National Park Service
U.S. Department of the Interior

Natural Resource Stewardship and Science

North Cascades National Park Complex Vegetation Classification and Mapping Project Report

Natural Resource Report NPS/NCCN/NRR—2021/2254

ON THIS PAGE
Buckner Mountain from the upper Park Creek Valley.
Image credit NPS

ON THE COVER
Fog fills the valley of the Chilliwack River below the heather meadows of Copper Ridge.
Image credit Eric Nielsen

North Cascades National Park Complex Vegetation Classification and Mapping Project Report

Natural Resource Report NPS/NCCN/NRR—2021/2254

Eric M. Nielsen,[1] Catharine Copass,[2] Rachel L. Brunner,[1] Lindsey K. Wise[1]

[1]Institute for Natural Resources
Portland State University
Portland, Oregon

[2]National Park Service
Olympic National Park
Port Angeles, Washington

May 2021

U.S. Department of the Interior
National Park Service
Natural Resource Stewardship and Science
Fort Collins, Colorado

The National Park Service, Natural Resource Stewardship and Science office in Fort Collins, Colorado, publishes a range of reports that address natural resource topics. These reports are of interest and applicability to a broad audience in the National Park Service and others in natural resource management, including scientists, conservation and environmental constituencies, and the public.

The Natural Resource Report Series is used to disseminate comprehensive information and analysis about natural resources and related topics concerning lands managed by the National Park Service. The series supports the advancement of science, informed decision-making, and the achievement of the National Park Service mission. The series also provides a forum for presenting more lengthy results that may not be accepted by publications with page limitations.

All manuscripts in the series receive the appropriate level of peer review to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and designed and published in a professional manner.

This report received informal peer review by subject-matter experts who were not directly involved in the collection, analysis, or reporting of the data.

Views, statements, findings, conclusions, recommendations, and data in this report do not necessarily reflect views and policies of the National Park Service, U.S. Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government.

This report is available in digital format from the NPS North Coast and Cascades Network website and the Natural Resource Publications Management website. If you have difficulty accessing information in this publication, particularly if using assistive technology, please email irma@nps.gov.

Please cite this publication as:

Nielsen, E. M., C. Copass, R. L. Brunner, and L. K. Wise. 2021. North Cascades National Park Complex vegetation classification and mapping project report. Natural Resource Report NPS/NCCN/NRR—2021/2254. National Park Service, Fort Collins, Colorado. https://doi.org/10.36967/nrr-2286418.

NPS 168/175301, May 2021

Contents

Contents (continued)

Contents (continued)

Contents (continued)

Figures

Page

Tables

Page

Supplements

Available as separate PDFs through the NPS Data Store at https://irma.nps.gov, or via other media upon request to the NPS North Coast and Cascades Network Data Manager.

Supplement 1: Vegetation map

Nielsen, E. M., C. Copass, R. L. Brunner, and K. Braun. 2021. North Cascades National Park Complex vegetation map. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Supplement 2: Map class descriptions

Nielsen, E. M., R. L. Brunner, C. Copass, and L. K. Wise. 2021. North Cascades National Park Complex map class descriptions. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Supplement 3: Map class key

Brunner, R. L., C. Copass, L. K. Wise, and E. M. Nielsen. 2021. North Cascades National Park Complex map class key. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Supplement 4: Accuracy assessment contingency tables

Institute for Natural Resources. 2021. North Cascades National Park Complex accuracy assessment contingency tables. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Supplement 5: Taxonomic tables

Institute for Natural Resources. 2021. Taxonomic tables for Mount Rainier, Olympic and North Cascades National Parks. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2283943.

Supplement 6: Mapping association descriptions

Nielsen, E. M., and R. L. Brunner. 2021. Vegetation associations for mapping Pacific Northwest national parks. Institute for Natural Resources, Portland State University, Portland, Oregon. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2283945 (available June 2021).

Executive Summary

The Vegetation Mapping Inventory (VMI) is an effort by the National Park Service (NPS) to classify, describe, and map vegetation communities present on NPS units across the United States. The Institute for Natural Resources, working in cooperation with the NPS North Coast and Cascades Network (NCCN), has completed a VMI project for the vegetation communities of North Cascades National Park Complex (NOCA), including Ross Lake and Lake Chelan National Recreation Areas (NRA).

The map is based on a vegetation classification developed during the project and was created using a inductive modeling approach. Data used to construct the classification were collected between 2005 and 2015, and included plots from Mount Rainier National Park and Olympic National Park. These plots were used to develop and refine the association-level National Vegetation Classification (NVC). The associations were combined into map classes based roughly on the NVC alliance-level classification, but updated to allow improved map detail and accuracy. Model training data relied only on plots from NOCA, also collected during the same years. Independent field accuracy assessment data were collected in 2016 and applied to the final map generated later.

The map development process was organized around the random forests machine learning algorithm. The modeling used 2,980 plots representing 151 vegetation associations and 46 map classes. Imagery from the National Agriculture Imagery Program and the Sentinel-2 and Landsat 8 satellites, airborne lidar bare earth and canopy height data, elevation data from the U.S. Geological Survey 3D Elevation Program, and climate normals from the PRISM Climate Group were used to develop a variety of predictor metrics. The predictors and the map class calls at each plot were input to a process in which each map class was modeled against every other map class in a factorial random forests scheme. We used plot-level modeling outcomes and species composition data to adjust the crosswalk between association and map class so that floristic consistency and model accuracy were jointly optimized across all classes. The map was produced by predicting the factorial models and selecting the overall best-performing class at each 3-meter pixel.

The final vegetation map, including a buffer surrounding the park complex, contains 41 natural vegetated classes, five mostly unvegetated natural classes, and five classes representing burned areas or anthropogenic disturbance. Conifers, some recently burned, cover about half of the complex. Upper montane forests dominated by mountain hemlock (*Tsuga mertensiana*) and silver fir (*Abies amabilis*) are most abundant in the National Park units. Forests at Ross Lake NRA are characterized by Douglas-fir (*Pseudotsuga menziesii*) and western hemlock (*Tsuga heterophylla*), and Lake Chelan NRA is notable for large expanses of burned forests and Douglas-fir and ponderosa pine (*Pinus ponderosa*) woodlands. Broadleaf and mixed forests occupy only three percent of the complex. Riparian broadleaf forests are the most widespread, but mixed paper birch (*Betula papyrifera*) forests are common around Ross Lake, and mixed bigleaf maple (*Acer macrophyllum*) forests are very abundant on debris aprons near the Stehekin River. Shrublands cover fifteen percent of the complex and are especially concentrated in the Park units, where they are dominated by Sitka alder (*Alnus viridis*) and mountain-heathers (*Phyllodoce empetriformis, Cassiope mertensiana*). Snowbrush

(*Ceanothus velutinus*) shrublands, associated with past fire, are common in Lake Chelan NRA. Herbaceous vegetation occupies only five percent of the overall landscape, with sparse alpine vegetation most common in the Park units, bedrock balds frequent in both NRAs, and green fescue (*Festuca viridis*) mostly associated with Lake Chelan NRA. Sparsely vegetated and entirely bare rock cover nearly one-fifth of the complex; bedrock barrens are most common and occupy fully sixteen percent. Exposed snow and ice (in the Park units) and lake surfaces (in the NRAs) round out the remaining six percent.

The accuracy assessment (AA) was based on 941 independent field-collected plots representing all the vegetated classes, as well as alluvial, colluvial and bedrock barrens and recently burned forests, which also often host vegetation communities. They were gathered from an inference area covering 10.4% of the complex. The overall map accuracy based on this sample was 84.5%. After correcting for map class prevalence in the inference area, the overall accuracy was 82.0%. In the accuracy assessment section, we review the mapping of all classes failing to meet NPS accuracy standards. Possible remedies for each mapping error are considered, and recommendations are provided to NPS for possible modifications to the map product in response to the issues identified.

Many new methodologies for mapping and floristic analysis were developed during this project. These innovations were also applied in mapping the other large NCCN national parks. In addition to allowing the development of this series of maps, these methods should be useful to the NCCN and VMI for other mapping projects and purposes. Products resulting from this project include (a) this report, (b) the report supplements listed above, (c) a geodatabase with map polygon attributes, plot locations, and project boundaries, (d) training and accuracy assessment plot field forms and data, including ground photography, (e) hard copy vegetation maps and (f) metadata for digital products. Geospatial products are provided in the Universal Transverse Mercator (UTM) Zone 10 projection using the North American Datum of 1983.

Acknowledgments

Many skilled individuals came together to complete this project over its fifteen-year duration. We are especially grateful to John Boetsch, Katherine Braun, Mark Huff, Matt Lee, Stacy McDonough and Tynan Ramm-Granberg, whose efforts over many years were truly instrumental to the project's success. The names and roles of all contributors are listed below; we sincerely apologize to anyone we have overlooked.

National Park Service

Noah Anderson — mapping crew member
William Arneson — classification crew lead
Patrick Barnes — accuracy assessment (AA) crew member
Charles Batey — AA crew member
Katherine Braun, OLYM GIS specialist — production of map sheets for mapping and AA fieldwork, production of final maps, assistance with GPS and metadata
Mignonne Biven, NOCA supervisory botanist — assistance with fieldwork, inputs to draft and final map products, report peer review
John Boetsch, network data manager/ecologist — creation and maintenance of field plot database
Anne Braaten, NOCA GIS specialist — project support
Karl Brown, national VIP manager — programmatic guidance
Robby Bryson — mapping crew member
Erin Burke — mapping crew member
William Clark — classification crew member
Tammy Cook, biologist — programmatic guidance
Catharine Copass, network VIP projects manager — project oversight and coordination
Evan Derickson — mapping crew member
Jacob DeGuzman — mapping crew member
Casey Duncan — AA crew member
Lise Grace, NCCN publication manager — report editing and publication assistance
Rich Gwodz — classification crew lead
Thomas Hender — mapping crew member
Mark Huff, *NCCN Program Coordinator* — programmatic guidance, project support
Sacha Johnson Buller — classification crew member
Sophia Kast — mapping crew member
Seth Keena-Levin — AA crew member
Karen Kopper, regional fire ecologist — assistance with fieldwork, inputs to draft and final map products
Allen McCoy, regional GIS specialist — data procurement
Andrew Nemeth — mapping crew member
Kelly O'Neil — classification crew member
Emily Pruiksma — classification crew member
Tynan Ramm-Granberg — mapping crew leader, mapping data QC assistance, vegetation classification assistance

Regina Rochefort, NOCA plant ecologist & science advisor — assistance with fieldwork, inputs to draft and final map products
Callie Schmidt — mapping crew member, AA crew member
Dorothy Wallace-Senft — classification crew member

Institute for Natural Resources (INR), Portland State University

Rachel Brunner, vegetation ecologist — field data QC lead, vegetation classification and image processing assistance, mapping and AA crew member
John Christy, wetlands ecologist — vegetation classification assistance
Eleanor Gaines, INR-Portland director — project management assistance
Jimmy Kagan, ecologist — vegetation classification assistance
Matt Lee, field data manager — AA crew leader, field data QC, mapping crew member, vegetation classification assistance
Eric Nielsen, geospatial scientist — INR PI, vegetation classification, image processing, modeling, statistical analysis, sample and field protocol designs, AA crew member
Sage Stowell — AA crew member
Lindsey Koepke Wise, biodiversity data manager — report editing, vegetation taxonomy assistance

NatureServe

Gwen Kittel, vegetation ecologist — NVC alliance development for NCCN
Marion Reid, vegetation ecologist — NVC alliance development for NCCN

Funding for this work was provided by the I&M VIP program, through the Pacific Northwest Cooperative Ecosystem Study Unit, Task Agreements P12AC10432 and P17AC01143.

Acronyms and Abbreviations

3DEP	3D Elevation Program (USGS), provider of DEMs
AA	accuracy assessment (see following glossary)
AIA	attempted inference area (see following glossary)
CE	commission error (see following glossary)
DBH	diameter at breast height, the diameter of a tree measured at 4.5 feet
DEM	digital elevation model; digitally represented elevation data
DIT	differential indicators tool, developed at INR
EBLA	Ebey's Landing National Historical Reserve
FGDC	Federal Geographic Data Committee
GPS	Global Positioning System
HALT	hybrid assemblage labeling tool, developed at INR
I&M	Inventory and Monitoring Program (NPS)
INR	Institute for Natural Resources (Portland State University)
LCNRA	Lake Chelan National Recreation Area
LEWI	Lewis and Clark National and State Historical Parks
MORA	Mount Rainier National Park
MMU	minimum mapping unit (see following glossary)
NAD83	North American Datum of 1983; the geographic datum for this project
NAIP	National Agriculture Imagery Program (USDA)
NCCN	NPS North Coast and Cascades Network
NOCA	North Cascades National Park
NPS	National Park Service
NTM	nested texture metric (see following glossary)
NVC	National Vegetation Classification
NVCS	National Vegetation Classification Standard
OE	omission error (see following glossary)
OLI	Operational Land Imager (Landsat 8, 2013–present)
OLYM	Olympic National Park
PA	producer's accuracy (see following glossary)
PCT	population contingency table (see following glossary)
PI	photo-interpretation, use of imagery to ascertain land cover characteristics
PNV	potential natural vegetation (see following glossary)
PNW	Pacific Northwest region
PRISM	Parameter-elevation Regressions on Independent Slopes Model (PRISM Climate Group 2019)
PSU	Portland State University
QC	quality control (see following glossary)
RF	random forests (see following glossary)
RIA	reached inference area (see following glossary)

