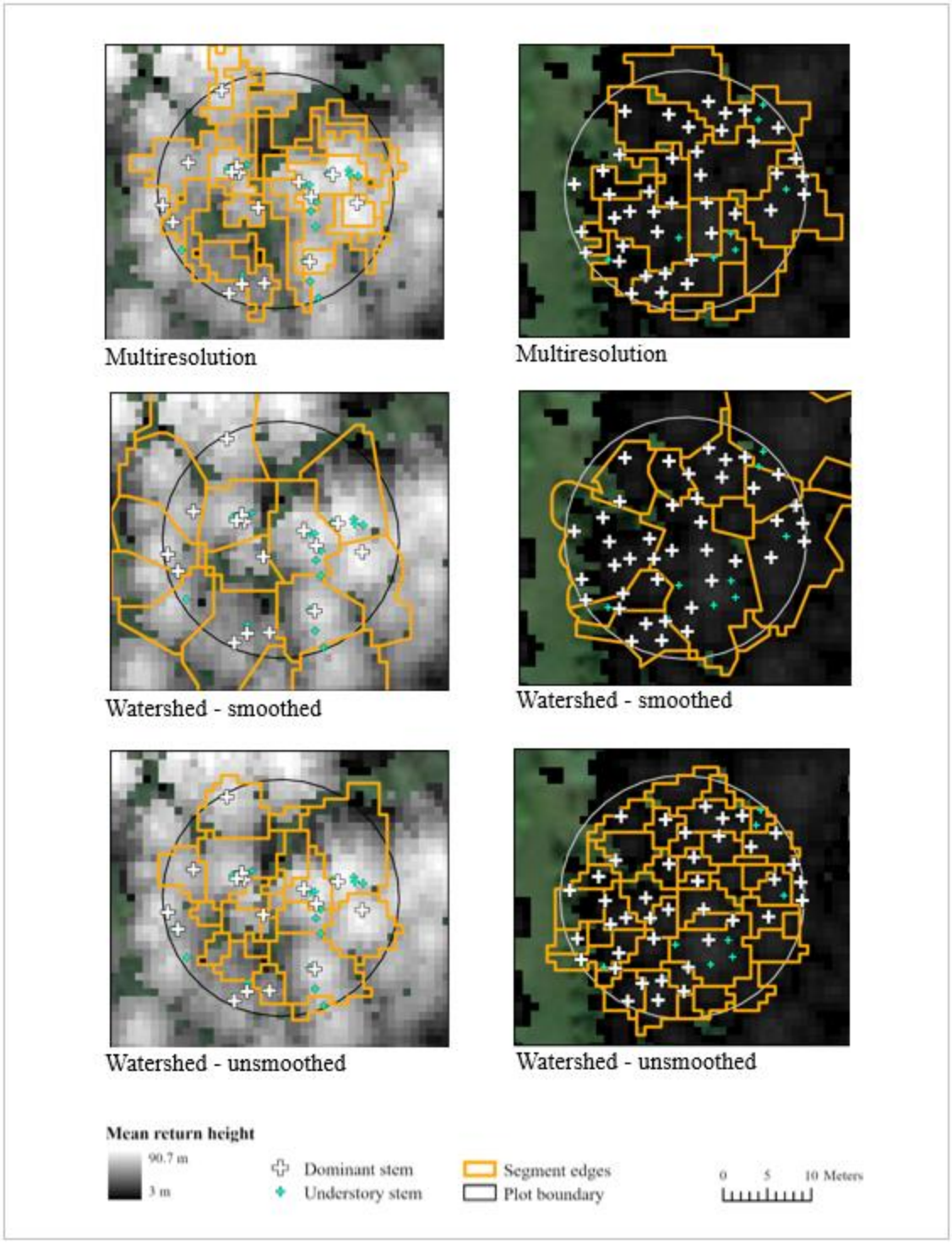

Figure 13: Examples of eCognition segmentation outputs for a plot with large coniferous trees in the upper North Fork (left panels) and a plot with densely spaced small conifers on the lower main fork (right panels). Segments produced using eCognition cannot overlap.

3.3.2. Height, Diameter, and Coniferous/Deciduous Classification

Models based on the two segmentation approaches that I selected for further processing varied in their usefulness for predicting ground-truth conditions (Table 16). These two approaches were the Dalponte2016 algorithm, which I selected as the best-performing approach, and the smoothed watershed approach, which I selected because it was the best-performing of the eCognition methods and it gave less-ambiguous results than the Li2012 algorithm. Segment-level height predictions were slightly more accurate for the watershed segmentation, but diameter estimates were more accurate for the Dalponte2016 method. Error from false positives accumulated faster at the 30-meter scale for the Dalponte2016 method than it did for the smoothed watershed method. In general, error was higher at the 30-meter scale, but the DBH estimates for the smoothed watershed segmentation were actually more accurate at the less detailed scale. In all cases, the most useful input parameter was a metric equivalent to the tallest point of the tree in the lidar point cloud.

Classification into coniferous or deciduous categories was similar for both segmentation methods. Both methods had an overall accuracy of 91% (with Cohen's kappa = 0.80), but the Dalponte2016 method had a more balanced performance. User's accuracy based on the Dalponte2016 method was the same for both conifers and deciduous trees, but the user's accuracy for the model based on the smoothed watershed segments showed that it was better at identifying deciduous trees than conifers (Table 16).

Table 16: Results of individual-tree-based models of forest structure. Ground truth data was averaged within each segment for the watershed method, but the Dalponte2016 method was ground-truthed with one dominant tree per segment. $n_{watershed}$ = 330 from 91 plots; $n_{Dalponte}$ = 1030 from 97 plots. Accuracies of segment-based analysis when restricted to segments with ground-truth data are in parentheses; other values represent segmentation accuracy when scaled up to the same level as the area-based analysis (i.e. field-validated accuracy of model predictions when evaluated at the 30-meter scale). Mixed models were calculated using R packages "nlme" (Pinheiro et al. 2020) and "lme4" (Bates et al. 2015).

SEGMENT-BASED ANALYSIS – SMOOTHED WATERSHED METHOD

	RMSE ± std. dev.	*Model*	*Parameters*
Mean height *(dominant)*	2.26 ± 1.75 m (2.87 ± 2.70 m)	Linear mixed-effect model *(lme4 package)*	95th percentile lidar return height
Mean height *(all trees)*	2.38 ± 1.80 m (3.14 ± 3.02 m)	Linear mixed-effect model *(lme4 package)*	95th percentile lidar return height
Maximum height	7.16 ± 5.30 m (3.30 ± 2.92 m)	Linear mixed-effect model *(lme4 package)*	95th percentile lidar return height
Mean DBH *(dominant)*	19.60 ± 13.58 cm (6.73 ± 6.71 cm)	Linear mixed-effect model *(nlme package)* Fixed variances	95th percentile lidar return height
Mean DBH *(all trees)*	13.08 ± 10.25 cm (6.32 ± 6.32 cm)	Linear mixed-effect model *(nlme package)* Fixed variances	95th percentile lidar return height

SEGMENT-BASED ANALYSIS – DALPONTE 2016 METHOD

	RMSE ± std. dev.	*Model*	*Parameters*
Tree height	2.50 ± 2.46 m (6.63 ± 5.22 m)	Linear mixed-effect model *(lme4 package)*	Height of lidar-derived tree top
Tree DBH	9.10 ± 10.56 cm (13.13 ± 9.21 cm)	Linear mixed-effect model *(nlme package)* Fixed variances	Height of lidar-derived tree top

Table 17: Confusion matrix for individual tree-based composition models for dominant trees. For both segmentation methods, composition was modeled with a generalized linear mixed model (Bates et al. 2015) on mean intensity (scaled by 1/1000) with random intercepts and plot as a random effect.

SEGMENT-BASED ANALYSIS – SMOOTHED WATERSHED METHOD

	Field control		
	Conifer	Deciduous	Row total
Classification			
Conifer	**97**	13	110
Deciduous	16	**204**	220
Column total	113	217	330

Producer's accuracy	**User's accuracy**
Conifer = 97/113 = 86%	Conifer = 97/110 = 88%
Deciduous = 204/217 = 94%	Deciduous = 204/220 = 93%
Overall accuracy = (97+204)/330 = 91%	
Kappa = 0.80	

SEGMENT-BASED ANALYSIS – DALPONTE 2016 METHOD

	Field control		
	Conifer	Deciduous	Row total
Classification			
Conifer	**295**	30	325
Deciduous	65	**682**	747
Column total	360	712	1072

Producer's accuracy	**User's accuracy**
Conifer = 295/360 = 82%	Conifer = 295/325 = 91%
Deciduous = 682/712 = 96%	Deciduous = 682/747 = 91%
Overall accuracy = (295+682)/1072 = 91%	
Kappa = 0.80	

3.3.3. Species Classification

For models based on the smoothed watershed segmentation, classification accuracies were highest for predicting cottonwood, with a user's accuracy of 86% (Table 18). Douglas fir, (which was classified after cottonwood trees had already been removed), followed with a user's accuracy of 79% (Table 19). Red alder also had a user's accuracy of 79%, but its overall accuracy was lower (Table 20). However, unlike the Douglas fir model, it wasn't dependent on another model being run first.

Species classification accuracies were higher for models based on the Dalponte2016 segmentation than for the watershed segmentation, but kappa values were in the same range. Cottonwood could be predicted with a user's accuracy of 89% (Table 21), Douglas fir had a user's accuracy of 83% (Table 22), and red alder had a user's accuracy of 83% (Table 23). Each of the models built on the Dalponte2016 segmentation could be run separately: none of them needed another species to be classified and removed first.

Table 18: Confusion matrix for cottonwood trees from the smoothed watershed segmentation. Species was modeled from canopy steepness and variation in canopy steepness. Segments were classified using a logistic generalized linear mixed model with plot held as a random effect.

	Field control		
	Cottonwood	Other species	Row total
Classification			
Cottonwood	**67**	11	78
Other species	25	**211**	236
Column total	92	222	314

Producer's accuracy	**User's accuracy**
Cottonwood = 67/92 = 73%	Cottonwood = 67/78 = 86%
Other species = 211/222 = 95%	Other species = 211/236 = 89%
Overall accuracy = (67+211)/314 = 89%	
Kappa = 0.71	

Table 19: Confusion matrix for Douglas fir trees from the smoothed watershed segmentation. Species was modeled using canopy steepness and percent cover. Segments were classified using a logistic generalized linear mixed model with plot held as a random effect. Cottonwood trees were classified and removed from the dataset before this model was run; therefore, cottonwoods are excluded from the "other species" category.

	Field control		
	Douglas fir	Other species	Row total
Classification			
Douglas fir	**26**	7	33
Other species	14	**189**	203
Column total	40	196	236

Producer's accuracy	**User's accuracy**
Douglas fir = 26/40 = 65%	Douglas fir = 26/33 = 79%
Other species = 189/196 = 96%	Other species = 189/203 = 93%
Overall accuracy = (26+189)/236 = 91%	
Kappa = 0.66	

Table 20: Confusion matrix for red alder trees using the smoothed watershed segmentation. Species was modeled with variability in canopy steepness and mean return intensity. Segments were classified using a logistic generalized linear mixed model with plot held as a random effect.

	Field control		
	Red alder	Other species	Row total
Classification			
Red alder	**46**	12	58
Other species	42	**214**	256
Column total	88	226	314
Producer's accuracy		**User's accuracy**	
Red alder = 46/88 = 52%		Red alder = 46/58 = 79%	
Other species = 214/226 = 95%		Other species = 214/256 = 84%	
Overall accuracy = (46+214)/314 = 83%			
Kappa = 0.52			

Table 21: Confusion matrix for cottonwood trees based on the Dalponte2016 segmentation model. Species was modeled using canopy steepness and variation in canopy steepness. Segments were classified using a logistic generalized linear mixed model with plot held as a random effect.

	Field control		
	Cottonwood	Other species	Row total
Classification			
Cottonwood	**189**	23	212
Other species	61	**757**	818
Column total	250	780	1030
Producer's accuracy		**User's accuracy**	
Cottonwood = 189/250 = 76%		Cottonwood = 189/212 = 89%	
Other species = 757/780 = 97%		Other species = 757/818 = 93%	
Overall accuracy = (189+757)/1030 = 92%			
Kappa = 0.66			

Table 22: Confusion matrix for Douglas fir trees based on the Dalponte2016 segmentation method. Species was modeled based on mean return intensity. Segments were classified using a logistic generalized linear mixed model with plot held as a random effect.

	Field control		
	Douglas fir	Other species	Row total
Classification			
Douglas fir	**119**	25	144
Other species	64	**822**	886
Column total	183	847	1030
Producer's accuracy		**User's accuracy**	
Douglas fir = 119/183 = 65%		Douglas fir = 119/144 = 83%	
Other species = 822/847 = 97%		Other species = 822/886 = 93%	
Overall accuracy = (119+822)/1030 = 91%			
Kappa = 0.68			

Table 23: Confusion matrix for red alder trees based on the Dalponte2016 segmentation method. Species was modeled based on mean return intensity. Segments were classified using a logistic generalized linear mixed model with plot held as a random effect.

	Field control		
	Red alder	Other species	Row total
Classification			
Red alder	**214**	43	257
Other species	121	**652**	773
Column total	335	695	1030
Producer's accuracy		**User's accuracy**	
Red alder = 214/335 = 64%		Red alder = 214/257 = 83%	
Other species = 652/695 = 94%		Other species = 652/773 = 84%	
Overall accuracy = (214+652)/1030 = 84%			
Kappa = 0.61			

3.4. DISCUSSION

In general, my segmentation model results were consistent with the results of previous studies in the Pacific Northwest despite the comparatively high structural heterogeneity in my study area. My tree detection rate for most algorithms was approximately 50%, which was comparable to previous studies (Heurich et al. 2004; Korpela et al. 2007). This seems to have been mostly due to high stand densities in my study area (Jeronimo et al. 2018). On average there were 34 stems in each 0.07 ha ground-truth plot (486 stems/ha), but the maximum was 87 stems per plot (1243 stems/ha) (Appendix F). Recall may also have been influenced by stand composition. Deciduous forests tend to lead to low detection rates (usually in the 50-60s range) compared to conifer stands which routinely support recall rates of 70% or more (Vauhkonen et al 2012).

Precision (the correctness of detected trees) was much more variable between approaches, but the maximum precision was around 80% for the Dalponte2016 and Silva2016 algorithms. For comparison, Pirotti et al. (2017) reported recall = 0.68 and precision = 0.72 when they compared the Li2012, Dalponte2016, and a watershed method in a mixed-forest in Slovenia, and Li et al. (2012) reported a precision of 0.94 for a Sierra Nevada study area primarily dominated by conifers. For all approaches, my overall segmentation accuracy tended to be low because of the low recall, which was consistent with a study by Jeronimo et al. (2018). Jeronimo et al. (2018) reported F-scores of 0.5 for Pacific Northwest stands, and they noted that recall was worse in high-density areas and precision was worse in low-density, tall, complex stands. Both of these patterns were apparent in my data as well.

To the best of my knowledge, my species classification study was unique in the Pacific Northwest for the complexity of forest types and terrain it encompassed. Most lidar-based species classification studies in the Pacific Northwest and worldwide have focused on relatively small, homogeneous study areas with gentle terrain, comparatively low variability in age classes, and good accessibility (Fassnacht et al. 2016; White et al. 2016). My species classification accuracies were comparable to results obtained under much simpler conditions with more complex lidar data. In a recent Pacific Northwest study that is the most directly comparable to mine, Vaughn et al. (2012) used a combination of extremely dense (104 points/m2) discrete point and waveform lidar data to classify five species (black cottonwood, bigleaf maple (*Acer macrophylum),* Douglas fir, red alder, and western redcedar) from 130 canopy segments in the Seattle Arboretum. They discarded segments that were found to contain multiple trees. When discrete point and waveform data were combined, the researchers reported user's accuracies of about 85% for four of the species and 95% accuracy for bigleaf maple. My user's accuracies for red alder, cottonwood, and Douglas fir were in a similar range, although western redcedar and bigleaf maple were too rare in my ground-truth data to support building a model.

Sample size relative to landscape variability may have been key. My classification accuracies for separating deciduous trees from conifers were higher than those reported for a study in the Seattle Arboretum by Kim et al. (2011), despite their study being predominantly based on semi-isolated individual trees in open, flat areas. It seems likely that the difference in accuracy was due to respective sample sizes, as well as the fact that their study area had a much higher biodiversity than mine: they sampled 233 individual trees from over 40 species in 15 genera, whereas my individual tree models were based on over 1110 trees from only 13 species. Although these results are encouraging for the application of individual tree analyses to dense

forests, they also underline the importance of understanding the potential effects that sample size and varying species composition can have.

I chose the smoothed watershed segmentation and the Dalponte2016 segmentation methods for further processing because the Dalponte2016 method was the most accurate overall, and the smoothed watershed method was the best of the eCognition methods. (I also calculated some individual metrics for the Silva2016 algorithm. As expected, accuracies were close to but slightly lower than for the Dalponte2016 algorithm. These results are included in Appendix E). Although the Dalponte2016 method better represented individual trees, the smoothed watershed method was surprisingly more effective at predicting tree height. This may have been because the ground truth data was averaged across the segment, thus downplaying potential errors caused by various factors including inaccurate ground-truth to segment pairing, the three-year temporal disjoint between the lidar data and the ground-truth data collection, or other inaccuracies. Any inaccuracy in either the lidar tree top height detection (tree tops can be difficult to sense with lidar (Gatziolis et al. 2010)) or ground-truth height measurements would have been smoothed out by the averaging approach taken with the watershed method. Errors in ground-truth height measurement are unquantifiable and may have been worse in areas with steep slopes. Interestingly, almost all ground truth trees that were paired with the Dalponte2016 segments had heights that had been directly measured in the field, suggesting that trees that were easier to pick out and measure in the field were also more likely to be picked out and segmented by the algorithm.

