



Forest Service
U.S. DEPARTMENT OF AGRICULTURE

Geospatial Technology and Applications Center Prince of Wales Existing Vegetation Map Project



Geospatial Technology and Applications Center
Salt Lake City, Utah

GTAC 10230-RPT1
May 2021



USDA Non-Discrimination Statement

In accordance with Federal civil rights law and U.S. Department of Agriculture (USDA) civil rights regulations and policies, the USDA, its Agencies, offices, and employees, and institutions participating in or administering USDA programs are prohibited from discriminating based on race, color, national origin, religion, sex, gender identity (including gender expression), sexual orientation, disability, age, marital status, family/parental status, income derived from a public assistance program, political beliefs, or reprisal or retaliation for prior civil rights activity, in any program or activity conducted or funded by USDA (not all bases apply to all programs). Remedies and complaint filing deadlines vary by program or incident.

Persons with disabilities who require alternative means of communication for program information (e.g., Braille, large print, audiotope, American Sign Language, etc.) should contact the responsible Agency or USDA's TARGET Center at (202) 720-2600 (voice and TTY) or contact USDA through the Federal Relay Service at (800) 877-8339. Additionally, program information may be made available in languages other than English.

To file a program discrimination complaint, complete the USDA Program Discrimination Complaint Form, AD-3027, found online at [How to File a Program Discrimination Complaint](#) and at any USDA office or write a letter addressed to USDA and provide in the letter all of the information requested in the form. To request a copy of the complaint form, call (866) 632-9992. Submit your completed form or letter to USDA by: (1) mail: U.S. Department of Agriculture, Office of the Assistant Secretary for Civil Rights, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410; (2) fax: (202) 690-7442; or (3) email: program.intake@usda.gov.

USDA is an equal opportunity provider, employer, and lender.

Bellante, G.; Pan, C.; Day, T.; Dillman, K.; Homan, K.; Krosse, P.; Schrader, B.; Wittwer, D.; Goetz, W.; Pugh, N.; Megown, K; Sheets, R. 2021. Prince of Wales Existing Vegetation Map Project. GTAC-10230-RPT1. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Geospatial Technology and Applications Center, pp 41.

Cover photo credit: USDA Forest Service, Tongass National Forest

Abstract

Existing vegetation data are available for the Prince of Wales Island and surrounding area as a result of a collaborative effort between the Tongass National Forest, Alaska Regional Office, and Geospatial Technology and Applications Center (GTAC). These data were designed to be consistent with the standards established in the Existing Vegetation Classification and Technical Guide (Nelson et al. 2015), and to provide baseline information to support project planning and inform land management of the Prince of Wales and surrounding islands. The geodatabase comprises seven distinct, integrated feature layers: 1) vegetation type; 2) tree canopy cover; 3) trees per acre (TPA) for trees $\geq 1'$ tall; 4) trees per acre for trees $\geq 6"$ diameter at breast height (dbh); 5) quadratic mean diameter (QMD) for trees $\geq 2"$ dbh; 6) quadratic mean diameter for trees $\geq 9"$ dbh; and 7) thematic tree size. The dominance type map consists of 18 classes, including 15 vegetation classes and three other land cover types. Continuous tree canopy cover, TPA, QMD, and thematic tree size were developed for areas classified as *forest* on the final vegetation type map layer. Geospatial data, including remotely sensed imagery, topographic data, and climate information, were assembled to classify vegetation and produce the maps. A semi-automated image segmentation process was used to develop the modeling units (mapping polygons), which delineate homogeneous areas of land cover. Field plots containing thematic vegetation type and tree size information were used as reference for random forest prediction models. Important model drivers for vegetation type prediction included 30 cm orthoimagery collected during the height of the 2019 growing season, as well as Sentinel 2 and Landsat 8 satellite imagery. Additionally, detailed tree inventory data was collected at precise field locations to develop forest metrics from Quality Level 1 (QL1) Light Detection and Ranging (LiDAR) data. LiDAR information was acquired across approximately 75% of the project's land area. Continuous tree canopy cover and second order forest metrics (TPA and QMD) were modeled across the LiDAR coverage area, and subsequently, extrapolated to the full project extent using Interferometric Synthetic Aperture Radar (IfSAR) as the primary topographic data source. The mapping process utilized various Forest Service Enterprise software, adopting the most contemporary methods and technology. Most of the reference information was collected during the 2018 growing season. Image interpretation allowed the high resolution orthoimagery acquired during the summer of 2019 to be used as the standard by which modeling results were evaluated and modifications to the maps were made. Consequently, the map products are indicative of the existing vegetation conditions found on Prince of Wales and surrounding islands at that time. Upon completion of the existing vegetation data products, an accuracy assessment was conducted to reveal individual class confusion and provide additional insight into the reliability of the final maps for resource applications.



Authors

Gabriel Bellante is a Geospatial Project Manager, Remote Sensing Analyst and the Lead Report Author employed by RedCastle Resources, Inc. at the Geospatial Technology & Applications Center (GTAC) in Salt Lake City, UT.

Caleb Pan is a Remote Sensing analyst employed by RedCastle Resources, Inc. at GTAC in Salt Lake City, UT.

Thomas Day is a Remote Sensing analyst employed by RedCastle Resources, Inc. at GTAC in Salt Lake City, UT.

Karen Dillman is the National Air Monitoring Coordinator for the Washington Office in Petersburg, AK.

Kim Homan is the Geospatial Program Manager for the USFS Alaska region, Engineering and Information Management Office in Juneau, AK.

Patti C. Krosse, Forest Ecologist, Tongass National Forest (retired).

Barb Schrader is the Regional Ecologist for the USFS Alaska region, Landscape and Vegetation Ecology Program in Juneau, AK.

Dustin Wittwer is the Remote Sensing Coordinator for the USFS Alaska region, Engineering and Information Management Office in Juneau, AK.

Wendy Goetz is a Vegetation Mapping Team Leader employed by RedCastle Resources, Inc. at GTAC in Salt Lake City, UT.

Nathan Pugh is an Assistant Program Leader for the Resource Inventory Monitoring and Mapping (RMIM) program at GTAC in Salt Lake City, UT.

Kevin Megown is the Program Leader for the Resource Inventory Monitoring and Mapping (RMIM) program at GTAC in Salt Lake City, UT.

Robert Sheets is the Tongass Young Growth Coordinator for the Tongass National Forest in Craig, AK.



Acknowledgements

We would like to acknowledge Jacquie Foss who provided local ecological expertise in the field and interpreted reference data for the shrub classes. Conor Reynolds of The Nature Conservancy led the data collection effort for LiDAR calibration and developed the LiDAR forest metrics for the phase 1 extent. Thanks to Bob McGaughey of the Pacific Northwest Research Station for producing the LiDAR first order metrics and for his consultation.

Katie Baer's help was greatly appreciated in classifying the Forest Inventory and Analysis plots utilized by the project. Thanks are owed to Andy Klimek who developed the Survey 123 data collection form used for field work and Jackie DeMontigny (retired) for her leadership in collecting field data critical to the project's success. Additional thanks to the Tongass National Forest field crews for their indispensable contributions in collecting the empirical field data and contributing team members who provided expertise throughout the length of the project, including:

Joni Johnson, Ryan Vosbigian, Riley McFarland, Abby Gnewuch, Dennis Landwehr, Thomas Heutte, Gregory Dunn, Melinda Lamb, Kelsey Aho, Jennifer Richter, Karen Hutten, and Paul Maus

We thank you all.

Partnerships

Our partners played a vital role in providing large datasets, resources toward classification and analysis, field crews and field surveys, mapping needs and product uses, expertise in legacy data, designing and testing applications, training, review of draft maps, and subject matter expertise in the development of the classification key. These partnerships are critical to ensure the highest level of integrity, objectivity, and usefulness for internal application and external consumption.

Federal Partners:

USDA Forest Service

- Alaska Regional Office
- Tongass National Forest
- Geospatial Technology & Applications Center (GTAC)
- Forest Inventory and Analysis (FIA)

US Geological Survey (USGS)

State Partner:

- State of Alaska

Private Partners:

- The Nature Conservancy
- Sealaska Corporation



Table of Contents

Abstract.....	ii
Authors	iii
Acknowledgements	iv
Partnerships.....	iv
Federal Partners:.....	iv
State Partner:	iv
Private Partners:	iv
Introduction	1
Project Area	1
Project Planning.....	1
Mapping Methods	2
Geospatial Data Acquisition.....	4
Image Segmentation	7
Reference Data Collection	8
Classification and Regression.....	11
Map Revision.....	17
Final Data Products.....	19
Vegetation Type	19
Tree Canopy Cover	24
Trees Per Acre	25
Quadratic Mean Diameter	28
Tree Size	29
Conclusion	31
References	34
Appendix A: Dichotomous Key Containing the Class Definitions for the Prince of Wales	35
Appendix B: Prince of Wales Vegetation Type Error Matrix.....	41

Introduction

Maps of existing vegetation support resource managers by informing project and landscape-level planning efforts with vegetation data that can be used in numerous applications. Use of existing vegetation maps can save time and money by eliminating work redundancies and informing a multitude of future management activities. Mission-critical Forest Service goals necessitate vegetation information for forest planning, ecological assessment, forest health monitoring, and wildlife habitat management. Additionally, existing vegetation maps are commonly employed for silviculture, fire risk assessment, natural resource inventories, rare and sensitive species monitoring, invasive species modeling, recreation management, disturbance susceptibility evaluations, and climate change analyses. This project implemented consistent methodologies, which used empirical data and leveraged contemporary technology, to develop defensible map products that utilized the best available science. The resultant map products establish a baseline of landscape ecological condition through the depiction of vegetation types, tree canopy cover, trees per acre (TPA), quadratic mean diameter (QMD), and thematic tree size.

Authority and funding for the Prince of Wales existing mapping project was provided by the Tongass National Forest and the Alaska Regional Office. The Geospatial Technology and Applications Center (GTAC) produced existing vegetation maps using contemporary mapping methods, adhering to the standards established in the Existing Vegetation Classification, Mapping, and Inventory Technical Guide (Nelson et al. 2015), and using the most current data available. This project provides land managers with a vegetation map to inform planning and management decisions pertinent to the Prince of Wales and surrounding islands.

Project Area

The Prince of Wales mapping project encompasses over 4.2 million acres of Southeastern Alaska—2.3 million acres of which are terrestrial. Surrounded by deep channels and fjords, the islands within the project area are a part of the Alexander Archipelago ecoregion in the Alaskan Panhandle (Nowacki et al. 2001). These islands form an outer barrier that protects the Inside Passage from approaching storms originating in the Gulf of Alaska and the Pacific Ocean. This periglacial maritime landscape is characterized by rugged mountains and dense forest. Mountains on Prince of Wales Island reach nearly 4,000 feet above sea-level, with high meadows and krummholz patches in exposed subalpine ecotones. The islands within the project area contain numerous lakes, rivers, wetlands, and bogs. Timber extraction, fishing, and ecotourism are the mainstays of the local economy. Most of Prince of Wales Island and surrounding islands are inhabited by a matrix of productive temperate rainforest and less productive forest characterized by poorly drained soils. Many of the productive forests are, or have been, actively managed and extensive timber harvests have been conducted across the landscape, resulting in large areas containing forests in various stages of regrowth.

Project Planning

In 2017, personnel from the Tongass National Forest met with partners from the Alaska Region and GTAC to identify objectives and outline a strategy for the Prince of Wales existing vegetation mapping project. This partnership discussed map unit design in order to develop a vegetation classification system that was both ecologically meaningful and realistic with respect to technology and the data available for the area. Vegetation map units share a common definition based on their physiognomic,

floristic, or structural characteristics. The map unit design process establishes the rules that define the map classes found in the classification key (Appendix A). This dichotomous key establishes the discrete absolute and relative vegetation cover percentages, as well as the height definitions that classify every vegetation community encountered on the ground. Although class assignment in the field may be difficult, especially when threshold cover and height determinations are approached, the class definitions themselves must be clear and unambiguous.

The classification key met four critical standards: 1) be exhaustive to describe the full range of environmental conditions that are to be mapped across the Prince of Wales project area; 2) be mutually exclusive to contain classes with no overlap or have any ambiguity in their respective definitions; 3) describe vegetation that is readily observed in the field; and 4) contain classes that are capable of being mapped and are congruent with the scale and scope of the project.

For this project, vegetation types and structure classes were identified to address the information needs of the land management agency and partners. GTAC was tasked to develop a set of mid-level existing vegetation maps for Prince of Wales Island and surrounding islands. Existing vegetation is the plant cover, or floristic composition and vegetation structure, occurring at a given location at the current time (Nelson et al. 2015). Some vegetation types are a combination of species which describe a vegetation community (e.g., Wet Herbaceous) while others identify specific species (e.g., Sitka Spruce). Certain vegetation type classes may include multiple dominance types when there is insufficient reference data, the dominance type occurs only rarely on the landscape, or because of the inability to differentiate certain dominance types with the available predictor data. Ultimately, there were a total of 18 vegetation type classes—15 vegetation types and three non-vegetated cover types—all of which can be found in the *Prince of Wales Dominance Type Key* (Appendix A). Additionally, six structure map products are included with the final deliverables—one continuous tree canopy cover layer; two Trees Per Acre (TPA) layers; two Quadratic Mean Diameter (QMD) layers; and one thematic tree size layer (Table 1).

These products were developed to provide up-to-date, comprehensive information about the vegetation communities, and their structure, across Prince of Wales Island and surrounding islands (hereafter, referred to as the ‘*Prince of Wales*’ project). Over 4.2 million acres, including other federal, state, local, native, and private land inholdings, were mapped. It is important to remember that the vegetation characteristics being described on the final maps is from a synoptic, overhead, bird’s-eye perspective (Figure 1). Therefore, understory vegetation that is not visible from above, is not being depicted.

Mapping Methods

The map products for this project were developed using remotely sensed multispectral imagery, topographic Light Detection and Ranging (LiDAR) and Interferometric Synthetic Aperture Radar (IfSAR) data, field and photo-interpreted reference sites, and object-oriented classification models. The fundamental modeling units (segments) were produced using a semi-automated image segmentation process that considers the shape, size, and spectral content of spatially contiguous pixels across the landscape. Random Forest, an ensemble classifier, was then used to characterize these modeling units and assign map class labels, which ultimately produced the final vegetation maps for the Prince of Wales project area.

The major mapping phases, which are discussed in more depth below, include: geospatial data acquisition, image segmentation, reference data collection, classification, draft map review, final map development, and map validation.

Table 1. List of vegetation types and structure layers for the Prince of Wales existing vegetation map project—(a) Vegetation types; and (b) Structure metrics. Note, some dominance types were not mapped because either the type was rare on the landscape, and consequently, there were insufficient reference data (e.g., Cottonwood class), or it was too difficult to distinguish types adequately given the available predictor data (e.g., red cedar vs. yellow cedar).