RLNRA	Ross Lake National Recreation Area
RMSE	root-mean-square error
RRRF	round robin random forests (see following glossary)
SAGA	System for Automated Geoscientific Analyses (Conrad et al. 2015)
SCM	species cover–association match tool, developed at INR (see following glossary)
SCT	sample contingency table (see following glossary)
TM	Thematic Mapper (Landsat 4 and 5, 1984–2011)
TNC	The Nature Conservancy
UA	user's accuracy (see following glossary)
USDA	United States Department of Agriculture
USFS	United States Forest Service
USGS	United States Geological Survey
UTM10	Universal Transverse Mercator zone 10; standard geographic projection for the project
UW	University of Washington
VIP	Vegetation Inventory Program (NPS)
VMI	Vegetation Mapping Inventory (NPS)
WNHP	Washington Natural Heritage Program

Glossary

accuracy assessment	Statistical analysis to determine the degree to which a map correctly represents on-the-ground conditions.
accuracy assessment plots	Field plots collected during the third major phase of field sampling, used as ground truth in the AA.
attempted inference area	The spatial region within which plots were targeted in the AA sample design.
classification plots	Field plots collected during the first major phase of field sampling, used to define the initial vegetation associations (Crawford et al. 2009). Supplemented by the mapping plots, they were also used to define the mapping associations and map classes, and to create model training data.
commission error	The frequency with which a map assigns a class where it is not present.
contingency table	An AA error matrix documenting the extent of class-specific confusion between mapped and ground-truth data (often called a "confusion matrix").
floristic similarity	The degree of species composition resemblance between two plots, associations or map classes.
full-ocular plot	A field sample including reasonably complete species cover data.
inductive model	A predictive representation of reality built from provided examples.
Landsat	Mid-resolution U.S. remote sensing satellites, active from 1972–present. The Landsat data used were at 30-meter resolution.
lidar	Light detection and ranging; a laser-based technology for measuring elevation.
map classes	The thematic units to which map polygons are labeled; formed by merging similar mapping associations.
mapping associations	The fundamental classification units on which the map classes and therefore the NCCN maps are based; formed by revising the NVC associations (Crawford et al. 2009, Ramm-Granberg et al. 2021) for increased floristic and modeling consistency.
mapping plots	Field plots collected during the second major phase of field sampling. Supplemented by the classification plots, they were used to define the mapping associations and map classes, and to create model training data.
minimum mapping unit	The smallest homogeneous area intended for representation in the map; for this project, nominally 500 square meters (0.05 hectares).
modeling similarity	The susceptibility of two plots, associations or map classes to incorrect labeling in inductive modeling; in other words, their degree of similarity in predictor data.

nested texture metrics	A method for extracting multi-resolution spatial patterning information from imagery, developed at INR.
omission error	The frequency with which a map neglects to show a class where it is actually present.
partial-ocular plot	A field sample with incomplete species cover data.
patch	A fairly homogeneous and contiguous area of land cover discernible on the ground, typically composed of a single vegetation or abiotic land cover type.
photointerpreted plots	Plots assigned to map class based on an assessment of imagery and other data sources available in the office, mostly used for training and AA of abiotic classes.
producer's accuracy	The estimated probability that a map is correct where a particular map class is found on the ground.
population contingency table	An AA error matrix scaled to the mapped extent of each class in the inference area.
potential natural vegetation	The vegetation type that would hypothetically exist at a location under a natural disturbance regime.
predictor data	Independent data (e.g., variables derived from imagery, topography, climate, etc.) provided to an induction model for prediction of a dependent variable (e.g., a map class).
Python	The programming language used for most project geoprocessing.
quality control	Process of improving the quality of data collected and/or entered.
R	The programming language used for most project statistical analyses
reached inference area	The portion of the AIA reached by AA field crews and from which accuracy conclusions were drawn.
random forests	An outlier-resistant inductive modeling algorithm (Breiman 2001).
round robin random forests	An extension to random forests developed at INR for modeling a large number of classes with reduced sample size-related bias
sample contingency table	An AA error matrix based on raw numbers of samples.
SCM taxa	The botanical taxa on which floristic similarity between plots, associations and map classes was determined; mostly species, but also including some genera and some sub-genus groupings of species.
Sentinel-2	Mid-resolution European remote sensing satellites, active from 2015–present. The Sentinel-2 data used were at 10- to 20-meter resolution.
species cover match	A set of tools for evaluating the degree of fit between a plot and an association or map class. Variants were created for use with full-ocular and partial-ocular plots.
training data	Locations confidently assigned to a particular map class and used to build inductive models connecting patterns in predictor data to that map class.
user's accuracy	The estimated probability that a map is correct where a particular map class is mapped.

1. Introduction

1.1. Background

1.1.1. NPS Vegetation Mapping Inventory and National Vegetation Classification

The Vegetation Mapping Inventory (VMI) was created to classify, map and describe vegetation communities on National Park Service units across the United States (NPS 2018). The resulting classifications, maps and reports contribute to the inventory of NPS resources and inform management and planning decisions. NPS has provided guidelines for vegetation classification (Lea 2011) and map accuracy assessment (Lea and Curtis 2010).

VMI maps are based on the National Vegetation Classification (NVC), a collaborative effort to classify the vegetation communities of the U.S. in a consistent manner. The NVC grew out of work by The Nature Conservancy (TNC), NatureServe, and the Natural Heritage Program network (Grossman et al. 1998). It is an evolving classification to which several federal agencies and non-profit organizations—including NPS, the U.S. Fish and Wildlife Service, the U.S. Geological Survey, TNC, and the Ecological Society of America—have contributed.

The National Vegetation Classification Standard (NVCS) provides the hierarchical structure for the NVC. Based in part on an earlier international classification (UNESCO 1973), it was originally adopted by the Federal Geographic Data Committee (FGDC) in 1997 and updated substantially by FGDC (2008). The upper levels of the hierarchy define classes based on broad-scale physiognomic and ecological factors (e.g., climate regimes, continentality), the middle levels incorporate floristic and additional physiognomic factors based on finer scale variation, and the lower levels are based entirely on floristics, including dominant and diagnostic overstory and understory species. The hierarchy for natural vegetation and the classification for an association found in Pacific Northwest (PNW) montane forests is shown in **Table 1**. The most recent revision of the NVC was published as USNVC (2019).

Table 1. National Vegetation Classification System hierarchy (version 2, FGDC 2008), and names of all levels for an example association.

Hierarchy level	Name	Code
Level 1–Class	Forest & Woodland	C01
Level 2–Subclass	Temperate & Boreal Forest & Woodland	S15
Level 3–Formation	Cool Temperate Forest & Woodland	F008
Level 4–Division	Vancouverian Forest & Woodland	D192
Level 5–Macrogroup	Vancouverian Subalpine-High Montane Forest	M025
Level 6–Group	North-Central Pacific Mountain Hemlock-Silver Fir Woodland	G849
Level 7–Alliance	Tsuga mertensiana-Abies amabilis Forest & Woodland	A3723
Level 8–Association	Abies amabilis/Rhododendron albiflorum Forest	CEGL000225

1.1.2. NCCN vegetation inventory project

The North Coast and Cascades Network (NCCN) vegetation inventory project (VIP) began in 2005. The first several years were primarily devoted to developing the regional association-level NVC (Crawford et al. 2009). The Institute for Natural Resources (INR) joined the project in 2008 to assist with the vegetation mapping portion of the project. In addition to the large parks—Mount Rainier National Park (MORA, 956 km^2), Olympic National Park (OLYM, 3734 km^2) and the North Cascades National Park Complex (NOCA, 2769 km^2)—INR and NPS also worked cooperatively to complete two other mapping projects, the Lewis and Clark National and State Historical Parks (LEWI, 38 km^2; Kagan et al. 2012) and Ebey's Landing National Historical Reserve (EBLA, 78 km^2; Copass and Ramm-Granberg 2016a).

1.1.3. NOCA vegetation classification and mapping project

The three large NCCN parks were treated as a single mapping endeavor, but delivered as three distinct projects and reports (see also Nielsen et al. 2021a, Nielsen et al. 2021b). Although much of the classification and mapping work proceeded concurrently, the fieldwork focus moved from one park to another during the map training and accuracy assessment phases. The NOCA project was the last of the parks to be sampled in each phase, with training data collection from 2012 to 2014 and accuracy assessment fieldwork in 2016. The work at NOCA benefitted substantially from the development of a consistent field sampling strategy and the cultivation of an experienced field crew. Given the magnitude and complexity of the challenges posed by sampling this very large and diverse park, those advantages were key to the project's success.

1.2. Approach

1.2.1. Classification

Mountainous environments in the Pacific Northwest present interlocking challenges for vegetation classification. First, the environmental envelopes of most species are largely determined by local climate, which responds in spatially continuous[1] and often complex patterns to elevation, aspect, and characteristics of the surrounding terrain. Competition results in gradual changes in species prominence along gradients that make field assessment of breaks based on thresholds of species cover difficult, and chance variation adds to that unreliability. Second, species succession is often drawn out over centuries, and its rate varies over both coarse and fine spatial scales. For example, in the montane zone, the most characteristic successional process is the gradual establishment and increase in cover of silver fir (*Abies amabilis*). Coarse-scale limitations on seed availability, germination and establishment are posed by the prominence of silver fir in the surrounding area and by interannual climate variability. At subalpine elevations, successional processes are similarly drawn out, but here they also vary at fine spatial scales as micro-habitats differ in their suitability for plant establishment in a given year. Succession results in vegetation classification ambiguity because most vegetation follows a gradual trajectory with no clear and repeatable breaks between stages. Third, at high elevations, the spatial grain of available habitat for individual species is so fine that it becomes impractical to delineate all distinct assemblages of species. Here, the vegetation might

[1] As contrasted with the more discontinuous influence exerted by factors such as soil chemistry.

better be described as a variable mosaic in which assemblage dominance tilts across more coarsely scaled gradients. This results in field interpretation challenges because of lack of clarity about the minimum patch homogeneity and size needed to constitute a sampling unit. At these elevations each species in fact exploits micro-habitat niches that become available to it on an individual basis.

The starting point for the NCCN map classification was an early draft of the interim NCCN alliances later presented in NatureServe (2012). These draft interim alliances were defined by a crosswalk from the vegetation associations presented in Crawford et al. (2009). Those associations, in turn, were defined by using *classification plots*—collected from 2005–07 at all NCCN parks—to refine and provide context to several previous regional classifications. Collection of *mapping plots*, which were field-assigned to vegetation association using the keys in Crawford et al. (2009), began in 2008, using draft versions of those keys. The original plan was to train map models based solely on mapping plots, but it quickly became clear that not enough mapping plots had been collected—particularly at MORA—to adequately train models, so the previously collected classification plots were added to the map training pool. The possibility that these plots were assigned to associations based on criteria other than the keys in Crawford et al. (2009) presented a potential downside to their use for this purpose.

Early map modeling results found a significant degree of mismatch between the assigned alliance of many plots and their modeling tendencies. Through experimentation, it became clear that many of the errors were a consequence of dissimilar plots, particularly in conifer forests, being assigned to the same alliance due to key breaks that resulted in artificial boundaries between types. Although the associations were originally derived via a multivariate cluster analyses and are generally "bloblike" in *n*-dimensional space, the keys carved straight lines through these concepts, squaring them off with hard breaks such as "*Oplopanax horridus* > 5%." Using key-based calls lowered floristic cohesion within the resulting associations and resulted in many plots that modeled poorly as the class to which they had been assigned. We addressed this problem by moving to a multivariate clustering approach for determining the best floristic match for a given plot.[2]

Field data from all plots collected after 2008 included reasonably complete species cover data, which allowed us to retroactively reassign association calls if the overall species composition warranted that. When this dataset was completed for all three parks, it also enabled us to revise the associations themselves in order to correct a variety of pre-existing issues. We termed the revised concepts *mapping associations*, as we were unsure whether the NVC would be adjusted to incorporate them (many have in fact been included in Ramm-Granberg et al. 2021). Regardless, the revisions were a necessary step in rolling the plot-level data up into a mappable classification. The mapping associations were combined into map classes based on their joint floristic and modeling similarities.

[2] Although we have provided a dichotomous key for field use in identifying the final map classes (in keeping with NPS required deliverables), dropping the use of keys for assigning vegetation associations was an essential step in deriving map classes with floristic and modeling cohesion.

Despite additional steps involved in their production, many of the mapped classes bear strong resemblances to the original concepts presented in NatureServe (2012).

Another classification challenge we encountered was that areas recently disturbed by fire, flooding or mass movement sometimes fit awkwardly into the original NVC associations. We moved away from strict floristics-based labeling of plots affected by disturbance and considered their setting and site history as well. Several iterations of plot-level examination and reassignment followed by association-level floristic recalculation resulted in convergence of the classification on several floristically consistent associations often connected with natural disturbance. Past anthropogenic disturbance, such as selective logging of Sitka spruce (*Picea sitchensis*) near the coast[3], also occasionally resulted in ambiguities. Moving away from key-based assignment of association calls helped considerably here: in the case of logged Sitka spruce, enough floristic signals persisted from the natural vegetation community that the correct call was evident, even if the spruce cover was now well below ten percent. Finally, areas experiencing ongoing change (e.g., conifer encroachment into established meadows) also present a challenge as the combination of species may not have been well-represented in the past. Some flexibility is necessary in such areas, which are likely harbingers of greater dilemmas to come.

1.2.2. Mapping

In contrast to many VMI products, we mapped the large NCCN parks using automated model-based methods rather than photo-interpretation (PI). This decision was originally made because of the size of the parks and the indistinct appearance of many of the map classes in imagery. For example, two-thirds of the NCCN parks are covered by coniferous forests and woodlands of 24 different map classes. The component tree species generally cannot be visually distinguished, and many are recognized in the field based as much on their understory composition. In addition, the gradual change of species prominence along climatic gradients in PNW forests, and the variable and patchy species composition characteristic of many non-forest patches, result in a landscape that is not easily divided by hand into discrete patches.[4] Despite this, and the inherent classification challenges discussed above, sites within the parks can be broken repeatably into map classes based on their full species composition, and these classes can be reliably mapped using model-based techniques.