One of the best features of the Dalponte2016 algorithm was its ability to avoid region-growing into non-canopy area. This probably contributed to its success at predicting conifer vs deciduous and the species of trees, since it avoided interference from gaps and non-canopy area

included in the "canopy" segments. All the other top contenders, the Silva2016, Li2012, and smoothed watershed algorithm, had problems with including too much of the ground next to a tree canopy. This almost certainly influenced many of the input metrics, especially measures of intensity and of canopy height roughness. The Li2012 algorithm produced segments that overlapped far too much and often had unrealistic boundaries, especially in low-rumple deciduous stands. This algorithm seemed to do better in areas where the trees were spaced farther apart, which is not surprising because it was developed for coniferous stands (Li et al. 2012). The unsmoothed watershed and the multiresolution segmentation methods were the least successful because they were not adaptable enough to handle the wide variety of conditions present in the study area.

The biggest problem with the Dalponte2016 segmentation was the amount of overlap between segments. With 10% of the segments overlapping another segment to some degree, dominant ground-truth stems occasionally were located in places where they had ambiguous membership in either segment. Overlapping tree canopies are realistic, but they are challenging from a geoprocessing standpoint, and were part of the reason I decided it was necessary to manually supervise the assigning of ground-truth stems to segments for the Dalponte2016 method. Manual oversight of segment/ground-truth stem pairing removed one source of potential error from the species classification models, but it significantly increased the processing time.

I chose to manually pair stems to segments because I knew there were widespread problems with leaning trees. When setting out to do fieldwork, I originally intended to keep track of all leaning trees in the ground-truth data, but I quickly realized that the majority of trees were displaced from their stem bases by several meters. In the end, trees tagged as "leaning" in the ground-truth data were only those that were so dramatically offset that I didn't think it would be

possible to match them to a canopy – for example, stems that were leaning at a 45-degree angle or more. Almost 5% of stems were tagged as "leaning" in the ground truth data, even though I didn't start recording the leaning characteristic until the second week of sampling. I observed that riparian trees tended to lean more than their upland counterparts, which I attributed to: a) a higher proportion of deciduous trees in riparian zones (deciduous trees often have a less upright growth structure than conifers), b) unstable slopes gradually shifting over time (this was mostly seen affecting older, larger trees that would normally grow fairly straight), and c) wet riparian conditions making trees less strongly rooted and less stable (this was especially noted in red alder trees that were growing in wetland areas). Leaning trees greatly complicated the process of matching ground truth stems to lidar canopies. Attempting to match stems to canopies in a mountainous riparian area was probably a worst-case scenario. Forked trees were another major problem, and they were similarly hard to quantify. Cottonwood and bigleaf maple trees were particularly likely to have multiple distinct canopies, but when one was working in dense, closed stands, it wasn't always easy to see how distinct the multi-part canopies were.

Many researchers have assumed that species model accuracy would be very dependent on accurate segmentation of individual trees (Fassnecht et al. 2016). Consequently, previous studies on classifying tree species (Kim et al. 2009; Kim et al. 2011; Vaughn et al. 2012) have generally favored basing their models on trees that were relatively isolated from other canopies. However, at least for detecting tree species that have distinctive morphology and reflectivity, my results suggest that the presence of additional smaller trees in a canopy segment may not be crippling to species classification even if the additional trees are dominant (present in the upper canopy) and theoretically visible from above. If this is true, it removes a major barrier to the classification of dense, complex stands, and it would be worthwhile to quantify the extent to which multiple trees

in a segment affect species classification accuracies. Many studies have documented the inherent difficulties in achieving true individual tree segmentation in complex forest stands (Richardson and Moskal 2011; North et al. 2017; Jeronimo et al. 2018) and some researchers have suggested the use of the term "tree-approximate object" to reflect these difficulties (North et al. 2017). However, my study intentionally did not focus on a specific subset of trees because I wanted to determine the practicality of species classification across a realistic landscape. I was not able to identify as many different species as other researchers (Kim et al. 2009; Kim et al. 2011; Vaughn et al. 2012), but investigating the ways the models broke down in a complex forest provided a useful insight into the system and suggested possible avenues for future improvement.

I identified several factors that caused difficulties in my models but that, if properly addressed, could potentially improve species identification in future work. One problem, related to the intrinsic limitations of monochromatic aerial lidar systems, was that most lidar-derived input variables were correlated with each other. This restricted the options for model building, and it made it difficult to consider more than one or two unique identifying characteristics (shape, texture, color/reflectivity, etc.) of a species at a time. A possible way around this might be to put more emphasis on measures of canopy morphology, such as canopy slope, canopy roughness, branching habits, and return densities at different strata, that are less likely to correlate with each other. The lidR package has a number of functions that I was unable to explore in this study, but which might be very useful for deriving such metrics. I also think it is important to think carefully about what makes each species unique to human eyes, and then consider how best to measure those attributes from the lidar's standpoint. In particular, it would be helpful to pay close attention to the scale of the attributes that need to be detected and include input variables calculated at multiple resolutions. Canopy roughness, for example, might benefit

from a multiple-scale approach, with a fine-textured scale to try and capture branch patterns, and a smoothed-over scale to capture the overall shape of the entire crown. However, it is important to note that relying more heavily on shape-based attributes might lead to unexpected consequences, especially in areas with extreme topography.

Differences in topography may have affected my model accuracies, even without widespread use of shape-based variables. Steep slopes almost certainly affected the accuracy of the ground truth data, both stem location and tree height measurements, and they may have reduced the accuracy of the lidar height data as well (Gatziolis et al. 2010). Steep slopes also resulted in a certain degree of distortion to the lidar point cloud and the lidar crown shapes, but I do not know if this had any significant effect on the model outcomes. This distortion occurred because when the point cloud was normalized, the angles of the tree trunks relative to the ground were altered as the "ground" was re-projected onto a flat plane. The shape distortion in the canopy was minimal, but it was visible.

Steep slopes were underrepresented in the ground truth data used for individual tree analyses. Out of the 104 plots collected during summer 2019, seven had no stem location data because the sites were too steep to access with the total station and were too steep for other methods of measuring location to be practical. Furthermore, the watershed contained many sites that were too steep for any sampling to be carried out at all. If the models developed in this study were applied across the entire study area, these very steep sites would probably have worse accuracy outcomes than the surrounding riparian forest, but there would be no way to quantify the difference.

My analysis was probably both helped and hindered by the size of my study area. The forests along the Nooksack are extremely varied, including everything from dense early-seral cottonwood stands to a few scattered patches of old growth trees. Parameters that worked well for young deciduous stands usually produced unwanted results in stands with high-relief canopies, and vice versa. I expect that better classification accuracies would be possible if the study area were split into subsections based on structural characteristics and separate segmentation model parameters were used for each. Subdividing the study area would need to be done with care in order to avoid increasing the extent of edge effects. The end result would be more complicated than the simple models I ran, but I think this is a worthwhile direction for future research. A definite advantage of the size of my study area and the number of ground-truth plots I sampled (97 plots with stem location data) was that I had large enough sample sizes to support building species models, despite the large amount of intra-species variability in riparian tree morphology. Sample size seemed to be very important: The only three species that were frequent enough in the data to support building a species classification model were cottonwood (29% of the watershed segments; 24% of the Dalponte2016 segments), red alder (28% of the watershed segments; 32% of the Dalponte2016 segments), and Douglas fir (18% of the watershed segments; 17% of the Dalponte2016 segments). The remaining ten species present in the watershed did not have distinctive enough characteristics to be detected with the available sample size.

When applied at the landscape scale, the primary purpose of individual tree segmentation is to derive metrics such as tree species that are not practical to detect from aggregated data. Individual tree segmentation is not always the best choice for all metrics because it is extremely computationally expensive and does not always produce results that are superior to area-based

approaches, especially when used to predict metrics such as average tree height (compare Table 16 to Table 9 in chapter 2). Persistent problems with complex canopies and with separating understory trees from their overstory counterparts during segmentation mean that errors at the individual tree level have the potential to propagate when results are aggregated to make stand-level decisions. Because individual tree data is rarely considered on an individual tree basis for planning purposes, I thought it was important to get a feel for how much error might creep in if the results were aggregated to a larger scale. Hopefully the disjoints exposed by the increased (and sometimes decreased!) errors in my 30-meter analysis relative to the segment-level errors (Table 16) can be useful to land managers when deciding how best to interpret the GIS layers that I produced from this lidar data.

3.5. CONCLUSION

Out of the algorithms I tested, the Dalponte2016 segmentation method performed the best in terms of accurate tree crown segmentation, stem density, and individual tree species identification for the three species successfully modeled in this study: Douglas fir, cottonwood, and red alder. The smoothed watershed segmentation method was more accurate when predicting localized averages of tree height. I primarily attributed this to individual variability in growth rates between the time of lidar data acquisition and ground-truth data collection. Both methods performed approximately equally well at classifying segments as either deciduous or coniferous, with overall accuracies at 91% and user's accuracies in the same range. Segmentation and coniferous/deciduous classification accuracies that I produced with the Dalponte2016 algorithm were similar to those reported by previous research (Kim et al. 2011). Species classification accuracies were lower than those reported by the limited prior research in the Pacific Northwest (Kim et al. 2009; Kim et al. 2011; Vaughn et al. 2012) which was probably due to the size of my

study area, the density of the stands, and the amount of variation in forest structure across the watershed. Future research should investigate using a greater variety of canopy morphology metrics as input variables and should explicitly consider multiple scales and resolutions when deriving input lidar metrics used to classify tree species.

4.0 - Works Cited

Anderson, S. W., and C. P. Konrad. 2019. Downstream-propagating channel responses to decadal-scale climate variability in a glaciated river basin. Journal of Geophysical Research: Earth Surface 124:902-919.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1-48.

Beechie, T. J., and T. H. Sibley. 1997. Relationships between channel characteristics, woody debris, and fish habitat in northwestern Washington streams. Transactions of the American Fisheries Society 126:217-229.

Beechie, T., G. Pess, S. Morley, L. Butler, P. Downs, A. Maltby, P. Skidmore, S. Clayton, C. Muhlfeld, and K. Hanson. 2013a. "Watershed assessments and identification of restoration needs" in Stream and Watershed Restoration: A Guide to Restoring Riverine Processes and Habitats. Ed. Roni, P., and T. Beechie. Wiley-Blackwell, Chichester, United Kingdom.

Beechie, T., J. S. Richardson, A. M. Gurnell, and J. Negishi. 2013b. "Watershed processes, human impacts, and process-based restoration" in Stream and Watershed Restoration: A Guide to Restoring Riverine Processes and Habitats. Ed. Roni, P., and T. Beechie. Wiley-Blackwell, Chichester, United Kingdom.

Bilby, R. E. and J. W. Ward. 1989. Changes in characteristics and function of woody debris with increasing size of streams in Western Washington. Transactions of the American Fisheries Society 118:368-378.

Bilby, R. E. and J. W. Ward. 1991. Characteristics and function of large woody debris in streams draining old-growth, clear-cut, and second-growth forests in Southwestern Washington. Canadian Journal of Fisheries and Aquatic Sciences 48:2499-2508.

Bisson, P. A., R. E. Bilby, M. D. Bryant, C. A. Dolloff, G. B. Grette, R. A. House, M. L. Murphy, K. V. Koski, and J. R. Sedell. 1987. "Large woody debris in forested streams in the Pacific Northwest: Past, present, and future" in Streamside management: Forestry and fishery interactions. Ed. Salo, E. O., and T. W. Cundy. College of Forest Resources, University of Washington, Seattle, Washington.

Bode, C. A., M. P. Limm, M. E. Power, and J. C. Finlay. 2014. Subcanopy solar radiation model: Predicting solar radiation across a heavily vegetated landscape using LiDAR and GIS solar radiation models. Remote Sensing of Environment 154:387-397.

Brown, M., and M. Maudlin. 2007. Upper South Fork Nooksack River habitat assessment. Lummi Nation, Natural Resources Department, Bellingham, Washington.

Burnett, K. M., G. H. Reeves, D. J. Miller, S. Clarke, K. Vance-Borland, K. Christiansen. 2007. Distribution of salmon-habitat potential relative to landscape characteristics and implications for conservation. Ecological Applications 17: 66-80.

Capuana, E. M. 2013. Assessment of riparian conditions in the Nooksack River basin with the combination of lidar, multi-spectral imagery and GIS. M.S. Thesis, Western Washington University, Bellingham, Washington.

Chen, W., H. Xiang, and K. Moriya. 2020. Individual tree position extraction and structural parameter retrieval based on airborne LiDAR data: Performance evaluation and comparison of four algorithms. Remote Sensing 12:571-591.

City of Bellingham. 2021. Middle Fork Nooksack river fish passage project. Available at https://www.cob.org/services/environment/restoration/middlefork (accessed 23 January 2021).

Cochran, W. G. 1977. Sampling Techniques, 3rd edition. John Wiley and Sons, New York.

Coe, T. 2001. Nooksack River watershed riparian function assessment. Natural Resources Department, Nooksack Indian Tribe. Available at http://salmonwria1.org/webfm_send/51 (accessed 31 May 2019).

Collins, B. D., and A. J. Sheikh. 2004a. Historical riverine dynamics of the Nooksack River. Report to Nooksack Indian Tribe Natural Resources Department, Bellingham, Washington.

Collins, B. D., and A. J. Sheikh. 2004b. Historical channel locations of the Nooksack River. Report to Whatcom County Public Works department, Bellingham, Washington.

Crozier, L. 2016. Impacts of climate change on salmon of the Pacific Northwest: A review of the scientific literature published in 2015. Northwest Fisheries Science Center, Fish Ecology Division, Seattle.

Dai, W., B. Yang, Z. Dong, and A. Shaker. 2018. A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 144:400-411.

Dalponte, M., and D. A. Coomes. 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods in Ecology and Evolution 7:1236-1245.

Dickerson-Lange, S. E., and R. Mitchell. 2014. Modeling the effects of climate change projections on streamflow in the Nooksack River basin, Northwest Washington. Hydrological Processes. DOI: 10.1002/hyp.10012

DNR (Washington State Department of Natural Resources). 2018. Washington LiDAR portal. Available at http://lidarportal.dnr.wa.gov/ (accessed 8 March 2018).

ECOS (Environmental Conservation Online System). 2020. Listed species believed to or known to occur in Washington. U.S. Fish and Wildlife Service. Available at https://ecos.fws.gov/ecp0/reports/species-listed-by-state-report?state=WA&status=listed (accessed 8 March 2020).

Eitel, J. U. H., B. Höfle, L. A. Vierling, A. Abellán, G. P. Asner, J. S. Deems, C. L. Glennie, P. C. Joerg, A. L. LeWinter, T. S. Magney, G. Mandlburger, D. C. Morton, J. Müller, and K. T. Vierling. 2016. Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences. Remote Sensing of Environment 186:372-392.

EPA (Environmental Protection Agency). 2003. EPA Region 10: guidance for Pacific Northwest state and tribal temperature water quality standards. EPA 910-B-03-002. Region 10 Office of Water, Seattle.

Esri, Maxar, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, AeroGRID, IGN, and the GIS User Community. 2020. Esri Basemap: World Imagery (satellite data, 0.5m resolution). Available at https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9 (accessed August 27, 2020).

Fassnacht, F. E., H. Latifi, K. Sterenczak, A. Modzelewska, M. Lefsky, L. T. Waser, C. Straub, and A. Ghosh. 2016. Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment 186:64-87.

Fox, M., and S. Bolton. 2007. A regional and geomorphic reference for quantities and volumes of instream wood in unmanaged forested basins of Washington State. North American Journal of Fisheries Management 27:342-359.

Frazer, G. W., S. Magnussen, M. A. Wulder, and K. O. Niemann. 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sensing of Environment 115:636-649.

Fu, P., and P. M. Rich. 2002. A geometric solar radiation model with applications in agriculture and forestry. Computers and Electronics in Agriculture 37:25-35.

Fullerton, A. H., T. J. Beechie, S. E. Baker, J. E. Hall, and K. A. Barnas. 2006. Regional patterns of riparian characteristics in the interior Columbia River basin, Northwestern USA: applications for restoration planning. Landscape Ecology 21:1347-1360.

Gatziolis, D., and H. Andersen. 2008. A guide to LIDAR data acquisition and processing for the forests of the Pacific Northwest. Pacific Northwest Research Station General Technical Report PNW-GTR-768. Portland, Oregon.

Gatziolis, D., J. S. Fried, and V. S. Monleon. 2010. Challenges to estimating tree height via LiDAR in closed-canopy forests: a parable from Western Oregon. Forest Science 56: 139-155.

Grah, O., and J. Beaulieu. 2013. The effect of climate change on glacier ablation and baseflow support in the Nooksack River basin and implications on Pacific salmonid protection and recovery. Climatic Change [online]. Doi: 10.1007/s10584-013-0747-y

Grähler, B., E. Pebesma, and G. Heuvelink. 2016. Spatio-temporal interpolation using gstat. The R Journal 8:204-218.

Gregory, S. V., F. J. Swanson, W. A. McKee, and K. W. Cummins. 1991. An ecosystem perspective of riparian zones. BioScience 41:540-551.

Gregory, S. V., and P. A. Bisson. 1997. "Degradation and loss of anadromous salmonid habitat in the Pacific Northwest" in Pacific Salmon and their Ecosystems: Status and Future Options. Ed. Strouder, D. J., P. A. Bisson, and R. J. Naiman. Chapman and Hall, New York.

Heurich, M., Å. Persson, J. Holmgren, and E. Kennel. 2004. Detecting and measuring individual trees with laser scanning in mixed mountain forest of central Europe using an algorithm developed for Swedish boreal forest conditions. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36:307-312.

Hijams, R. J. 2020. raster: Geographic data analysis and modeling. R package version 3.1-5. Available at https://CRAN.R-project.org/package=raster (accessed 30 January 2021).

Holmgren, J., and Å. Persson. 2004. Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment 90:415-423.

Hyatt, T. 2007. Lower North Fork Nooksack River: Reach Assessment and Restoration Recommendations. Natural Resources Department, Nooksack Indian Tribe, Deming, Washington.

Haytt, T., T. Z. Waldo, and T. J. Beechie. 2004. A watershed scale assessment of riparian forests, with implications for restoration. Restoration Ecology 12:175-183.

Hyatt, T., and R. J. Naiman. 2001. The residence time of large woody debris in the Queets River, Washington, USA. Ecological Applications 11:191-202.

Hyyppä, J., H. Hyyppä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo. 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. International Journal of Remote Sensing 29:1339-1366.