(a)

Map Group	Vegetation Types	Map Unit Abbreviation
Conifer Forest	Sitka Spruce	SS
	Sitka Spruce-Western Hemlock	SS-WH
	Western Hemlock	WH
	Cedar	CE
	Mountain Hemlock Mix	MHmix
	Dwarf Conifer	DC
	Mixed Conifer	MC
	Mixed Species	MS
Broadleaf Forest	Red Alder	RA
Mixed Forest	Sitka Spruce-Red Alder	SS-RA
Shrub	Alder Shrub	AS
	Tall Shrubs	TS
	Low Shrubs	LS
Herbaceous	Wet Herbaceous	WHB
	Aquatic Herbaceous	AHB
Other	Water	WA
	Barren/Sparse Vegetation	BR/SV
	Developed	DEV

(b)

Structure Metrics	Description
Tree Canopy Cover	Continuous tree canopy cover from 0 – 100%
Trees Per Acre, ≥ 1' tall	Number of live trees ≥ 1' tall per acre
Trees Per Acre, ≥ 6" diameter at breast height (dbh)	Number of live trees ≥ 6" dbh per acre
Quadratic Mean Diameter, ≥ 2" dbh	Quadratic mean diameter of live trees ≥ 2" dbh
Quadratic Mean Diameter, ≥ 9" dbh	Quadratic mean diameter of live trees ≥ 9" dbh
Tree Size	Thematic classes: 1. TS1 Sapling (< 5" dbh); 2. TS2 Pole (5-8.9" dbh); 3. TS3 Medium (9-20.9" dbh); 4. TS4 Large (≥ 21" dbh); 5. Non-Tree

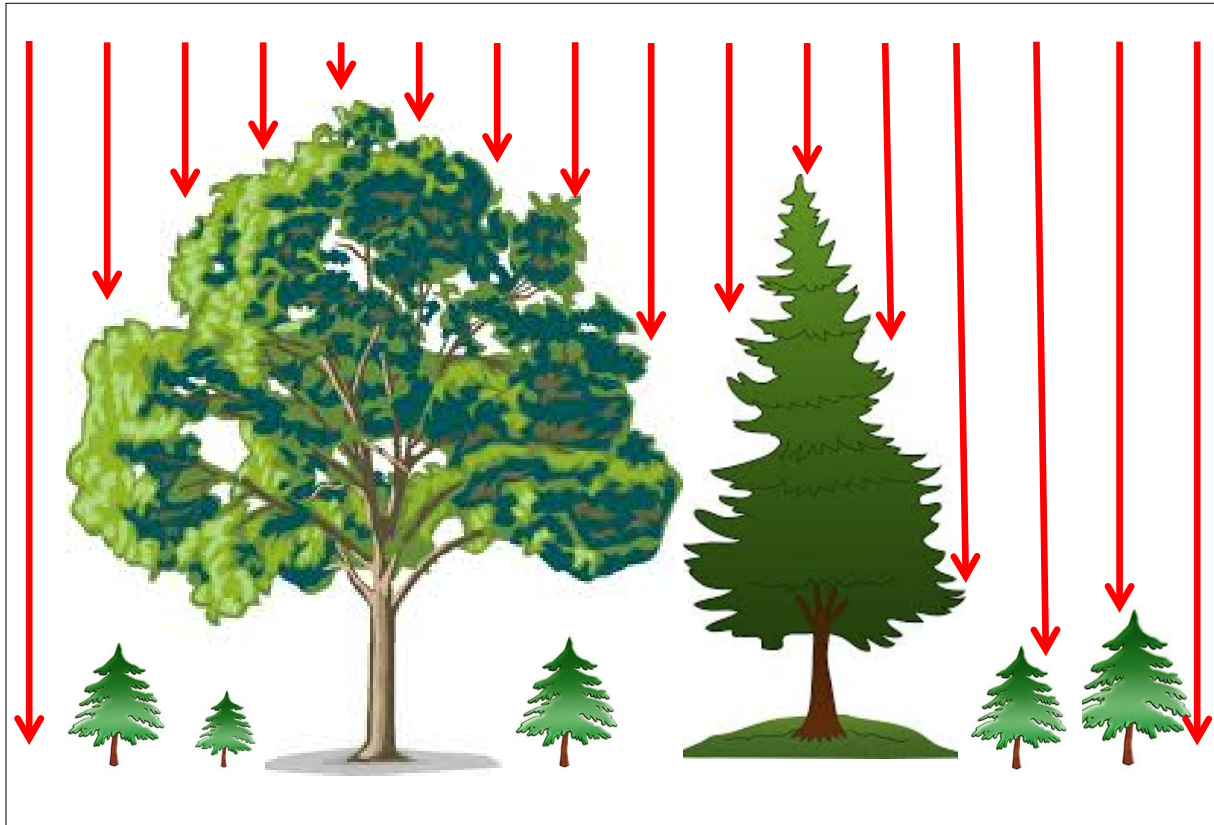


Figure 1. An example of the synoptic remote sensing perspective when viewing the landscape from above. The red arrows illustrate the vegetation that would be detected from an overhead sensor.

Geospatial Data Acquisition

This project utilized remotely sensed imagery acquired from various sensors on both satellite and airborne platforms. Each image sensor has a unique set of qualities that, along with the imaging geometry, determines the spectral, spatial, and radiometric resolutions of the data that is collected. Multiple sources of imagery were acquired for this project in order to utilize the unique information afforded different sensors and to maximize the range of data used in the computational modeling. Image mosaics were developed from SPOT, Landsat 8, and Sentinel 2 satellite image archives. Imagery from the SPOT satellite system was collected between 2010 and 2016 as part of the Alaska Statewide Digital Mapping Initiative. Landsat 8 and Sentinel 2 images were reviewed and prioritized in order to reflect current ground conditions, limit cloud obscenity, and capture variations in vegetation phenology. Selected individual scenes were then mosaicked together in Google Earth Engine to remove clouds and aggregate adjacent image swaths for the purpose of developing seamless image mosaics of the entire project area for each sensor. Independent image mosaics are ideally generated for the spring, summer, and fall seasons in order to depict phenological conditions throughout the growing season to better distinguish vegetation types. Ultimately, three seasonal mosaics (spring, summer, and fall) were generated for both Sentinel 2 and Landsat 8. Spring mosaics depict early growing season conditions;

summer mosaics capture the height of the growing season at the climax of vegetation development for most communities; and fall mosaics attempt to capture the end of the growing season when vegetation senesces, and chlorophyll degrades. A Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) were produced for each image mosaic. Also, a Tasseled-Cap Transformation was developed for each Landsat 8 mosaic (Crist and Cicone 1984), while Principal Component Analysis (PCA) was performed on the Sentinel 2 image mosaics.

High-resolution imagery is critical for photo interpretation, which allows an analyst to evaluate and modify model outputs and is instrumental for developing relatively fine-scaled segments for a mapping project. Resource imagery was acquired during the summer of 2019 and covers over 99% of the project area. Only a few islands in the northeastern portion of the study area were not captured. This imagery is ortho corrected, contains 4-spectral bands (red, green, blue, and near infrared), has a 30 cm resolution, and is cloud-free.

Elevation data for the project area came from LiDAR and IfSAR data sources. The LiDAR data was collected in 2017 and 2018, in two phases. The mission was funded by the US Forest Service and additional partners, including The Nature Conservancy, Sealaska Corporation, and a matching grant from the US Geological Survey (USGS) 3D Elevation Program. These LiDAR data acquisitions were flown by Quantum Spatial and designated to meet USGS base specification quality level 1 with an accuracy of 10.0 cm root mean square error in z (RMSE $_z$) and a density of 8 pulses per square meter (pls/m²) (Heidemann 2014). This LiDAR data covers over 2,700 square miles (over 1.7 million acres) of Prince of Wales Island and surrounding islands, 75% of the land area within the project boundary. LiDAR data was the foremost model driver and most accurate predictor of vegetation structure for this project. Where LiDAR data was not available, IfSAR data was the topographic data source used for vegetation structure prediction. IfSAR covered the full extent of the Prince of Wales project area and had three 5 m resolution components: 1) a digital terrain model (DTM), derived from the P-band, which penetrates through vegetation to provide a bare-earth approximation; 2) a digital surface model (DSM) derived from the X-band, which reflects higher canopy vegetation and provides an estimate of canopy surface elevation; and 3) a canopy height model (CHM), which is an approximation of vegetation height by taking the elevation difference between the DSM and DTM. Topographic derivatives including slope, aspect, heat load, a topographic position index (TPI), and a topographic wetness index (TWI) were produced from the IfSAR DTM.

Ancillary data, including climate and ownership spatial layers, were also used in the mapping process. Daymet climate data—including shortwave radiation, min/max temperature, water vapor pressure, precipitation, and continentality—were used in the vegetation type modeling. The ownership layer assisted in the development of an access layer that was used for field reference site placement and informed the editing of harvest areas on non-Forest Service lands, where harvest unit boundary information was limited. All final data layers were co-registered and projected to the State Plane coordinate system, NAD83, Alaska Zone 1. The data were resampled to 5 meters to maintain consistency in spatial resolution across all data layers. A complete list of geospatial data used in the project can be found in Table 2.

Table 2. List of data sources with a quick description of each layer, associated spatial resolution, and the corresponding application of the data source in the mapping process.

Geospatial Data Source		Product Description	Spatial Resolution	Purpose
Spectral Data	2019 Orthorectified Resource Imagery	Bands: Blue, Green, Red, Near Infrared (NIR) Indices: Normalized Difference Vegetation Index (NDVI)	30 cm	Segmentation, Vegetation Type & Structure Modeling, Photointerpretation
	Sentinel 2 Mosaics - Spring April 1, 2019 - Summer Aug. 29, 2019 - Fall Oct. 8, 2019	Bands: Blue, Green, Red, NIR Indices: NDVI	10 m	Segmentation, Vegetation Type & Structure Modeling
		Bands: Red Edge (RE) 1, RE 2, RE 3, Shortwave Infrared (SWIR) 1, SWIR 2 Indices: Normalized Difference Moisture Index (NDMI), Principle Component (PC)1, PC2, PC3	20 m	
	Landsat 8 OLI Mosaics - Spring April 27, 2018 May 17, 2017 May 14, 2016 - Summer June 18, 2017 Aug. 7, 2017 Aug. 27, 2016 - Fall Sept. 16, 2018 Nov. 2, 2018	Bands: Blue, Green, Red, NIR, SWIR 1, SWIR 2 Indices: NDVI, NDMI, Tasseled Cap Transformation (brightness, greenness, wetness)	30 m	Vegetation Type & Structure Modeling
	SPOT 5, 6, & 7 Collected between 2011 and 2016	Bands: Green, Red, NIR Indices: NDVI, PC1, PC2, PC3	Pansharpened 2.5 m (Native Resolution 5 m)	Segmentation, Vegetation Type & Structure Modeling
Topographic Data	LiDAR - Collected during the summers of 2017 and 2018 - USGS Quality Level 1 (QL1) 8 pls/m²	Digital Terrain Model (DTM) Digital Surface Model (DSM) Canopy Height Model (CHM)	1 m	Segmentation, Vegetation Type & Structure Modeling, Reference Data
		First Order Metrics (87 metrics)	30 m	
	IfSAR Collected during the summer of 2012	Digital Terrain Model (DTM) Digital Surface Model (DSM) Canopy Height Model (CHM)	5 m	Segmentation, Vegetation Type & Structure Modeling
	Sentinel 1 Composites generated for April, 2020 and August, 2019	C-Band Polarizations: VV (ascending), VV (descending), VH (ascending), VH (descending) Indices: VV to VH (ascending), VV to VH (descending) * V=Vertical; H=Horizontal	10 m	Vegetation Type & Structure Modeling

Geospatial Data Source		Product Description	Spatial Resolution	Purpose
Ancillary Data	Daymet Climate Data 30-year means were calculated for each climate metric	Daily incident shortwave radiation (W/m ²) Daily maximum temperature (C) Daily minimum temperature (C) Daily water vapor pressure (Pa) Annual sum precipitation (mm) Continental: July - January daily max temperature (C)	1 km	Vegetation Type & Structure Modeling
	Cover Type Map	GIS Vector Data	na	Timber Harvest Delineation
	Ownership	GIS Vector Data	na	Reference Data Site Selection & Access
	Transportation (Roads & Trails)	GIS Vector Data	na	Reference Data Site Selection & Access

Image Segmentation

Image segmentation is the process of partitioning digital imagery into spatially cohesive modeling units (mapping polygons) that represent discrete areas or objects on a landscape (Ryherd and Woodcock 1996). The goal is to develop homogenous segments that delineate vegetation of similar physiognomic, floristic, and structural characteristics to serve as the fundamental modeling units. High spatial resolution imagery and vegetation structure data excels at portraying vegetation patterns across the landscape during the segmentation process, such as delineating forest edges or isolating patches of shrub. Therefore, the 2019 orthoimagery and LiDAR canopy height model were the most important data sources for generating the mapping segments. The 2019 orthoimagery covered 99% of the area of interest—all but a few islands in the northeastern corner of the project area. Where the 2019 imagery was not available, SPOT imagery was used in its place. LiDAR data was acquired for approximately 75% of the project area and IfSAR data were used as the topographic data source where LiDAR was not available. Additionally, Sentinel 2 satellite imagery was used because it covers the entire project area and serves as a consistent moderate resolution data input to refine the segmentation. All data was resampled to 5-meters to make data processing more efficient and avoid over-segmentation of the complex landscape of the Prince of Wales and surrounding islands.

Development of the Prince of Wales segments was an iterative process which utilized a variety of algorithms and a combination of data sources structured into an eCognition ruleset. eCognition is a FS enterprise software that is used to partition the landscape into homogenous units in a semi-automated process. Coarse segments were initially generated to delineate the ocean, large waterbodies, and intertidal areas. The segments were incrementally refined to more finely delineate landscape features until the final segments were achieved and were commensurate with the scale and scope of the project. The final database included over 1 million segments, which had a mean size of 3.87 acres. Median segment size was 1.64 acres since large bodies of water were classified and merged during the

segmentation process, creating very large segments that increased the mean as compared to the median. The final segments were filtered and smoothed to ensure that the smallest segment was 0.25 acres or larger to prevent segments from capturing landscape features too small to adequately model with the available geospatial predictor data (Figure 2).

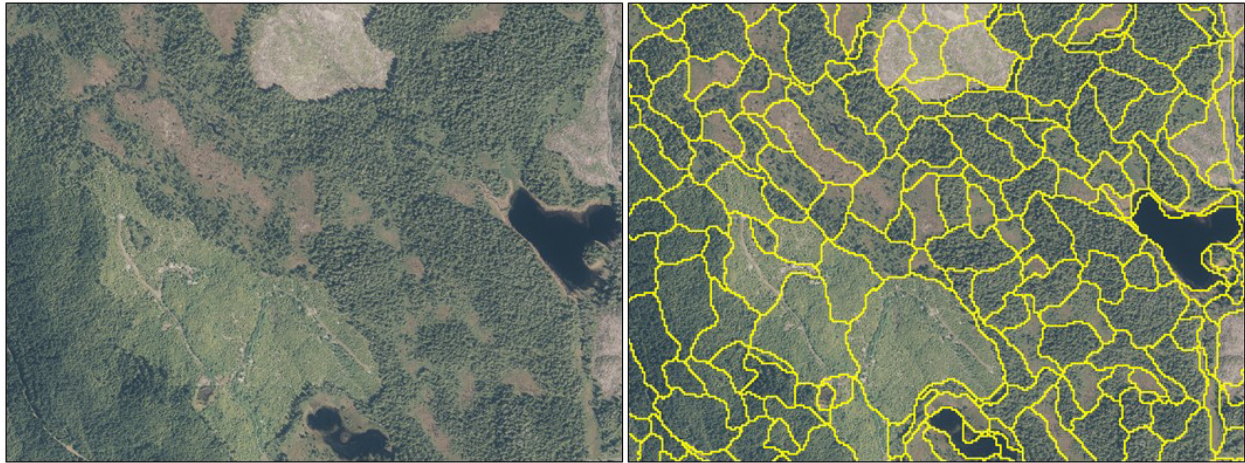


Figure 2. Example of the final segments generated as the fundamental modeling units for the mapping project using Trimble eCognition. This is a snapshot of the 30 cm 2019 orthoimagery just north of Thorne Bay (left) and overlaid with the final segments (right). Scale 1:6,000

Reference Data Collection

Consistent and precise reference information is imperative to successfully map existing vegetation. Personnel from the Tongass National Forest worked with GTAC to identify areas accessible to field sampling. Field crews from the Tongass, State of Alaska, and The Nature Conservancy collected the reference data required for modeling vegetation across the diverse landscape of the Prince of Wales project area. Ecologists from the Tongass National Forest, Alaska Regional Office, and GTAC collaborated to produce a vegetation classification system for the Tongass National Forest as a whole and identified the desired map units (vegetation type map classes) to be depicted for the Prince of Wales existing vegetation mapping project (Appendix A).