We used an inductive modeling process, in which a computer learns how to distinguish map classes by the examples provided from field plots. We used the *random forests* algorithm (RF; Breiman 2001), as adapted for the R language (Liaw and Wiener 2002). The large number and unequal abundance of map classes proved to be a challenge to multi-class models, which were unable to simultaneously perform well at the prediction of all classes. To address this, we decomposed each

[3] Because the same classification pertains to MORA, OLYM and NOCA, throughout the text we have chosen examples to illustrate our approach from across the three parks

[4] Many non-forest classes might have been mapped by hand with ideally timed high-resolution aerial imagery. However, we did not feel that the expense of collecting such imagery over the large expanse of the parks could be justified for an uncertain outcome that still would have left the forested lands unmappable via photo-interpretation.

multi-class model into many *one versus one* binary models, in which individual map classes were modeled directly against each other (see Bishop 2006, p. 339). We used a novel predictor selection scheme that reduced prediction time, limited collinearity in the predictive variables, and co-optimized model accuracy and effective spatial resolution. The model-based results were manually edited where needed[5] and then subjected to a map accuracy assessment based on independent field data collected based on a stratified random sample.

1.2.3. Spatial resolution and minimum mapping unit

Several map classes often occur in patches of 100 m^2 or smaller. We attempted to capture these occurrences, and to produce a map resembling manually delineated VMI maps, by modeling on 3-meter pixels (9 m^2). We smoothed the raw model outputs and filtered to a class-specific minimum patch size ranging from 81–441 m^2. Many occurrences above those thresholds likely remained undetected, because some essential predictors were derived from coarser resolution sources. Based on the average resolution of the predictors selected across all models, a typical minimum mapping unit (MMU) of 500 m^2 can be assumed, although many patches smaller than that are mapped.

1.3. Project area

1.3.1. Geography

North Cascades National Park Complex is located in the rugged northern Cascade Range on the U.S./Canada border in the state of Washington, lying about 150 km (95 miles) northeast of Seattle. It is the second largest of the North Coast and Cascades Network parks, with which it is shown in **Figure 1**. It is surrounded by other protected lands, including large wilderness areas in the Mount Baker and Okanogan National Forests and several provincial parks in British Columbia. Mount Baker, a 3,300-meter (10,800-foot) Cascades stratovolcano, lies nearby to the west, but does not create a strong rain shadow within the park.

[5] Polygons of some distinct yet rare vegetation types (e.g., ruderal meadows at old farm sites at OLYM) modeled poorly due to insufficient training data, but were easily reassigned by hand. Strips adjacent to roads also frequently mapped poorly and were reassigned manually.

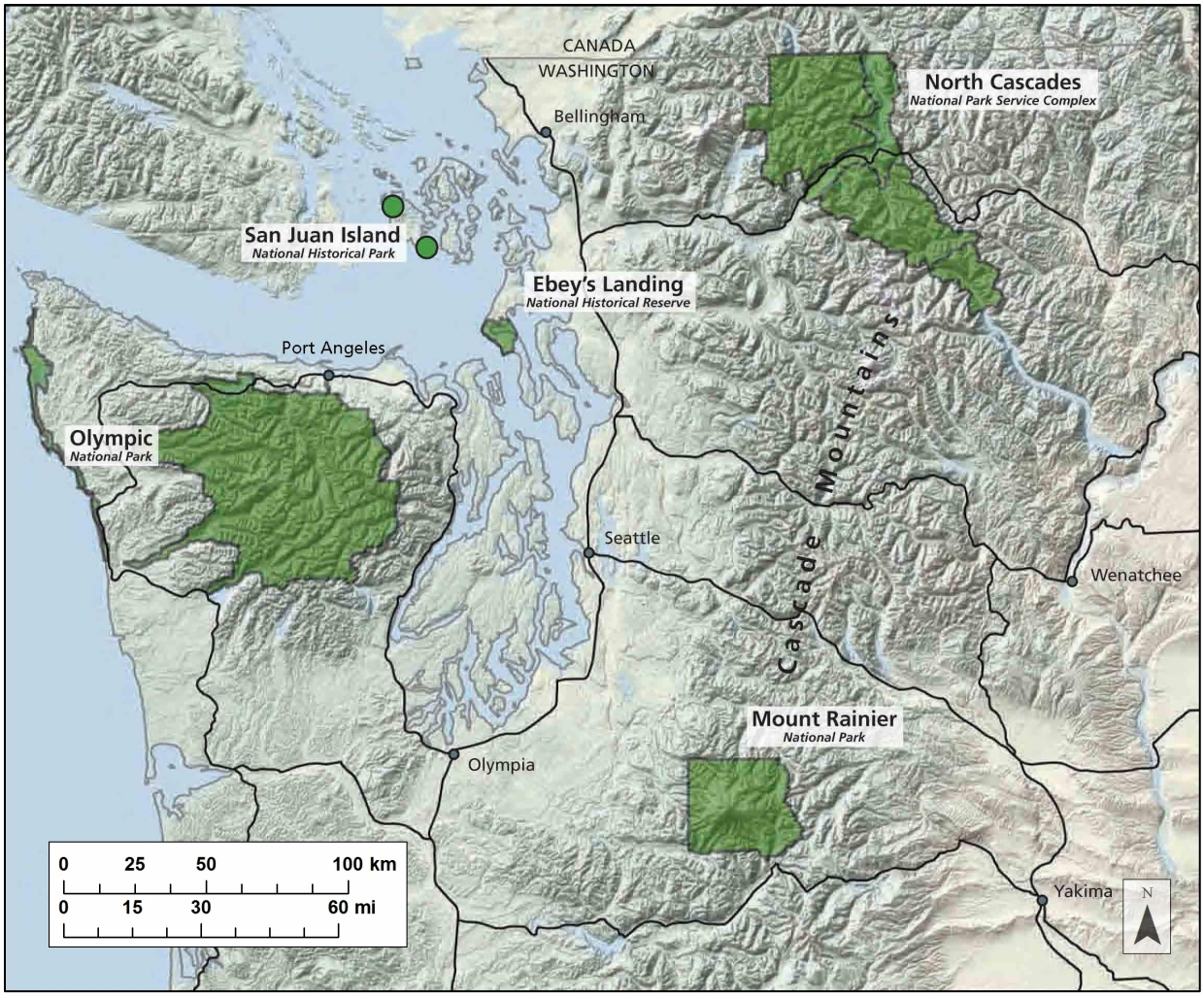


Figure 1. Map of North Coast and Cascades Network National Parks, from Copass and Ramm-Granberg (2016b). Fort Vancouver National Historic Site and Lewis & Clark National Historical Park lie farther south and are not shown.

The complex (**Figure 2**) is comprised of three units: North Cascades National Park, Ross Lake National Recreation Area (RLNRA) and Lake Chelan National Recreation Area (LCNRA). The Park itself contains a north (NPSN) and south (NPSS) unit. RLNRA mostly hosts lowland conifer and broadleaf forests, while LCNRA also includes dry montane and subalpine habitats. The NPS units are mostly characterized by montane and subalpine coniferous forests and woodlands, subalpine and alpine meadows and shrublands, steep rocky peaks, glaciers and other permanent snow and ice. Elevations range from 110 meters (350 feet) to well over 2,750 meters (9,000 feet), with high peaks distributed throughout the NPS units.

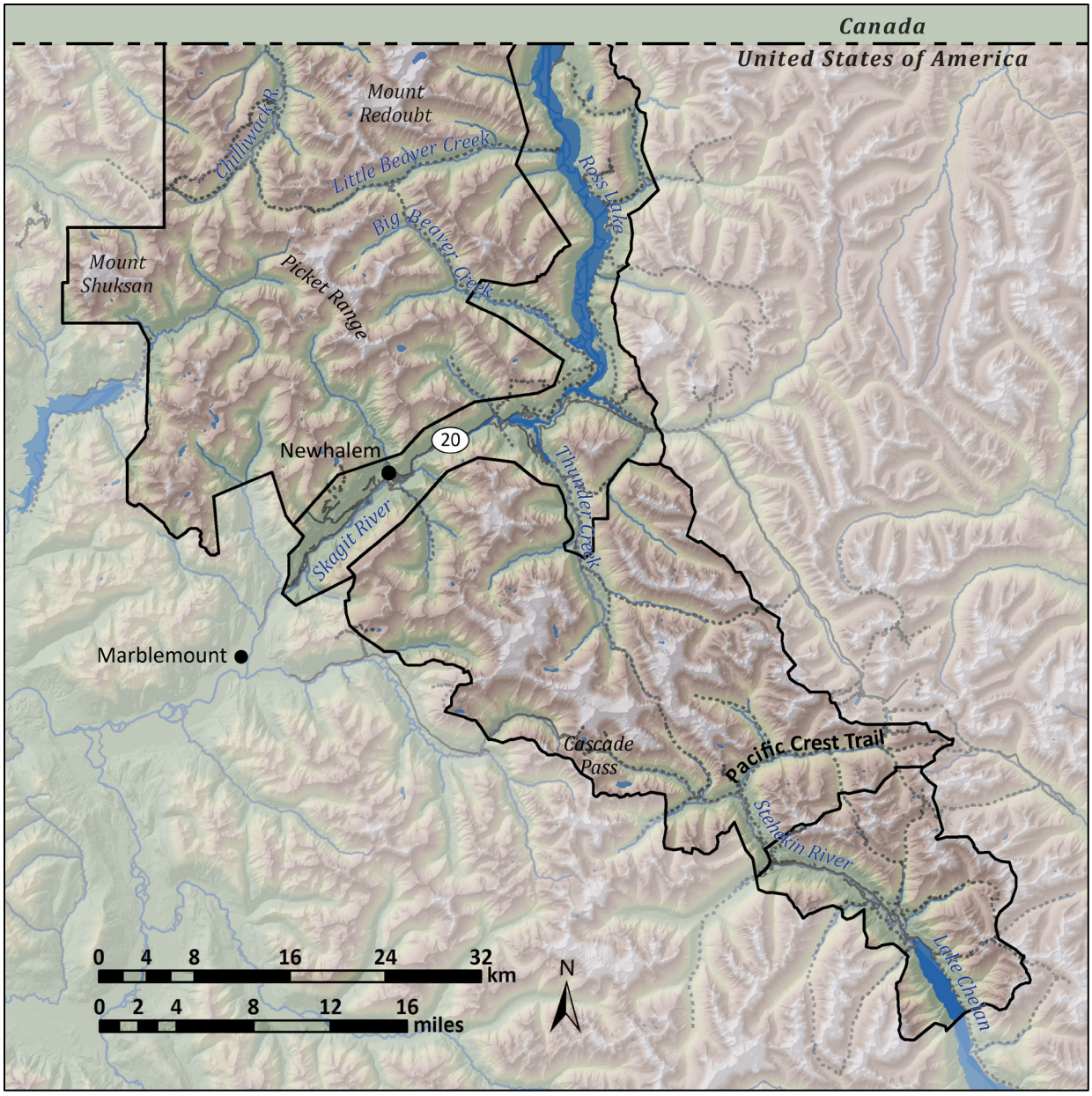


Figure 2. Map of North Cascades National Park Complex, illustrating topography, rivers, roads and other features. Units are delineated by solid black lines; from northwest to southeast they are the NPS North Unit (NPSN), Ross Lake National Recreation Area (RLNRA), NPS South Unit (NPSS), and Lake Chelan National Recreation Area (LCNRA).

The entire landscape has been extensively shaped by both alpine and continental glaciation, as well as by major rivers, including the Skagit, Nooksack, Baker, Cascade and Stehekin. The upper Skagit River was drowned beneath Diablo Lake and Ross Lake reservoirs, but Lake Chelan, fed by the Stehekin, is a naturally occurring lake whose great depth was increased only slightly by damming. The east side of the complex lies in an intense rain shadow cast by the mountains to its west; this results in quite distinct vegetation communities, especially in LCNRA where elevations are quite low.

We defined the project area as the complex (275,942 hectares; 681,867 acres), in addition to a variable-width buffer determined by availability of key predictor geospatial data. The surrounding buffer area, which was not assessed for map accuracy, accounts for 34% of the total project extent of 420,604 hectares (1,039,335 acres).

1.3.2. Environmental setting, bioclimatic zones and major vegetation types

The North Cascades are a jumbled mélange of folded, faulted, and uplifted terranes, stitched together with intrusive plutons and then deeply eroded (Tabor and Haugerud 1999). The range was heavily impacted by both continental and alpine glaciation during the Pleistocene. Large glaciers filled the river valleys, leaving the highest peaks as nunataks in a sea of ice, and creating broad valley bottoms and steep valley walls. Following the retreat of the Cordilleran ice sheet, alpine glaciation continued to shape the spires, benches, cirques, and moraines of the higher peaks.

Vertical relief within the complex spans over 2,700 meters (8,850 feet); the tallest peaks rise half their total height in the last two horizontal kilometers. The steep slopes result in the compression of a wide range of climates into a short distance. The capture of moisture from Pacific air masses further increases climate diversity by superimposing a strong precipitation decrease from west to east. Western lowlands and slopes receive about 400 cm (160 in) of precipitation per year, with snow accumulations of up to 46 feet (Fleischner and Weisberg 1992), but eastern slopes receive only 150–200 cm (60–80 in). Although the Cascades hydrologic divide runs well to the east of RLNRA, the rugged Picket Range—in the center of the north NPS unit—results in a strong rain shadow around Ross Lake, with dry slopes and balds resembling east-side environments (Franklin and Dyrness 1988). To the south, LCNRA is drier yet.