Jeronimo, S. M. A., V. R. Kane, D. J. Churchill, R. J. McGaughey, and J. F. Franklin. 2018. Applying LiDAR individual tree detection to management of structurally diverse forest landscapes. Journal of Forestry 116:336-346.

Johnson, M. F., and R. L. Wilby. 2015. Seeing the landscape for the trees: Metrics to guide riparian shade management in river catchments. Water Resources Research 51:3754-3769.

Kemp, P., D. Sear, A. Collins, P. Naden, and J. Iwan. 2011. The impacts of fine sediment on riverine fish. Hydrological Processes 25:1800-1821.

Kennedy, J. T., H. Nicholas, S. Job, J. Butcher, S. Hood, and T. Mohamedali. 2020. South Fork Nooksack River temperature total maximum daily load: Water quality improvement report and implementation plan. Washington State Department of Ecology, Olympia, Washington. Available at https://apps.ecology.wa.gov/publications/SummaryPages/2010007.html (accessed 21 February 2021).

Kim, S., T. Hinckley, and D. Briggs. 2011. Classifying individual tree genera using stepwise cluster analysis based on height and intensity metrics derived from airborne laser scanner data. Remote Sensing of Environment 115:3329-3342.

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

Korpela, I., B. Dahlin, H. Schäfer, E. Bruun, F. Haapaniemi, J. Honkasalo, S. Ilvesniemi, V. Kuutti, M. Linkosalmi, J. Mustonen, M. Salo, O. Suomi, and H. Virtanen. 2007. Single-tree forest inventory using lidar and aerial images for 3D treetop positioning, species recognition, height and crown width estimation. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36:227-233.

Li, W., Q. Guo, M. K. Jakubowski, and M. Kelly. 2012. A new method for segmenting individual trees from the lidar point cloud. Photogrammetric Engineering & Remote Sensing 78:75-84.

LNR (Lummi Nation Department of Natural Resources). 2011. Middle Fork Nooksack River habitat assessment. Available at http://hwsconnect.ekosystem.us/Project/360/8575 (accessed 31 May 2019).

Loicq, P., F. Moatar, Y. Julian, S. J. Dugdale, and D. M. Hannah. 2018. Improving representation of riparian vegetation shading in a regional stream temperature model using LiDAR data. Science of the Total Environment 642: 480-490.

Lowe, A., T. Selko, and E. P. Silvia. 2017. North Puget 2017: Western Washington 3DEP LiDAR technical data report. USGS and Quantum Spatial, Corvallis Oregon. Available at http://lidarportal.dnr.wa.gov (accessed 21 May 2018).

Lunetta, R. S., B. L. Cosentino, D. R. Montgomery, E. M. Beamer, and T. J. Beechie. 1997. GIS-Based evaluation of salmon habitat in the Pacific Northwest. Photogrammetric Engineering and Remote Sensing 63: 1219-1229.

Maudlin, M., T. Coe, N. Currence, and J. Hansen. 2002. South Fork Nooksack River Acme-Saxon reach restoration planning: Analysis of existing information and preliminary

recommendations. Available at http://salmonwria1.org/webfm_send/48 (accessed 31 May 2019).

Mazza, R., and D. Gatziolis. 2013. Demystifying lidar technologies for temperate rain forests in the Pacific Northwest. Pacific Northwest Research Station, USDA Forest Service.

McGaughey, R. 2018. FUSION/LDV: Software for LiDAR analysis and visualization, FUSION version 3.78. Available at https://forsys.cfr.washinton.edu.fusion.html/ (accessed May 26, 2018).

Mehtätalo, L., S. de Miguel, and T. Gregoire. 2015. Modeling height-diameter curves for prediction. Canadian Journal of Forest Research 45:826-837.

Meixer, M. S., and M. B. Bain. 2010. Landscape scale assessment of stream channel and riparian habitat restoration needs. Landscape and Ecological Engineering 6:235-245.

Meng, R., P. E. Dennison, F. Zhao, I. Shendryk, A. Rickert, R. P. Hanavan, B. D. Cook, and S. P. Serbin. 2018. Mapping canopy defoliation by herbivorous insects at the individual tree level using bi-temporal airborne imaging spectroscopy and LiDAR measurements. Remote Sensing of Environment 15:170-183.

Mollot, L. A., and R. E. Bilby. 2008. The use of geographic information system, remote sensing, and suitability modeling to identify conifer restoration sites with high biological potential for anadromous fish at the Cedar River municipal watershed in western Washington, USA. Restoration Ecology 16:336-347.

Mongus, D. and B Žalik. 2015. An efficient approach to 3D single tree-crown delineation in LiDAR data. ISPRS Journal of Photogrammetry and Remote Sensing 108:219-233.

Moskal, M. L., and A. Cooke. 2015. Feasibility of applying remote sensing to a riparian stand conditions assessment. Prepared for Washington Department of Natural Resources, Sedro Woolley, Washington.

Moskal, M. L., A. Cooke, T. Axe, and J. M. Comnick. 2017. Extensive riparian vegetation monitoring – Remote sensing pilot study, Agreement No. IAA 16-205. Prepared for Washington Department of Natural Resources, Sedro Woolley, Washington.

Naiman, R. J., R. E. Bilby, and P. A. Bisson. 2000. Riparian ecology and management in the Pacific coastal rain forest. BioScience 50:996-1011.

Naiman, R. J., and J. J. Latterell. 2005. Principles for linking fish habitat to fisheries management and conservation. Journal of Fish Biology 67:166-185.

Nakano, S., H. Miyasaka, and N. Kuhara. 1999. Terrestrial-aquatic linkages: riparian arthropod inputs alter trophic cascades in a stream food web. Ecology 80:2435-2441.

NNR (Nooksack Indian Tribe Natural Resources Department). 2015. South Fork Nooksack (Nuxw7iyem) Nesset Phase 1 Restoration. Salmon Recovery Funding Board Project #15-1283. Washington State Recreation and Conservation Office, Olympia, Washington.

NNR (Nooksack Indian Tribe Natural Resources Department). 2018. North Fork Nooksack (Xwqélém) Farmhouse Phase 4 Restoration. Salmon Recovery Funding Board Project #18-1685. Washington State Recreation and Conservation Office, Olympia, Washington.

NOAA. 2020. NOWData – NOAA Online Weather Data, Bellingham Intl., WA. National Weather Service Forecast Office, Seattle, WA. Available at https://w2.weather.gov/climate/xmacis.php?wfo=sew (accessed 18 June 2020).

North, M. P., J. T. Kane, V. R. Kane, G. P. Asner, W. Berigan, D. J. Churchill, S. Conway, R. J. Gutiérrez, S. Jeronimo, J. Keane, A. Koltunov, T. Mark, M. Moskal, T. Munton, Z.

Peery, C. Ramierez, R. Sollmann, A. White, and S. Whitmore. 2017. Cover of tall trees best predicts California spotted owl habitat. Forest Ecological Management 405:166-178.

Oregon Land Cover Standard. 2006. Oregon Land Cover Standard. Oregon Geographic Information Council. Available at https://www.oregon.gov/geo/standards/Oregon%20Land%20Cover%20Standard,%20v1.0.pdf. (accessed 29 July 2020).

Paulsen, C. M., and T. R. Fisher. 2001. Statistical relationship between parr-to-smolt survival of Snake River Spring-Summer Chinook Salmon and indices of land use. Transactions of the American Fisheries Society 130:347-358.

Pebesma, E. J. 2004. Multivariable geostatistics in S: the gstat package. Computers and Geosciences 30:683-691.

Pinheiro J., D. Bates, S. DebRoy, D. Sarkar, and the R Core Team (2020). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-144. Available at https://CRAN.R-project.org/package=nlme> (accessed 15 September 2020).

Pirotti, F., M. Kobal, and J. R. Roussel. 2017. A comparison of tree segmentation methods using very high density airborne laser scanning data. The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences, Vol. XLII-2/W7, 2017.

Puget Sound Partnership. 2018. 2018-2022 Action Agenda for Puget Sound Implementation Plan. Available at https://www.psp.wa.gov/action_agenda_center.php (accessed June 3, 2019).

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/ (accessed 30 January 2021).

Richardson, J. J., L. M. Moskal, and S. H. Kim. 2009. Modeling approaches to estimate effective leaf area index from aerial discrete return lidar. Agricultural and Forest Meteorology 149:1152-1160.

Richardson, J. J., and L. M. Moskal. 2011. Strengths and limitations of assessing forest density and spatial configuration with aerial LiDAR. Remote Sensing of Environment 115:2640-2651.

Richardson, J. J., and L. M. Moskal. 2016. An integrated approach for monitoring contemporary and recruitable large woody debris. Remote Sensing 8:778-791.

Richardson, J. J., C. E. Torgersen, and L. M. Moskal. 2019. Lidar-based approaches for estimating solar insolation in heavily forested streams. Hydrology and Earth System Sciences 23:2813-2822.

Roni, P., T. J. Beechie, R. E. Bilby, F. E. Leonetti, M. M. Pollock, and G. R. Pess. 2002. A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in Pacific Northwest watersheds. North American Journal of Fisheries Management 22:1-20.

Roni, P., P. J. Anders, T. J. Beechie, and D. J. Kaplowe. 2018. Review of tools for identifying, planning, and implementing habitat restoration for pacific salmon and steelhead. North American Journal of Fisheries Management 38:355-376.

Roussel, J. R., and D. Auty. 2020. lidR: Airborne LiDAR data manipulation and visualization for forestry applications. R package version 3.0.3. Available at https://CRAN.R-project.org/package=lidR (accessed 17 September 2020).

Roussel, J. R., T. R. H. Goodbody, and P. Tompalski. 2020. The lidR package. Available at https://jean-romain.github.io/lidRbook/ (accessed 18 September 2020).

Seixas, G. B., T. J. Beechie, C. Fogel, and P. M. Kiffney. 2018. Historical and future stream temperature change predicted by a lidar-based assessment of riparian condition and channel width. Journal of the American Water Resources Association 54:974-991.

Silva, C. A., A. T. Hudak, L. A. Vierling, E. L. Loudermilk, J. J. O'Brian, J. K. Hiers, S. B. Jack, C. Gonzalez-Benecke, H. Lee, M. J. Falkowski, and A. Khosravipour. 2016. Imputation of individual Longleaf Pine (*Pinus palustris* Mill.) tree attributes from field and LiDAR data. Canadian Journal of Remote Sensing 42:554-573.

Smith, C. J. 2002. Salmon and steelhead limiting factors in WRIA 1, the Nooksack Basin. Washington State Conservation Commission, Lacey, Washington.

Soicher, A., T. Coe, and N. Currence. 2006. South Fork Nooksack River Acme-Confluence reach restoration planning: Analysis of existing information and preliminary restoration strategies. Report to Salmon Recovery Funding Board (SRFB). Available at http://salmonwria1.org/webfm_send/52 (accessed 31 May 2019).

SRFB (Salmon Recovery Funding Board). 2020. 2020 SRFB Grant restoration and protection strategy matrices. Available at https://salmonwria1.org/sites/default/files/2019-09/2019%20SRFB%20Grant%20Restoration%20and%20Protection%20Strategy%20Matrices__February%2020-2019f.pdf (accessed 30 January 2021).

Strunk, J. L., S. E. Reutebuch, H. E. Anderson, P. J. Gould, and R. J. McGaughy. 2012. Model-assisted forest yield estimation with light detection and ranging. Western Journal of Applied Forestry 27:53-59.

Suttle, K. B., M. E. Power, J. M. Levine, and C. McNeely. 2004. How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. Ecological Applications 14:969-974.

Terrapoint. 2005a. Project Report: Nooksack North Fork – Nooksack Natural Resources, Contract #2309-H. Ottawa, Ontario. Available at https://lidarportal.dnr.wa.gov (accessed 11 September 2020).

Terrapoint 2005b. Project Report: Nooksack South Fork – Lummi Indian Nation, Contract #2291-H. Ottawa, Ontario. Available at https://lidarportal.dnr.wa.gov (accessed 11 September 2020).

Tompalski, P., N. C. Coops, J. C. White, M. A. Wulder, and A. Yuill. 2017. Characterizing streams and riparian areas with airborne laser scanning data. Remote Sensing of Environment 192:73-86.

Trimble. 2020. eCognition Developer 9.5. Available at https://geospatial.trimble.com/products-and-solutions/ecognition (accessed 30 January 2021).

USGS (United States Geologic Survey). 2018. National Hydrography Dataset. Available at https://www.usgs.gov/core-science-systems/ngp/national-hydrography (accessed 10 May 2018).

Vaughn, N. R., L. M. Moskal, and E. C. Turnblom. 2012. Tree species detection accuracies using discrete point lidar and airborne waveform lidar. Remote Sensing 4:377-403.

Vauhkonen, J. L. Ene, S. Gupta, J. Heinzel, J. Holmgren, J. Pitäkanen, S. Solberg, Y. Wang, H. Weinacker, K. M. Hauglin, V. Lien, P. Packalén, T. Gobakken, B. Koch, E. Næsset, T.

Tokola, and M. Maltamo. 2012. Comparative testing of single-tree detection algorithms under different types of forest. Forestry 85:27-40.

Vauhkonen, J., H. O. Orka, J. Holmgren, M. Dalponte, J. Heinzel, and B. Koch. 2014. "Tree species recognition based on airborne laser scanning and complementary data sources" in Forestry Applications of Airborne Laser Scanning. Ed. Maltamo, M., E. Næsset and J. Vauhkonen. Springer, Dordrechet.

WDFW (Washington Department of Fish and Wildlife). 2019. Salmonscape (web map). Available at: http://apps.wdfw.wa.gov/salmonscape/map.html (accessed 18 February 2019).

White, J. C., N. C. Coops, M. A. Wulder, M. Vastaranta, T. Hilker, and P. Tompalski. 2016. Remote sensing technologies for enhancing forest inventories: A review. Canadian Journal of Remote Sensing 42: 619-641.

WRIA 1 Salmon Recovery Board (SRB). 2005. WRIA 1 Salmonid Recovery Plan. Bellingham, WA. Available at: http://salmonwria1.org/webfm_send/23 (accessed 16 May 2019).

WSI (Watershed Sciences, Inc.) 2009. LiDAR Remote Sensing Data Collection: Mainstem, North, & South Nooksack Rivers, WA. Corvallis, Oregon. Available at https://lidarportal.dnr.wa.gov (accessed September 11, 2020).

WSI (Watershed Sciences, Inc.) 2013. Nooksack River Basin LiDAR Technical Data Report – Delivery 2. Watershed Sciences, Inc., Corvallis, Oregon. Available at: https://lidarportal.dnr.wa.gov (accessed 11 September 2020).

Wulder, M. A., J. C. White, R. F. Nelson, E. Næsset, H. O. Ørka, N. C. Coops, T. Hilker, C. W. Bater, and T. Gobakken. 2012. Lidar sampling for large-area forest characterization: A review. Remote Sensing of Environment 121:196-209.

Zlinszky, A., H. Heilmeier, H. Balzter, B. Czucz, and N. Pfeifer. 2015. Remote sensing and GIS for habitat quality monitoring: new approaches and future research. Remote Sensing 7:7987-7994.

5.0 - Appendices

5.1. Appendix A: k-means clustering on random vs. non-random ground truth plots

True random selection of all plot locations was impractical due to land access constraints and safety concerns. I used a cluster analysis to check whether this had introduced any detectable bias to the plot-level data. I used k-means clustering with 50 iterations to cluster ten plot-level variables into two groups. Input variables included measures such as average diameter, stem density, percent conifer, and basal area. Then, I compared the k-means-generated groups to the real-world group membership (random vs non-random location selection). If the method used to locate plots had an effect, then one would expect the k-means-generated groups to correspond to whether each plot location had been randomly selected or not.

K-means clustering did not show any difference between randomly selected plots and manually-selected locations. Of the ten variables that were assessed, percent dominant and percent conifer were the most different between k-means-generated groups. However, even along these variables there was no clear separation into two groups, and k-means group membership did not correspond to random vs non-random group membership (Figure 1-A).

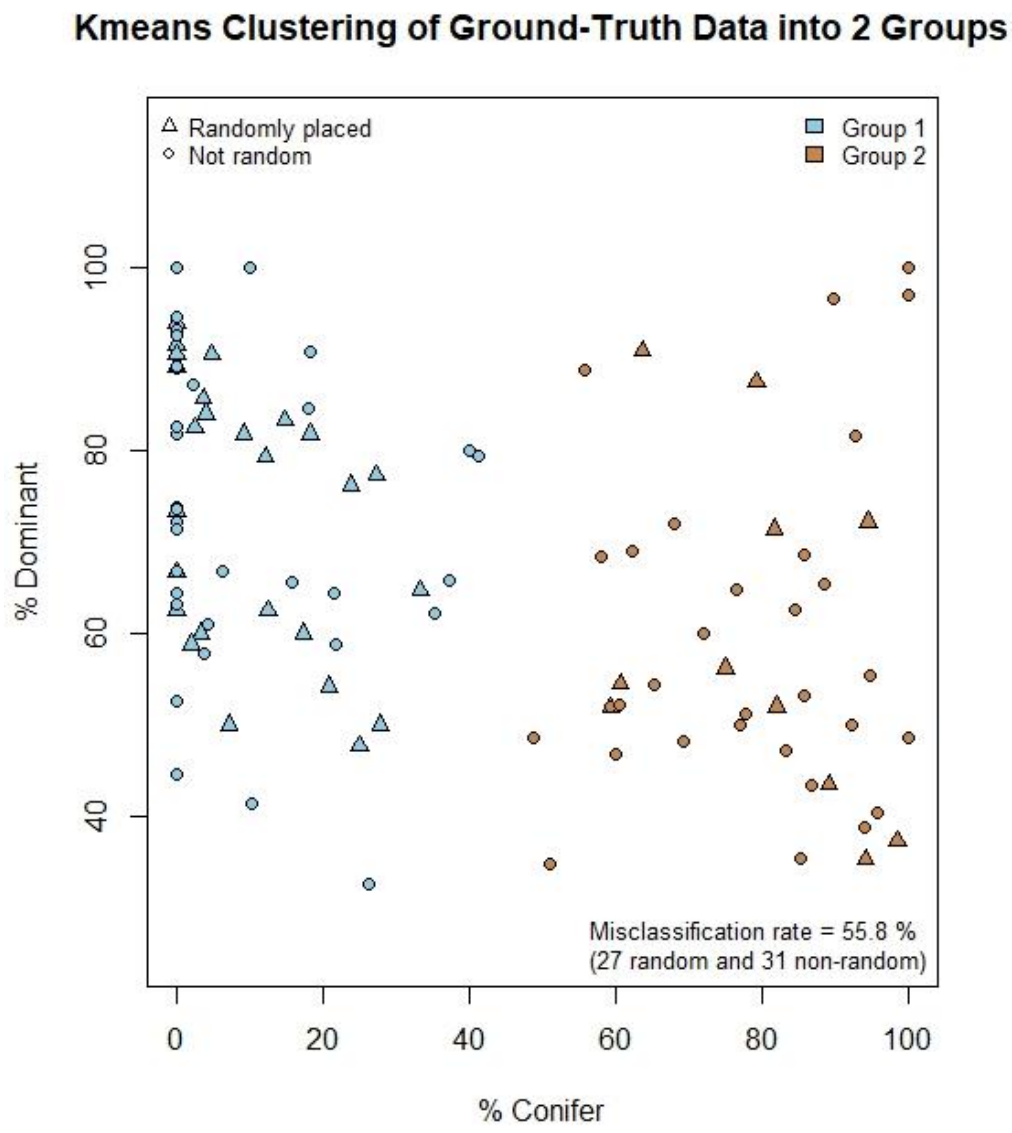

Figure A-1: Results of k-means clustering on plot-level variables (n=104). Plots were sorted into two stable clusters over 50 iterations. The k-means-generated clusters did not correspond to whether the plot location was random or not. If there had been an effect, most triangles in the plot above would have been blue and most circles would have been orange (or vice versa). The misclassification rate was extremely high, with the majority of the random plots misclassified.