Reference data for this project came from numerous sources, including: 1) field crews collecting vegetation information specific to this project; 2) Young Growth Inventory data; 3) legacy data from former ecology plots and the Forest Inventory and Analysis (FIA) program; and 4) photo interpretation (Figure 3). Tongass National Forest personnel collected most of the ground data that was targeted for this mapping effort using a variety of access means—such as, by helicopter, floatplane, boat, or by foot from existing trail and road infrastructure. The Nature Conservancy and state of Alaska field crews collected the necessary LiDAR calibration data to develop the vegetation structure models and were used to bolster the reference sample for modeling vegetation type. Young growth inventory information was leveraged as reference for forests that are currently, or have been, actively managed in the past. These data were a result of the Challenge Cost Share Agreement between the Tongass National Forest and State of Alaska Division of Forestry. FIA data were cross-referenced with the Tongass National Forest existing vegetation mapping classification key in order to label each systematic plot with a vegetation type class. Photo interpretation techniques were used to bolster vegetation types with

relatively low sample numbers, especially the shrub classes, and in order to modify or improve interim map models.

All these reference data from the various sources were consolidated into a single database, and each site was reviewed within the context of their corresponding map segment using high-resolution imagery and topographic data sources. The final reference database included 10,294 sites—360 of these sites were from the FIA program (not shown). Inevitably, the more abundant vegetation and structure types were sampled at a higher frequency. It can be difficult to obtain an adequate sample for rarer classes and some of the dominance types were dropped as a result.

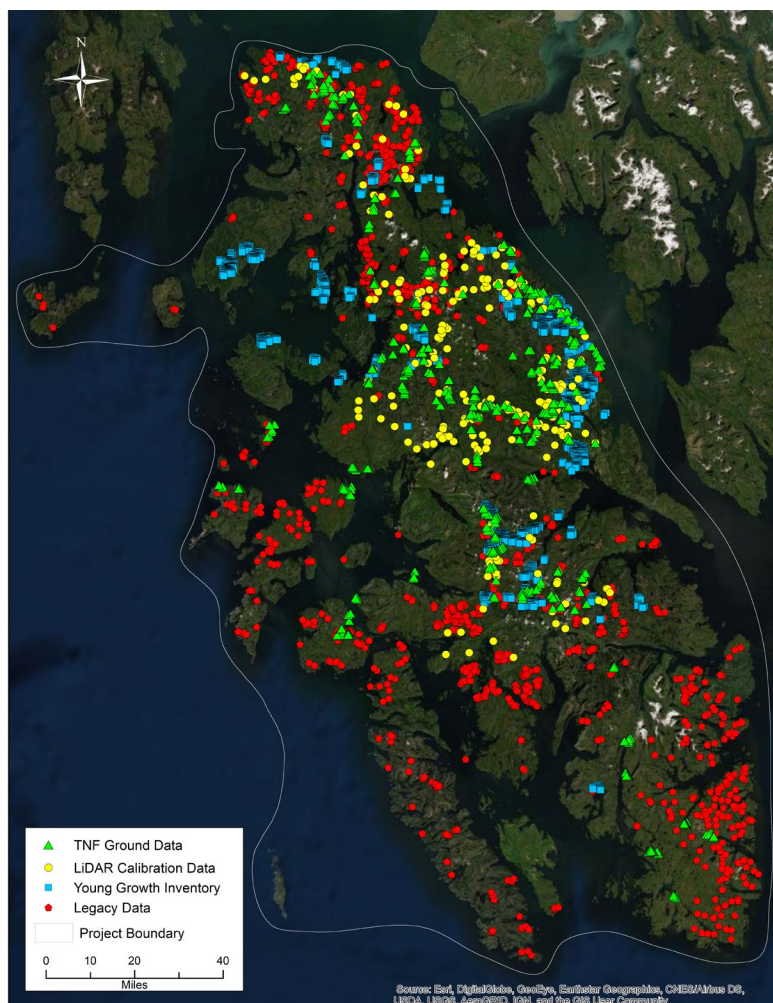


Figure 3. Illustration showing the distribution of reference data, from the various sources, across the Prince of Wales project area.

Ground Data

A total of 804 sites were collected on the ground by Forest Service personnel during summer 2018. These sites were primarily pre-selected using an image stratification of a Landsat 8 summer mosaic to help distribute the sites equally across the full range of ground conditions within the project area. This stratification was confined within the bounds of an access layer that identified areas accessible with

respect to ownership, terrain slope, avoiding dangerous river or stream crossings, and remoteness. Additional consideration was used to cluster targeted sites to maximize sampling efficiency and to minimize travel between sampling locations. The immediate vicinity of all pre-selected sites was reviewed for homogeneity and representativeness using high-resolution imagery.

Two types of field sites were collected by field crews—descriptive and observation. A 50' radius plot was evaluated at each sampled location. Descriptive sites contain highly detailed, comprehensive plot information on species and structure, whereas an observational plot is a quick method by which a field crew can make dominance type and structure determinations without collecting discrete plot data. For descriptive plots, detailed plant cover information was collected, including ocular estimates of vegetation cover by species, along with height and diameter information for tall shrubs and trees, respectively. Total absolute cover equaled 100% for every reference plot. After vegetation cover, tree diameter, and shrub heights were measured, final dominance type and structure determinations were made using the dichotomous key (see Appendix A). For observational plots, dominance type and associated structure determinations were made after a brief assessment of the plot area. Estimates were made from a 'bird's-eye' perspective to mimic that of a remote sensing instrument from above, discounting vegetation that is overtopped. Approximately 40% of the total number of the ground plots collected specifically for this project were the observational type. Figure 4 is a picture taken on a plot above Clarence Strait, near Coffman Cove during the 2018 field data collection campaign.

Additionally, 249 plots were collected by the Nature Conservancy and state of Alaska to calibrate the LiDAR data that was collected in 2017 and 2018. Discrete forest structure metrics—tree canopy cover, Trees Per Acre (TPA), Quadratic Mean Diameter (QMD), and tree size—could then be modeled using the detailed tree inventory data obtained from high-accuracy (sub-meter accuracy) GPS plot locations. The plot information was also cross-walked to provide additional vegetation type reference data.

Young Growth Inventory Data

The Tongass National Forest recently completed a young growth inventory to evaluate the current status and to analyze site productivity of previously harvested stands on Forest Service

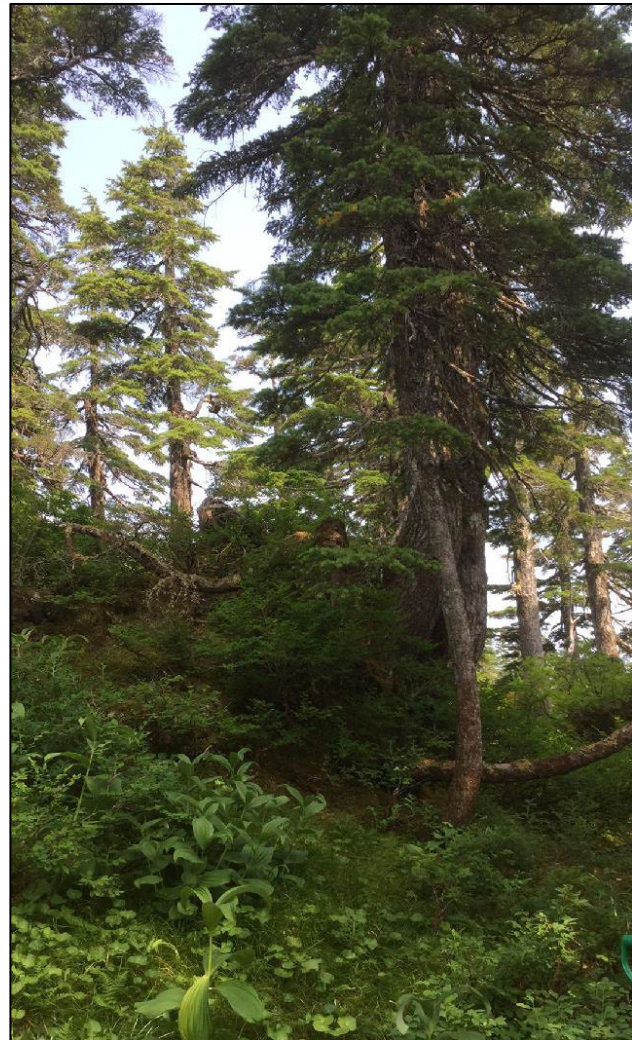


Figure 4. Sampled reference field site of a Mountain Hemlock forest during the 2018 ground data collection campaign.

lands, as the agency transitions from focusing harvest activities in young growth as opposed to old growth timber. These plots used a systematic design and had a density of one plot for every 2.5 acres. Over 8,000 young growth plots fell within the Prince of Wales project boundary and were intersected with the mapping segments. Relative tree canopy cover by species was calculated for all the young growth inventory plots and subsequently used to assign vegetation type labels to the corresponding segments according to the definitions set forth in the classification key. In some cases, multiple plots intersected a single segment and therefore plot data was summarized in these instances. This process yielded a total of 6,475 reference sites that were used in the vegetation type classification models.

Legacy Data

Additional reference sites were derived from alternate field data sources, collected previously for other Forest Service projects. A total of 2,149 plots, comprised mostly of ecology and soils plots, were cross-walked to the vegetation classification system used for this mapping effort and photo interpreted to determine if each site accurately represented vegetation within the corresponding mapping segment. Data collected as part of the FIA program are often used in vegetation mapping projects because of the program's statistically robust, systematic random sampling design, which utilizes fixed-radius plots to inventory forest resources across all ownerships. This project classified 360 FIA plots according to the Prince of Wales vegetation type map unit definitions and were utilized as reference in the prediction modeling.

All legacy data underwent a rigorous quality assurance/quality control process using high resolution imagery and topographic data. Each site was reviewed for adequate representation and homogeneity within the context of the mapping segments. If a site contained relatively uniform vegetation characteristics and the vegetation type map unit could accurately be ascertained, then it was utilized as a reference or validation site in the mapping process.

Photo Interpreted Data

Local experts working for the Tongass National Forest photo interpreted 249 shrubland sites because the project was notably deficient in shrub samples. Interpretation of different shrubland vegetation types were focused in certain geomorphology settings. For the *alder shrub* class, sites were targeted in avalanche chutes and in depositional landscapes, whereas *tall shrub* sites were interpreted on erosional surfaces and areas of recent harvest. Multiple imagery datasets were used for the purpose of interpretation. The high-resolution resource imagery collected in 2019, an imagery mosaic from Digital Globe acquired between 2010 and 2013 (40 cm; visible bands: red, green, and blue), and the Maxar Esri basemap imagery were the most frequently utilized data sources.

Classification and Regression

The 2019 orthoimagery and QL1 LiDAR data were the most important drivers for the vegetation type and structure models. Given that these sources had different coverages, it mandated three modeling extents to accommodate for these discrepancies—1) *LiDAR*; 2) *No LiDAR*; and 3) *No Ortho* (Figure 5). Recently acquired LiDAR data covers approximately 75% of the project's land area and was coincident with all other data sources. Areas outside of the *LiDAR* extent (referred to as the "*No LiDAR*" extent), relied on IfSAR for topographic information, and was coincident with all other data sources, except for a small cluster of islands in the northeastern portion of the project area. Recent orthoimagery was not acquired for these islands, and therefore, SPOT satellite data was substituted (referred to as the "*No*

Ortho” extent). Zonal statistics were generated according to the available geospatial data for these three modeling extents.

Two distinct classification methodologies were used to model vegetation type and vegetation structure, respectively. Field data containing dominance type information was used as reference data in developing the vegetation type random forest classification models, whereas the LiDAR calibration data was used to develop the second order metrics that were subsequently utilized to model vegetation structure. The first order LiDAR metrics were developed using FUSION LiDAR processing software (McGaughey 2009). One of these, the all-returns proportion above 2-meters, was used to model continuous tree canopy cover. A Pearson’s correlation analysis was used to optimize a selection of first order LiDAR metrics to be used as explanatory variables for target second order forest metrics using linear regression (Reynolds 2019). After the forest metrics were generated for the LiDAR extent, they were extrapolated to the remaining project area (*No LiDAR* and *No Ortho* extents) using a random sample of 10,000 coincident sites to relate the LiDAR-generated metrics to zonal statistics derived from IfSAR and spectral data sources. The data mining technique, random forest, was then used to predict and assign vegetation attributes to the mapping polygons for vegetation type and extrapolate the LiDAR model results (Breiman 2001, Cutler et al. 2007). Random forest is an ensemble classifier that uses the plurality vote in the case of classification, or the average of continuous predictions in the case of regression, for the multitude of individual decision trees that make up the ‘forest’ to determine final class assignment or regression output. Final map features, containing both the vegetation type and structure information, were filtered to meet the minimum mapping unit (MMU) of 0.25 acres. Tree canopy cover was derived using the proportion of returns from the point cloud above a 2-meter height threshold within the *LiDAR* extent. Other first order metrics were used to derive meaningful second order metrics, including trees per acre and quadratic mean diameter products. Subsequently, second order LiDAR metrics and spectral data were used to model thematic tree size.

Model prediction utilized summary statistics derived from the satellite imagery, topographic data, and climate information (Figure 5). Zonal statistics, including minimum, maximum, range, standard deviation, mean, and median, were generated for the mapping segments using these layers. This equated to nearly 400 statistics being generated for each of the one million plus mapping segments. Subsequently, these statistics were compiled into a single dataset to be used in the computational modeling. Zonal statistics associated with the reference mapping segments were then used to predict and characterize vegetation across the Prince of Wales project area. A vegetation mask was implemented to restrict the individual vegetation structure predictions to the *forest* map groups only. This mask enforced consistency among the various vegetation map layers and the *Prince of Wales Classification Key*.