A range of disturbance processes reset successional pathways and increase landscape scale vegetation complexity. Between 2001 and 2018, wildland fires burned over an estimated 13,800 hectares (34,000 acres) of the complex, including about 4,900 hectares (12,200 acres) in the year 2015 (K. Kopper, pers. comm.). Insect outbreaks—primarily of natives such as mountain pine beetle (*Dendroctonus ponderosae*) and western spruce budworm (*Choristoneura occidentalis*)—habitually damage Douglas-fir, lodgepole pine, whitebark pine, silver fir and subalpine fir stands (Hoffman et al. 2015). These episodic disturbances primarily impact the east side. Avalanches and landslides are disturbance agents which are more equitably distributed across the complex (Antonova et al. 2013).

Topography, climate, and disturbance dynamics structure the vegetation of the complex, and the convergence of three major phytogeographic provinces—the Vancouverian flora from the west, the Rocky Mountain flora from the east, and the Canadian Circumboreal province from the north—further drive significant plant diversity. A total of 1,158 native vascular taxa[6] are thought to occur within the park (Biven and Rochefort 2010). The major vegetation zones, which are often

[6] Although Biven and Rochefort (2010) state that this is the number of native *species*, the lists provided clearly include subspecies and varieties in the count.

represented by distinct variants in the northwest, northeast, and southeast portions of the complex, are discussed below.

Lowland forests

In all but the warmest and driest parts of the park, lowland forests dominated by Douglas-fir (*Pseudotsuga menziesii*) and western hemlock (*Tsuga heterophylla*) occupy most of the landscape up to about 750 meters elevation. Western redcedar (*Thuja plicata*) is important in these forests, and grand fir (*Abies grandis*) is occasionally prominent. Understories vary depending on site moisture and stand history. Broadleaf forests in riparian areas are dominated by red alder (*Alnus rubra*) and black cottonwood (*Populus trichocarpa*), while in uplands they typically represent recovery from disturbance and are dominated by bigleaf maple (*Acer macrophyllum*) and/or paper birch (*Betula papyrifera*), usually with substantial Douglas-fir and other conifers. Talus slopes and the lower portions of avalanche tracks are occupied by tall vine maple (*Acer circinatum*) shrublands which extend up into the montane zone. In the rain shadow of the Picket Range, east of Ross Lake, western hemlock is substantially less prominent and paper birch is increasingly important. To the southeast, in the Stehekin watershed and adjacent areas, the lowland forest zone extends to higher elevations. Douglas-fir remains the most dominant tree, but it is accompanied in conifer stands primarily by ponderosa pine (*Pinus ponderosa*), and in mixed broadleaf stands by bigleaf maple.

Lower montane forests

The lower montane zone is defined primarily based on the codominance of silver fir (*Abies amabilis*) with western hemlock and the rarity of tree species associated with higher elevations. The lowest occurrences of silver fir are in moist forests on valley bottoms and north-facing slopes. The bulk of the zone, however, occurs in valley wall settings. On the west side, mesic forests characterized by silver fir and western hemlock are predominant up to about 1,150 meters elevation. In the area around Ross Lake, many areas with bedrock-limited soils are occupied by lodgepole pine (*Pinus contorta*) stands, which occupy lower montane elevations despite the relative scarcity of silver fir there. The zone is increasingly constricted toward the southeast as lowland forests extend higher. It is essentially eliminated in the Stehekin watershed, where lowland forests contact the upper montane zone. Though widespread, the lower montane zone occupies a smaller proportion of NOCA than the other NCCN parks[7] due to the steeper and generally drier conditions.

Upper montane forests

The upper montane zone is characterized by closed forests with substantial cover of higher elevation species such as mountain hemlock (*Tsuga mertensiana*), subalpine fir (*Abies lasiocarpa*) and Alaska-cedar (*Callitropsis nootkatensis*). It ranges from about 1,150 to 1,500 meters on the west side, where silver fir remains at least codominant, and from 900 to 1,600 meters on the east side, where Douglas-fir extends upward from lowland forests and generally codominates with subalpine fir. Tall shrublands of Sitka alder (*Alnus viridis*), often with Alaska-cedar, are frequent in avalanche tracks and other disturbed areas. On the east side, and particularly in the Stehekin watershed, dense post-fire

[7] The lower montane zone is especially abundant at OLYM.

shrublands of snowbrush (*Ceanothus velutinus*) and Scouler's willow (*Salix scouleriana*) are common.

Subalpine environments

Above the closed forest zone, upper montane forests transition into subalpine woodlands and tree islands. Subalpine *parklands*, mosaics of wooded areas with dwarf shrublands and meadows, are predominant from about 1,500 to 1,700 meters elevation on the west side, where mountain hemlock and Alaska-cedar are the dominant trees, and from 1,600 to 1,900 meters on the east side, where subalpine fir is more significant. On the driest east-side ridges, whitebark pine (*Pinus albicaulis*) and subalpine larch (*Larix lyallii*) join subalpine fir. These woodlands and tree islands are interspersed with shrublands and meadows at a range of spatial scales. Subalpine shrublands are characterized by dwarf ericaceous shrubs such as pink mountain-heather (*Phyllodoce empetriformis*), white mountain-heather (*Cassiope mertensiana*), Cascade blueberry (*Vaccinium deliciosum*) and grouse whortleberry (*Vaccinium scoparium*), with some taller shrubs such as big huckleberry (*Vaccinium membranaceum*) and Sitka mountain-ash (*Sorbus sitchensis*) in protected areas. Herbaceous subalpine vegetation is represented by several meadow types, with common species including subalpine lupine (*Lupinus latifolius*), wandering daisy (*Erigeron glacialis*), green fescue (*Festuca viridula*), showy sedge (*Carex spectabilis*), mountain sandwort (*Eremogone capillaris*), scarlet Indian paintbrush (*Castilleja miniata*) and spreading phlox (*Phlox diffusa*).

Alpine environments

In the alpine zone, which generally ranges upward from about 1,700 meters on the west side or 1,900 meters on the east side, tree cover is reduced to stunted krummholz of mountain hemlock and subalpine fir. Dwarf shrublands and meadows transition into sparser alpine variants with shorter vegetation and fewer species. Common species include pink and white mountain-heather, partridgefoot (*Luetkea pectinata*), Piper's woodrush (*Luzula piperi*), black alpine sedge (*Carex nigricans*), Tolmie's saxifrage (*Micranthes tolmiei*), Parry's and Drummond's rushes (*Juncus parryi, J. drummondii*), and mountain hairgrass (*Vahlodea atropurpurea*). Eventually vegetation gives way almost completely to barren bedrock, talus, permanent snowfields and glaciers.

1.3.3. Human history

The rugged passes of the North Cascades were used by indigenous peoples to move between the northwest coast and interior plateau regions of the Pacific Northwest. For at least 10,000 years, people living in the lowlands traversed the passes and used the high country to fish, hunt, gather plants, conduct religious practices and connect with other tribes from all over the region. The landscapes that are now included in the park are intrinsically linked to the cultural identity of these tribal groups, which include the Nooksack, Sto:lo, Sauk-Suiattle, Upper Skagit, Chelan, Methow, Entiat, Wenatchi and Nlakapamux (Boxberger 1996).

A majority of the current park was set aside in 1897 as the Washington Forest Reserve, before being transferred to the United States Forest Service (USFS). Multiple use management and extraction, including grazing, mining, logging, hydroelectric projects and recreation developments occurred throughout the first half of the twentieth century. The Seattle City Light Skagit Hydroelectric Project, including the Gorge, Diablo and Ross Dams, started in the 1920s and was completed in 1953 (Wilma

2003). During this period, failed efforts to create a national park in the region provoked a longstanding jurisdictional dispute between the USFS and the NPS. At the time of dam completion, the North Cascades remained an extensive, but increasingly isolated pocket of wilderness in a region undergoing rapid post-war growth. The conservation movement of the 1960s shifted public support in favor of wilderness protection; this sentiment was further strengthened in reaction to increased timber extraction pressure. During the early 1960s, there was conflict between disparate park visions: some favored wilderness, others a developed natural area with access by roads, trams and helicopters (Louter 1998).

Proposals, negotiations, battles, counterattacks and compromises eventually gave way to legislation resolving the competing plans. The long political process culminated with the 1968 establishment of one of the largest U.S. wilderness parks created in the latter half of the century. The resulting park complex covers a total land and water area of 275,942 ha (684,000 ac).[8] Ross Lake National Recreation Area (47,275 ha; 17.1% of the full complex) includes Ross Lake itself in addition to hydroelectric infrastructure, Highway 20, and associated development. RLNRA and the highway split the NPS portion of the complex into the North Unit (118,554 ha; 43.0%) and South Unit (84,328 ha; 30.6%). The creation of Lake Chelan National Recreation Area (25,784 ha; 9.3%) reduced the access concerns of property owners and other stakeholders in the lower Stehekin and upper Lake Chelan areas. Lands bordering the complex to the east were retained by the USFS as the Glacier Peak and Pasayten Wilderness areas.

Nearly 95% of the complex is protected as the Stephen Mather Wilderness, which spans the units. Development remains minimal, with substantial areas 15–25 linear kilometers (10–15 miles) or more from the nearest road. The extensive trail system has few access points, posing challenges to fieldwork logistics.

1.3.4. Previous vegetation studies

Compared to other parts of the Cascades, the relative inaccessibility of the project area has limited the number of vegetation studies. Botanic collections were made starting as early as 1892, but even in 2005 collections were generating new voucher records for the park (Biven and Rochefort 2010).[9] Douglas (1972) and Douglas and Bliss (1977) sampled and classified plant communities in the park's subalpine and alpine zones, and Risvold and Fonda (2001) sampled wetlands. Other studies took place in specific areas, including the Chilliwack Valley (Comulada 1981), Chowder Ridge (Taylor and Douglas 1977), Stetattle Creek (Wagstaff and Taylor 1980), Nooksack Cirque (Oliver et al. 1985), Big Beaver Valley (Vanbianchi and Wagstaff 1988) and the Lake Chelan area (Alverson and Arnett 1986). Fire dynamics research has been done in the Stehekin and Thunder Creek watersheds (Prichard 2003, Cwynar 1987).

---

[8] This and all other area figures were computed from a GIS layer provided by NPS. They include water surfaces, including that of Ross Lake and Lake Chelan. They may disagree with other published figures which have been determined differently.

[9] The inventory conducted by Biven and Rochefort (2010) lists 1,351 vascular plant taxa; of these 1,158 are native.

Studies in the broader region and in adjacent areas have also contributed to the understanding of vegetation patterns. For example, Franklin and Trapper (1963) classified subalpine communities throughout the North Cascades, and extensive sampling and classification development were performed in support of a regional analysis of grizzly bear habitat (Almack et al. 1993, Wooten and Morrison 2006). Detailed and relevant vegetation community descriptions were developed for the adjacent Okanagan (Williams and Lillybridge 1983), Wenatchee (Lillybridge et al. 1995) and Mount Baker-Snoqualmie (Henderson et al. 1992) National Forests. Many of these classification results, as well as other regional studies, were incorporated into the NVC revisions that have occurred as part of this project (Crawford et al. 2009, updated by Ramm-Granberg et al. 2021).

Vegetation mapping efforts have proceeded in fits and starts over the last 85 years. Mapping of the Douglas-fir zone of the entire Pacific Northwest produced detailed maps of dominant forest canopy species and disturbance impacts (Andrews and Cowlin 1936). The next spate of mapping came with the increasing availability and ease of use of satellite imagery in the 1980s. Agee and Pickford (1985) developed a vegetation and fuels map, identifying 22 cover types via ordination and two-way species indicator analysis,[10] and mapping the types using 60-meter resolution Landsat MSS and climate data (Agee and Kertis 1987). Almack et al. (1993) describe the mapping of the classification developed above for the Grizzly Bear Ecosystem Mapping Project, also using mid-1980s Landsat MSS data. Finally, Pacific Meridian Resources used a modified supervised classification method to map 17 vegetated and three unvegetated classes, in addition to some elements of vegetation structure, from 30-meter resolution Landsat TM imagery collected in 1988 (PMR 1997). The classification focused on forests; all herbaceous and shrub-dominated vegetation were lumped into a single class each.

1.4. Project timeline

The following timeline describes the primary activities during each year of the 15-year project. Only activities at NOCA are described; activities were focused on other NCCN parks during several years.

2005 — NPS project initiation, planning and scoping, fieldwork for accuracy assessment of previous generation vegetation map (PMR 1997), database development.
2006 — Classification fieldwork, planning and scoping, database development.
2007 — Classification fieldwork, development of association-level NVC and database.
2008 — INR joins project. Development of map training fieldwork protocols, development of association-level and higher-level NVC.
2009 — Development of map training fieldwork protocols, development of higher level NVC, development of predictor metrics methodologies.
2010 — Development of predictor metrics methodologies.
2011 — Development of predictor metrics methodologies, NAIP and satellite image collection and processing.
2012 — Map training sample design, map training fieldwork.

[10] Eight coniferous forest types were identified, each with an open and closed canopy variant. The other six types were broadleaf forests, tall shrublands on avalanche tracks, disturbed lowland herbaceous vegetation, green fescue meadows, lush subalpine herbaceous meadows, and mountain-heather dwarf shrublands.

2013 — Map training fieldwork, NAIP and satellite image collection and processing.
2014 — Map training fieldwork.
2015 — NPS gives INR go-ahead to approach NCCN projects as a single entity and to work on the classification as needed for successful mapping. Training data quality control, floristics methods development, satellite image collection and processing.
2016 — Production of draft vegetation map for use in stratification of accuracy assessment sampling, AA sample design, AA fieldwork, training data quality control, floristics quality control, AA data quality control, mapping associations development, NAIP image collection and processing.
2017 — Training data quality control, floristics quality control, AA data quality control, mapping associations development, map classification development, development of new topographic predictor metrics, refinement of nested texture metrics methodology.
2018 — Training data quality control, mapping associations development, map classification development, satellite image collection and processing, draft map production.
2019 — NPS draft map review. Training data quality control, mapping associations completion, map classification completion, AA data quality control, development of shadow correction methods for NAIP imagery.
2020 — Production of final maps, AA analysis, report.