5.2. Appendix B: Height/Diameter curves

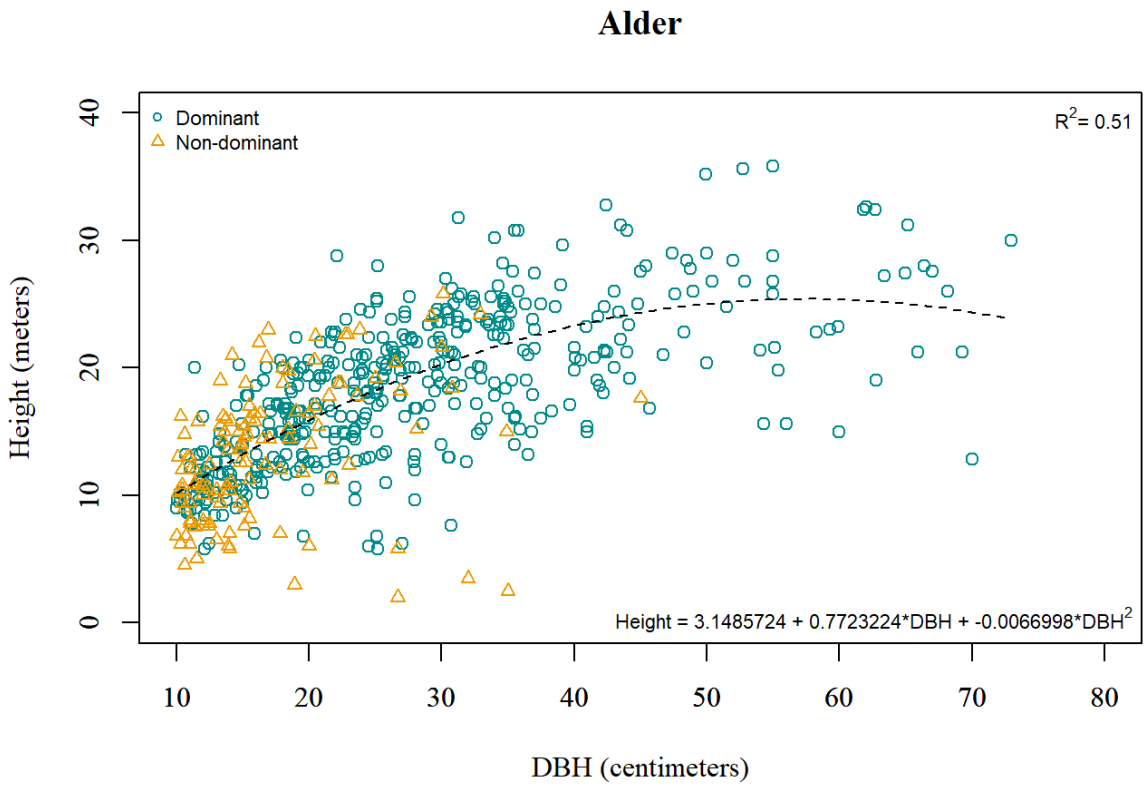

Figure B1: Height prediction curve for red alder (*Alnus Rubra).* n = 635 observations from 104 plots.

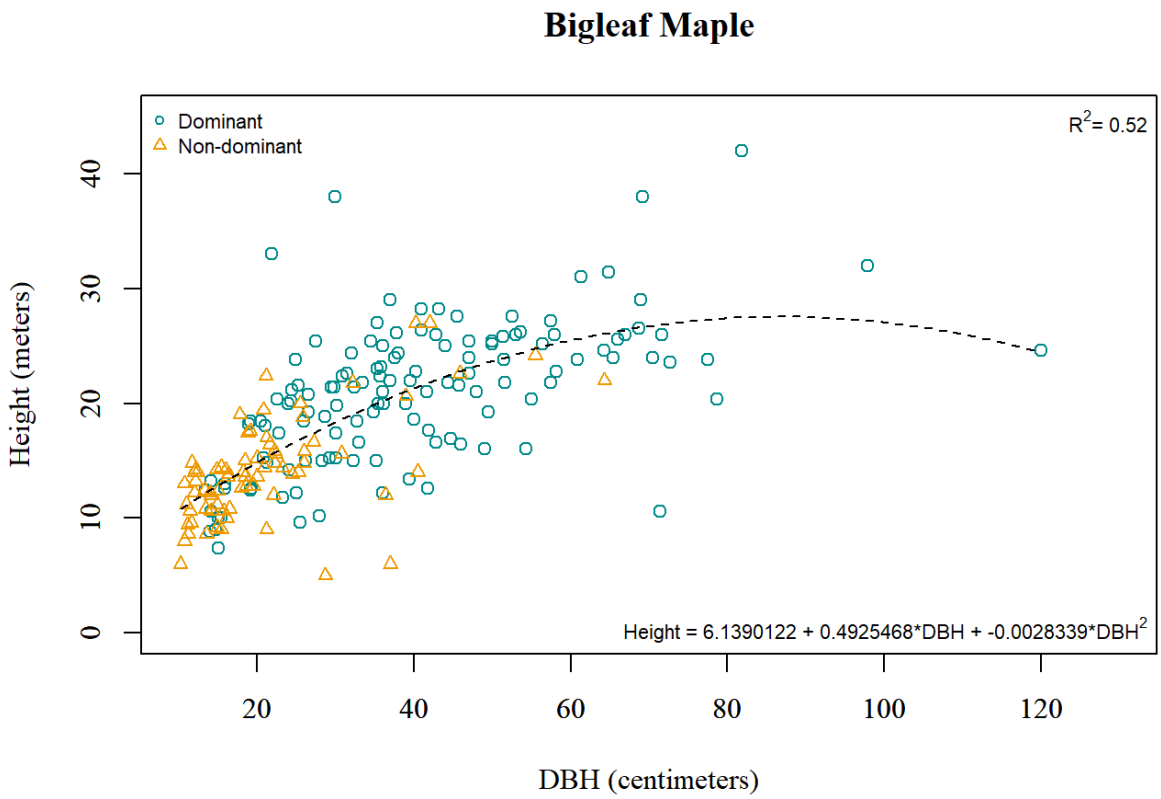

Figure B2: Height prediction curve for bigleaf maple (*Acer macrophyllum).* n = 209 observations from 104 plots.

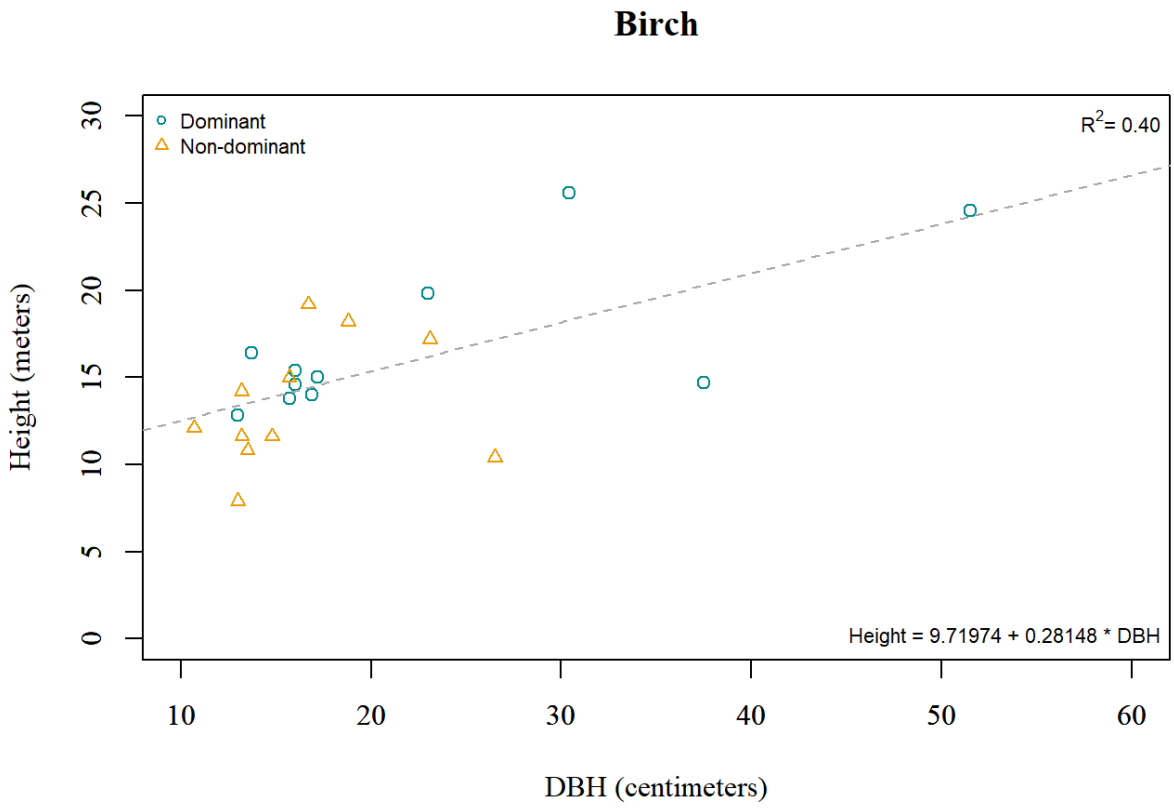

Figure B3: Height prediction line for paper birch (*Betula papyrifera*). n = 22 observations from 104 plots.

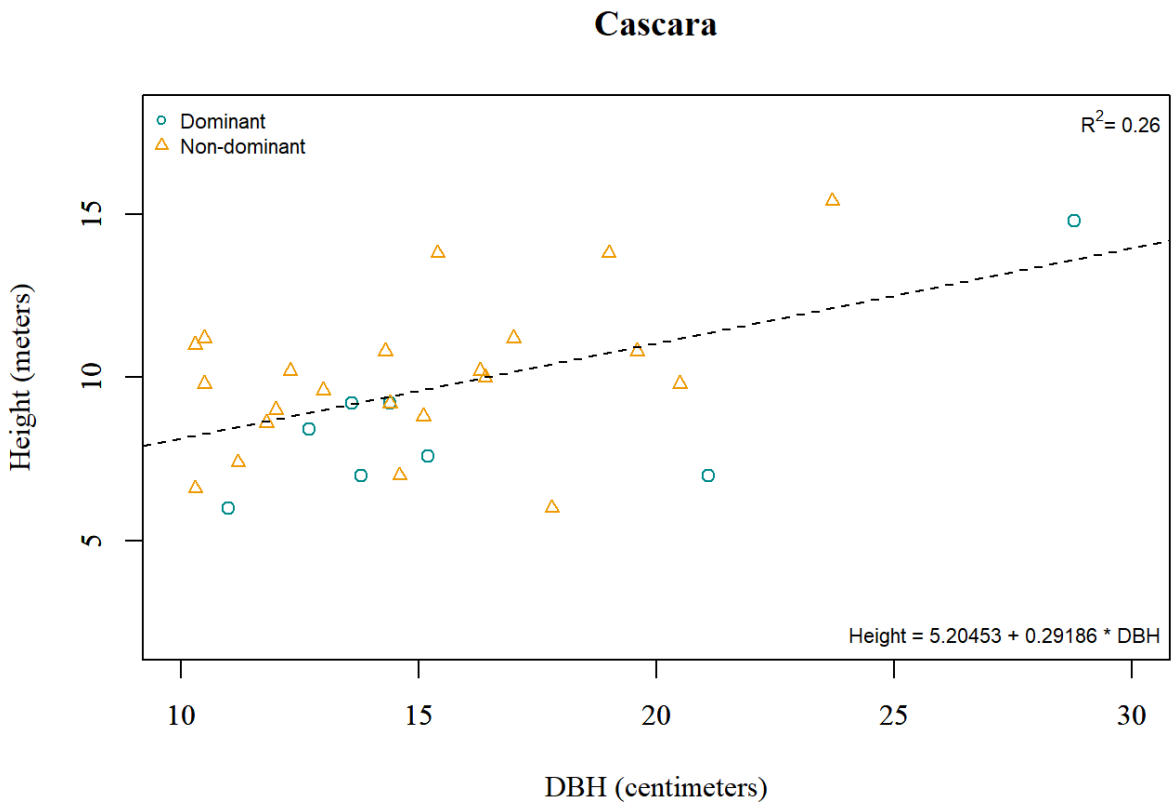

Figure B4: Height prediction line for cascara (*Rhamnus purshiana).* n = 31 observations from 104 plots.

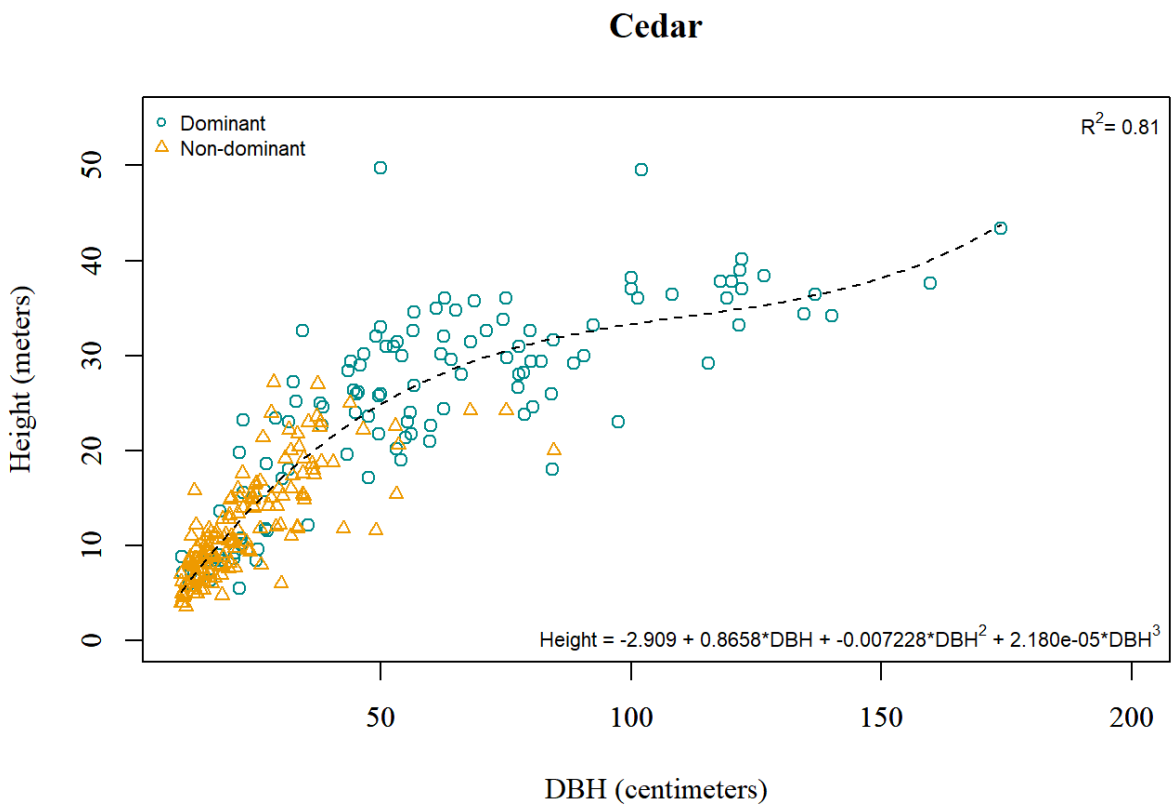

Figure B5: Height prediction curve for western redcedar (*Thuja plicata).* n = 283 observations from 104 plots.