Vegetation Type

A separability analysis was performed to indicate which classes were most readily discernible. This informed a mapping hierarchy that grouped classes based on data similarity. The mapping hierarchy determined the sequence in which models were run. Spectrally distinct classes were mapped first, while classes that were more difficult to distinguish were grouped together and subsequently modeled further down the hierarchy (Figure 6). This iterative process of evaluating and re-running classification models at each level of the mapping hierarchy is a sequential operation in which broad vegetation groupings are subsequently further divided until all vegetation types are sufficiently modeled. There are advantages to

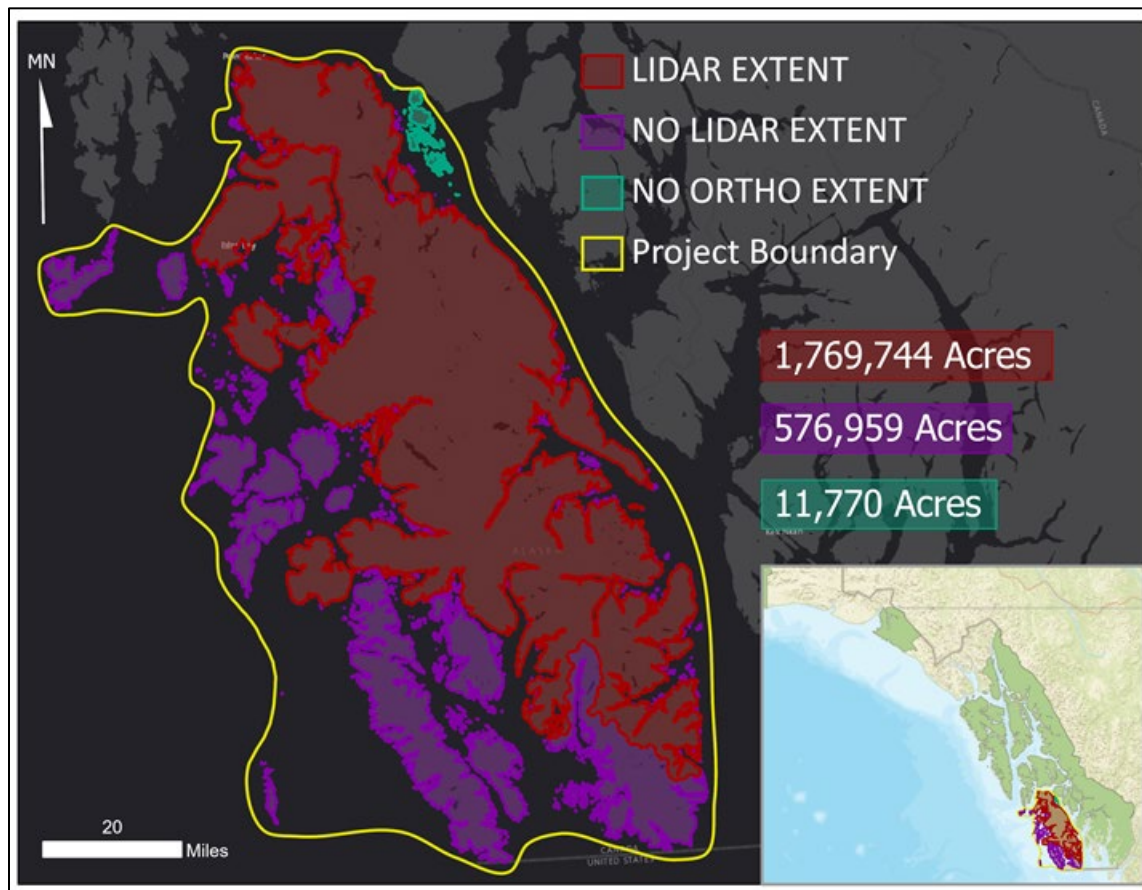


Figure 5. The three modeling extents for the Prince of Wales existing vegetation mapping project: a) LiDAR; b) No LiDAR; and c) No Ortho.

using this hierarchical modeling approach because it enables a targeted review of map outputs at each level, where conspicuous errors can be addressed. For example, the first level of the hierarchy discriminates between *barren/sparse vegetation*, aquatic areas, and vegetated land cover. Model outputs were evaluated and optimized using photo interpretation at each stage of the mapping hierarchy to reduce model confusion and improve overall map accuracy. Note, the *Developed* class was added to the map manually since permanent infrastructure is mostly confined to urban centers and anthropogenic sprawl is difficult to adequately delineate with segmentation in a project of this scale. The *aquatic herbaceous* class was also manually digitized because many of the emergent vegetation beds were absorbed into adjacent water bodies during the segmentation process. Significant linear hydrographic features that were not adequately captured during the segmentation process and had an approximate width of 20-meters or more, were manually digitized to be included as continuous map features classed as *Water*.

The three modeling regions dictated which reference sites could be utilized depending on the data extent being classified. Only the field reference data that intersected the *LiDAR* extent could be used in the classification of *vegetation type* for the LiDAR coverage area. Not only did the LiDAR data cover most of the project area, it also contributed highly detailed topographic and forest metrics that were unique

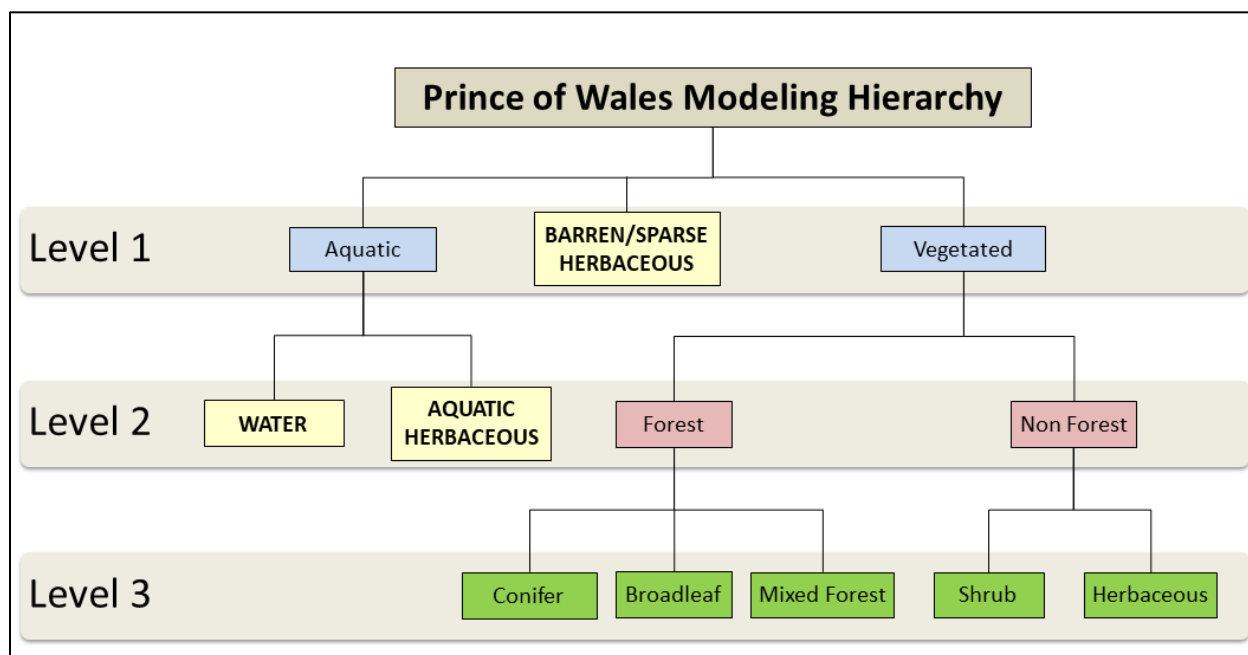


Figure 6. A diagram of the highest levels of the vegetation type modeling hierarchy for the Prince of Wales existing vegetation mapping project. An independent classification model was developed at every node within each level of the modeling hierarchy. For example, two independent classification models were developed for level 3—1. Differentiating the three types of forest; and 2. Distinguishing shrub from herbaceous. Note, Yellow highlighted boxes with emboldened class names in all caps indicate final classes. Other colored boxes indicate similar groupings of vegetation that are further refined at subsequent modeling levels.

and critical for classification. Areas outside of the *LiDAR* extent were able to be modeled with all the available field reference data, since the other geospatial predictor data mostly came from sources that covered the entirety of the project extent. The only exception were the few islands not covered by the 2019 orthoimagery. However, all field reference information was still utilized in this scenario because no reference data were collected on these islands. The final *Vegetation Type* map layer was produced by aggregating the model results for the three data extents. Results were prioritized using a ranking system based on the quality of geospatial data afforded the various modeling extents. The *LiDAR* region was given highest priority, followed by the *No LiDAR* region, which included the recent orthoimagery, and lowest priority was given to the *No Ortho* area that did not contain either the high-resolution *LiDAR* or 2019 orthoimagery.

Structure Modeling

Structure products including Tree Canopy Cover (TCC), Quadratic Mean Diameter (QMD), Trees Per Acre (TPA), and Tree Size were produced for the Prince of Wales study area and enumerated to the project segments. The principal source for these metrics were two phases of *LiDAR* acquisitions from 2017 and 2018, covering 75% of the project's land area. Using these *LiDAR* datasets, structure metrics were extrapolated to the entire project extent using a non-parametric random forest using numerous geospatial data layers as predictors. A random selection of 10,000 polygons with the targeted *LiDAR* metric of interest were used as reference in structure model development. The following sections outline the methods, reference data, and predictor variables used in the derivation of each structure metric.

Tree Canopy Cover

Tree canopy cover was modeled continuously, from 0-100%, using a random forest regression model. Continuous tree canopy cover values were then assigned to map polygons classified as *forest* on the final dominance type map. *Forest* is defined by the dominance type key as any area containing at least 10% tree cover when viewed from above, discounting over-topped trees.

Traditionally, tree canopy cover has been derived from LiDAR data using a variety of modeling approaches, including: 1) thresholding the CHM to derive a binary canopy cover layer; 2) the all-return proportion (Equation 1); or 3) the first-return proportion (Equation 2) (Arumäe and Lang 2018, Hopkinson and Chasmer 2009, van Leeuwen and Nieuwenhuis 2010, Smith et al. 2009, Wasser et al. 2013). All three of these approaches were explored, and ultimately, the all-return proportion most accurately depicted tree canopy cover over the range of ground conditions. Additionally, a height threshold must be identified that best differentiates trees from surrounding vegetation. An inverse relationship exists between canopy cover and threshold height. As threshold height increases, canopy cover decreases due to a smaller proportion of total returns, or cells, within a given segment occurring above the identified threshold value, and therefore, less are classified as canopy (Figure 7). A 2-meter height threshold was chosen to best make this distinction, however, inevitably some *tall shrubs* were errantly counted, and some stunted trees were discounted in mixed polygons mapped as *forest*.

Equation 1.

$$\text{all-return proportion} = \frac{\# \text{ all returns } > 2\text{m}}{\# \text{ all returns (total)}}$$

Equation 2.

$$\text{1st-return proportion} = \frac{\# \text{ 1st returns } > 2\text{m}}{\# \text{ 1st returns (total)}}$$

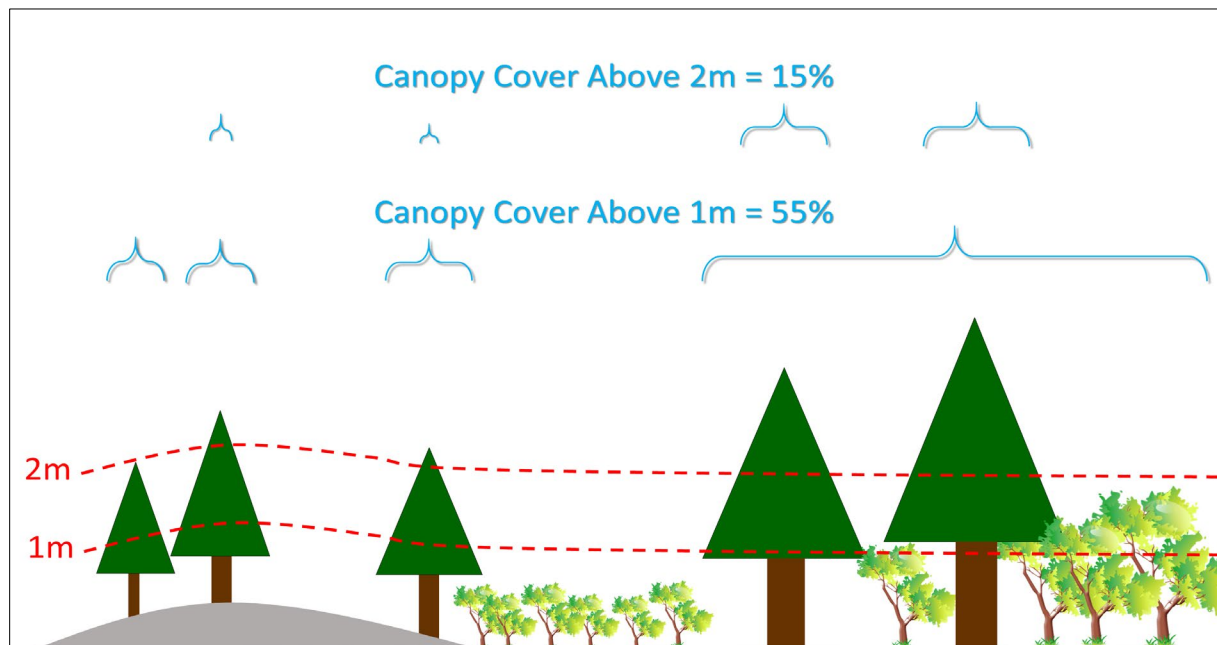


Figure 7. The top brackets represent areas considered cover as estimated by a 2-meter CHM threshold (top dashed line). The bottom set of brackets represent areas considered cover as estimated by a 1-meter CHM threshold (bottom dashed line).

After a qualitative examination of each tree canopy cover approach, the all-return proportion cover metric, obtained from the first order LiDAR metrics, was found to be the best method for determining tree canopy cover within the *LiDAR* extent. Comparisons showed the first return proportion approach overestimated tree canopy cover and skewed values to 100%. Counter to the first-return proportion, which utilizes only the first return of each beam, the all-returns proportion utilizes all returns above and below a specified threshold.

A random forest classifier was used to extend the tree canopy cover model into regions of the Prince of Wales that did not have LiDAR coverage. The random forest used a random sample of 10,000 segments extracted within the *LiDAR* extent. Each polygon sample included the derived tree canopy cover value, which was used to train and validate the random forest model. In the *No Lidar* extent, 2019 ortho imagery and the IfSAR CHM were used as important predictors for tree canopy cover along with numerous other data sources. However, in the area of Prince of Wales, where ortho imagery was not acquired, SPOT imagery was used in its place. Results for the three modeling extents were mosaicked together to create a seamless, continuous (0-100%) tree canopy cover product for the entire Prince of Wales project area.

Trees Per Acre

The forest metric, Trees Per Acre (TPA), is a good approximation for the number of trees that meet a certain size requirement within a given area. Regions with a relatively high TPA, often indicates a high number of small trees such as saplings, whereas a treed area with a low TPA indicates a low density of large trees. TPA was calculated under two conditions, the first included TPA for all trees greater than or equal to 1-foot tall ($\geq 1'$ tall) and the second included TPA for all trees greater than or equal to six inches in diameter at breast height ($\geq 6''$ dbh). Using first order LiDAR metrics, the following equation was used to calculate TPA $\geq 1'$ tall:

$$\text{TPA } (\geq 1' \text{ tall}) = e^{(9.002736 + (0.013568 * \text{Percent_first_returns_above_2}) + (-1.544788 * \ln_elev_p20))}$$

For larger trees, where TPA $\geq 6''$ dbh, the following equation was used:

$$\text{TPA } (\geq 6'' \text{ dbh}) = (6.18864 + (0.19670 * \text{Percent_first_returns_above_mean}) + (4.06816 * \ln_elev_CV))^2$$

After the models were applied, a random selection of 10,000 polygons within the *LiDAR* extent were used as reference to extend the two TPA models to the *no LiDAR* and *no Ortho* modeling extents. TPA values were assigned to mapping polygons classified as *forest* on the final vegetation type map.

Quadratic Mean Diameter

Quadratic Mean Diameter (QMD) is a measure of central tendency for characterizing a group of trees that have been measured and is an excellent proxy for tree size and volume in an area. QMD was calculated for all trees greater than or equal to two inches in diameter at breast height ($\geq 2''$ dbh) and for merchantable timber greater than or equal to nine inches in diameter at breast height ($\geq 9''$ dbh) using first order LiDAR metrics coupled with *in situ* plots. There was a total of 237 plots collected that contained live trees at least two inches in diameter and 194 plots that contained live trees that at least nine inches in diameter.

QMD ($\geq 2''$ dbh) used Percent Cover Above Mean Height and the 50th percentile height (elev_p50) to predict QMD with the following final equation:

$$QMD (\geq 2'' \text{ dbh}) = e^{(0.44478 + (-0.01236 * \text{Percent_all_return_above_mean}) + (0.094337 * \ln_elev_p50))}$$

Using the 194 validation plots, QMD ($\geq 9''$ dbh) found the Percent Cover Above Mean Height and the 75% percentile (elev_p75) to be the best first order LiDAR metric predictors with the following equation:

$$QMD (\geq 9'' \text{ dbh}) = e^{(1.76810 + (-0.00320 * \text{Percent_all_returns_above_mean}) + (0.27402 * \sqrt{\text{elev_p75}}))}$$

After the models were applied, a random selection of 10,000 polygons within the *LiDAR* extent were used as reference to extend the two QMD models to the *no LiDAR* and *no Ortho* modeling extents. QMD values were assigned to mapping polygons classified as *forest* on the final vegetation type map.