2. Methods and Results

Most NPS VMI maps have been produced by photo-interpretation (PI). We used model-based methods instead, because of the large size of the NCCN parks and the visual similarity of many of the key plant communities. Machine learning methods were used to extrapolate from a large set of classified field plots to the full extent of the park. The mapped vegetation units were 3-meter pixels rather than polygons, because we found that pixel-based modeling was the only reliable method for boundary detection between visually similar map classes. We invented and developed a variety of innovative image processing and modeling techniques to achieve finer spatial resolution and greater accuracy than is typical of model-based vegetation maps. The primary phases of the mapping process—many of which occurred concurrently—were collection and basic quality control of training field data (**Section 2.1**), floristics data treatment and associated plot QC (**Section 2.2**), development of mapping associations and associated plot QC (**Section 2.3**), crosswalking associations to map classes (**Section 2.4**), acquisition and pre-processing of predictive data sources (**Section 2.5**), development and creation of predictive metrics (**Section 2.6**), machine learning-based modeling (**Section 2.7**) and post-processing and editing (**Section 2.8**).

2.1. Field data

Field data were used to develop an association classification (**Section 2.3**) and map classification (**Section 2.4**) and to provide training data for the machine learning processes used to create the map (**Section 2.7.2**). The same classification was used at each of the large NCCN parks; its development drew from 4,110 plots collected at MORA, OLYM and NOCA. Because the classification relied on data collected from all parks, each of those protocols is reviewed here. The map training data included 2,198 field plots, all collected at NOCA, in addition to the PI plots discussed later.

2.1.1. Sample collection

We trained predictive models using plots from multiple sampling efforts with distinct sample designs and field protocols. Most were collected during the mapping phase of the project, between 2008 and 2015. Although field protocols evolved over this time, the fundamentals were in place by 2010.[11] We also used many plots collected in 2006–07, during the initial NVC development phase of the project (see Crawford et al. 2009). Many of these plots included full floristic data and suited our needs well.

Training data for inadequately sampled vegetation types were supplemented by incorporating plots from a variety of other field efforts in the parks between 2005 and 2015. Although the protocols varied widely for these plots, through the quality control process (**Section 2.1.2**) we converted all data to a standardized format: a circle of known radius georeferenced to aerial imagery collected in 2015 (**Section 2.5.1**) and a species list with cover estimates to the nearest one percent (or *trace* if

[11] The primary requirements were (a) plot dimensions adjusted to match a homogeneous vegetation patch, up to a maximum 40-m radius circle; (b) documented plot center location and radius in four cardinal directions; (c) diagram illustrating landmarks and land cover transitions, for spatial QC; (d) reasonably complete floristic data including visual percent cover estimates; and (e) photos at cardinal directions from center, for later spatial and floristic QC.

present in smaller amounts). Plots with reasonably complete species occurrence data were designated *full-ocular* plots, while those with incomplete data were designated as *partial-ocular*.

The following sections outline the sample designs, field protocols, and data QC procedures for all plots used to create the map classification or the NOCA vegetation map. **Table 2** summarizes this information. The table does not include photo-interpreted plots, which were mainly used as training for unvegetated classes.

Table 2. Total number of field plots used to create the map classification ("full floristics plots") and the NOCA vegetation map ("model plots"), categorized by park and collection effort. "Ocular type" specifies whether documenting full species cover data was an objective of the protocol. Photo-interpreted plots were also used for modeling; those are not included here.

Collection effort	Collection years	Collected by	Ocular type	Full Floristic plots			Model plots [A]
				MORA	OLYM	NOCA	NOCA
VIP classification	2006–07	NPS	Full	186	228	79	88
VIP mapping protocol X	2008	NPS	Partial [D]	44	0	0	0
VIP mapping protocol M [B]	2009–11	NPS	Partial [D]	33	10	0	0
VIP mapping protocol Y	2009–11	NPS	Full	151	1,094	0	0
VIP mapping protocol Z [C]	2012–15	NPS	Full	0	233	1,612	1881
VIP mapping protocol Q	2014, 2019	INR, NPS	Partial [D]	1	0	0	0
PMR accuracy assessment	2005–06	NPS	Full	61	18	10	134
Monitoring reconnaissance	2005–14	NPS	Full	91	46	48	95
UW forest community	2015	UW	Full	165	0	0	0
Total	**–**	**–**	**–**	**732**	**1,629**	**1,749**	**2,198**

[A] There is a large amount of overlap with full floristics plots collected at NOCA. Totals include additional training data created in adjacent or included patches based on plot notes.

[B] Totals reflect number of individual patches from subdivision of original mosaic plots.

[C] Including verification plots (revisits and updates to previously collected plots using revised protocols).

[D] Some plots had floristics supplemented later by inspection of field photos and were treated as full ocular plots.

NCCN VIP plot types

The following plot types were collected especially for the NCCN VIP and are listed chronologically.

Classification (2006–07)

These plots were intended primarily to support development of the NVC for the NCCN (Crawford et al. 2009, Ramm-Granberg et al. 2021) and were collected by NPS at the three major parks. Crews sampled a broad range of types.[12] The protocol included collection of a comprehensive species list

[12] Data collection was particularly focused on vegetation types known to be undersampled in the existing draft classification, such as shrub-dominated avalanche chutes. Circular plots were located opportunistically in homogenous patches that were large enough to meet plot size recommendations. Forested plots were sampled over

with cover estimates and several field photos. Plots were assigned to an association from an early draft of the NVC, or to a provisional association if no good match could be determined.

Mapping protocol X (2008)

Collected only at MORA, these plots represented the earliest sampling implemented during the VIP mapping phase, before significant improvements were made in spring 2009. Polygons segmented from true color aerial imagery were targeted for opportunistic sampling. Estimates of crown cover were made for up to three tree species, but understory species were generally not documented. At least two photos were taken. A vegetation association was selected from an early draft of the NVC.[13]

Mapping protocol Y (2009–11)

A variety of changes to the sample design and field protocols were implemented beginning in the 2009 field season. A stratified sample design was implemented to guide the effort.[14] Map sheets produced from resolution-merged aerial and satellite imagery allowed crews to navigate more efficiently, locate plots more accurately, and document vegetation patches for later use; similar paper maps were used in all subsequent field efforts (see **Figure 3** for an example from the NOCA fieldwork in 2013). Beginning with protocol Y, plot dimensions were determined by the extent of the homogeneous vegetation patch present at plot center, up to a maximum 40-meter radius circle. Vegetation transitions along four perpendicular transects from the plot center were documented, and the plots were drawn by hand on the map sheet and documented in greater detail in a field diagram on the data sheet. Species cover was visually estimated for most plant species present.[15]. Photos were taken at cardinal directions from plot center, and the best-fit vegetation association was selected from the newly published NCCN NVC classification (Crawford et al. 2009).

---

an 11.3-meter radius (400 m^2), woodlands and shrublands over an 8.0-meter radius (200 m^2), and herbaceous and sparsely vegetated plots over a 5.6-meter radius (100 m^2). Notes on soil conditions and fire history were taken.

[13] A plot center was selected within a representative homogeneous area for assessment of the vegetation association and canopy composition, which was documented within a 20-meter radius circle around the point. A secondary association was listed if there was a clearly distinct association located nearby, but no corresponding location information was documented.

[14] Sample sites were targeted by using an unsupervised classification technique to break parks into fifty distinct strata based on Landsat reflectance data, topographic metrics, and geographic blocks. Within each stratum, targets were established in the most homogeneous Landsat pixel clumps within several hundred meters of trails. Field crews navigated to these locations and also established opportunistic plots in homogeneous areas, occurrences of vegetation types that had been poorly represented in the targeted sampling.

[15] Over the years, field crews were increasingly comprised of returning, experienced members, and the capability to collect complete species composition data increased. The OLYM protocol Y data (collected primarily in 2010–11) therefore have considerably greater completeness than that from MORA (collected primarily in 2009). NPS's original aim had been to produce a completed map of MORA before the other parks, so the MORA fieldwork was done in "hurry-up" mode, and a return there with the more experienced crews was never realized.

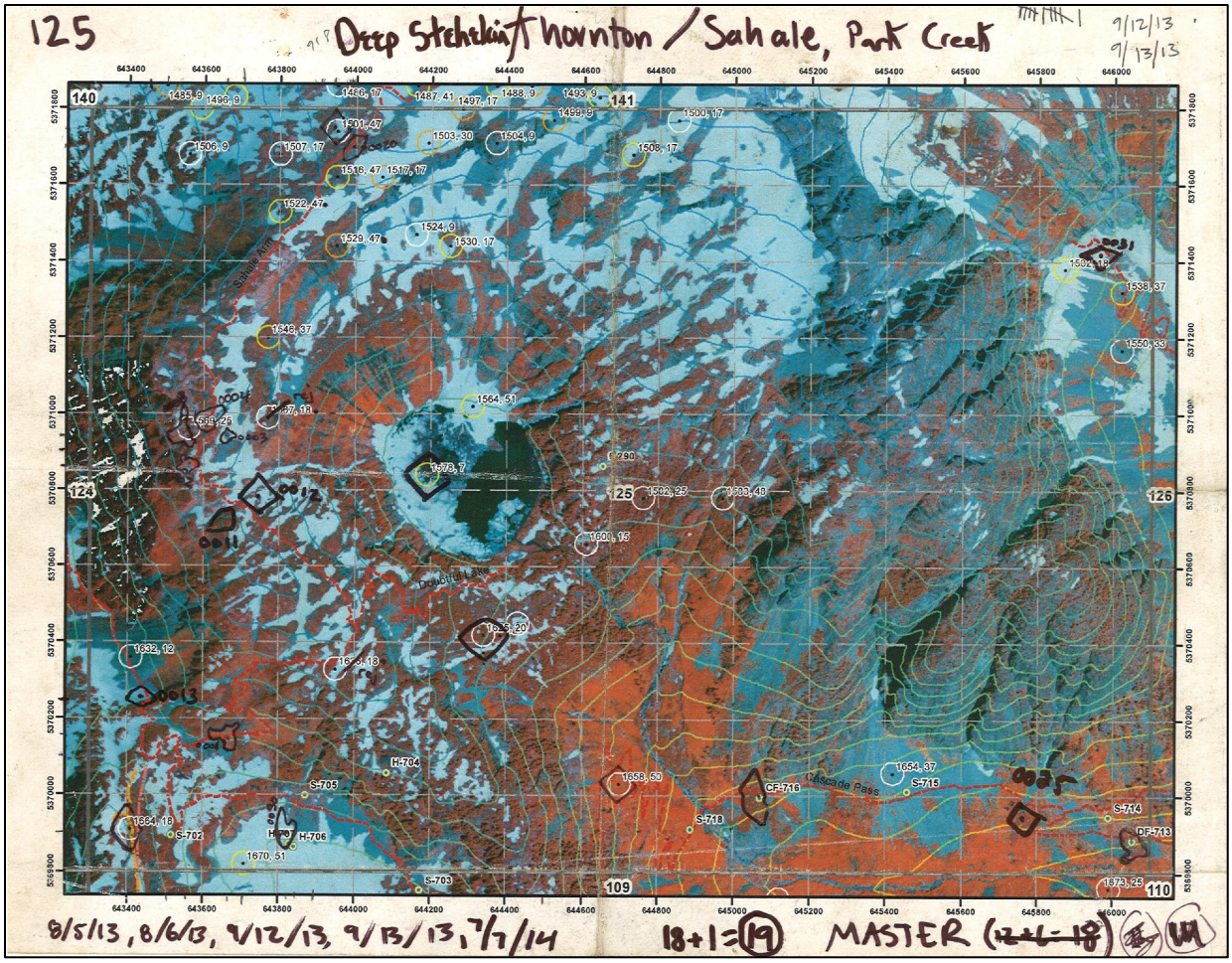


Figure 3. Completed mapping phase map sheet (protocol Z, 2013).

Mapping protocol M (2009–11)

Field crews targeted subalpine and alpine areas to determine whether fine-scale mosaics of distinct alliance level vegetation might be combined into recurring mappable types.[16] This effort targeted mosaics of vegetation patches, each of which was smaller than those considered for sampling during prior efforts. Species cover was estimated for the most significant species in each distinct patch. and an association was chosen from Crawford et al. (2009). The data were later analyzed to assess whether the patches might be combined into consistent coarser-scale vegetation types, but patterns were not consistent enough to allow this.

[16] Early in the project, we had planned to map to the default NPS vegetation inventory MMU of one half-hectare (Lea and Curtis 2010). Most subalpine alliance-level vegetation occurs at considerably finer scales than this.

Mapping protocol Z (2012–14)

Field sampling was guided by a revised stratified sample design.[17] The field protocol was similar to protocol Y, except that association transitions along each transect were documented in great detail. This allowed additional plots to be generated later if needed. Field crews were remarkably stable during this period, allowing the collection of more complete species cover data at nearly all plots. The data collected during this time period was integral to the refinement of the mapping associations (**Section 2.3**). Protocol Z was the main source of map training data at NOCA; a completed fieldsheet is shown in **Figure 4**.

Mapping protocol Q (2014)

The primary aim of this brief sampling effort was to collect data in accessible but unsampled regions of MORA, which had remained the most poorly sampled park. Early draft maps had difficulties separating forest types containing *Tsuga mertensiana* from those lacking it. To address this problem, we created a species distribution model for *T. mertensiana* and used it to target locations with an intermediate likelihood of presence. We also targeted sites exhibiting a high degree of draft map class uncertainty. Plots were 11-meter radius circles. All species with significant presence were documented, but cover was only estimated for tree species. Understory plants were simply listed in descending order of prominence.