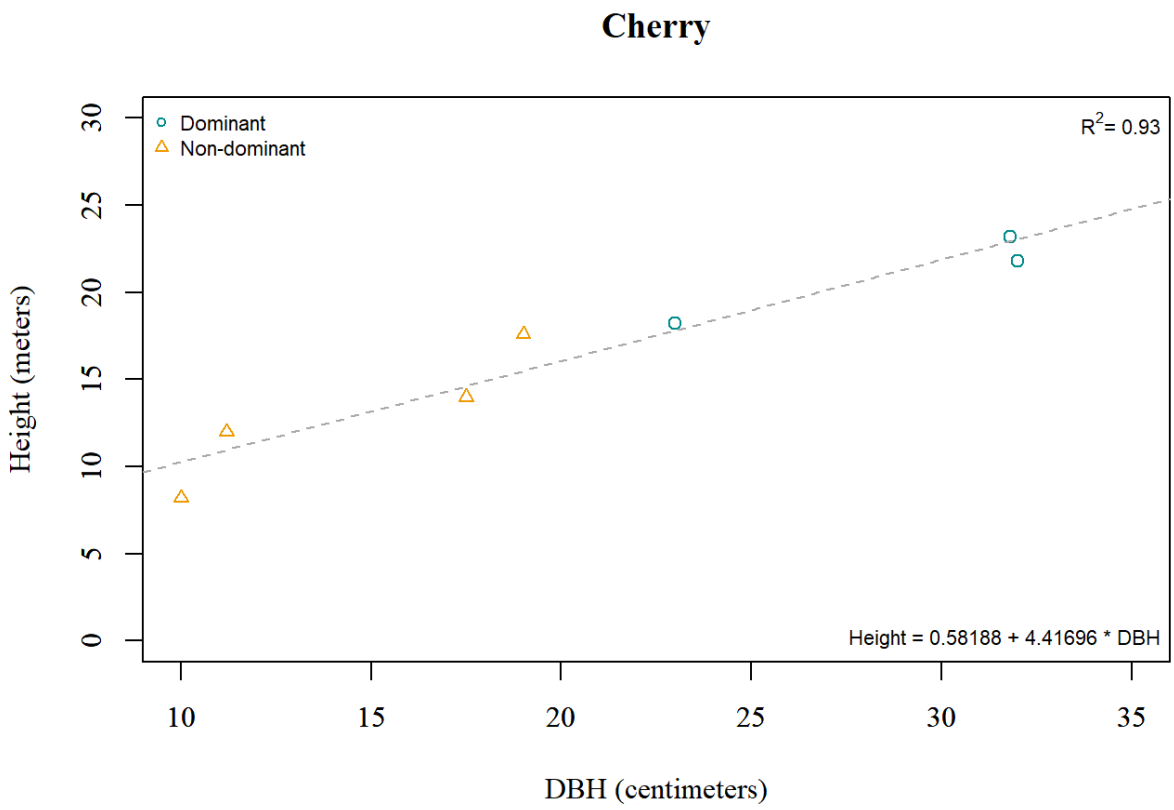

Figure B6: Height prediction curve for bitter cherry (*Prunus emarginata*). n = 8 observations from 104 plots.

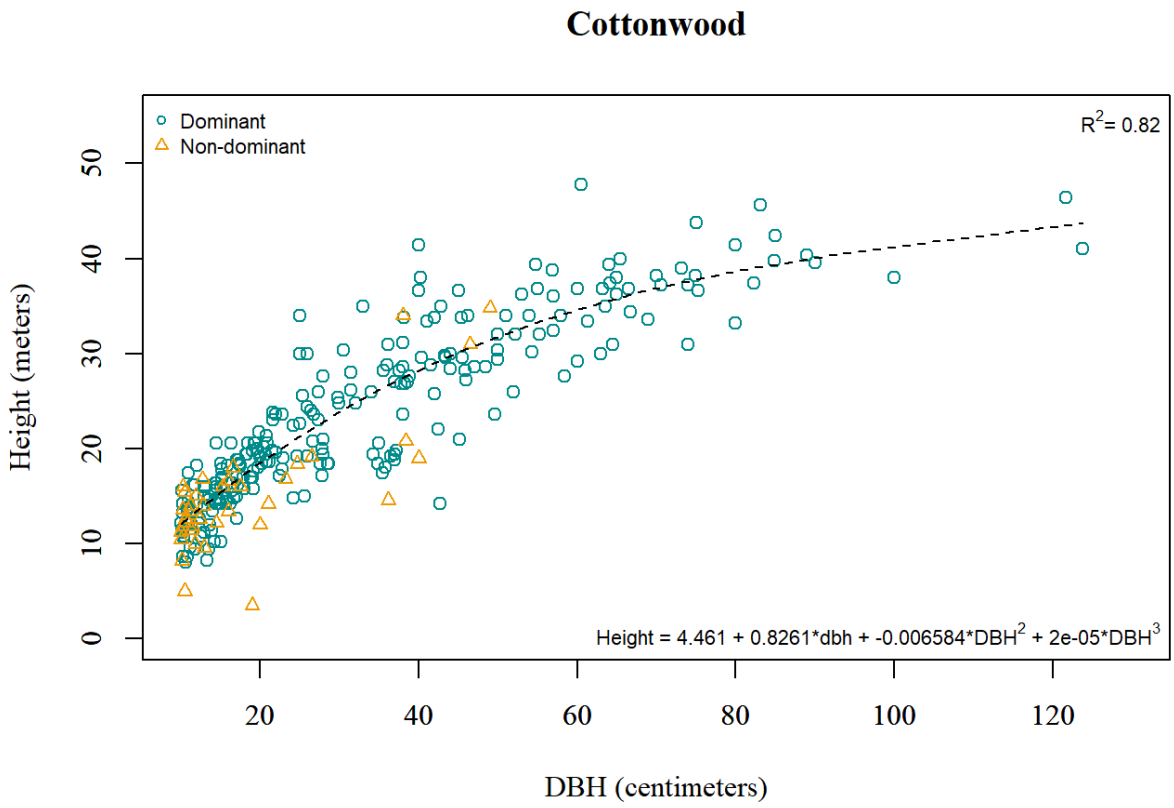

Figure B7: Height prediction curve for black cottonwood (*Populus balsamifera ssp. trichocarpa*). n = 297 observations from 104 plots.

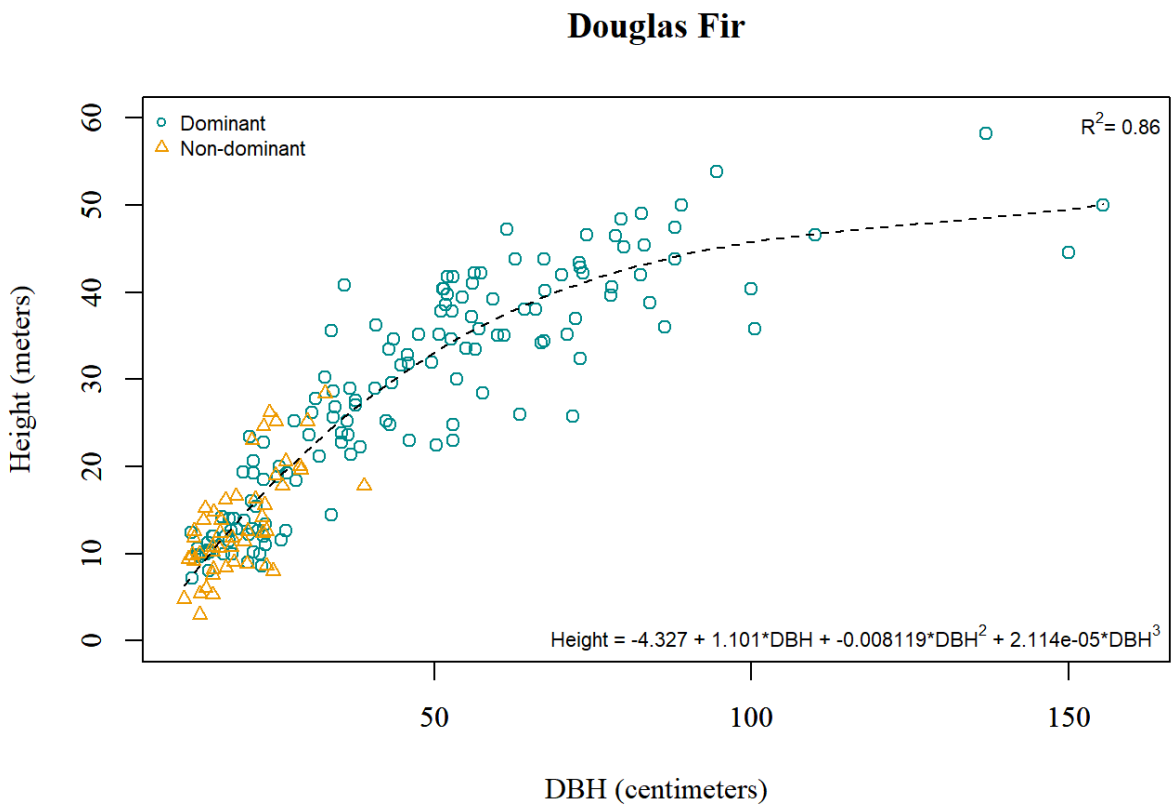

Figure B8: Height prediction curve for Douglas fir (*Pseudotsuga menziesii).* n = 201 observations from 104 plots.

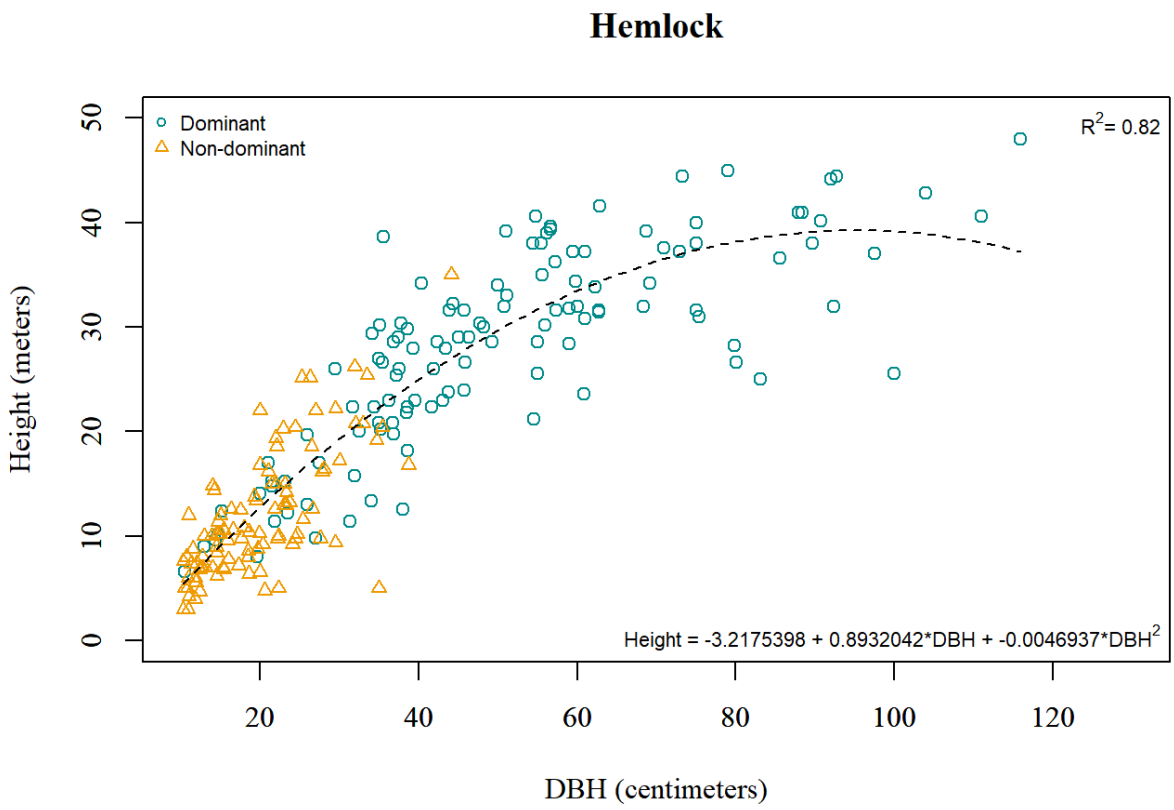

Figure B9: Height prediction curve for western hemlock (*Tsuga heterophylla*). n = 241 observations from 104 plots.

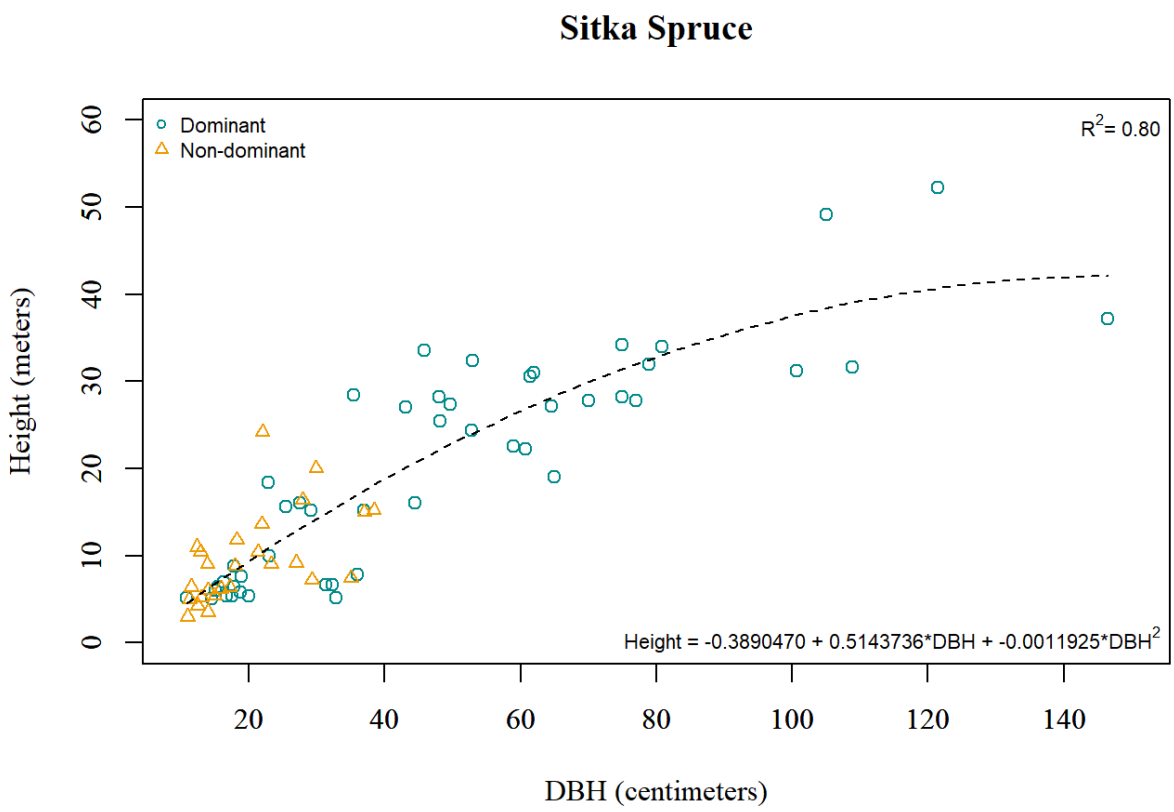

Figure B10: Height prediction curve for Sitka spruce (*Picea sitchensis).* n = 79 observations from 104 plots.

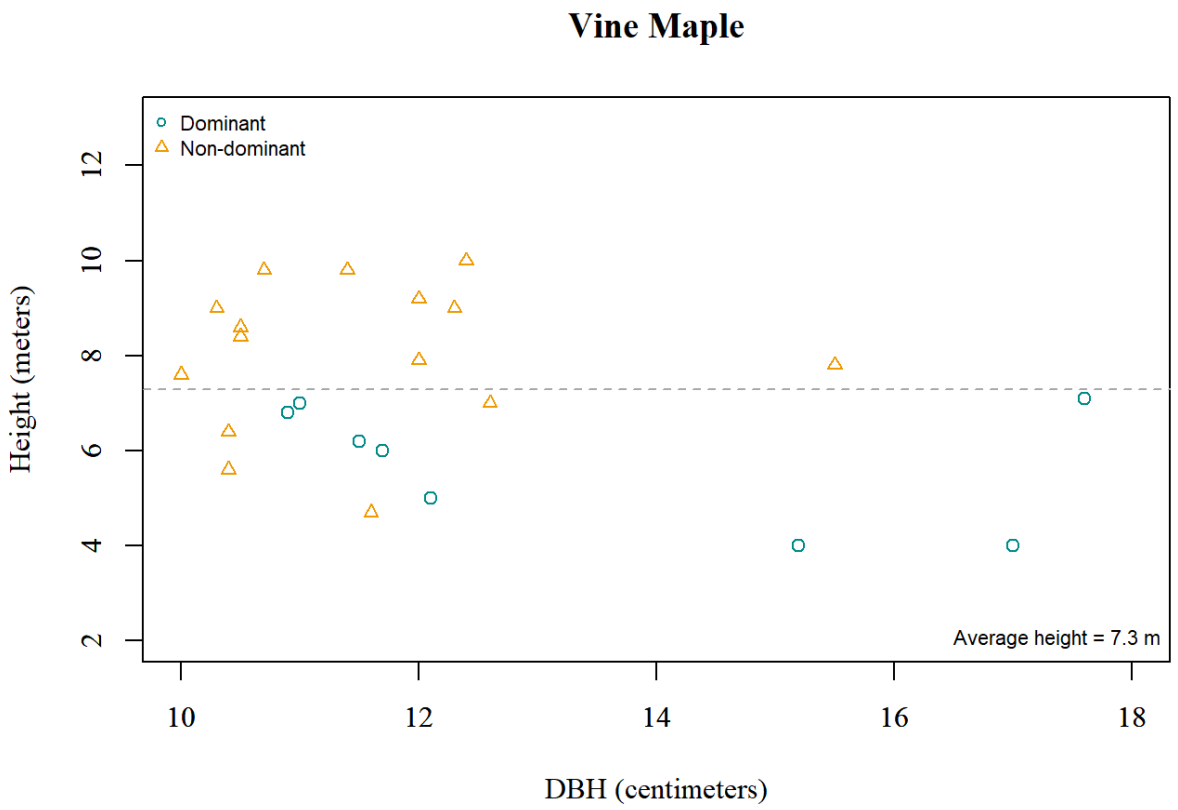

Figure B11: Mean height for vine maple (*Acer circinatum).* n = 23 observations from 104 plots. There was no significant relationship between height and diameter, and the mean height was used for all predictions.