Tree Size

The thematic tree size (TS) layer depicts four categorical classes: 1) TS1 Sapling (<4.5' tall - 4.9" dbh); 2) TS2 Pole (5" - 8.9" dbh); 3) TS3 Medium (9" - 20.9" dbh); and 4) TS4 Large ($\geq 21''$ dbh). Areas where tree cover was < 10% the class "Non-Tree" was assigned. Tree size is defined as the plurality diameter class forming the uppermost canopy layer when viewed from above, discounting over-topped trees. Tree diameter was measured at breast height (DBH), 4.5 feet above the ground. The thematic tree size product is different from QMD and TPA in that it is not derived directly from the first order LiDAR metrics, but rather is modeled using 469 reference field sites collected for this project. Second order LiDAR metrics (TPA, basal area, biomass, and QMD), that were extrapolated to the entire project area, were used in conjunction with the resource imagery, Sentinel 1 SAR, SPOT, Sentinel 2, and IfSAR data to produce a random forest tree size classification model for all three data extents.

Map Revision

After initial models were reviewed and optimized by GTAC personnel, draft versions of five map layers were created: 1) vegetation type; 2) tree canopy cover; 3) trees per acre ($\geq 1'$ tall); 4) quadratic mean diameter ($\geq 2''$ dbh); and 5) tree size. These layers were provided to local and regional experts for review within a web application that provided a platform by which any edits and feedback could be submitted.

In order to provide a more accurate mid-level existing vegetation map describing overall vegetation patterns, during the draft map review process, a web map of the draft map was published on ArcGIS Online for collaborators to review. Consistent with *vegetation type* definitions provided by the *Prince of Wales Dominance Type Key*, collaborators were asked to identify and delineate major discrepancies between what was depicted on the maps and existing ground conditions. Specific attention was paid to systemic misclassification and over/under classification of specific vegetation types. In addition, collaborators were asked to focus on general vegetation distribution and structural patterns in order to determine if the overall community types and structure information that occur throughout the project area were accurately represented. Using a polygon drawing tool within the web service application, reviewers spatially delineated and described the reason for a proposed edit within an associated comment field for areas of concern. Edits ranged from small-scale discrepancies, like a rock pit being misclassified as *Western Hemlock* forest, to more systemic issues, such as the noted overabundance of *Sitka Spruce*. After all web edits were completed and uploaded, an analyst then addressed each of the 171 edits made by Forest personnel.

Feedback on two, broad-scale, landscape pattern observations were provided by the local experts after their review of the vegetation type map layer. These impactful observations were: 1) the *Sitka Spruce*

vegetation type was broadly over classified, especially at the expense of the *Sitka Spruce-Western Hemlock* vegetation type; and 2) areas of recent harvest that were not being captured by the map layers or being misclassified as *Dwarf Conifer*.

In order to affect change that significantly altered the amount of *Sitka Spruce* mapped, the balance of reference data needed to shift to provide a proportionally higher number of spruce sites. *Young growth inventory plots* containing relative canopy cover values for individual tree species provided analysts with detailed information to review these sites within the context of their associated mapping segments. *Sitka Spruce young growth* reference segments, interpreted to contain substantial *Western Hemlock* during the reference data review process, were reassigned to the *Sitka Spruce-Western Hemlock* class. Subsequently, these sites changed the balance of reference data for these classes and were used to update the forest model that discriminated the various types of conifers. The updated model resulted in substantially less *Sitka Spruce* being mapped as compared to prior model iterations. Reviewers also noted that many north-facing, shadowed slopes were being erroneously classified as *Sitka Spruce*. A GIS ruleset was developed to select spruce stands that had a northerly aspect (ranging from 337.5 degrees Northwest to 22.5 degrees Northeast) and had an average slope steeper than 30 degrees. Subsequently, these areas classified as *Sitka Spruce* were reclassified as *Sitka Spruce-Western Hemlock*. These changes reduced overall *Sitka Spruce* acreage by more than 80,000 acres and resulted in a net increase in *Sitka Spruce-Western Hemlock* by approximately 100,000 acres in the final map as compared to the draft.

Accurately depicting dynamic landscapes, like areas of active management and recent harvests, is extremely challenging because the vegetation models rely on data that was acquired on different days and different years. Consequently, if there are landscape changes following a specific data acquisition, such as a harvest occurring after a LiDAR acquisition, current conditions are not captured, and other methods are needed to reconcile these inconsistencies. These map products attempt to depict ground conditions during the summer of 2019, consistent with what can be observed in the 2019 30-cm orthoimagery. Ground conditions at the height of the 2019 growing season are the project standard by which all map layers attempt to accurately depict. Recent harvest areas were identified to account for temporal discrepancies between geospatial datasets, namely the LiDAR and the 2019 orthoimagery. A model was developed for the purpose of detecting post-LiDAR harvests in the Prince of Wales Island study area, where trees were harvested after the acquisition of the QL1 LiDAR (acquisition date range: 2017-05-26 to 2018-10-18) and prior to the acquisition of the 30 cm orthorectified imagery (acquisition date range: 2019-06-08 to 2019-08-31). The model was sensitive to areas where the LiDAR data indicated a standing forest but the orthoimagery spectral data reflected a barren or sparsely vegetated ground condition, indicating a recent clear cut. These recent harvests were then updated to either the *Barren/Sparse Vegetation* class, when there was low NIR reflectance, or to the *Tall Shrubs* class, when NIR reflectance indicated new vegetation. It is known that blueberry and other *tall shrub* species commonly are the first successional inhabitants after a harvest.

Few edits were applied to the structure layers, aside from the zeroing of structure attributes to depict the current *non-forest* condition in recent harvest units, because few corrections were provided by local experts and there was a high degree of confidence in these products, especially within the *LiDAR* extent. Most updates to structure attributes involved reconciling the changes made to the *vegetation type* layer to maintain consistency between attribute relationships. Recent harvest areas misclassified as *dwarf*

conifer, where the short stature of forest regeneration was confused with stunted conifers that inhabit low-productivity sites, were manually changed to match forest types in adjacent areas. Most of this confusion was localized because the spectral predictors were usually sensitive to peatlands indicative of poor soil drainage and outweighed the short structure signals of the forest metrics used in the model.

Final Data Products

The following sections describe each of the final data products. Acreage summaries and validation statistics are reported to provide a more comprehensive understanding of the final map products for the Prince of Wales existing vegetation mapping project. These product descriptions should assist map users in how to best leverage these map products for future work and justify their use.

Vegetation Type

The final vegetation type map consisted of 18 land cover classes contained within six map groups: eight *conifer forest* types; one *broadleaf forest* type; one *mixed forest* type; three *shrub* types; two *herbaceous* types; and three *other* types (Figure 8). Of the total vegetated area, the *forest* map groups encompass 96% of the vegetation on the Prince of Wales, while *shrub* covers 3% and *herbaceous* types cover 1% of the vegetated area. A list of the vegetation type map classes with their associated areal extent across the Prince of Wales is tabulated in Table 3.

Vegetation Type and Map Group Validation

An accuracy assessment was conducted to validate the existing *vegetation type* data from a spatial standpoint and helps to reveal details of individual class confusion. The fundamental modeling units for this project were the segments (mapping polygons), therefore that is the unit by which the map was validated. Prior to building the predictive classification models, a random subset of the reference data provided for this project were withheld to later serve as an independent validation sample for accuracy assessment purposes. Thirty reference sites were withheld for each vegetation type class, except for the *low shrub* class, which only had 20 sites withheld because of low sample numbers. No reference data were withheld for the *barren/sparse vegetation* and *water* classes since they could be accurately photo interpreted. These classes were modeled and validated using photo interpreted reference data. Only inland waterbodies were evaluated as a part of this accuracy assessment because of the large amount of ocean within the project boundary. Note, the *developed* class was not evaluated because it is inherently difficult to model, therefore major roadways and urban centers were manually digitized into the final map and includes a level of subjectivity by an analyst. Ultimately, a total of 500 reference sites were randomly selected, regardless of reference data source, and used to develop error matrices at the *map group* and *vegetation type* map levels (Table 4). It is critical to understand that although accuracy assessment provides important insights and can inform further investigations, the field sampling used for validation was confined to areas with proximity to roadways, shorelines, easily accessed foot trails, and old plot designs. Therefore, accuracy measures are biased and better thought of as an agreement that elucidates class relationships.

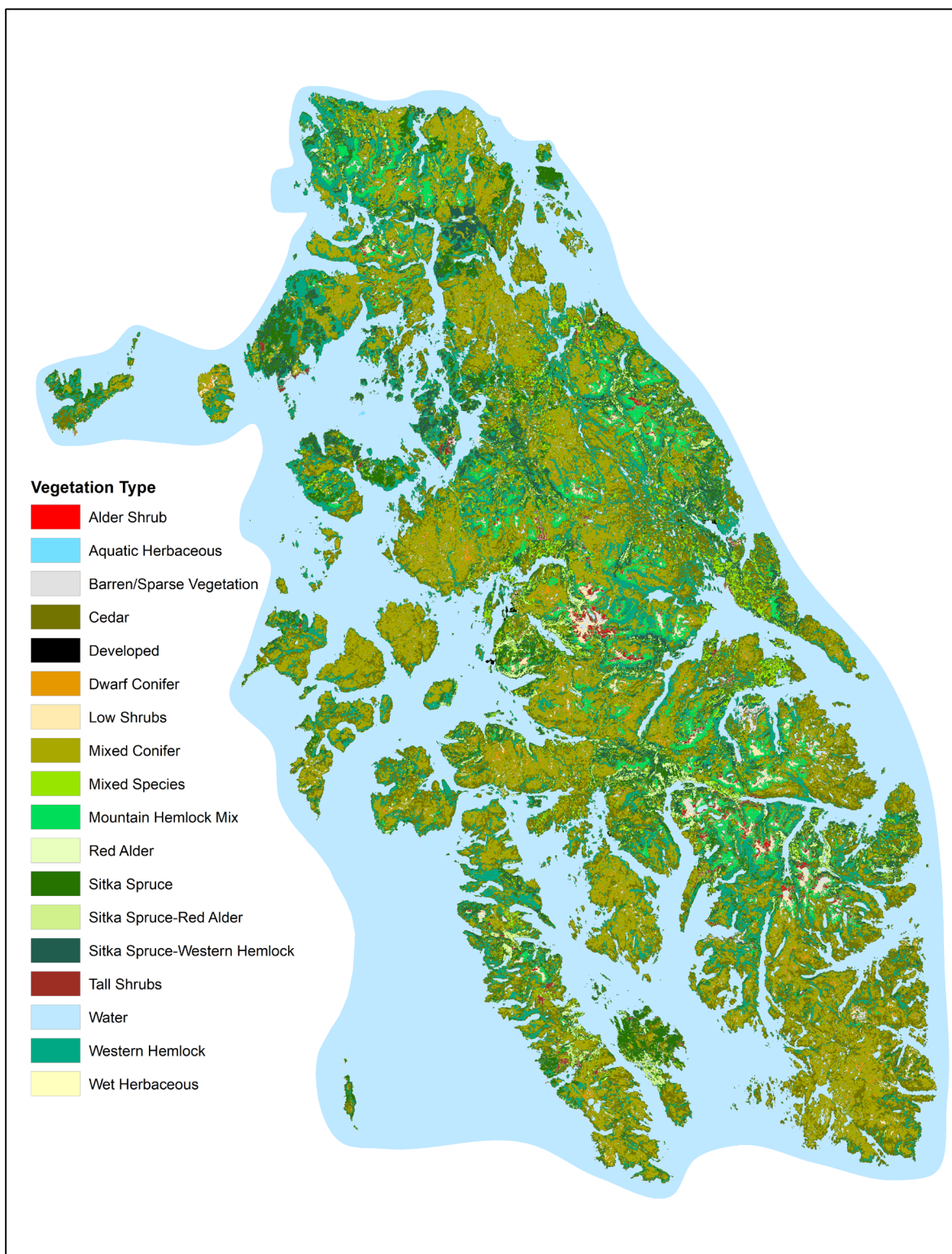


Figure 8. Map of the vegetation types across the Prince of Wales project area.

Table 3. Map group and dominance type acreage summary for the Prince of Wales existing vegetation map.

Map Group	Area (ac)	Percent of Area	Vegetation Type	Acres	Percent of Area	Percent of Vegetated Area
Conifer Forest	2,111,565	50.2%	Mixed Conifer	648,365	15.4%	28.6%
			Cedar	474,462	11.3%	21.0%
			Western Hemlock	410,456	9.8%	18.1%
			Sitka Spruce	197,647	4.7%	8.7%
			Sitka Spruce-Western Hemlock	160,925	3.8%	7.1%
			Mountain Hemlock Mix	92,042	2.2%	4.1%
			Mixed Species	69,183	1.6%	3.1%
			Dwarf Conifer	58,485	1.4%	2.6%
Broadleaf Forest	8,127	0.2%	Red Alder	8,127	0.2%	0.4%
Mixed Forest	54,859	1.3%	Sitka Spruce-Red Alder	54,859	1.3%	2.4%
Shrub	74,505	1.7%	Low Shrubs	38,190	0.9%	1.7%
			Tall Shrubs	31,166	0.7%	1.4%
			Alder Shrub	5,149	0.1%	0.2%
Herbaceous	15,040	0.4%	Wet Herbaceous	12,115	0.3%	0.5%
			Aquatic Herbaceous	2,925	0.1%	0.1%
Total Vegetated Area				2,264,097	53.8%	100%
Other	1,945,478	46.2%	Water	1,902,527	45.2%	
			Barren/Sparse Vegetation	39,089	0.9%	
			Developed	3,862	0.1%	
Total Area	4,209,574	100%	Total Area	4,209,574	100%	

Table 4. Source of validation data used in map group and vegetation type error matrices.

Reference Data Source	Count (n)
FIA	6
Lidar Calibration	8
Legacy	74
Ground	95
Photo Interpreted	154
Young Growth	163
Total	500

The 500 validation sites were used to develop two error matrices to evaluate thematic cover type agreement—the first for Map Group (Table 5) and the other for final Vegetation Type (Appendix B).

These tables display the agreement between the reference data (ground truth) and the map data (model prediction). Overall accuracy is expressed as a percentage and is calculated by dividing the number of sites that agreed with the reference data classification by the total number of reference data sites. The kappa statistic is a measure of how well the classification performed in comparison to a random classification assignment. Overall accuracy was 89% at the map group-level (Table 5).

Table 5. Error matrix of the Prince of Wales existing vegetation map at the map group-level.

Map Group		Reference Data						User's Accuracy	Commission Error
		Broadleaf Forest	Conifer Forest	Mixed Forest	Shrub	Herbaceous	Other		
Map Data	Broadleaf Forest	20	2	1	1	0	0	83%	17%
	Conifer Forest	1	221	8	0	2	0	95%	5%
	Mixed Forest	9	4	21	0	0	0	62%	38%
	Shrub	0	13	0	78	9	0	78%	22%
	Herbaceous	0	0	0	0	48	0	100%	0%
	Other	0	0	0	1	1	57	95%	5%
Producer's Accuracy		67%	92%	70%	96%	80%	100%	Kappa	0.85
Omission Error		33%	8%	30%	4%	20%	0%	Overall Accuracy	89%
								Area-Weighted Accuracy	95%

Overall accuracy is the most comprehensive statistic when it comes to understanding the underlying reliability of a map product. It is calculated by taking the proportion of sites classified correctly divided by the total number of sites assessed for each product. Numerous factors impact classification accuracy, including: 1) classification complexity; 2) landscape complexity; and 3) quality of the reference and geospatial data that is available for a project. Map accuracy has an inverse relationship with classification complexity, meaning that the more classes you have the less accurate your classification output will be. Considering this, the overall vegetation type class accuracy was 64% (Appendix B). This level of agreement is consistent with results from other mid-level vegetation mapping projects.