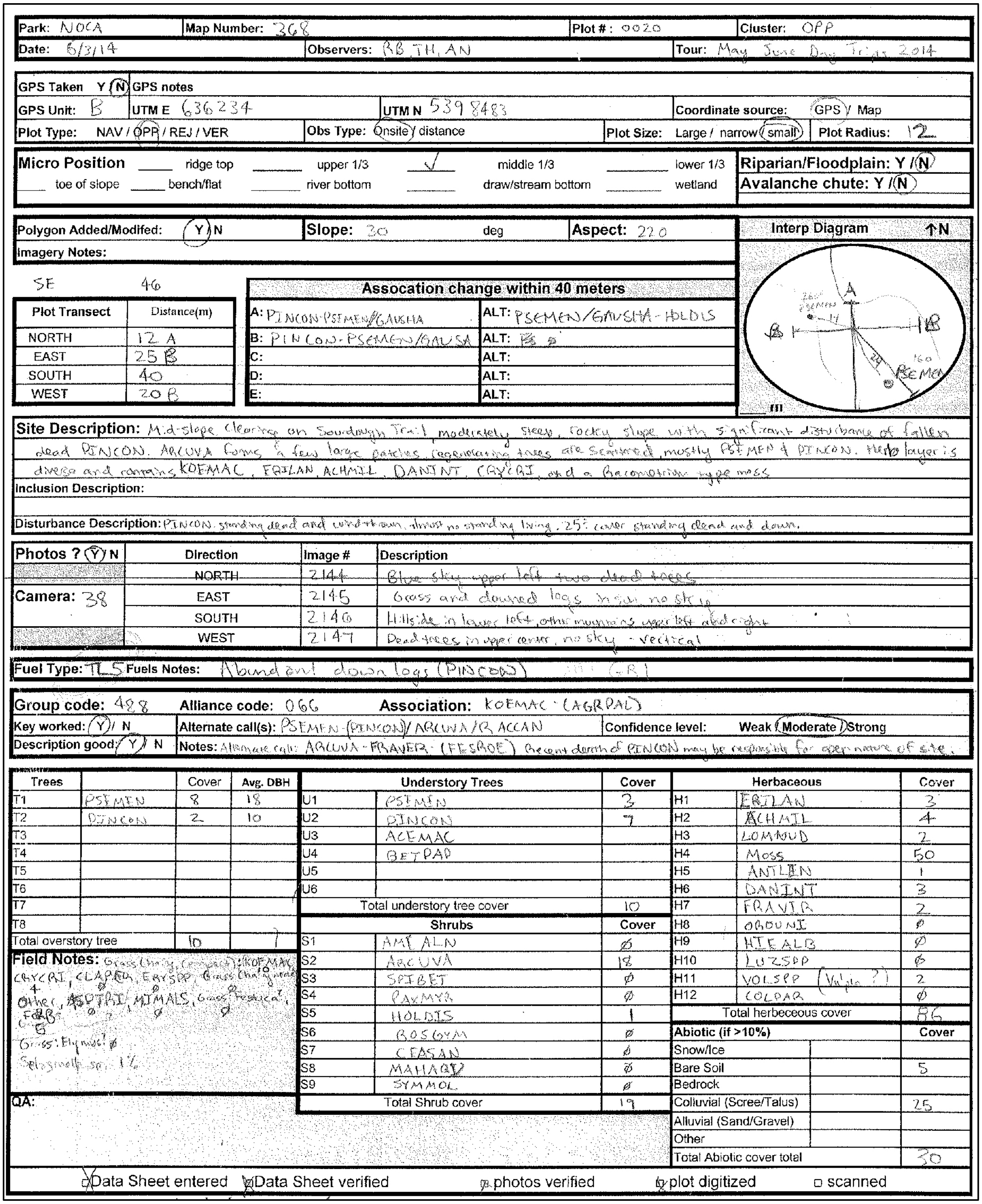
Other plot types

We used plots collected for several other projects to provide data for vegetation types that would have otherwise been inadequately sampled. During the quality control process (**Section 2.1.2**) we adapted the available information to our purposes, making use of field notes, photos, and imagery.

PMR accuracy assessment plots (2005–06)

These plots were collected to assess the accuracy of the previous generation of NCCN vegetation maps (Pacific Meridian Resources 1997). Sampling locations were stratified across the mapped classes. Plots were 28.5-meter radius circles and were labeled with an association from an early NVC draft or with a provisional call. Cover of the top three species in various height strata was collected.

[17] Combined unsupervised-supervised classification was used to break the landscape into 52 unique strata of Landsat spectral reflectance, climate metrics, and topographic curvature. Again, the most homogeneous accessible areas were determined using an automated procedure, and these locations were targeted on an as-needed basis by crews, in addition to sampling opportunistically in homogeneous vegetation encountered along the routes.



Park: NOCA | Map Number: 368 | Plot #: 0020 | Cluster: OPP

Date: 6/3/14 | Observers: RB, TH, AN | Tour: May June Day Trips 2014

GPS Taken Y / (N) GPS notes

GPS Unit: B | UTM E 636234 | UTM N 5398483 | Coordinate source: (GPS) / Map

Plot Type: NAV / (OPP) / REJ / VER | Obs Type: (Onsite) / distance | Plot Size: Large / narrow / (small) | Plot Radius: 12

Micro Position ____ ridge top ____ upper 1/3 __✓__ middle 1/3 ____ lower 1/3 ____ toe of slope ____ bench/flat ____ river bottom ____ draw/stream bottom ____ wetland

Riparian/Floodplain: Y / (N)

Avalanche chute: Y / (N)

Polygon Added/Modifed: (Y) / N | **Slope:** 30 deg | **Aspect:** 220 | Interp Diagram ↑N

Imagery Notes:

SE 46

Plot Transect	Distance(m)
NORTH	12 A
EAST	25 B
SOUTH	40
WEST	20 B

Assocation change within 40 meters

A: PINCON-PSEMEN/GAUSHA	ALT: PSEMEN/GAUSHA-HOLDIS
B: PINCON-PSEMEN/GAUSHA	ALT:
C:	ALT:
D:	ALT:
E:	ALT:

Site Description: Mid-slope clearing on Sourdough Trail moderately steep, rocky slope with significant disturbance of fallen dead PINCON. ARCUVA forms a few large patches. Regenerating trees are scattered, mostly PSEMEN & PINCON. Herb layer is diverse and contains KOEMAC, ERILAN, ACHMIL, DANINT, CRYCRI, and a Racomitrium type moss

Inclusion Description:

Disturbance Description: PINCON standing dead and windthrown, almost no standing living. 25% cover standing dead and down.

Photos ? (Y) N	Direction	Image #	Description
	~~NORTH~~	~~2144~~	~~Blue sky upper left two dead trees~~
Camera: 38	EAST	2145	Grass and downed logs in sun, no sky
	SOUTH	2146	Hillside in lower left, other mountains upper left and right
	WEST	2147	Dead trees in upper center, no sky - vertical

Fuel Type: TL5 | Fuels Notes: Abundant down logs (PINCON)

Group code: 488 | **Alliance code:** 066 | **Association:** KOEMAC-(AGRPAL)

Key worked: (Y) / N | Alternate call(s): PSEMEN-(PINCON)/ARCUVA/R ACCAN | Confidence level: Weak (Moderate) Strong

Description good: (Y) / N | Notes: Alternate call: ARCUVA-FRAVER-(FESROE). Recent death of PINCON may be responsible for open nature of site.

Trees		Cover	Avg. DBH
T1	PSEMEN	8	18
T2	PINCON	2	10
T3			
T4			
T5			
T6			
T7			
T8			
Total overstory tree		10	

Understory Trees		Cover
U1	PSEMEN	3
U2	PINCON	7
U3	ACEMAC	
U4	BETPAP	
U5		
U6		
Total understory tree cover		10

Shrubs		Cover
S1	AMEALN	ø
S2	ARCUVA	18
S3	SPIBET	ø
S4	PAXMYR	ø
S5	HOLDIS	1
S6	ROSGYM	ø
S7	CEASAN	ø
S8	MAHAQU	ø
S9	SYMMOL	ø
Total Shrub cover		19

Herbaceous		Cover
H1	ERILAN	3
H2	ACHMIL	4
H3	LOMNUD	2
H4	Moss	50
H5	ANTLAN	1
H6	DANINT	3
H7	FRAVIR	2
H8	OROUNI	ø
H9	HIEALB	ø
H10	LUZSPP	ø
H11	VOLSPP (Viola ?)	2
H12	COLPAR	ø
Total herbeceous cover		86

Abiotic (if >10%)		Cover
Snow/Ice		
Bare Soil		5
Bedrock		
Colluvial (Scree/Talus)		25
Alluvial (Sand/Gravel)		
Other		
Total Abiotic cover totel		30

Field Notes: Grass (Carex, Juncaceae): KOEMAC, CRYCRI, CLAPER, CARSPP, Grass (Carex, Juncaceae) other, AGPTRI, MIMALS, Grass: Festuca?, FORB: ...; Grass: Elymus? ø; Selaginella sp. 1%

QA:

☒ Data Sheet entered ☒ Data Sheet verified ☐ photos verified ☐ plot digitized ☐ scanned

Figure 4. Completed mapping phase data sheet (protocol Z, 2014).

Monitoring reconnaissance plots (2005–15)

Plots were collected in forests and subalpine areas to assess the suitability of randomly selected locations for long-term monitoring plots. Forest plots were 50x50-meter squares; subalpine plots varied in size. Cover was estimated for dominant species and the surrounding area was coded to an association from the most recent available NCCN classification.

Forest legacy plots (2015)

Plots collected at MORA in the 1970s and 1980s (Franklin et al. 1988) were revisited in a project of Dr. Hille Ris Lambers at the University of Washington. The cover of understory vegetation was estimated over several small subplots, but tree species were documented by counting the number of stems in distinct bole diameter classes rather than by cover.[18] The field notes allowed us to convert these estimates into cover estimates that were reasonably compatible with other plots, and the fairly complete ocular data collected at these plots were critical in providing reference floristics data at MORA, which had been generally undersampled in this regard in earlier efforts. No vegetation type was assigned in the field.

Photo-interpreted plots (2014–19)

We supplemented the field-collected data for several structurally-defined, abiotic, and otherwise distinctive map classes[19] by assigning PI locations where it was possible to do so confidently. Generally, we approached this as an iterative process, using previous map generations to assign additional data in areas that appeared to map poorly. We also used targeted absence plots to improve mapping in areas where we could only narrow down the correct answer to one of several map classes but were confident that draft maps were in error. These plots were used only in the binary models (**Section 2.7.3**) which pitted the specified possibly correct classes against the clearly incorrect classes.[20]

---

[18] Understory plants were assessed via a cover estimate in four 1x1-meter quadrats and presence/absence in a 4-meter radius circle; these were converted into an average percent cover for each species. Trees taller than breast height were individually measured in a 12.6-meter radius circle and were summarized in m^2/ha. We converted the stem counts to rough percent cover estimates assuming that crown area was proportional to the square of bole diameter at the individual tree level.

[19] The map classes that received PI plots at NOCA were C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND (distinctive signature and structure), C21–MOUNTAIN HEMLOCK, SUBALPINE FIR AND HEATHER WOODLAND (distinctive structure), C22–SUBALPINE LARCH WOODLAND (larch with fall foliage), C26–CONIFER KRUMMHOLZ AND TREED CLIFF, B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND, S40W–LOW ELEVATION SHRUB-DOMINATED WETLAND, S43–SITKA ALDER SHRUBLAND (especially on shaded, steep north-facing slopes), S45–VINE MAPLE SHRUBLAND (vine maple leaves bright red in 2015 NAIP imagery), S48–SUBALPINE HEATHER SHRUBLAND, S49–ALPINE HEATHER SHRUBLAND, H51W–SUBALPINE HERBACEOUS WETLAND, H58–VEGETATED BALD, R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE, R72–COLLUVIAL BARREN, R73–BEDROCK BARREN, W81–FRESH WATER and W82–EXPOSED SNOW AND ICE.

[20] Examples included tall shrubland openings in forests (either S43–SITKA ALDER SHRUBLAND or S46–SNOWBRUSH AND SCOULER'S WILLOW SHRUBLAND) that had a tendency to map as C16–NORTH CASCADES DOUGLAS-FIR AND SUBALPINE FIR WOODLAND, and conifer forests in very flat areas (either C16–NORTH CASCADES DOUGLAS-FIR AND

The efforts above resulted in the collection of about 6,500 field plots across the three parks. The NOCA field plots used for map training are shown along with photo-interpreted plots in **Figure 5**.