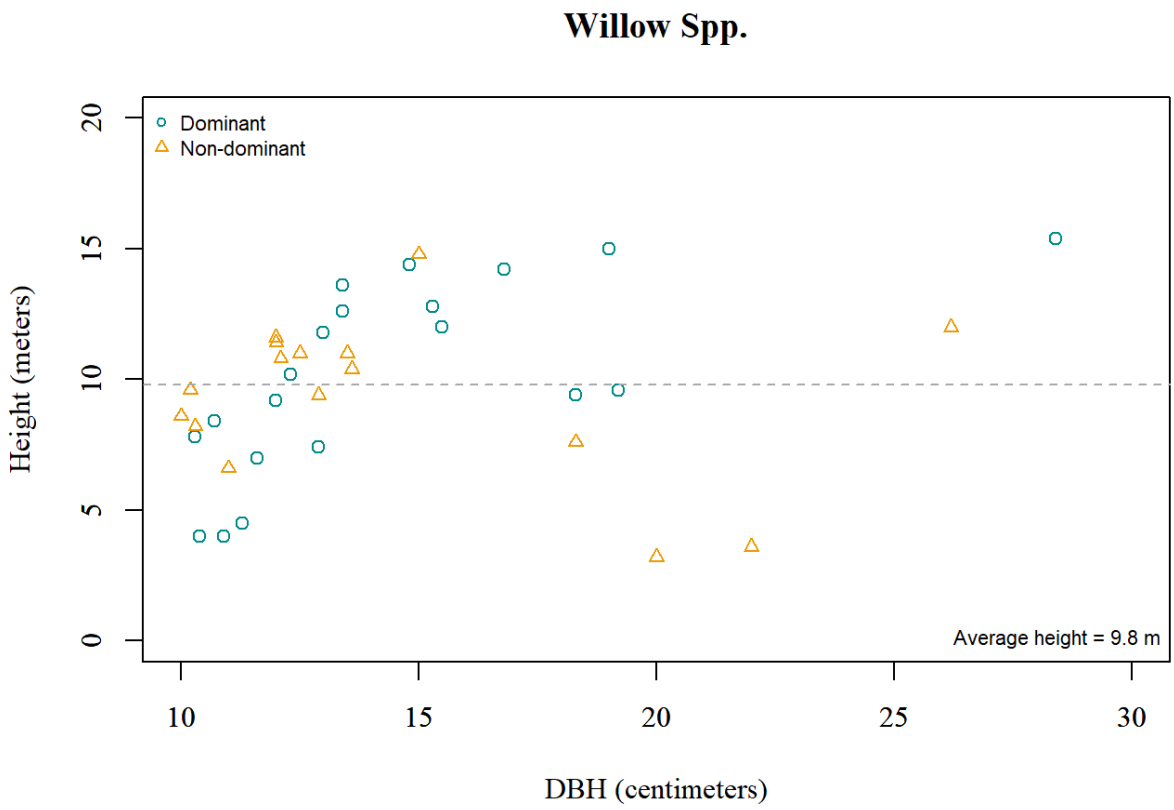

Figure B12: Height mean for willow (*Salix spp.).* n = 39 observations from 104 plots. There was no significant relationship between height and diameter, so the mean height was used for all predictions.

5.3. Appendix C: Watershed-scale shade model script for current conditions

I created a workflow (two scripts and some manual geoprocessing) that uses lidar-derived leaf area index to correct outputs from the Area Solar Radiation (Spatial Analyst) tool in ArcGIS Pro (Figure C1) for a limited area (it processes the river, but not the surrounding gravel bars). The output is a raster showing the estimated amount of solar radiation on the water surface taking into account both topographic shading and semi-transparent canopy shading. This 2-part script was written in Python 3.6 using ArcPy tools and is available for download from my github repository (https://github.com/Julia7/NooksackShade2).

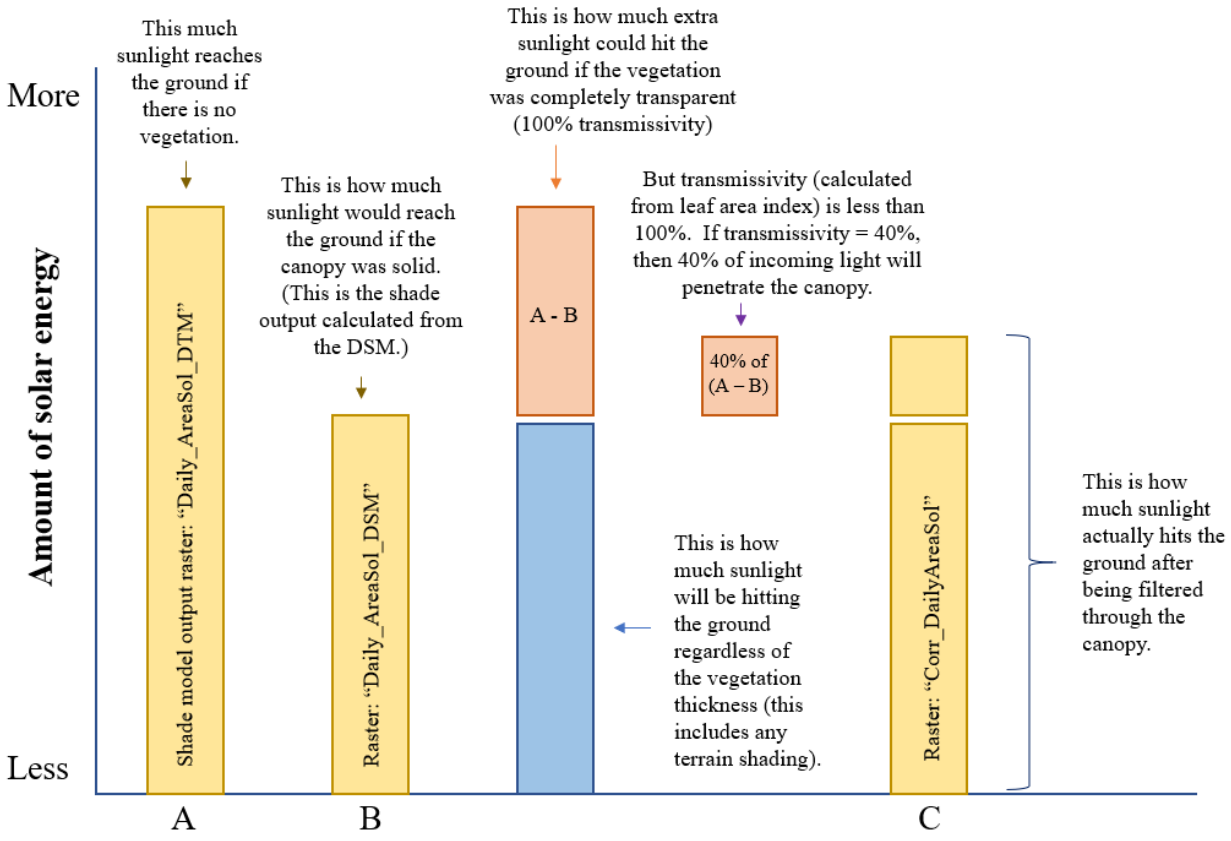

Figure C1: Conceptual process for correcting shade outputs using leaf area index. Raster A is the output from Esri's Area Solar Radiation tool run on the digital terrain model (bare ground), raster B is the output from the solar radiation tool run on a digital surface model (canopy surface), and raster C is the corrected output shade model.

Workflow Description:

For the inundated area of the river, part 1 of the script determines which direction the shade is coming from, calculates the mean leaf area index for the bank casting the shadows, and then estimates the proportion of sunlight that would reach the water surface based on the modeled transmissivity of the vegetation. The output is a series of partially overlapping tiff rasters in the GoodOutputs folder (Figure C2) and shapefiles in the ErrorOutputs folder. The files in the GoodOutputs folder need to be mosaiced together into a single raster, with the "mean" option used for the overlapping areas. The ErrorOutputs file contains any sub-reaches that the script was unable to process (this is usually related to complex geometry, which is a problem in braided channel areas). Once part 1 is finished, I would recommend merging the GoodOutputs in the GUI, then filling in any significant holes by hand (i.e. look at the leaf area index of the bank casting the deepest shadows, use equation 1 (from Richardson et al. 2009) to calculate the transmissivity of the canopy, and then assign the transmissivity value to the missing data area).

$$proportion\ of\ light\ transmitted = \frac{intensity\ below\ canopy}{intensity\ above\ canopy} = e^{-kL} \qquad \text{Eq.1}$$

where the extinction coefficient k = 0.47687, and L = leaf area index.

If you just want to add nodata values to all the Error areas, you can do this easily by adding a new field to the inundated area polygon ("RiverArea") filled with the nodata value, converting the inundated area polygon to raster, and then using the Mosaic to New Raster tool to combine it with the other rasters. Use the Mosaic Operator parameter to specify that the error

values should only be copied if the is nothing in the canopy transmissivity layer (with the good values).

Once you are satisfied with the coverage of the canopy transmissivity layer, save it as "canopy_transmissivity" in the GDB (Figure C2) (snap the cells to the "DTM_minus_DSM" layer) and then run Part 2 (which corrects the shade values).

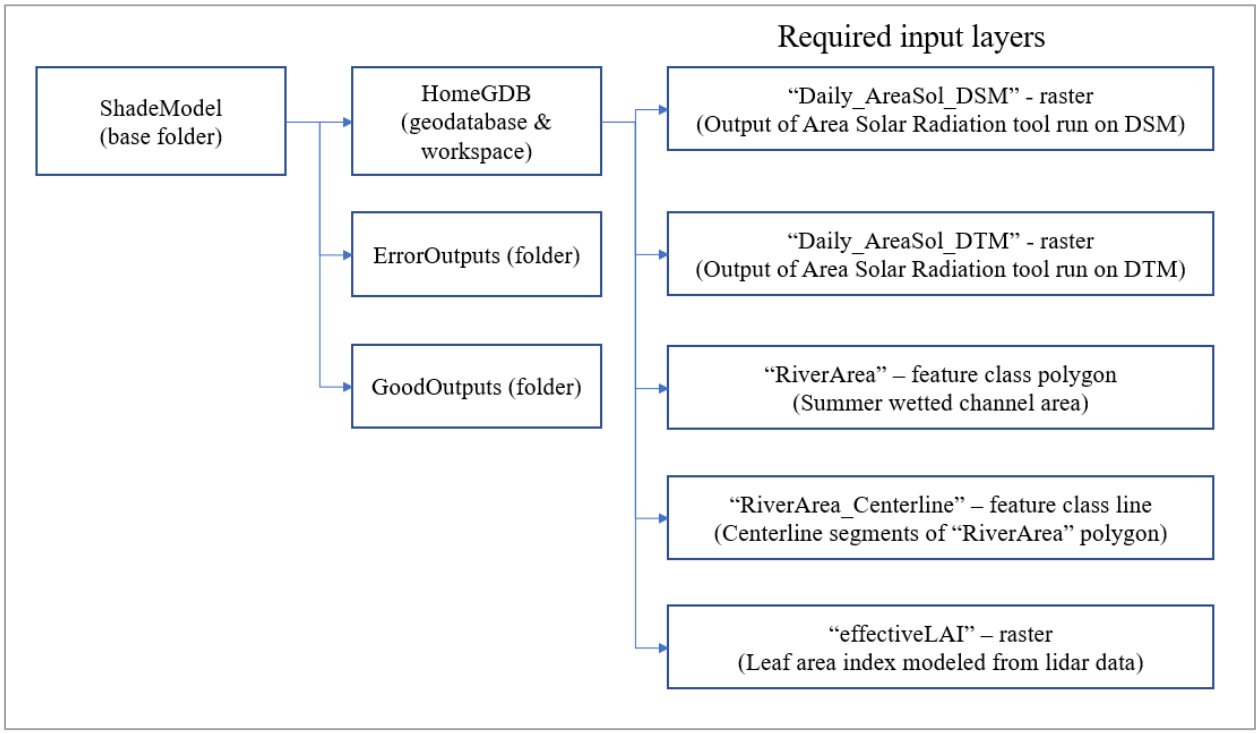

Figure C2: File hierarchy underlying scripts under starting conditions.

Limitations and appropriate use:

This script is intended to be used on watershed-scale analyses where a more detailed manual approach would be cost-prohibitive. If you are only working with one or two reaches, it is probably better to assign transmissivity modifiers by hand. This script is most likely to mis-

calculate transmissivity or give error values for sub-reaches with a north/south orientation and on sub-reaches that are very curved (U-shaped) relative to their width (usually these are short side channels).

Appendix D: Priority matrices and associated data

Table D1: Reach priority rankings, overall priority score, and location

Reach ID	**Reach center coordinates (NAD 1983 UTM 10N)**		**Priority rankings:**					**Overall score**	**Top scores**
	Easting	*Northing*	*Catchment size ranking*	*Impairment score*	*Shade*	*Proximity to good habitat*	*South fork?*		
0	561976	5409492	0.75	2	0.5	0	0	3.25	
71	565437	5386101	1	2	2	0	1	6	*
83	565698	5385809	1	2	2	0	1	6	*
84	565030	5388100	1	1	0	1	1	4	
90	565644	5386526	1	1	0	1	1	4	
97	564838	5387637	1	1	0	1	1	4	
103	577166	5416464	1	1.5	0	2	0	4.5	
107	566217	5385076	1	1.5	0.5	0	1	4	
110	558379	5396774	1	2	0	0	1	4	
112	566096	5385547	1	1.5	0	0	1	3.5	
115	558817	5404423	1	2	0	0	1	4	
117	566914	5384549	1	2	1	1	1	6	*
119	583511	5417311	1	2	0	2	0	5	
123	557567	5400488	1	2	1	1	1	6	*
125	562535	5412997	0.75	2	0	1	0	3.75	

Reach ID	**Reach center coordinates (NAD 1983 UTM 10N)**		**Priority rankings:**					**Overall score**	**Top scores**
	Easting	*Northing*	*Catchment size ranking*	*Impairment score*	*Shade*	*Proximity to good habitat*	*South fork?*		
129	558317	5403122	1	2	0	0	1	4	
130	557742	5401870	1	2	0	0	1	4	
132	558387	5397240	1	2	0	0	1	4	
133	566492	5384746	1	2	2.5	0	1	6.5	**
136	558811	5396564	1	2	1	1	1	6	*
141	558806	5403917	1	2	0	0	1	4	
143	558312	5407142	0.25	1	0	1	0	2.25	
145	557654	5401403	1	2	0	0	1	4	
146	567017	5417678	0.75	1	0	1	0	2.75	
147	557586	5400939	1	2	0	0	1	4	
148	561279	5392209	1	1	0	1	1	4	
149	572803	5418500	1	2	0.5	2	0	5.5	
150	551849	5410016	0.25	2	0	0	0	2.25	
152	558434	5397660	1	2	1	0	1	5	
153	558534	5403534	1	2	0	0	1	4	
154	559503	5395931	1	1	0.5	1	1	4.5	
155	572140	5419134	1	1.5	1.5	2	0	6	*
159	562203	5415084	0.75	2	0	1	0	3.75	

Reach ID	Reach center coordinates (NAD 1983 UTM 10N)		Priority rankings:					Overall score	Top scores
	Easting	*Northing*	*Catchment size ranking*	*Impairment score*	*Shade*	*Proximity to good habitat*	*South fork?*		
162	563037	5408798	1	1.5	1	0	0	3.5	
166	562184	5414181	0.75	2	0.5	1	0	4.25	
167	561242	5394020	1	1.5	0	1	1	4.5	
169	562678	5415940	0.75	2	1	1	0	4.75	
172	563229	5416607	0.75	2	0	0	0	2.75	
174	557972	5397824	1	2	1	0	1	5	
175	553239	5409195	0.25	2	0	0	0	2.25	
176	569907	5419562	0.75	1.5	0	2	0	4.25	
177	566556	5417601	0.75	1	0	0	0	1.75	
178	558809	5405179	1	1.5	0	1	1	4.5	
179	551641	5410505	2.5	1.5	0	0	0	4	
182	559058	5404921	1	2	0	0	1	4	
184	557775	5402370	1	1	0.5	0	1	3.5	
185	563472	5408597	1	2	2	1	0	6	*
186	553153	5409704	0.25	2	0	0	0	2.25	
187	573324	5417648	1	2	0.5	2	0	5.5	
190	548648	5415485	0.25	1.5	0.5	0	0	2.25	
191	567610	5418312	0.75	1.5	0	2	0	4.25	

Reach ID	Reach center coordinates (NAD 1983 UTM 10N)		Priority rankings:					Overall score	Top scores
	Easting	*Northing*	*Catchment size ranking*	*Impairment score*	*Shade*	*Proximity to good habitat*	*South fork?*		
192	557925	5398316	1	2	0	1	1	5	
193	552222	5409848	0.25	2	0	0	0	2.25	
194	564057	5405340	1	2	0.5	1	0	4.5	
195	562400	5412511	0.75	2	0	0	0	2.75	
196	557856	5399161	1	1.5	0	2	1	5.5	
197	560042	5395291	1	1.5	1	1	1	5.5	
198	563847	5406295	1	1.5	2	1	0	5.5	
199	552709	5409855	0.25	1.5	0	0	0	1.75	
200	558067	5402678	1	2	0	0	1	4	
201	562145	5414645	0.75	2	2	0	0	4.75	
202	563496	5416992	0.75	2	2	0	0	4.75	
203	559648	5395501	1	2	0	0	1	4	
205	561457	5393583	1	2	0.5	0	1	4.5	
208	563951	5417093	0.75	2	0	0	0	2.75	
209	564298	5404946	1	2	0.5	0	0	3.5	
211	563046	5416175	0.75	2	0.5	0	0	3.25	
212	554456	5408810	0.25	2	0	0	0	2.25	
213	564450	5417016	0.75	2	0	0	0	2.75	