Error matrices with individual class accuracies were computed for the *vegetation type* and thematic *tree size* map products. There are two ways to analyze individual class accuracy: 1) producer's accuracy, which is the proportion of sites correctly mapped for that class to the total number of sites of that class as determined by the reference data (i.e., the column total); and 2) user's accuracy, which is the proportion of sites correctly mapped for that class to the total number of sites assigned that particular class (i.e., the row total) (Congalton 1991). Producer's accuracy provides a measure of omission error that describes the probability that an area on the ground is mapped correctly. User's accuracy provides a measure of commission error that describes the probability that a mapped class reflects reality on the ground. For example, *dwarf conifer* had a producer's accuracy of 30% but had a user's accuracy of 90% (Appendix B). This indicates that this class was under-mapped because of the relatively high omission error. Confusion with the *mixed conifer* and *low shrubs* classes comes at the expense of the *dwarf*

conifer class. This makes sense in the context of the local environment, since all three of these classes occur in similar ecological niches that are characterized by poorly drained soils. Within these highly transitional peatland areas, classification readily changes depending on subtle differences in soil drainage and productivity governing local tree density and tree size. This illustrates how studying the error matrices can provide insight not only into the reliability of an individual map class, but also into how and where confusion occurs. Information like this could be leveraged by map users to further investigate questions of interest through a deeper understanding of the classification and the final map products.

Calculating an area-weighted accuracy that considers the relative proportion, or abundance, of the individual classes that comprise the maps provides a more representative measure of overall map quality. The assessment discussed in the previous paragraph utilized a sample that was stratified, in order to adequately sample each cover type, and biased by accessibility, and therefore the distribution of assessment sites did not correspond to the relative proportions of the cover types found across the project area. This means that overall accuracy could be disproportionately influenced by rarer classes or by classes more easily accessed. To account for this, overall area-weighted accuracies were calculated by taking the proportion of correctly classified accuracy assessment sites for each class (the individual class user's accuracies) and multiplying them by the proportion of the total area that the class occupies on the final map (the area weight factor) and summing across every mapped class. Although the true relative abundance of each class across the mapped area cannot be known, the user's accuracy is the best proxy to estimate the distributions of the various classes. Both overall accuracy measures were reported since the area-weighted measure is going to be comparatively inflated since the most common classes are usually modeled more accurately and don't necessarily contain vegetation, such as *water*. For example, the overall area-weighted accuracy was 68% at the dominance type-level and 89% at the map group-level, as opposed to 64% and 95%, respectively (Appendix B and Table 5).

When studying the error matrices, even classes with relatively low accuracies may still provide valuable spatial information regarding vegetation assemblages of interest. Correct interpretation of the error matrices allows a user to apply expert knowledge of known plant associations to discriminate between errors caused by completely erroneous model associations and those that were logical confusions. For example, the *sitka spruce-western hemlock* class can be confused, however nearly all the class confusion occurs with other *conifer forest* types. Further inspection reveals that errors of omission are due to misclassification of *sitka spruce-western hemlock* forest as either *sitka spruce* or *western hemlock*, since only narrow differences in relative tree cover of these species distinguish the three classes. Most errors of commission for the *sitka spruce-western hemlock* class occur within these classes and confusion with the *mixed species* class. The *mixed species* class commonly contains a component of *sitka spruce* and *western hemlock* trees and is defined by the *Prince of Wales Classification Key* as being "actively managed", which implies regenerating vegetation that is highly transitional and difficult to classify. Depending on a user's needs, perhaps these classes could be strategically combined to still answer questions of interest while minimizing confusion. Such confusion is common when discrete decision rules are applied to a continuous landscape. Although critical thinking may be necessary to tease out meaningful information and gain a comprehensive understanding of class relationships, individual class accuracy numbers, when taken by themselves, do not tell the whole story.

Tree Canopy Cover

Continuous tree canopy cover was mapped across the entire project area and indicates cover for all areas classified as *forest*, as defined by the '*Prince of Wales Vegetation Dominance Type Key*' (Figure 9). Tree canopy cover is assessed as the total tree cover as viewed from above, discounting overtopped trees. Continuous canopy cover was binned into three general categories in order to provide an acreage summary across the project extent: 1) Woodland 10-24%; 2) Open 25-59%; and 3) Closed 60-100% (Table 6). Note that the tree canopy cover map itself depicts continuous tree canopy cover values from 10 to 100%, so there is highly detailed information on the map that is not included in the thematic acreage summary provided below. All areas containing less than 10% tree canopy cover during the summer of 2019 are assigned a value of zero because these areas are considered *non-forest* by our existing vegetation classification definitions.

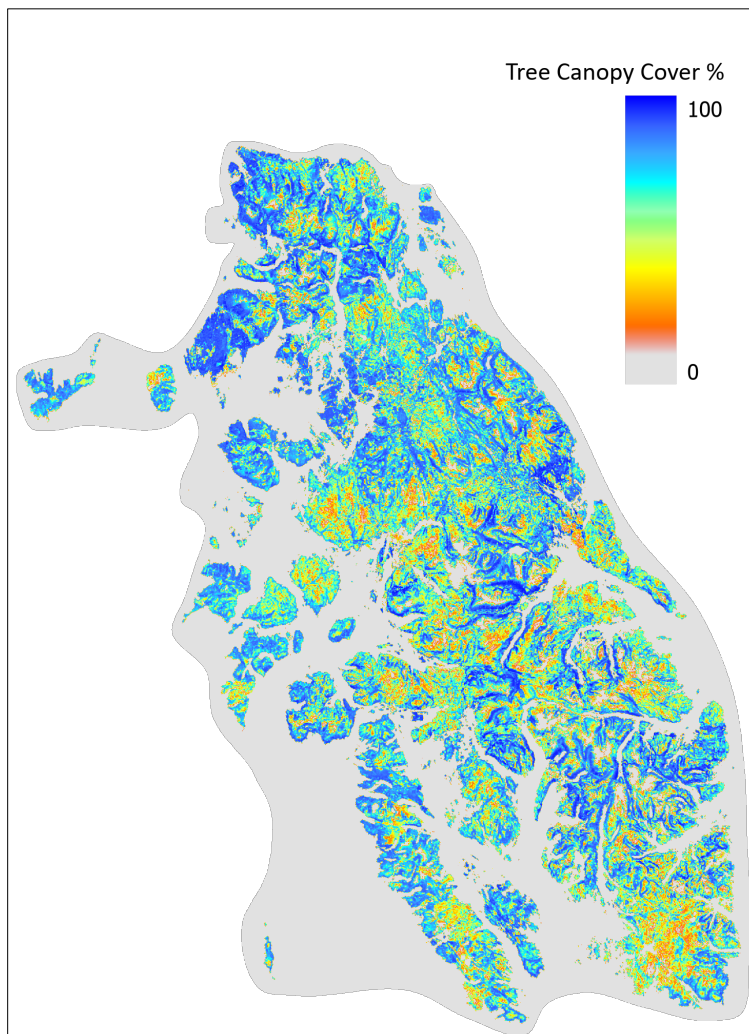


Figure 9. Tree canopy cover across the Prince of Wales project area.

Table 6. Tree canopy cover acreage summary for the Prince of Wales existing vegetation map.

Tree Cover Class	Area (ac)	% Forest Area
Woodland (10 - 24%)	135,425	6.2%
Open (25 - 59%)	500,805	23.0%
Closed (60 - 100%)	1,538,321	70.7%
Total	2,174,551	100%

Tree Canopy Cover Validation

The conglomeration of the various vegetation structure outputs for each of the different modeling scenarios resulted in seamless data products that depicted vegetation structure patterns across the entirety of the *Prince of Wales* study area. The all-return proportion at a 2-meter height threshold was the first order LiDAR metric used to determine tree canopy cover within the *LiDAR* extent. A random decision forest model was subsequently produced using a random selection from the LiDAR model to extrapolate the LiDAR tree canopy cover model over the remaining project extent. A total of 10,000 random points was generated within the *LiDAR* extent to train the random forest classifier. An independent random sample of 500 points was used to compare the predicted tree canopy cover for the *no LiDAR* and *no Ortho* models with the observed tree canopy cover from the *LiDAR* model (Figure 10). Note that because tree canopy cover calculated from LiDAR data is known to be reliable, the all-return proportion model result was used as the observed tree canopy cover value or truth when evaluating models predicting tree canopy cover outside of the *LiDAR* extent. The *no LiDAR* region produced an r^2 value of 0.94 and the *no Ortho* area produced an r^2 of 0.88. This means that 88% of the variance was explained by the *no Ortho* model. The final continuous tree canopy layer was assembled by combining the model results from the *LiDAR*, *no LiDAR*, and *no Ortho* extents such that priority was placed on the *LiDAR* output, followed by the *no LiDAR* output, and finally, the *no Ortho* output (see Figure 5). This order of priority was dictated by ranking model confidence given the available predictor data for each respective model extent.

Trees Per Acre

Two structure layers depicting trees per acre (TPA) at two size thresholds—all trees over 1-foot tall ($\geq 1'$ tall) and all trees greater than or equal to six inches in diameter at breast height ($\geq 6''$ dbh)—were developed for the *Prince of Wales* project area (Figure 11). TPA values were assigned to those map features classified as *forest* on the final *vegetation type* map. Trees per acre were binned for both TPA products to produce acreage summaries across the study area (Table 7).

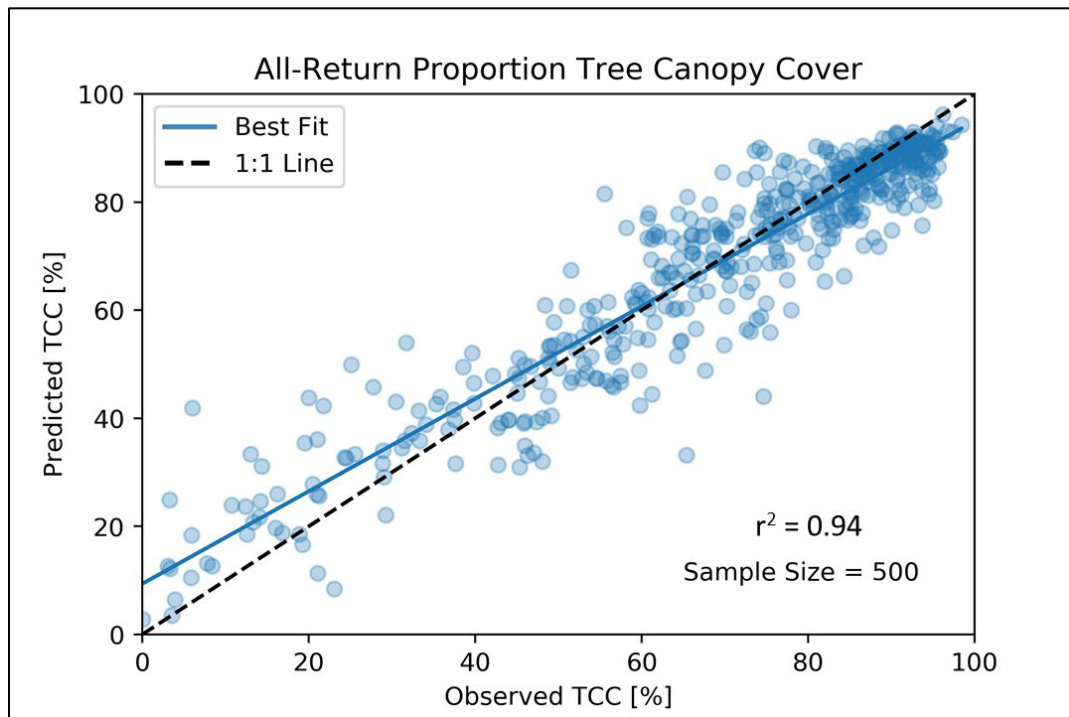


Figure 10. Scatter plot of observed versus predicted percent tree canopy cover for the no LiDAR data extent random forest model. The all-returns proportion, derived from the LiDAR point cloud, were considered to be truth and served as the 'observed' tree canopy cover values.

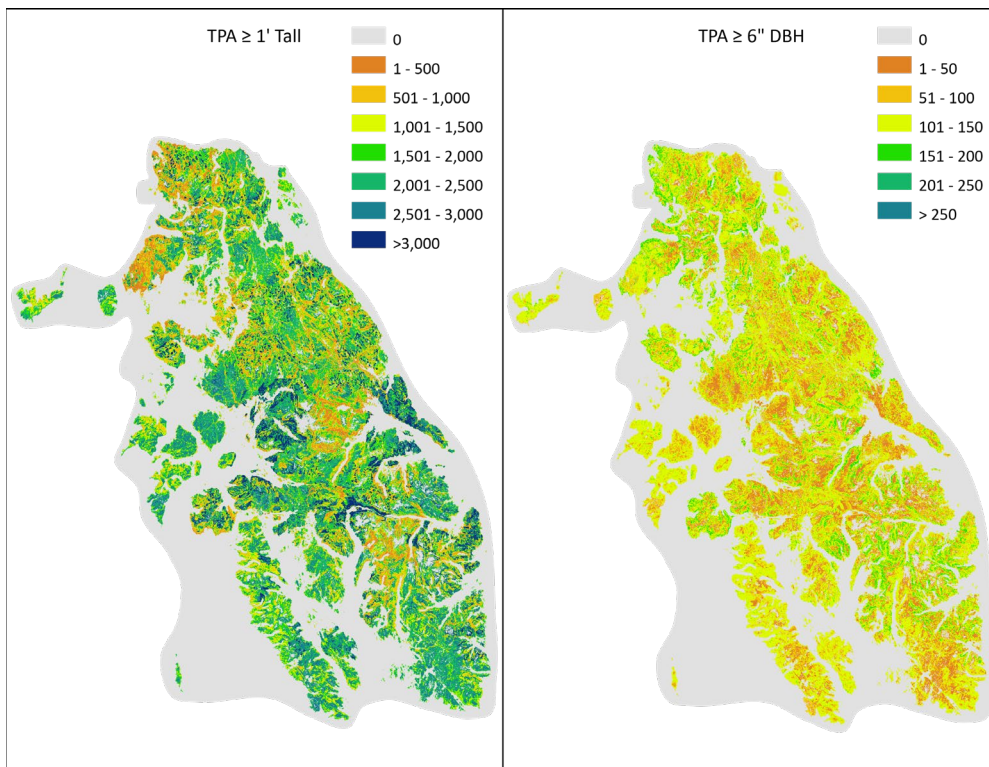


Figure 11. Trees per acre for all live trees $\geq 1'$ tall (left) and all live trees $\geq 6''$ dbh (right) across the Prince of Wales project area.

Table 7. Trees per acre acreage summary for the Prince of Wales existing vegetation map.