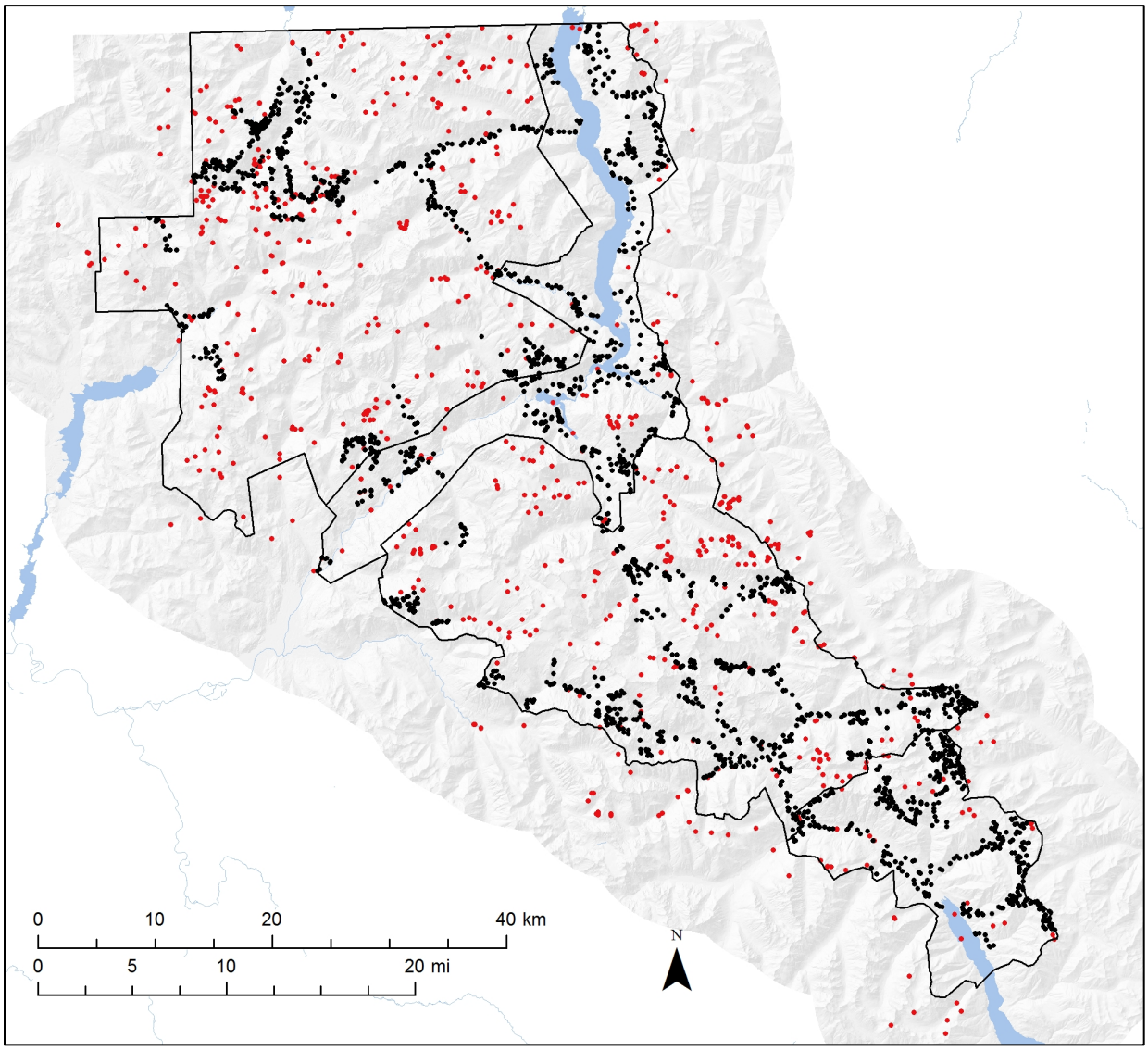


Figure 5. Model training plot locations. Of the 2,980 plots, 2,198 were collected in the field (shown in black), and 782 were photo-interpreted (shown in red).

2.1.2. Basic quality control

An extensive quality control process was necessary due to the many distinct field protocols used, the variable field effort applied at different plots (particularly regarding species ocular estimates), ambiguities in patch delineation and ocular estimates resulting from heterogeneous vegetation,

SUBALPINE FIR WOODLAND or C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND) that had previously mapped as W81–FRESH WATER.

spatial inaccuracies due to poor GPS reception, field call misassignments caused by key artifacts, updates to the NVC vegetation classification during the data collection process, occasional species misidentification and data entry errors. The basic quality control steps for training data plots are described below. Quality control of floristics data and vegetation association calls are discussed in **Section 2.2** and **Section 2.3**, respectively.

Spatial characteristics

Because of the fine-scale heterogeneity associated with many vegetation types in the park, we aimed to precisely and accurately locate each training plot with respect to the 2015 NAIP, which was the finest-resolution predictor dataset available. Minimizing spatial error was particularly important for non-forest plots in small patches.

The vegetation patches represented by field plots varied in size and shape. In order to simplify data management and modeling, we converted all plots to a circle throughout which the assigned call was applicable, excepting permitted inclusions of 81 m^2 or less (nine 3x3-meter pixels). This allowed us to represent plots simply as a center point and radius. We used a script to identify the center point and radius of the largest circle that would fit within any delineated field polygons. Initial locations and radii were set for other plots based on the protocol's assessment dimensions. For all plots, we verified the spatial characteristics by comparing field notes, plot diagrams, and field photos with NAIP and coarser-resolution satellite data.

Plots were flagged for additional review if the GPS center point taken in the field was more than 20 meters from the center point of the field-drawn polygon. We prioritized positioning the circle in the section of the plot closest to the GPS point, assuming that the area nearby was the most thoroughly surveyed portion of the plot. If there were signs of inconsistency between the GPS point, the plot description, and the appearance of the surroundings in imagery, we prioritized the plot diagram and field photos (if provided), repositioning the circle on the plot center as determined by that information.

Disturbance review

Various disturbances impacted the park in the years between field data collection and acquisition of the imagery used for final map production. To prevent the use of training data for which the field-assigned vegetation type no longer corresponded to a plot's condition in imagery, we identified and excluded plots that were disturbed between their sampling date and the acquisition date of the most recent imagery source used in modeling, August 11, 2017.[21] Fires were the main source of disturbance. We digitized the perimeters and entered the dates of all documented fires in the park since 1990. For plots that lay within these perimeters, the sample collection date was compared to the disturbance date; if the disturbance occurred after sampling, recent satellite imagery was used to

[21] Any disturbance impacting the park since this date will not be reflected in the map. Furthermore, areas disturbed within the time window spanned by the predictive imagery sources used (i.e., between summer 2015 and this date) may not be represented correctly.

assess the extent of disturbance. If conditions no longer resembled those present at the time of sampling, the plots were not used in modeling.[22]

2.2. Floristics

2.2.1. Debugging species lists

A fair number of inconsistencies in species nomenclature occurred in the ocular data, due to the use of at least two taxonomic references (Hitchcock and Cronquist 1973, Pojar et al. 2004). For the purpose of our analysis, we standardized the records by selecting the most frequently used name in the field datasets. We used Burke Herbarium (2020) to identify synonymies and to determine a standardized name for taxa where there was no prevalent name in the field datasets. Henceforth we refer to these standardized names, used in the field by NPS crews, as *field names*; we provide a crosswalk to Hitchcock and Cronquist (2018) in INR (2021b).

We systematically addressed problems involving confused taxa that resulted from consistent misidentifications during particular field collection efforts and from easily scrambled species names and codes. In some cases, resolving these and other thorny issues required that we refer to plot photos, field notes, location information and data from surrounding plots. Occasionally we fell back on the judgment of experienced botanists that a particular taxonomic record was unlikely. We considered the experience level of the field crew involved on such plots in making our decisions. Other cases were easier to resolve, such as recognizing that a *Eucephalus ledophyllus* record at NOCA was probably really *E. engelmannii*, based on their established range boundaries.

2.2.2. Expanding species lists

Plots collected under Protocols Y and Z emphasized collection of full-ocular data. Because these were the main protocols used for mapping plots at OLYM and NOCA, the relationship between floristics and map units (mapping associations or map classes) was very well characterized at those parks. In contrast, MORA had very few full-ocular plots, due to the use of the minimal Protocol X and the lower overall sampling effort beyond the classification phase. This created two related difficulties. First, it reduced the degree to which MORA plots were represented in the relationships developed between floristics and map units, which threatened to make the resulting map classification less applicable at MORA and thus less representative of the NCCN as a whole. Second, it made the MORA association calls—even at plots with full ocular data—less reliable, because the distinct character of the vegetation there was not well-captured in the data.

In an attempt to address these concerns, we made additional efforts to improve the floristic completeness of many partial-ocular plots at MORA. For targeted plots in poorly sampled portions of the classification, a field botanist examined the plot photos and field notes, adding observed species and adjusting cover estimates. We tested the consequences of creating species lists entirely in this manner, comparing results to lists generated in the field at full ocular plots. We found that on average three-quarters of the species with at least 1% cover were found. Automated placement of these

[22] These plots were still used to develop floristic characteristics for the classification, despite no longer persisting in that condition.

office-created ocular records into an association[23] matched the full-ocular result at the association level at about 35% frequency and at the map class level at nearly 70% frequency. While field-collected full-ocular data are clearly preferable, these results represent a significant improvement over the level of field detail provided with most partial-ocular plots at MORA. We assigned all plots to association or map class manually, based not only on species cover estimates but on the cumulative weight of all available evidence.

In addition, we lowered the standards for labeling MORA plots as full ocular, to better represent the park in map unit floristics. In general, if we determined that a plot's ocular data likely represented all prominent species and contained the most significant species in each vegetation layer,[24] we considered it full-ocular and used it in establishing the floristic characteristics of mapping associations and map classes. Despite these efforts, MORA remained poorly represented compared to the other parks, though the later incorporation of the forest legacy plots helped considerably (see **Table 2**).

2.2.3. Taxonomic treatment for floristic analyses

Plot-level species lists were used in all phases of this project. They were the primary source of data we used to assign plots to associations, to rework the mapping associations and form the map classes, and to describe the associations and map classes for users. However, the level of floristic detail captured varied by observer, collection effort, time of year and weather conditions, especially for uncommon, cryptic and ephemeral taxa. Additionally, the number of observations of many less common species fell short of the sample sizes needed to generate reliable statistics. To address these issues, we aggregated rarer species into groups to increase statistical strength and took other steps to reduce variability in floristic detail across plots. We defined a set of analysis taxa in which common and readily identifiable species were treated at species level, while less common or troublesome (cryptic or otherwise difficult) taxa were treated at genus level or as intermediate sub-genus groups defined by lumping species with similar habitats. Some infrequently observed taxa were dropped entirely from analyses involving plot-level comparison.

Troublesome species found in the field were often identified at the genus level.[25] Because our floristic analysis presupposed that the same taxonomic units were used across all plots, leaving these records at genus level would have required lumping the genus and sacrificing the species-level data collected across all other plots. To avoid this, we worked to link genus-level calls to a more specific taxon, particularly for common genera associated with diverse habitats. We accomplished this by creating sub-genus species groups with similar overall morphology, habitat requirements,

---

[23] Automated assignments were made using the species cover–match tools discussed in **Section 2.2.4**.

[24] We made this determination with reference to our own field experience and by comparing the species lists to available field photos and to other data collected nearby.

[25] A few records were identified at family or higher taxonomic levels; these were excluded from analysis.

distributional data and community affinities.[26] We then assigned genus-level field records to the sub-genus groups based on the weight of evidence at plots (e.g., elevation, topographic position, species co-occurrence matrices). Other less common species and genera were lumped to either the sub-genus or genus level, in order to gain necessary sample sizes for analysis.

Species of *Carex* and *Salix*, which are key indicators of several vegetation types, required the most attention. In these widespread yet difficult genera, the group formation process focused primarily on morphology and habitat requirements. As an example, unknown dwarfed alpine willows were coded as *Salix nivalis*+, which was defined to include *S. nivalis* as well as *S. petrophila* and *S. cascadensis*. This entailed losing the distinction between the three alpine species, but we deemed that far preferable to combining all of them with unlike lower elevation species such as *S. commutata* and *S. sitchensis.* Other genera, in which species were less clearly sortable by morphology and life zone, were treated primarily based on species affinity data, using the full plot database to develop co-occurrence relationships and sub-generic groupings.[27] Finally, some species that were often confused by field crews were also lumped (e.g., *Juncus parryi* and *J. drummondii* were lumped as *J. parryi*+).

Across 49 genera, 73 distinct sub-genus taxa were created in this manner, with genus-level occurrences assigned downwards to them and species-level occurrences lumped upwards into them. 65 other genera were treated at the genus level, lumping species-level occurrences up. These taxa, in addition to the species treated at species level, are cumulatively referred to as *SCM taxa* (see **Section 2.2.4**). **Table 2** in INR (2021b) identifies the SCM taxon used for each field-identified taxon.[28] Prior to publication, plant nomenclature was updated to match Hitchcock and Cronquist (2018); the resulting name changes are documented in **Table 1** of INR (2021b).

2.2.4. Floristic analysis tools

An enormous quality control effort was needed to bring consistency to the association calls across the more than 6,100 field plots that were available for use as model training data. We developed several floristic analysis tools to allow us to objectively evaluate and prioritize the review of association labels. The tools were also used to help guide the development of mapping associations (**Section 2.3**). They are briefly described below.

---

[26] Treatment at this level required that we merge the species-level data collected at other plots into the same sub-genus categories, so determining appropriate categories was critical. Former NPS field botanists Matt Lee, Tynan Ramm-Granberg and Rachel Brunner were instrumental in this step.

[27] For example, we treated the genus *Arnica* as three taxa for analysis: a sub-genus group *A. latifolia*+, containing *A. latifolia*, *A. longifolia*, and a taxon identified in the field as *A. alpina*; another sub-genus group *A. mollis*+, containing *A. mollis*, *A. parryi* and a taxon identified as *A. amplexicaulis*; and a distinct species *A. cordifolia*. Genus-level records were assigned to one of the three based on species co-occurrence data.

[28] Henceforth in this report, both actual taxonomic species and the sub-genus groups defined here may be referred to simply as species, for simplicity.