Reach ID	Reach center coordinates (NAD 1983 UTM 10N)		Priority rankings:					Overall score	Top scores
	Easting	*Northing*	*Catchment size ranking*	*Impairment score*	*Shade*	*Proximity to good habitat*	*South fork?*		
214	564898	5417121	0.75	2	2	0	0	4.75	
215	553569	5408815	0.25	2	0.5	0	0	2.75	
216	561431	5393103	1	2	0	0	1	4	
217	555286	5408356	0.25	1.5	0	0	0	1.75	
218	563568	5407679	1	1.5	2	2	0	6.5	**
219	564913	5404276	1	1.5	0.5	1	0	4	
220	562466	5411545	0.75	2	0	0	0	2.75	
221	562034	5410171	0.75	2	0.5	0	0	3.25	
222	562348	5411064	0.75	2	0	0	0	2.75	
223	565788	5417143	0.75	2	1	0	0	3.75	
224	558895	5406666	0.5	2	0	0	0	2.5	
225	562189	5410606	0.75	2	0	0	0	2.75	
226	561239	5392658	1	2	0.5	0	1	4.5	
227	556647	5407922	0.25	2	0	0	0	2.25	
228	559630	5407167	0.5	2	0	0	0	2.5	
229	566204	5417374	0.75	1	0	0	0	1.75	
230	555789	5408306	0.25	1.5	0	0	0	1.75	
231	563746	5406760	1	1.5	0.5	1	0	4	

Reach ID	Reach center coordinates (NAD 1983 UTM 10N)		Priority rankings:					Overall score	Top scores
	Easting	*Northing*	*Catchment size ranking*	*Impairment score*	*Shade*	*Proximity to good habitat*	*South fork?*		
232	554835	5408455	0.25	1.5	0	0	0	1.75	
233	557558	5407680	0.25	1	0	1	0	2.25	
234	550348	5412445	0.25	2	0	0	0	2.25	
235	562716	5409180	1	2	0	0	0	3	
237	550661	5412078	0.25	2	0	1	0	3.25	
238	565334	5417365	0.75	2	2	0	0	4.75	
239	561636	5409024	0.5	2	0	0	0	2.5	
240	562447	5412043	0.75	2	0	0	0	2.75	
242	557102	5407694	0.25	2	0	0	0	2.25	
243	551377	5410844	0.25	1.5	0	1	0	2.75	
244	564622	5404586	1	1.5	0.5	0	0	3	
246	559328	5406801	0.5	2	0.5	0	0	3	
247	554015	5408941	0.25	2	1	0	0	3.25	
248	550065	5412869	0.25	2	0	1	0	3.25	
249	556245	5408137	0.25	1	0	0	0	1.25	
250	558411	5406677	0.25	1.5	0	0	0	1.75	
251	560329	5407765	0.5	2	0	0	0	2.5	
252	559860	5407609	0.5	2	2	0	0	4.5	

Reach ID	Reach center coordinates (NAD 1983 UTM 10N)		Priority rankings:					Overall score	Top scores
	Easting	*Northing*	*Catchment size ranking*	*Impairment score*	*Shade*	*Proximity to good habitat*	*South fork?*		
253	561151	5408207	0.5	2	1	0	0	3.5	
254	560771	5407952	0.5	2	0	0	0	2.5	
256	561414	5408604	0.5	2	0	0	0	2.5	
257	548950	5413711	0.25	2	0	1	0	3.25	
260	548701	5414593	0.25	2	0	0	0	2.25	
261	549093	5414251	0.25	2	0	0	0	2.25	
262	548665	5415108	0.25	2	0	0	0	2.25	
264	558648	5406220	1	1	0	1	1	4	
265	562362	5409382	1	2	2	0	0	5	

Table D2: Reference key for shade model code and priority scores

REFERENCE KEY

CODE	DESCRIPTION	Priority rank (0-3)
0	Moderate increase in shade; affects sub-reaches > 30 m wide	1
1	Substantial increase in shade; affects sub-reaches > 30 m wide	1.5
2	Substantial increase in shade; affects sub-reaches 15 - 30 m wide	2.5
3	Moderate increase in shade; affects sub-reaches 15 - 30 m wide	2
4	Moderate or substantial increase in shade; affects sub-reaches < 15 m wide	3
5	Small increase in shade; affects sub-reaches < 30 m wide	0.5
6	Small increase in shade affecting sub-reaches > 30m wide, or no effect	0

"Small effect"	< 5% area of the river has input energy reductions of ≥1 kWh
"Moderate effect"	5-10% area of the river has input energy reductions of ≥1 kWh
"Large effect"	> 10 % area of the river has input energy reductions of ≥1 kWh

Note: Widths refer to the section that is shaded, not necessarily the overall reach width

Table D3: Reach-level riparian condition, surface area, length, width, catchment area, percent shade increase, shade code (see table D2) and shade priority score for each reach.

Reach ID	Riparian condition	All river area	Main channel area	All river length	Main channel length	Main channel width	All channel width	Catchment area	Percent area shaded up to each cutoff		Shade Code	Shade priority score
	% "good" or above	*sq. meters*	*sq. meters*	*meters*	*meters*	*meters*	*meters*	*sq. km*	*Percent <1kWh*	*Percent ≥ 1kWh*		
0	45	63935	40523	3121	1011	40.1	20.5	746.0534	6.66	1.32	5	0.5
71	27	23570	11332	1565	465	24.4	15.1	209.6356	20.02	6.22	3	2
83	34	17527	14709	814	451	32.6	21.5	209.28	12.32	5.68	3	2
84	45	22700	22700	606	537	42.3	37.5	226.0497	1.13	0.00	6	0
90	40	17802	16478	748	527	31.2	23.8	216.2284	0.11	0.01	6	0
97	47	25315	24197	727	560	43.2	34.8	225.0989	0.00	0.00	6	0
103	41	27058	24321	944	656	37.1	28.7	390.3303	4.11	2.14	6	0
107	45	17270	17270	551	551	31.4	31.4	205.9964	25.60	0.35	5	0.5
110	39	26389	24484	826	508	48.2	31.9	396.8188	26.26	1.87	6	0
112	41	21979	21774	685	548	39.7	32.1	208.0495	24.94	1.34	6	0
115	9	27572	26328	630	521	50.6	43.8	464.5726	33.90	0.18	6	0
117	35	16397	12617	930	580	21.8	17.6	186.7729	22.49	6.34	0	1
119	42	18497	7243	1485	441	16.4	12.5	275.6403	0.00	0.00	6	0
123	23	22948	22948	919	499	46.0	25.0	421.2272	55.84	5.08	0	1
125	46	33020	30551	909	517	59.0	36.3	719.6134	0.14	0.05	6	0
129	24	24957	24957	511	511	48.8	48.8	438.4377	53.13	1.42	6	0
130	43	31904	28323	981	533	53.1	32.5	424.2987	3.97	0.01	6	0
132	25	27274	22509	1269	630	35.7	21.5	403.4258	32.32	3.84	6	0

Reach ID	Riparian condition	All river area	Main channel area	All river length	Main channel length	Main channel width	All channel width	Catchment area	Percent area shaded up to each cutoff		Shade Code	Shade priority score
	% "good" or above	*sq. meters*	*sq. meters*	*meters*	*meters*	*meters*	*meters*	*sq. km*	*Percent <1kWh*	*Percent ≥ 1kWh*		
133	46	24763	15238	1500	525	29.0	16.5	205.087	14.03	14.37	2	2.5
136	23	22481	20200	842	606	33.3	26.7	396.2993	20.34	7.99	0	1
141	38	25262	25262	646	544	46.4	39.1	440.4764	15.74	0.04	6	0
143	32	33643	29719	932	527	56.4	36.1	1499.6069	6.03	0.00	6	0
145	18	30839	30009	734	549	54.6	42.0	424.2581	27.45	0.17	6	0
146	47	16641	16363	667	540	30.3	25.0	589.715	3.32	0.66	6	0
147	25	22795	22795	732	577	39.5	31.2	421.9435	44.25	0.84	6	0
148	46	26630	25000	663	520	48.1	40.2	332.1962	4.56	2.92	6	0
149	43	29236	21515	1590	772	27.9	18.4	506.9295	34.99	1.61	5	0.5
150	46	38143	38143	532	532	71.7	71.7	1528.9346	2.94	0.62	6	0
152	23	28192	20454	1230	704	29.1	22.9	406.0294	21.99	6.75	0	1
153	41	25035	25035	613	567	44.1	40.8	438.5922	25.91	2.28	6	0
154	43	23452	23452	572	572	41.0	41.0	388.9097	30.24	4.23	5	0.5
155	49	20686	17212	840	504	34.1	24.6	515.7075	22.85	15.65	1	1.5
159	47	41423	20768	2106	533	39.0	19.7	711.3443	4.63	0.84	6	0
162	43	21851	15931	1024	541	29.4	21.3	235.7069	10.82	5.95	0	1
166	45	31406	17387	1689	564	30.8	18.6	714.2974	11.20	0.05	5	0.5
167	46	24331	20474	799	470	43.6	30.4	341.027	6.81	1.02	6	0
169	25	28728	15864	1541	376	42.1	18.6	606.4231	17.69	7.20	0	1
172	31	44286	15812	3286	557	28.4	13.5	599.9717	1.40	0.05	6	0

Reach ID	Riparian condition	All river area	Main channel area	All river length	Main channel length	Main channel width	All channel width	Catchment area	Percent area shaded up to each cutoff		Shade Code	Shade priority score
	% "good" or above	*sq. meters*	*sq. meters*	*meters*	*meters*	*meters*	*meters*	*sq. km*	*Percent <1kWh*	*Percent ≥ 1kWh*		
174	40	30613	28521	881	625	45.6	34.8	408.4944	17.67	7.27	0	1
175	45	37804	36767	708	557	66.1	53.4	1523.1607	5.41	0.26	6	0
176	48	16962	14257	829	512	27.9	20.5	540.3606	0.29	0.00	6	0
177	46	21618	19471	834	465	41.9	25.9	591.6282	3.41	1.42	6	0
178	25	36660	33635	1146	804	41.9	32.0	469.5598	22.84	3.81	6	0
179	40	50341	46565	1065	487	95.6	47.3	1529.334	0.93	0.06	6	0
182	18	44735	37051	1297	615	60.3	34.5	465.7474	23.44	0.08	6	0
184	45	26810	26810	699	679	39.5	38.4	424.4423	48.08	0.03	5	0.5
185	17	30183	14056	1298	519	27.1	23.3	231.7136	34.65	7.92	3	2
186	34	47450	47450	586	586	81.0	81.0	1524.3364	0.00	0.00	6	0
187	45	33248	18673	2007	679	27.5	16.6	501.2258	9.47	2.32	5	0.5
190	22	40758	18064	1937	393	46.0	21.0	1603.3935	19.14	3.37	5	0.5
191	42	20063	19858	608	531	37.4	33.0	556.9893	0.86	0.00	6	0
192	40	33411	29583	1118	666	44.4	29.9	413.1627	6.52	0.44	6	0
193	38	41464	41464	618	521	79.6	67.0	1525.2613	3.07	3.90	6	0
194	41	17745	11857	1437	535	22.2	12.3	221.3938	20.25	4.07	5	0.5
195	40	35965	25725	1436	588	43.7	25.0	719.749	3.47	0.01	6	0
196	42	31376	27640	1053	608	45.5	29.8	415.7357	12.00	0.11	6	0
197	47	20664	18920	583	472	40.1	35.4	387.2954	13.54	7.23	0	1
198	35	26862	10331	1804	529	19.5	14.9	225.1051	26.11	6.13	3	2

Reach ID	Riparian condition	All river area	Main channel area	All river length	Main channel length	Main channel width	All channel width	Catchment area	Percent area shaded up to each cutoff		Shade Code	Shade priority score
	% "good" or above	*sq. meters*	*sq. meters*	*meters*	*meters*	*meters*	*meters*	*sq. km*	*Percent <1kWh*	*Percent ≥ 1kWh*		
199	24	36043	36043	505	505	71.4	71.4	1524.4605	2.39	1.94	6	0
200	33	24871	22552	922	790	28.5	27.0	432.0986	16.99	0.12	6	0
201	36	37081	20654	2035	519	39.8	18.2	714.2831	18.07	5.26	3	2
202	32	31963	16972	2055	470	36.1	15.6	599.6583	10.80	8.96	3	2
203	28	34719	28579	1474	795	36.0	23.5	387.7842	13.89	1.66	6	0
205	41	23823	20384	986	618	33.0	24.2	335.2612	17.03	2.21	5	0.5
208	41	50995	18026	3803	617	29.2	13.4	599.1463	0.79	0.81	6	0
209	39	20408	14153	1401	602	23.5	14.6	209.0683	14.89	3.05	5	0.5
211	20	46982	16780	3347	625	26.9	14.0	606.2734	7.44	3.04	5	0.5
212	41	41466	35768	893	537	66.6	46.5	1518.7245	7.96	1.94	6	0
213	31	25712	19227	1297	551	34.9	19.8	598.6564	6.27	3.50	6	0
214	31	34465	18111	2040	580	31.2	16.9	598.1773	11.34	5.82	3	2
215	16	39727	30483	1288	602	50.6	30.9	1522.4753	16.61	3.49	5	0.5
216	35	25004	20491	1129	590	34.7	22.2	333.0381	0.71	0.00	6	0
217	39	41545	41545	508	508	81.7	81.7	1517.0102	5.53	4.00	6	0
218	33	27012	16221	1718	663	24.5	15.7	228.5976	40.08	9.88	3	2
219	46	21214	14620	1099	560	26.1	19.3	196.0929	10.81	0.73	5	0.5
220	36	40723	20256	2329	491	41.3	17.5	720.684	6.60	1.02	6	0
221	36	33002	12853	1878	466	27.6	17.6	731.8751	14.47	2.05	5	0.5
222	23	41381	21482	1494	593	36.2	27.7	729.7949	0.38	0.00	6	0

Reach ID	Riparian condition	All river area	Main channel area	All river length	Main channel length	Main channel width	All channel width	Catchment area	Percent area shaded up to each cutoff		Shade Code	Shade priority score
	% "good" or above	*sq. meters*	*sq. meters*	*meters*	*meters*	*meters*	*meters*	*sq. km*	*Percent <1kWh*	*Percent ≥ 1kWh*		
223	22	38957	27618	1911	698	39.6	20.4	592.6235	15.29	6.76	0	1
224	38	39593	26133	1437	585	44.6	27.5	1020.7987	4.12	1.21	6	0
225	43	26684	19245	1059	537	35.9	25.2	731.6644	2.10	0.00	6	0
226	33	26835	18814	1283	598	31.4	20.9	332.5021	14.87	3.12	5	0.5
227	30	53470	41237	2036	701	58.8	26.3	1503.7797	0.14	0.03	6	0
228	31	22365	18282	760	454	40.3	29.4	1018.9463	0.00	0.00	6	0
229	36	19933	19346	606	476	40.7	32.9	591.716	3.95	2.12	6	0
230	32	37106	36118	667	587	61.5	55.6	1516.4647	4.45	0.01	6	0
231	48	29153	11423	2237	549	20.8	13.0	225.4178	38.74	3.48	5	0.5
232	30	38922	36315	962	542	67.0	40.5	1518.2769	0.00	0.00	6	0
233	31	34795	34795	458	458	75.9	75.9	1501.0367	4.06	4.93	6	0
234	41	61627	21810	2205	522	41.8	28.0	1562.2189	1.70	0.01	6	0
235	48	16042	13887	822	522	26.6	19.5	259.3208	0.12	0.00	6	0
237	17	72108	44577	2071	523	85.2	34.8	1558.6865	5.02	1.48	6	0
238	32	26556	9473	1506	366	25.9	17.6	597.6264	6.38	5.40	3	2
239	44	52963	25331	2031	574	44.1	26.1	1006.7838	2.64	1.83	6	0
240	34	41059	27684	2304	582	47.5	17.8	720.679	3.49	0.43	6	0
242	29	76861	39397	2380	615	64.0	32.3	1503.3946	2.02	1.00	6	0
243	47	28502	26914	875	462	58.2	32.6	1529.4236	5.68	1.47	6	0
244	44	14691	13340	756	521	25.6	19.4	208.0249	20.33	1.83	5	0.5

Reach ID	Riparian condition	All river area	Main channel area	All river length	Main channel length	Main channel width	All channel width	Catchment area	Percent area shaded up to each cutoff		Shade Code	Shade priority score
	% "good" or above	*sq. meters*	*sq. meters*	*meters*	*meters*	*meters*	*meters*	*sq. km*	*Percent <1kWh*	*Percent ≥ 1kWh*		
246	30	48507	30429	2479	616	49.4	19.6	1020.3365	12.86	1.73	5	0.5
247	18	36777	25328	1540	623	40.6	23.9	1519.4735	7.30	5.78	0	1
248	47	66214	53966	1494	716	75.4	44.3	1562.3539	1.43	0.00	6	0
249	45	41208	41012	887	674	60.9	46.4	1515.8241	0.00	0.00	6	0
250	45	37874	37511	772	727	51.6	49.1	1498.9515	2.60	0.00	6	0
251	27	49441	26702	2037	574	46.5	24.3	1017.0592	0.67	0.13	6	0
252	26	39316	25218	2025	510	49.5	19.4	1017.4361	7.81	5.42	3	2
253	38	38290	23301	1279	520	44.8	29.9	1008.1756	4.60	0.37	5	1
254	41	46741	19270	2069	495	38.9	22.6	1012.5538	0.48	0.24	6	0
256	33	37341	28204	1143	529	53.3	32.7	1006.9979	7.25	0.02	6	0
257	47	48352	32710	1748	488	67.0	27.7	1563.8003	4.05	0.24	6	0
260	28	77963	39978	2964	809	49.4	26.3	1602.0589	1.78	0.56	6	0
261	38	79395	34569	2583	742	46.6	30.7	1564.2383	2.09	0.81	6	0
262	17	82567	39516	2826	680	58.1	29.2	1603.2248	4.83	0.83	6	0
264	46	27027	25095	770	623	40.3	35.1	475.0034	1.83	2.28	6	0
265	42	23266	8074	1820	416	19.4	12.8	260.2887	9.75	8.78	3	2

Table D4: Anthropogenic influence over channel, determined based on basic satellite imagery (Esri...2020). The "main factor" and "secondary factor" columns are the main and secondary apparent causes of the low RCI score for these reaches.