Trees Per Acre (TPA) ≥ 1' tall	Area (ac)	% Forest Area
1 – 500	106,116	4.9%
501 – 1,000	356,996	16.4%
1,001 – 1,500	336,462	15.5%
1,501 – 2,000	368,474	16.9%
2,001 – 2,500	518,283	23.8%
2,501 – 3,000	332,048	15.3%
> 3,000	156,172	7.2%
Total	2,174,551	100%
Trees Per Acre (TPA) ≥ 6" DBH	Area (ac)	% Forest Area
0 - 50	293,241	13.5%
51 - 100	465,884	21.4%
101 - 150	1,180,235	54.3%
151 - 200	234,728	10.8%
201 - 250+	464	0.0%
Total	2,174,551	100%

Trees Per Acre Validation

The linear regression models for TPA ≥ 1' tall and TPA ≥ 6" dbh produced an overall r^2 of 0.35 and 0.59, respectively. No RMSE was reported for TPA ≥ 1' tall, but the TPA ≥ 6" dbh produced an RMSE of 58.8 trees. Despite lower model performance as compared to the other continuous forest metrics, the TPA models produced moderately correlated model results that do capture tree density patterns across the entire project area. *For more information on these and the QMD models, please reference [The Nature Conservancy report \(Reynolds 2019\)](#).* Once the linear regression models were finalized for the LiDAR extent, the model was extended to the *no LiDAR* and *no Ortho* areas using 10,000 samples from the LiDAR model as reference. TPA ≥ 1' tall produced an r^2 of 0.67 and 0.68 in the *no LiDAR* and *no Ortho* extents, respectively, when compared to the LiDAR linear regression models. Model performance improved for TPA ≥ 6" dbh with a 0.75 correlation statistic (r^2) for both *no LiDAR* and *no Ortho* extents (Table 8).

Table 8. The r^2 and RMSE values for the structure models outside of the LiDAR extent.

Forest Metric	No LiDAR		No Ortho	
	r^2 [%]	RMSE	r^2 [%]	RMSE
TPA ≥ 1' tall	67	541	68	541
TPA ≥ 6 DBH"	75	22	75	22

Quadratic Mean Diameter

Two different structure layers were developed for quadratic mean diameter (QMD) at two diameter thresholds—trees with at least a 2-inch diameter ($\geq 2''$ dbh) and for merchantable timber with at least a 9-inch diameter ($\geq 9''$ dbh) (Figure 12). QMD values were assigned to those map features classified as *forest* on the final *vegetation type* map. Quadratic mean diameter was binned for both products to produce acreage summaries across the study area (Table 9).

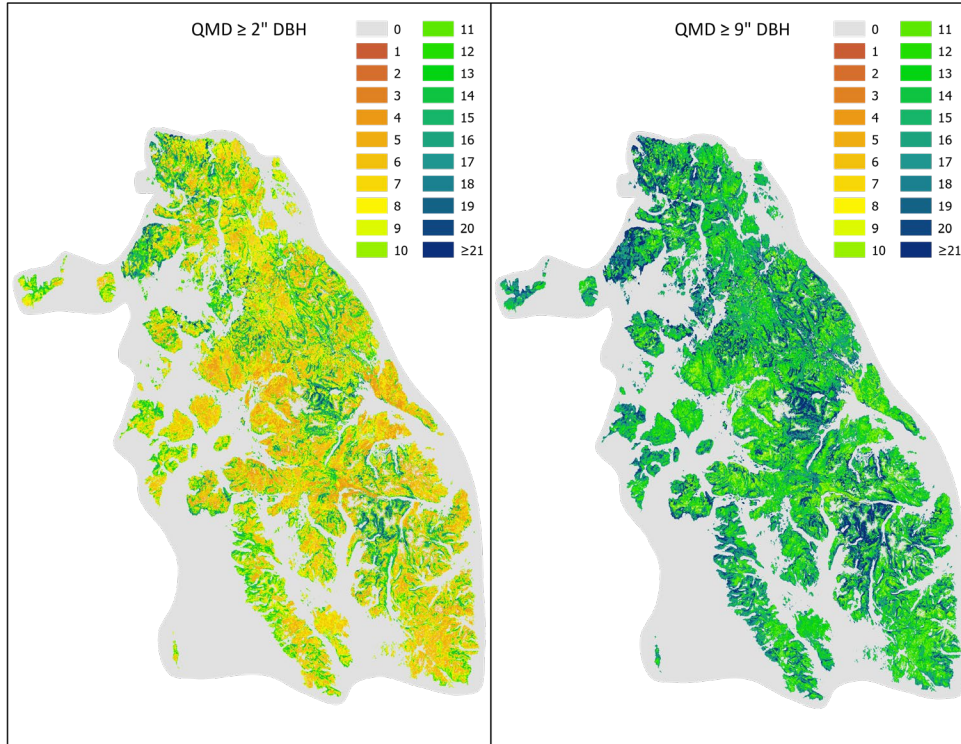


Figure 12. Quadratic mean diameter for all live trees $\geq 2''$ dbh (left) and all live trees $\geq 9''$ dbh (right) across the Prince of Wales project area.

Table 9. Quadratic mean diameter summary for the Prince of Wales existing vegetation map.

Quadratic Mean Diameter (QMD)	$\geq 2''$ DBH Area (ac)	$\geq 2''$ DBH % Forest Area	$\geq 9''$ DBH Area (ac)	$\geq 9''$ DBH % Forest Area
< 5"	93,646	4.3%	0	0.0%
5 - 8.9"	1,023,409	47.1%	5	0.0%
9 - 14.9"	844,532	38.8%	1,127,328	51.8%
15 - 20.9"	202,793	9.3%	852,306	39.2%
$\geq 21''$	10,172	0.5%	194,913	9.0%
Total	2,174,551	100%	2,174,551	100%

Quadratic Mean Diameter Validation

The QMD LiDAR linear regression models produced an overall accuracy where r^2 is 0.75 for QMD $\geq 2''$ dbh and 0.65 for QMD $\geq 9''$ dbh. The Root Mean Square Errors (RMSEs) were 2.7 inches and 5.3 inches for the 2-inch and 9-inch diameter thresholds, respectively. After calculating the QMD structure metrics within the LiDAR extent, the QMD models were extrapolated to areas not covered by the LiDAR acquisitions. Like the methods employed to extend the tree canopy cover and TPA models, QMD metrics were extrapolated to the *no LiDAR* and *no Ortho* regions of the *Prince of Wales*. An important distinction needs to be made between structure metrics derived within the LiDAR extent versus outside. Where recent LiDAR data was available, these forest metrics have a high degree of confidence given that LiDAR is the standard when it comes to topographic data sources. Outside the LiDAR extent, the models had to utilize less reliable IfSAR data paired with spectral information. Although areas modeled without LiDAR coverage are going to be less dependable, given the significant amount of reference data (10,000 sites) afforded data overlap across the large LiDAR coverage area, the forest metric models do capture structure patterns adequately outside the LiDAR extent. These extrapolation models yielded an r^2 of 0.73 for QMD $\geq 2''$ and 0.75 for QMD $\geq 9''$ in the *no LiDAR* extent. Model performance in the *no Ortho* extent remained the same at 0.73 for QMD $\geq 2''$ but slightly decreased to 0.72 for QMD $\geq 9''$ (Table 10).

Table 10. The r^2 and RMSE values for the structure models outside of the LiDAR extent.

Forest Metric	No LiDAR		No Ortho	
	r^2 [%]	RMSE	r^2 [%]	RMSE
QMD $\geq 2''$ DBH	73	2.21	73	2.22
QMD $\geq 9''$ DBH	75	2.12	72	1.12

Tree Size

A thematic tree size map with four categorical diameter classes was generated for all areas classified as *forest* on the final *vegetation type* map (Figure 13). Tree size is determined as the diameter class containing the plurality of cover within a given area or mapping polygon. Seedlings less than 4.5 feet tall are included in the smallest tree diameter class (*Tree Size 1*). Plurality of cover is determined by comparing the areal tree cover of individual diameter classes when viewed from above—discounting overtopped trees. For example, smaller trees that are obstructed from the synoptic perspective by larger trees are ignored and not counted in the diameter class estimate. Project-wide acreage summaries of the tree size classes are provided in Table 11.

Table 11. Thematic tree size acreage summary for the *Prince of Wales* existing vegetation map

Tree Size Class (Diameter at Breast Height)	Area (ac)	% Forest Area
TS1 (Sapling < 5" dbh)	350,916	16.1%
TS2 (Pole 5 - 8.9" dbh)	550,473	25.3%
TS3 (Medium 9 - 20.9" dbh)	1,057,173	48.6%
TS4 (Large $\geq 21''$ dbh)	215,990	9.9%
Total	2,174,551	100%

Tree Size Validation

For each class, 30% of the calls were withheld for validation and the remaining 70% were used to train the random forest classifier. Both the *sapling* and the *large* tree classes had the highest user's accuracies, whereas most of the confusion occurred in the *pole* and *medium* timber classes. Given a majority of the project area contains *forest* that is 5 to 20 inches in diameter, confusion is expected when canopy height and spectral similarity may exist between the *pole* and *medium* tree size classes. The overall accuracy for *tree size* was 62% (Table 12). The lower agreement is likely due to the high degree of subjectivity in assigning a particular *tree size* class to a given area while on the ground. Tree sizes commonly mix, especially in natural settings. Additionally, only 18 sites were used to evaluate the *large* tree size class and this relatively low sample size may not be adequate for class evaluation. Ultimately, all the reference data were used, meaning that data withheld for accuracy assessment were reintroduced to create the most accurate *tree size* map output possible.

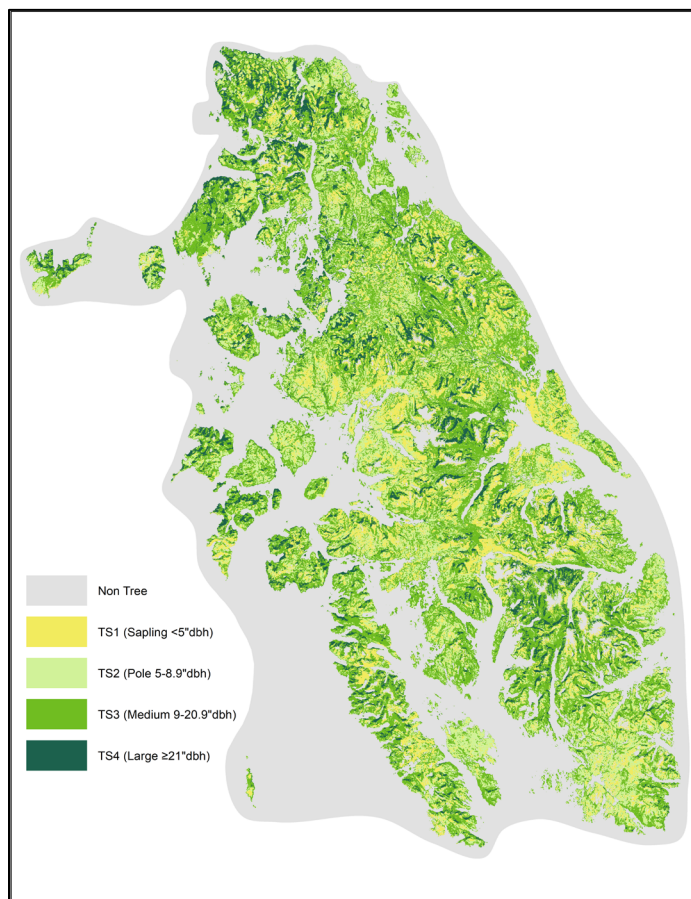


Figure 13. Tree size across the Prince of Wales project area.

Table 12. Error matrix of tree size for the Prince of Wales existing vegetation mapping project.

Tree Size		Reference Data				User's Accuracy	Commission Error
		TS1	TS2	TS3	TS4		
Map Data	TS1 (Sapling <5" dbh)	31	7	3	0	76%	24%
	TS2 (Pole 5 - 8.9" dbh)	1	15	15	0	48%	52%
	TS3 (Medium 9 - 20.9" dbh)	0	12	36	14	58%	42%
	TS4 (Large ≥ 21" dbh)	0	0	1	4	80%	20%
Producer's Accuracy		97%	44%	65%	22%	Kappa	0.45
Omission Error		3%	56%	35%	42%	Overall Accuracy	62%
						Area-Weighted Accuracy	45%

Conclusion

Existing vegetation was mapped through a partnership with the Tongass National Forest, Alaska Regional Office (Region 10), and the Geospatial Applications and Technology Center (GTAC). These map data were designed to be consistent with the standards established in the Existing Vegetation Classification and Technical Guide (Nelson et al. 2015), and to provide detailed forest structure information to support project planning and inform land management of the Prince of Wales and surrounding islands. The final map comprises seven distinct, integrated feature layers: 1) vegetation type; 2) tree canopy cover; 3) trees per acre (TPA) for trees ≥ 1' tall; 4) trees per acre for trees ≥ 6" diameter at breast height (dbh); 5) quadratic mean diameter (QMD) for trees ≥ 2" dbh; 6) quadratic mean diameter for trees ≥ 9" dbh; and 7) thematic tree size. The dominance type map consists of 18 classes, including 15 vegetation classes and 3 other land cover types. Continuous tree canopy cover, TPA, QMD, and thematic tree size were developed for areas classified as *forest* on the final vegetation type map layer. The final *Prince of Wales existing vegetation map* products provide a spatial depiction of vegetation floristics and structure in 2019. These products can be used in numerous ways to assist resource specialists and land managers. Existing vegetation maps can inform further project-level investigations, timber management, fire behavior, wildlife habitat modeling, and provide region-wide estimations of resource availability and status.

This project was made possible through a collaborative team effort that took dedicated work over a span of several years. Different mapping methods were employed based on the available data, desired map classes, and mapping objectives. These methods utilized the best available science and will inform future mapping efforts to make regionally consistent maps across coastal Alaska.

This project used an image object-oriented approach, and therefore, relied on an automated segmentation process to develop the mapping polygons to be used as the fundamental modeling units. Predictor data including remotely sensed imagery, topographic data, and climate information, were summarized as zonal statistics to these segments. Subsequently, reference data either collected in the field, derived from LiDAR, or photo interpreted were intersected with the corresponding segments to extract associated statistics and to produce the predictive classification models. Random decision forest, a machine learning data mining technique, was used to assign land cover and vegetation structure

attributes and produce the final map products. The 2019 orthoimagery and Sentinel 2 satellite imagery were the most important model drivers for *vegetation type* prediction, while vegetation structure models relied most heavily on LiDAR and IfSAR topographic data sources. Even though ground reference information was collected in 2018 and the invaluable LiDAR data was acquired in two phases during 2017 and 2018, the map products were developed to reflect the ground conditions of 2019. This was an effort to provide the most contemporary map products possible, which leveraged the 2019 high-resolution orthoimagery, acquired at the height of the 2019 growing season, as the standard for reconciling temporal differences between datasets. Consequently, the maps are considered to reflect existing vegetation conditions found on the *Prince of Wales* during the summer of 2019.

Although this map achieved relatively high accuracies, there were data limitations and other factors that made this project challenging. The climate of Southcentral Alaska makes obtaining cloud-free imagery difficult, especially when data acquisition has seasonal constraints and imaging sensors have infrequent revisit schedules. Widespread active management also causes additional complexity when it comes to modeling a dynamic landscape, like *Prince of Wales*. Reference data may capture ground conditions that have since changed when compared to the geospatial predictor data used for classification. Additionally, temporal discrepancies between geospatial data sources alters workflows, which requires the mapping process to accommodate disparities in data extents. Much work was required to reconcile these differences and strive for map consistency. Spectral distinction in areas of forest regeneration and young growth, where tree heights are comparable to other lifeforms and tree species can be artificially mixed, are especially difficult. These areas were ubiquitous across productive regions of the project area. Despite these challenges, disparate data sources were strategically utilized to best leverage the available data and achieve these data products.

These existing vegetation data products provide a reasonable portrayal of vegetation patterns across the *Prince of Wales* landscape. Overall accuracies were 89% at the map group-level and 64% for the vegetation type layer. Detailed forest structure metrics including tree canopy cover, trees per acre, and quadratic mean diameter, were developed using first order LiDAR metrics. These metrics are known to be highly reliable within the quality level 1 *LiDAR* data extent, which covers approximately 75% of the project's land area. These metrics were extrapolated outside of the *LiDAR* extent using IfSAR and spectral information. Additionally, a thematic tree size layer was created that covers all of the Prince of Wales Island and surrounding islands. Even though localized structure model accuracy may vary depending on the source data used, the overall forest composition patterns were captured effectively across the entire project area.

For more information please refer to the Alaska Region, Plants and Animals, Alaska Region Vegetation Mapping and Ecology [website](#) for links to the *Prince of Wales* Vegetation Mapping ArcGIS StoryMap. This Story Map contains interactive map applications, descriptions of the project, and links that enable the user to download associated project data. Downloadable data includes: the classification key, imagery, reference data, and final map products.

Currently, more mapping projects are being conducted within Coastal Alaska. Ongoing projects include mapping existing vegetation for the interior portion of the Cordova Ranger District and the Glacier Ranger District on the Chugach National Forest and the Ketchikan-Misty Ranger District on the Tongass National Forest. Additionally, multiple hydrographic projects have commenced across Coastal Alaska,



including *shoreline*, *2-Dimensional hydrographic feature*, and *glacier* mapping. The Alaska Regional office is working with the individual National Forests and other land management agency partners to coordinate these mapping efforts. This collaboration is critical to the identification of project objectives and designing strategies for achieving those objectives, which are necessary steps to adequately map these ecologically important areas.

References

- Arumäe, T.; Lang, M. 2018. Estimation of canopy cover in dense mixed-species forests using airborne LiDAR data. *European Journal of Remote Sensing*. 51(2): 132-131.
- Breiman, L. 2001. Random Forests. *Machine Learning*. 45(1): 5-32.
- Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*. 37(1): 35-46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
- Crist, E.P.; Cicone, R.C. 1984. A physically-based transformation of thematic mapper data—The TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing*. 108(4): 422-435.
- Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. 2007. Random forests for classification in ecology. *Ecology*. 88(11): 2783-2797.
- Heidemann, H.K. Lidar Base Specification (Version 1.2, November 2014). Available online: <https://pubs.usgs.gov/tm/11b4/pdf/tm11-B4.pdf>
- Hopkinson, C.; Chasmer, L. 2009. Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*. 113(1): 275-288. <http://dx.doi.org/10.1016/j.rse.2008.09.012>
- McGaughey, R. J. 2009. FUSION/LDV: Software for LiDAR data analysis and visualization. *U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station*. <http://forsys.cfr.washington.edu/FUSION/fusionlatest.html>
- Nelson, M.L.; Brewer, C.K.; Solem, S.L., eds. 2015. Existing vegetation classification, mapping, and inventory technical guide, version 2.0 Gen. Tech. Rep. WO-90. Washington, DC: *U.S. Department of Agriculture, Forest Service, Ecosystem Management Coordination Staff*. 210 p.
- Nowacki, G.; Spencer, P.; Brock, T.; Fleming, M.; & Jorgenson, T. 2001. Ecoregions of Alaska and neighboring territory (map). *U.S. Geological Survey*. <http://www.adfg.alaska.gov/index.cfm?adfg=ecosystems.ecoregions#prettyPhoto/0/>
- Reynolds, C. 2019. LiDAR-derived forest metric models for Prince of Wales Island, Alaska. *The Nature Conservancy*. pp.1-33.
- Ryherd, S.; Woodcock, C. 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering & Remote Sensing*. 62(2): 181-194.
- Smith, A.M.; Falkowski, M.J.; Hudak, A.T.; Evans, J.S.; Robinson, A.P.; Steele, C.M. 2009. A cross-comparison of field, spectral, and LiDAR estimates of forest canopy cover. *Canadian Journal of Remote Sensing*. 35(5): 447-459.
- van Leeuwen, M.; Nieuwenhuis, M. 2010. Retrieval of forest structural parameters using LiDAR remote sensing. *European Journal of Forest Research*. 129(4): 749-770.
- Wasser, L.; Day, R.; Chasmer, L.; Taylor, A. 2013. Influence of vegetation structure on LiDAR-derived canopy height and fractional cover in forested riparian buffers during leaf-off and leaf-on conditions. *PLOS One*. 8(1). <https://doi.org/10.1371/journal.pone.0054776>

Appendix A: Dichotomous Key Containing the Class Definitions for the Prince of Wales

Master Key		
Couplet	Vegetation Formations Key- Cover values in the key to formations are absolute and relative cover. Tree cover includes both regeneration and overstory trees, so that young stands are classified as forest. Some young regeneration stands may be also classified as shrublands if trees are less than 1.5 m tall. (<i>Absolute cover is the proportion of the plot's area included in the perpendicular downward projection of the species, or "Seen from Above". Relative cover of a species is the proportion it composes of the total plant cover of that lifeform on the plot. Relative cover values are calculated from absolute cover values</i>).	Lifeform Type
1a	Human Modified Vegetation: where vegetation displays a characteristic of intensive human disturbance or man-made structures that will not recover to a natural habitat i.e. Rock pits, roads, human settlements, LTF's.	Developed
1b	Natural Vegetation: where vegetation may or may not display some human or other disturbance, but if disturbed, will be left to recover to a natural habitat i.e. Clearcuts, landslides.	2
2a	Total <i>absolute</i> tree cover is $\geq 10\%$	4
2b	Total <i>absolute</i> tree cover is $< 10\%$	3
3a	Total vascular vegetation <i>absolute</i> cover is $< 25\%$ and not dominated by open water	Barren/Sparse Vegetation
3b	Total vascular vegetation <i>absolute</i> cover is $< 25\%$ and dominated by open water	Water
3c	Total vascular vegetation <i>absolute</i> cover is $\geq 25\%$	6
4a	Tree <i>relative</i> cover is $\geq 75\%$ conifer species	Conifer Forest
4b	Tree <i>relative</i> cover is $< 75\%$ conifer species	5
5a	Tree <i>relative</i> cover of broadleaf species is $\geq 75\%$ of tree cover	Broadleaf Forest
5b	Tree <i>relative</i> cover of broadleaf species is $< 75\%$ of tree cover	Mixed Forest
6a	Shrub <i>absolute</i> cover is $\geq 25\%$	Shrub
6b	Shrub <i>absolute</i> cover is $< 25\%$	7
7a	<i>Absolute</i> cover of herbaceous species is $\geq 25\%$ (includes graminoids and/or forbs, alone or could combine with $< 25\%$ shrubs)	Herbaceous
7b	<i>Absolute</i> cover of herbaceous species is $< 25\%$ (includes graminoid and/or forbs, alone or could combine with $< 25\%$ shrubs)	Barren/Sparse Vegetation

Couplet	Conifer Forest Key	Dominance Type	Map Unit	Map Unit Symbol
1a	Sitka spruce with $\geq 60\%$ <i>relative</i> canopy cover; broadleaf trees or tall shrubs with $<30\%$ <i>relative</i> canopy cover	Sitka Spruce	Sitka Spruce	SS
1b	Sitka spruce with $\geq 60\%$ <i>relative</i> canopy cover, broadleaf trees or tall shrubs with $\geq 30\%$ <i>relative</i> canopy cover	Go to Mixed Forest Key		
1c	Sitka spruce with $< 60\%$ <i>relative</i> canopy cover, no broadleaf trees present	2		
2a	Mountain Hemlock with $\geq 75\%$ <i>relative</i> canopy cover	Mountain Hemlock	Mountain Hemlock Mix	MHmix
2b	Mountain Hemlock with $<75\%$ <i>relative</i> canopy cover	3		
3a	Mountain Hemlock present with $< 75\%$ <i>relative</i> canopy cover. Sitka Spruce is present and can be codominant.	Sitka Spruce/Mountain Hemlock	Mountain Hemlock Mix	MHmix
3b	Mountain Hemlock if present is $< 75\%$ <i>relative</i> canopy cover. Sitka Spruce is not present or if present, not codominant and minor	4		
4a	Subalpine Fir with $\geq 75\%$ <i>relative</i> canopy cover.	Subalpine Fir	Mountain Hemlock Mix	MHmix
4b	Subalpine Fir with $<75\%$ <i>relative</i> canopy cover.	5		
5a	Shorepine with $\geq 60\%$ <i>relative</i> canopy cover	6		
5b	Shorepine with $<60\%$ <i>relative</i> canopy cover	7		
6a	Trees are stunted within a peatland	Dwarf Shorepine	Dwarf Conifer	DC
6b	Trees are not stunted within a peatland	Shorepine	Mixed Conifer	MC
7a	Red and yellow cedar are growing together and the combined <i>relative</i> canopy cover is $\geq 40\%$	Cedar	Cedar	CE
7b	Red and yellow cedar are not growing together. The <i>relative</i> canopy cover of either red or yellow is $\geq 40\%$	8		

Couplet	Conifer Forest Key (cont.)	Dominance Type	Map Unit	Map Unit Symbol
7c	Red or yellow cedar growing together or as a single species is <40% <i>relative</i> canopy cover	10		
8a	Red cedar with ≥ 40% <i>relative</i> canopy cover.	Red Cedar	Cedar	CE
8b	Red cedar < 40% <i>relative</i> canopy cover.	9		
9a	Yellow cedar ≥40% <i>relative</i> canopy cover, most trees stunted	Dwarf Yellow Cedar	Dwarf Conifer	DC
9b	Yellow cedar <40% <i>relative</i> canopy cover, most trees stunted	Dwarf Conifer	Dwarf Conifer	DC
9c	Yellow cedar ≥40 % <i>relative</i> canopy cover, most trees not stunted	Yellow Cedar	Cedar	CE
10a	Western Hemlock with ≥ 75% <i>relative</i> canopy cover.	Western Hemlock	Western Hemlock	WH
10b	Western Hemlock with <75% <i>relative</i> canopy cover.	11		
11a	Western Hemlock is <75% <i>relative</i> canopy cover but is always combined with other species in various <i>relative</i> canopy covers. Stand is mid to late-seral. Not actively managed.	Mixed Conifer	Mixed Conifer	MC
11b	Western Hemlock is <75% <i>relative</i> canopy cover but is always combined with other species in various <i>relative</i> canopy covers. Stand is early seral. Actively managed.	12		
12a	Sitka Spruce and/or Western Hemlock combined equals < 90% total <i>relative</i> canopy cover.	Mixed Species	Mixed Species	MS
12b	Sitka Spruce and Western Hemlock combined equals ≥ 90% total <i>relative</i> canopy cover.	Spruce-Hemlock	Spruce-Hemlock	SH
12c	Site is none of the above forest species conditions	Unnamed Conifer	Unnamed Conifer	UC

Couplet	Broadleaf Forest Key	Dominance Type	Map Unit	Map Unit Symbol
1a	Alder present with $\geq 75\%$ <i>relative</i> canopy cover.	Alder (any height)	Red Alder	RA
1b	Alder present or not, with $< 75\%$ <i>relative</i> canopy cover	2		
2a	Cottonwood with $\geq 75\%$ <i>relative</i> canopy cover	Cottonwood	Cottonwood	CW
2b	Cottonwood with $< 75\%$ <i>relative</i> canopy cover	Go to Mixed Forest Key		

Couplet	Mixed Forest Key	Dominance Type	Map Unit	Map Unit Symbol
1a	Cottonwood with $\geq 25\%$ <i>relative</i> canopy cover and together with Sitka spruce comprise $\geq 75\%$ <i>relative</i> canopy cover	Sitka Spruce-Cottonwood	Sitka Spruce-Cottonwood	SS-CW
1b	Cottonwood with $< 25\%$ <i>relative</i> canopy cover and together with Sitka spruce comprise $< 75\%$ <i>relative</i> canopy cover	2		
2a	Red Alder with $\geq 25\%$ <i>relative</i> canopy cover and together with Sitka spruce comprise $\geq 75\%$ <i>relative</i> canopy cover	Sitka Spruce- Red Alder	Sitka Spruce-Red Alder	SS-RA
2b	Red Alder with $< 25\%$ <i>relative</i> canopy cover and together with Sitka spruce comprise $< 75\%$ <i>relative</i> canopy cover	Undetermined mix of Hardwood and Conifer	Undetermined mix of Hardwood and Conifer	UHC



Couplet	Shrub Key	Dominance Type	Map Unit	Map Unit Symbol
1a	<i>Relative</i> canopy cover of Sitka alder is $\geq$ 75%	Alder Shrub	Alder Shrub	AS
1b	<i>Relative</i> canopy cover of Sitka alder is < 75%	2		
2a	<i>Relative</i> canopy cover of combined taller shrubs ($\geq 1.5\text{m}$) such as willow species, spirea, copperbush, crabapple, Elderberry, sweet gale, Sitka alder, Salmonberry, Devils club, Blueberry etc is, dominant at a site compared to low shrubs, if present (ex: heather, crowberry, bog blueberry etc)	Tall Shrubs	Tall Shrubs	TS
2b	<i>Relative</i> canopy cover of combined low and dwarf shrubs (<1.5m) is dominant at a site (ex: heather, crowberry, bog blueberry etc) compared to tall shrubs, if present.	Low Shrubs	Low Shrubs	LS



Couplet	Herbaceous Key	Dominance Type	Map Unit	Map Unit Symbol
1a	Site is aquatic with permanent standing water present	2		
1b	Site is wet to mesic, with little or no standing water	3		
2a	Site is permanent freshwater ponds or lakes with various emergent grasses and forbs (roots below the water such as <i>Potamogeton</i> and <i>Nuphar</i>) $\geq 25\%$ <i>absolute</i> cover (the rest is water)	Aquatic Herbaceous	Aquatic Herbaceous	AHB
2b	Site is not the above.	3		
3a	Site contains areas of higher water table and seasonal standing water, various non-emergent grasses, and forbs (such as buckbean, cottongrass, Sitka other sedges) dominate	Wet Herbaceous	Wet Herbaceous	WHB
3b	Site <i>does not</i> contain higher water table or seasonal standing water, various non-emergent grasses, and forbs (may see <i>Calamagrostis</i> , lupine, yarrow, paintbrush) dominate	Mesic Herbaceous	Wet Herbaceous	WHB



Appendix B: Prince of Wales Vegetation Type Error Matrix

Vegetation Type		Reference Data																	User's Accuracy	Commission Error
		SS	SS-WH	WH	CE	MHmix	DC	MC	MS	RA	SS-RA	AS	TS	LS	WH B	AHB	WA	BR/SV		
Map Data	SS	5	6	6	5	0	0	0	6	1	5	0	0	0	0	0	0	0	15%	85%
	SS-WH	14	24	9	4	0	0	1	17	0	3	0	0	0	0	0	0	0	33%	67%
	WH	1	7	14	5	0	0	1	4	0	0	0	0	0	0	0	0	0	44%	56%
	CE	0	0	0	7	2	1	13	0	0	0	0	0	0	1	0	0	0	29%	71%
	MHmix	0	0	0	0	21	3	0	0	0	0	0	0	0	0	0	0	0	88%	13%
	DC	0	0	0	0	0	9	1	0	0	0	0	0	0	0	0	0	0	90%	10%
	MC	0	0	1	9	3	7	14	0	0	0	0	0	0	1	0	0	0	40%	60%
	MS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	100%	0%
	RA	0	0	0	0	1	0	0	1	20	1	1	0	0	0	0	0	0	83%	17%
	SS-RA	3	0	0	0	0	0	0	1	9	21	0	0	0	0	0	0	0	62%	38%
	AS	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	100%	0%
	TS	0	0	0	0	3	0	0	0	0	0	0	28	0	4	0	0	0	80%	20%
	LS	0	0	0	0	0	10	0	0	0	0	0	1	20	5	0	0	0	56%	44%
	WHB	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	100%	0%
	AHB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	100%	0%
	WA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	30	0	97%	3%
	BR/SV	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	30	94%	6%
Producer's Accuracy		22%	65%	47%	23%	70%	30%	47%	3%	67%	70%	97%	90%	100%	63%	97%	100%	100%	Kappa	0.62
Omission Error		78%	35%	53%	77%	30%	70%	53%	97%	33%	30%	3%	10%	0%	37%	3%	0%	0%	Overall Accuracy	64%
																			Area Weighted Accuracy	68%