Reach ID	***Channel***			**Roads or RRDs**				**Nearest Impact**				**MAIN FACTOR**					**SECONDARY FACTOR**						**Impairment score**
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
0			x	x				x				x							x				2
71		x					x				x	x										x	2
83		x					x				x	x										x	2
84	x						x				x		x									x	1
90	x					x		x					x									x	1
97	x						x				x		x									x	1
103	x						x				x	x										x	1.5
107	x						x				x	x										x	1.5
110	x			x				x						x						x			2
112	x						x				x	x										x	1.5
115	x			x				x						x						x			2
117		x				x		x				x										x	2
119			x		x					x		x										x	2
123	x					x		x						x								x	2
125	x			x				x						x								x	2

Reach ID	*Channel*			Roads or RRDs				Nearest Impact				MAIN FACTOR					SECONDARY FACTOR						Impairment score
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
129	x				x			x						x				x					2
130	x						x	x						x			x						2
132		x		x				x						x			x						2
133			x				x				x	x										x	2
136		x		x				x						x						x			2
141	x						x	x						x				x					2
143	x			x				x							x			x					1
145	x					x		x						x				x					2
146	x						x				x		x									x	1
147	x				x			x						x								x	2
148	x			x				x					x							x			1
149		x		x				x				x							x				2
150	x			x				x						x			x						2
152		x		x				x						x			x						2
153	x					x		x						x				x					2
154	x					x		x					x						x				1

Reach ID	*Channel*			Roads or RRDs				Nearest Impact				MAIN FACTOR					SECONDARY FACTOR						Impairment score
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
155		x			x				x				x						x				1.5
159			x	x				x				x								x			2
162		x		x				x					x							x			1.5
166			x			x			x			x										x	2
167		x				x				x			x									x	1.5
169			x				x	x				x							x				2
172			x				x	x				x							x				2
174	x					x		x						x								x	2
175	x						x	x						x			x						2
176		x					x		x				x						x				1.5
177	x						x				x		x									x	1
178	x			x				x				x							x				1.5
179	x					x		x				x							x				1.5
182	x			x				x						x			x						2
184	x					x				x			x									x	1
185		x			x				x			x							x				2

Reach ID	*Channel*			Roads or RRDs				Nearest Impact				MAIN FACTOR					SECONDARY FACTOR						Impairment score
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
186	x					x		x						x				x					2
187			x		x				x			x										x	2
190			x				x			x			x									x	1.5
191	x						x				x	x										x	1.5
192		x				x		x						x				x					2
193	x						x	x						x			x						2
194			x				x				x	x										x	2
195			x	x				x				x							x				2
196		x					x	x					x						x				1.5
197		x					x		x				x						x				1.5
198			x			x		x					x						x				1.5
199	x					x		x				x							x				1.5
200	x					x		x						x			x						2
201			x	x				x				x							x				2
202			x			x		x				x							x				2
203		x				x		x						x			x						2

Reach ID	*Channel*			Roads or RRDs				Nearest Impact				MAIN FACTOR					SECONDARY FACTOR						Impairment score
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
205		x					x				x	x										x	2
208			x			x		x				x							x				2
209		x					x				x	x										x	2
211			x			x		x				x							x				2
212		x					x				x	x										x	2
213			x				x				x	x										x	2
214			x				x				x	x										x	2
215		x				x		x				x										x	2
216		x					x				x	x										x	2
217	x						x				x	x										x	1.5
218			x				x	x					x						x				1.5
219		x					x		x				x									x	1.5
220			x		x			x				x							x				2
221			x	x				x				x							x				2
222		x				x		x				x							x				2
223			x				x				x	x										x	2

Reach ID	*Channel*			Roads or RRDs				Nearest Impact				MAIN FACTOR					SECONDARY FACTOR						Impairment score
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
224			x	x				x				x								x			2
225		x					x			x		x							x				2
226		x					x	x				x							x				2
227		x				x			x			x							x				2
228		x				x		x				x										x	2
229	x						x				x		x									x	1
230	x						x				x	x										x	1.5
231			x				x				x		x									x	1.5
232	x					x				x		x										x	1.5
233	x					x				x			x						x				1
234			x			x		x				x							x				2
235		x				x		x				x							x				2
237		x				x		x				x							x				2
238			x				x				x	x										x	2
239			x	x				x				x										x	2
240			x		x			x				x							x				2

Reach ID	*Channel*			Roads or RRDs				Nearest Impact				MAIN FACTOR					SECONDARY FACTOR						Impairment score
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag, fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
242			x			x		x				x							x				2
243	x					x		x				x							x				1.5
244	x						x				x	x										x	1.5
246			x	x				x				x							x				2
247		x				x					x	x							x				2
248		x				x		x				x							x				2
249	x						x			x			x						x				1
250	x			x				x				x								x			1.5
251			x				x				x	x										x	2
252			x				x			x		x							x				2
253			x				x	x				x							x				2
254			x				x			x		x							x				2
256		x			x				x			x							x				2
257			x				x			x		x							x				2
260			x				x			x		x							x				2
261			x			x			x			x							x				2

Reach ID	*Channel*			Roads or RRDs				Nearest Impact				MAIN FACTOR					SECONDARY FACTOR						Impairment score
	Single	Multi.	Braided	Y, paved or RRD, <50m	Yes, paved, >50m	Yes, (gravel or skid rd.)	No	< 50 m	< 100 m	Study area	NA (gravel)	Bare alluvium	Gravel + early seral	Ag. fields, cleared areas	Buildings, roads, railroads	Landslides	Bare alluvium	Gravel + early seral	Ag. fields, cleared areas	Buildings, roads, railroads	Landslides	*NA*	
262			x				x	x				x							x				2
264	x				x			x					x						x				1
265			x			x			x			x							x				2

5.5. Appendix E: Results from Silva2016 individual tree models

Table E1: Results of individual-tree-based models of forest structure based on the Silva2016 algorithm in the lidR package. Segments were ground-truthed with one dominant tree per segment. n = 1037 segments from 97 plots. Mixed models were calculated using R packages "nlme" (Pinheiro et al. 2020) and "lme4" (Bates et al. 2015).

SEGMENT-BASED ANALYSIS – SILVA 2016 METHOD

	RMSE ± std. dev.	*Model*	*Parameters*
Tree height	2.50 ± 2.46 m (6.63 ± 5.22 m)	Linear mixed-effect model *(lme4 package)*	Height of lidar-derived tree top
Tree DBH	9.10 ± 10.56 cm (13.13 ± 9.21 cm)	Linear mixed-effect model *(nlme package)* Fixed variances	Height of lidar-derived tree top

Table E2: Confusion matrix for deciduous versus coniferous classification using the Silva2016 segments. Composition was modeled with a generalized linear mixed model (Bates et al. 2015) on mean intensity (scaled by 1/1000) with random intercepts and plot as a random effect.

SEGMENT-BASED ANALYSIS – SILVA 2016 METHOD

	Field control		
	Conifer	Deciduous	Row total
Classification			
Conifer	**291**	33	324
Deciduous	66	**684**	750
Column total	357	717	1074
Producer's accuracy		**User's accuracy**	
Conifer = 291/357 = 82%		Conifer = 291/324 = 90%	
Deciduous = 684/717 = 95%		Deciduous = 684/750 = 95%	
Overall accuracy = (291+684)/1074 = 91%			
Kappa = 0.79			

Appendix F: Plot-level ground-truth data

Table F: Plot-level ground-truth data for 104 plots along the Nooksack River, Washington. Plots were circular with a 15-meter radius. Data was collected during summer 2019. "Random" indicates whether the plot location was randomly chosen or not. "Stems" are defined as discrete trunks at breast height (1.37 meters); a tree could have multiple stems if it forked below breast height. Dominant trees were overstory trees that made up part of the upper canopy. Height units are given in meters, DBH units are in centimeters, basal area is in square meters per hectare, and stem density represents stems per hectare.

Plot ID	Random?	Mean height	SD height	Mean DBH	SD DBH	Stem count	Tree count	Snag count	Broken count	Forked count	Pct. Dominant	Pct. Conifer	Basal area	Stem density
28	Y	21.2	3.6	35.1	10.6	28	28	1	1	2	86	4	167.2	429
1	Y	19.4	6.7	33.0	16.3	30	28	2	0	1	73	0	179.1	300
7	Y	18.0	2.5	26.0	6.5	21	21	0	0	1	91	5	66.8	243
11	N	19.8	7.4	41.5	21.9	17	17	5	0	1	65	77	164.3	429
12	Y	21.2	10.8	30.4	16.4	30	30	18	5	0	60	3	157.5	314
15	Y	21.3	8.2	44.6	19.5	22	19	1	0	1	77	27	230.6	129
17	N	20.9	8.2	59.1	19.3	9	9	1	0	0	89	56	152.9	314
23	Y	21.8	7.6	34.5	14.6	22	20	1	1	0	82	9	136.3	400
33	N	19.0	7.3	26.1	13.5	72	71	21	0	0	40	96	275.1	1029
39	Y	28.2	9.6	85.9	45.4	11	10	NA	NA	0	91	64	452.1	157
46	Y	14.6	4.2	19.1	7.7	54	51	4	1	0	83	15	101.7	771
54	Y	17.6	2.7	26.7	8.2	33	33	1	1	0	94	0	114.0	471
57	Y	18.3	7.8	35.6	23.5	50	50	9	3	0	52	82	401.2	714
58	N	25.0	11.8	35.3	17.4	49	49	4	3	0	53	86	335.8	700
66	Y	14.7	9.0	46.6	27.4	11	10	0	0	0	82	18	139.4	157
68	Y	16.6	5.4	20.9	8.5	40	40	4	1	0	63	0	90.3	571
70	Y	25.7	11.2	50.4	33.6	9	5	NA	NA	0	67	0	141.9	129
72	Y	20.7	7.1	34.3	15.4	24	24	2	0	0	88	79	149.6	343
73	Y	15.4	6.8	21.2	9.3	51	49	10	5	1	59	2	121.0	729
75	Y	13.8	7.9	25.9	14.0	21	16	0	0	0	76	24	79.9	300
85	Y	16.8	4.6	26.7	11.7	40	40	6	0	0	83	3	150.5	571
86	Y	14.1	7.5	29.7	15.6	21	21	0	0	0	91	0	104.2	300
90	Y	20.2	11.7	30.6	16.9	44	43	7	2	0	48	25	237.8	629
91	Y	14.2	5.5	18.2	7.2	51	50	NA	NA	0	65	33	86.4	729
101	Y	22.8	12.4	40.5	29.3	18	18	3	1	0	50	28	195.6	257
102	Y	18.7	6.0	37.9	21.7	24	18	NA	NA	1	54	21	201.7	343
108	Y	10.7	8.2	26.8	21.2	25	24	2	2	0	84	4	127.8	357
110	Y	17.8	5.9	32.2	15.2	40	36	4	0	1	63	13	224.9	571

Plot ID	Random?	Mean height	SD height	Mean DBH	SD DBH	Stem count	Tree count	Snag count	Broken count	Forked count	Pct. Dominant	Pct. Conifer	Basal area	Stem density
116	Y	21.6	12.7	40.5	22.9	14	14	2	1	1	50	7	132.6	200
118	Y	16.3	7.0	20.6	8.7	75	69	7	7	0	60	17	165.9	1071
120	N	18.9	9.0	36.1	20.0	36	36	10	4	0	47	83	270.3	514
121	Y	9.1	1.1	13.4	3.2	37	37	0	0	0	89	0	31.3	529
122	Y	16.4	2.6	16.6	4.3	82	82	2	0	0	92	0	107.4	1171
125	Y	23.5	9.0	39.4	15.4	49	49	7	6	0	71	82	389.3	700
128	Y	18.7	11.1	37.2	21.6	46	46	12	5	0	44	89	376.6	657
133	Y	21.7	14.3	34.8	23.1	51	51	9	0	0	35	94	393.3	729
135	Y	19.9	8.9	35.4	19.9	33	32	6	1	0	55	61	240.2	471
136	Y	25.4	14.4	50.5	21.7	27	27	4	3	0	52	59	359.8	386
139	Y	23.0	17.3	71.9	57.4	16	16	2	1	3	56	75	587.9	229
140	Y	20.8	11.7	36.5	21.5	67	64	5	2	0	37	99	532.9	957
144	Y	33.7	12.9	58.3	23.7	18	18	0	0	0	72	94	314.2	257
153	N	17.4	5.5	27.1	8.0	35	34	3	0	0	66	37	123.8	500
161	N	16.8	7.9	30.6	14.2	37	37	4	4	0	49	49	186.2	529
162	N	18.8	8.4	31.8	18.2	19	15	6	1	3	74	0	112.1	271
163	N	20.0	9.9	37.1	22.3	26	24	2	1	1	58	4	214.7	371
164	N	12.0	4.5	16.9	5.8	57	57	2	5	0	63	0	81.1	814
167	Y	15.4	4.7	20.6	8.8	21	19	1	0	0	91	0	46.5	300
200	N	21.5	12.8	44.0	29.1	35	35	4	3	0	69	86	429.0	500
201	N	13.7	3.0	16.7	4.1	43	43	19	4	0	72	0	56.5	614
202	N	16.4	9.4	37.3	25.9	25	24	4	2	1	72	68	225.9	357
203	N	14.2	5.7	21.0	10.0	51	51	4	2	0	59	22	122.2	729
204	N	16.2	3.9	19.9	6.9	60	60	5	1	0	82	0	118.3	857
205	N	17.7	6.9	23.9	12.0	56	46	7	3	0	45	0	176.9	800
206	N	13.1	3.0	17.8	7.8	47	47	4	1	1	87	2	78.5	671
207	N	12.5	2.3	15.6	4.1	44	43	7	0	0	93	0	51.0	629
208	N	20.3	10.7	31.4	19.1	46	45	11	2	0	33	26	274.3	657
209	N	18.8	8.7	34.8	18.4	46	46	14	9	0	54	65	315.5	657
210	N	20.4	7.7	27.9	14.5	52	50	9	0	1	48	69	227.3	743
211	N	10.6	5.4	26.3	8.8	15	14	0	0	0	47	60	51.0	214
212	N	14.0	4.4	21.0	7.1	39	31	8	4	0	85	18	84.7	557
213	N	17.7	6.2	27.1	10.2	32	32	3	2	2	66	16	119.0	457
214	N	5.9	0.7	27.0	7.0	5	4	0	0	0	100	100	17.1	71
215	N	16.7	6.7	31.2	19.8	25	24	4	1	0	80	40	149.8	357
216	N	18.6	8.1	30.2	12.5	27	27	3	2	0	52	59	127.2	386
217	N	24.5	11.9	46.1	27.5	26	26	2	0	0	65	89	330.1	371
218	N	17.8	7.9	42.9	21.0	14	14	6	4	0	64	21	140.2	200
219	N	19.9	10.3	33.8	18.4	49	45	7	4	0	35	51	322.0	700
220	N	17.8	11.4	30.9	25.4	32	32	3	2	0	63	84	224.7	457

Plot ID	Random?	Mean height	SD height	Mean DBH	SD DBH	Stem count	Tree count	Snag count	Broken count	Forked count	Pct. Dominant	Pct. Conifer	Basal area	Stem density
221	N	25.1	12.3	37.2	17.8	33	33	3	0	0	49	100	248.2	471
222	N	15.9	9.7	27.0	19.6	45	45	5	0	0	51	78	220.7	643
223	N	23.8	9.9	50.3	23.5	29	29	5	3	1	69	62	394.5	414
224	N	18.2	6.7	28.2	26.2	23	17	0	0	0	61	4	148.4	329
225	N	16.8	10.1	26.6	15.1	67	66	16	1	0	39	94	277.0	957
226	N	17.2	6.4	28.8	23.7	37	36	6	2	0	62	35	226.1	529
227	N	23.7	14.1	48.0	29.9	26	26	3	3	0	50	92	366.0	371
228	N	16.4	6.3	21.8	10.5	40	39	6	5	0	53	0	103.9	571
229	N	21.2	13.3	47.4	30.9	30	30	6	4	0	43	87	422.8	429
230	N	11.6	2.2	20.3	5.8	54	52	2	0	0	82	93	106.4	771
231	N	21.0	13.4	42.9	38.2	34	34	4	0	1	35	85	491.6	486
232	N	15.7	7.4	22.4	10.5	48	48	5	0	0	52	60	130.2	686
233	N	11.4	3.8	18.9	11.8	10	9	3	0	0	100	10	21.4	143
234	N	8.6	2.0	19.0	3.9	35	35	0	0	0	97	100	58.2	500
235	N	19.5	10.8	39.2	31.2	29	27	4	7	1	41	10	318.7	414
236	N	9.8	2.9	17.4	6.7	29	28	0	0	0	97	90	44.4	414
237	N	10.7	2.4	12.6	1.8	11	11	0	0	0	100	0	7.9	157
238	N	28.6	9.9	55.9	22.2	14	14	2	0	0	71	0	223.3	200
239	N	12.7	5.9	19.6	8.4	12	12	2	1	0	67	0	24.1	171
240	N	16.4	3.9	15.3	4.6	41	41	1	0	0	93	0	46.3	586
241	N	15.3	8.3	29.9	33.5	34	33	7	4	0	79	41	299.8	486
242	N	14.0	3.8	15.4	5.8	46	45	1	0	0	89	0	55.1	657
243	N	18.6	10.2	29.7	14.7	38	38	5	2	0	55	95	184.9	543
244	N	16.6	4.0	22.5	5.7	28	28	2	4	0	89	0	67.0	400
245	N	5.5	2.5	11.1	0.9	3	3	0	0	0	100	0	1.7	43
246	Y	16.2	3.9	16.7	5.3	58	58	1	0	0	79	12	78.8	829
247	N	13.1	6.5	17.9	8.4	19	19	0	2	0	95	0	32.6	271
248	N	12.9	3.3	17.1	5.3	46	46	1	0	0	83	0	65.8	657
249	N	22.2	13.3	42.2	21.1	26	25	9	6	0	50	77	255.3	371
250	N	15.6	3.0	15.1	4.3	87	87	3	1	0	74	0	95.6	1243
251	N	22.7	10.7	46.2	16.5	25	25	7	7	1	60	72	265.7	357
252	N	21.2	5.0	26.7	9.4	45	45	5	1	1	64	0	159.7	643
253	N	20.8	5.8	31.6	12.5	33	33	2	0	0	67	6	168.4	471
254	N	10.8	8.8	32.3	9.6	11	11	7	6	0	91	18	55.2	157
255	N	18.9	9.1	46.4	23.7	19	19	4	1	1	68	58	226.9	271
256	N	12.8	2.7	16.4	4.7	28	28	3	0	0	89	0	36.2	400