Species cover match tool

Associations, and the map classes we developed from them, are defined by their floristics, their physiognomic structure, and their position along multiple environmental gradients. We developed a tool called *species cover match* (SCM) to provide a quantitative representation of the degree of fit of a plot to the floristic and (to a lesser extent) structural aspects of a class, and to flag plots that were outliers within the class to which they were assigned. Generally, these resulted from field crews having encountered vegetation communities that had not been treated in Crawford et al. (2009), from mixed species lists due to heterogeneous plots combining multiple vegetation patches, from artifacts relating to hard breaks in the keys, or from differing crew interpretations of how significantly to weight different components of the association descriptions.[29]

SCM used the R vegclust package (De Cáceres et al. 2010) to compute the multivariate floristic distance of each of the 4,100+ full ocular plots from the centroid of each class, as defined by the plot labels.30 In order to more closely align the analysis with the emphasis on vertical stratification in the NVC, we weighted the cover values of each SCM taxon by a lifeform-specific[31] multiplier, applied to the transformed and standardized cover values.[32] To approximate the NVC's structural emphasis, we calculated total cover for each lifeform and for all vascular vegetation,[33] and incorporated those in the analysis as if they were additional taxa.

---

[29] For example, the description for *Alnus rubra/Polystichum munitum* stated that "the herb layer is always dominated by *Polystichum munitum*," and also that it "occurs on upland slopes" and "is [a] result of succession after [disturbance]." On encountering a plot on an upland slope initiated by disturbance that lacked *Polystichum munitum* dominance but otherwise matched the description, some crews would emphasize the setting and decide it was a good enough match, while others would put more emphasis on the insufficient *Polystichum munitum* and choose another alternative. Since the key required 5% or more cover of *Polystichum munitum*, that might often have been used to resolve the question. Using the full species list to make these decisions results in many fewer such ambiguities.

[30] For this analysis, we transformed percent cover via a modified exponential equation (resulting in rapid changes of the transformed value in the indicative 2–10% cover range) to mimic breaks in the original association keys and allow the multivariate data-driven results to maintain as much compatibility with the keys as possible. We then standardized with respect to the mean and standard deviation of each species across all plots.

[31] We assigned all taxa to the following lifeform categories: broadleaf tree, conifer, tall shrub, standard shrub, dwarf shrub, forb, grass, sedge, rush, fern, fern ally, bryophyte and lichen (see INR 2021b).

[32] We used a multiplier of 2.0 for conifer and broadleaf tree species, and 1.5 for tall and standard shrubs. All other lifeforms had a multiplier of 1.0. In order to give more weight to taxa that were instrumental in defining the Crawford et al. (2009) associations, the multiplier for each SCM taxon was increased from its lifeform default proportionally to its maximum constancy in any mapping association, up to a maximum of 0.5 for taxa that were always present in an association.

[33] For purposes of lifeform totals, broadleaf trees and conifers were split into two vertical categories, GT5 (height over five meters) and LT5 (height less than five meters, or *regen*). Lifeform and total vascular cover were transformed using a sigmoidal curve to emphasize change in the region of 10% cover, in keeping with treatment of these thresholds in the original association keys.

Partial species cover match tool

Class labels on partial ocular plots could not be evaluated reliably using SCM and vegclust, because they didn't include true absence data (i.e., crews may have simply omitted a species). To evaluate these plots, we developed the *partial species cover match tool* (pSCM), which compared the SCM taxon cover estimates for partial-ocular plots to expectations derived from the class constancy and cover tables computed from full-ocular plots. The tool output a similarity metric between each partial ocular plot and each class and could be used in several different modes.

Three options were available to control the functioning of pSCM: *full mode* versus *partial mode*, *cover mode* versus *presence mode*, and *lifeform mode* versus *no-lifeform mode*. In full mode, pSCM penalized absences of taxa that were characteristically present in a vegetation class, while partial mode ignored these and so allowed for more missing information. In cover mode, cover estimates for a taxon that were significantly greater or less than the average cover for the class were penalized, while in presence mode only the presence or absence of a taxon was considered. In lifeform mode, lifeform totals were used in the similarity estimate, in addition to taxon cover estimates. Any combination of modes from the three options could be selected, allowing tailoring of the assessment to the amount of information available at a plot.

Differential indicators tool

Finally, we also developed a *differential indicators tool* (DIT) which we used to determine which of two classes was a better fit to a plot based only on the presence of the documented SCM taxa. For each present taxon, DIT calculated the ratio between its constancy in two selected classes. Each ratio was clamped at a maximum value of 10 before taking its square root. The transformed ratios were averaged across all present taxa and compared between the two classes, with the class giving the highest average ratio favored.

SCM, pSCM and DIT were all used to assist in determining the best calls at plots, depending on the sampling effort at the plot. SCM was primarily used during the earlier plot QC stages while we were still ironing out the mapping associations, while pSCM and DIT were used more in the later phases, especially at MORA where full-ocular plots were in short supply. The ability to label partial-ocular plots confidently was extremely helpful at increasing the available training data for modeling less common map classes at all parks.

2.3. Mapping associations and plot label QC

Although the following steps are written in sequential order, the processes occurred in tandem. The development of mapping associations and the quality control of plot association calls were strongly iterative processes. We have attempted to describe the steps with a minimal number of references to other parts of the process, but to some extent that has been unavoidable.

2.3.1. Mapping associations definition and floristics-based plot QC

Early drafts of the vegetation maps were based, with minor adjustments, on the vegetation alliances defined by NatureServe (2012), which in turn were based on associations defined by Crawford et al. (2009). Model error rates (see **Section 2.7.5**) and preliminary comparison of draft maps from MORA (in 2011) and OLYM (in 2013) to independent accuracy assessment data (see **Section 3**) indicated

that the maps were falling well short of accuracy goals. As discussed in **Section 1.2.1**, it became evident through working with the training plots that making field calls based on dichotomous keys had resulted in a noisy dataset that may not always have correctly responded to the intentions of Crawford et al. (2009). Another source of error may have been the ongoing evolution of the classification itself during the fieldwork. Regardless, in some portions of the classification, the associations—as defined by the groups of plots assigned to them—lacked the needed floristic cohesion to support repeatable field identification and accurate mapping.

By early 2015, when mapping plot data collected at NOCA were delivered to INR, more than 4,100 field plots with reasonably complete species composition data were available across the three parks for use in floristic calibration. This significantly exceeded the information that had been available for the development of the earlier classification. We used the cumulative dataset to enhance the classification for floristic consistency and mapping purposes, creating a set of *mapping associations*. Despite their differences, the NCCN parks share many dominant plants and plant communities. We took advantage of this commonality, so each park benefitted from plots collected across the network.

We began by reviewing full ocular plots with SCM. Plots that were significantly more similar to a different association than that to which they were assigned were examined to determine why their floristics differed from expectations. We checked field photos, plot descriptions, imagery and environmental setting, and changed the call to the association suggested by SCM if the balance of evidence supported that. For classes that were strongly defined by their vegetation structure (e.g., krummholz), we were more lenient in allowing floristic outliers to persist.

The process was applied iteratively: as plot QC continued, the analysis was occasionally updated, tightening the floristic groupings as the number of outliers was reduced in each cluster. In this manner, we refined the Crawford et al. (2009) associations while minimizing changes to their essential character. SCM was also used to suggest a best call at plots for which no confident call had been previously made.

We continued the revision process by eliminating problematic types from the mapping associations. Beginning with the original 311 upland and 50 wetland types, we removed (a) associations with fewer than two floristic calibration plots;[34] (b) associations distinguished from others based solely on total vegetative cover, either cumulative or in a single layer;[35] (c) associations named and defined based on the presence of a single common species (often a dwarf shrub such as *Vaccinium*

[34] These had often been included in Crawford et al. (2009) based on literature from areas adjacent to the NCCN parks. We retained one association with only one plot, *Populus tremuloides/Cornus nuttallii*, because of its distinctiveness and the clear range limitation that prevents it from being more widespread in the parks.

[35] Two examples in Crawford et al. (2009) are the associations labeled as "bryophyte and lithomorphic sparse vegetation," keyed under a break based on the total vascular cover, and the three depauperate understory forest associations, keyed on overstory species and low understory vascular plant cover.

deliciosum or *Juniperus communis*), regardless of the other vegetation present;[36] and (d) associations that were excessively heterogeneous in species composition (as represented in the floristic calibration plots), occurring in a variety of settings.[37]

We used SCM to reassign affected plots to the next most similar vegetated association, which was usually a very good fit. In addition to eliminating associations and merging their plots with similar types, we developed new mapping associations for groups of plots that were either poorly represented in Crawford et al. (2009) or had become badly tangled in the floristic calibration plots. These groups included dry shrublands, dry subalpine and alpine meadows, vegetation of talus slopes and avalanche chutes, riparian and wetland shrublands, and seral post-fire vegetation. We created the new associations by clustering all plots assigned to an association in each of the groups with the R **vegclust** package (De Cáceres et al. 2010).[38]

Forests with depauperate understories provide a good example of the sorts of changes we made to the classification. These plots—usually in seral stands, but occasionally in older forests on valley bottoms—were originally lumped into associations based solely on the tree canopy species present, but we found these often did not model and map well together.[39] DIT and pSCM were helpful in making the best use of the understory floristic data, even if plants only occurred in trace amounts. For mapping purposes, the identities of the species present were much more important than how much ground they cover. For instance, we found that a trace amount of *Orthilia secunda* was a consistent indicator of the most common mid-slope successional silver fir association. Silver fir plots with equally sparse understories that lacked *O. secunda* typically had moist site indicators instead, and had closer floristic and modeling similarities to lush silver fir associations found on lower slopes. The plots simply represented unusually sparse manifestations of those usually lusher types.

---

[36] Most plots assigned to these calls were small and represented a localized patch of the species in question. Generally these patches did not correspond to any meaningful landscape pattern, but simply reflected the stochastic dispersal and establishment processes of the single species, superimposed on a variety of background vegetation types.

[37] Typically, these associations—which were termed *catchalls* by field crews—resulted from key artifacts. They were recognized by their tendency to model with a variety of map classes, depending on the other vegetation present in addition to the species on which the key had focused.

[38] We log-transformed raw percent cover data for each SCM taxon and normalized across sites using the decostand function in the R **vegan** package (Oksanen et al. 2019) before using k-means clustering in **vegclust**. We experimented with the number of output clusters until the results captured a similar level of detail to that used elsewhere in the classification.

[39] Depauperate conditions occur in the stem-exclusion phase of a range of successional forest types, and can persist for over a century in the Pacific Northwest (Agee 1993, p.193). Thus, seral forests can be impossible to place definitively into classifications relying on understory species composition, and can be easily confused with very distinct valley bottom stands that are similarly depauperate (Franklin et al. 1988, p.126).

Our classification efforts resulted in a total of 228 mapping associations in the large NCCN parks. Nielsen and Brunner (2021) provide descriptions, including floristic and distribution details, as well as more information about the process of creating the associations from the original classification.

2.3.2. Mapping associations refinement and model-based plot QC

We also prioritized examination of individual plots using model results to identify plots that modeled better as an association different than their current assignment.[40] Plots that modeled poorly had often been noted as problematic by the field crew and were generally in heterogeneous areas, in very small patches, or had mismatched structure and floristics (frequently due to disturbance; e.g., a forest that had experienced a blowdown event and was now dominated by shrubs, but with understory species more typical of a forest). Other plots that modeled poorly had been mislocated due either to extreme GPS error or data entry errors; there was considerable feedback between association-level modeling and the spatial QC described in **Section 2.1.2**.

An occasional outcome of plot-level model-based QC was a decision that a plot should not be used in modeling because of a poor match to any association, an uncertain location, or both. These plots were often still useful in refining the classification's approach to disturbance or in identifying range extensions of associations known primarily from another park. Throughout the process, we incorporated these observations into refined descriptions of the structure, setting and range of each mapping association.

2.3.3. Final plot check with a hybrid assemblage labeling tool

In the above QC steps, we considered floristics and modeling similarities separately and only examined plots that failed to pass some test by a significant threshold. After development of the mapping associations had been completed, we used a final check—the *hybrid assemblage labeling tool* (HALT)—which considered the floristics and modeling analyses simultaneously to spot instances where both pointed in the same direction, but perhaps at a lower level of certainty. HALT enabled us to detect and reassign about 50 plots to an association that was a better overall fit.

The QC process, while lengthy, accomplished several critical steps toward development of the map classification and the associated map: (a) development of an association-level classification with high internal cohesion in both floristics and modeling tendencies; (b) development of clear descriptions of floristics, structure and setting for those associations; and (c) allowing maximum use of all plot data by improving the consistency of association calls on all plots, and particularly by assigning reliable calls to partial-ocular plots.

[40] We did this by creating random forests models (see **Section 2.7.5**) at the association level, examining the cumulative out-of-bag error associated with each plot, and noting the alternate associations with which it was most frequently confused.

2.4. Map classification

2.4.1. Development of vegetated map classes

Building crosswalk

The low accuracies of early draft maps indicated that changes to the alliance concepts were needed. A crosswalk to combine mapping associations into floristically cohesive and mappable entities provided the structure around which revisions were organized. We used the draft alliances from NatureServe (2012) and their relationship to the associations in Crawford et al. (2009) as a reference point during the revision process.

Our goal in this process—described in greater detail in Brunner et al. (2017)—was to minimize class confusion, both during field interpretation and in the map. Our approach was data-driven, using a quantitative proxy for each of these confusion types. As a proxy for field confusion, we used floristic similarity, since the more floristically similar two classes are, the less likely field observers will agree on the correct label for a plot. SCM, described in **Section 2.2.4**, provided an easy way of quantifying this at the plot level. To represent map confusion, random forests model confusion was clearly the appropriate proxy, as that was the means by which we planned to produce the map.[41] The main constraint we placed on the process was to follow the NVCS protocol of a many-to-one crosswalk between mapping associations and map classes, in which each association was a member of a single map class. In order to foster consistent map class definitions across NCCN parks, we aimed to use the same crosswalk for each of the mapping projects.

We began by identifying common and distinct mapping associations, emphasizing those that represented the cores of alliance concepts from NatureServe (2012). We used these as seeds for initializing map classes. If possible, we selected associations that were present at all NCCN parks in order to provide a common thread. If this was not possible, and we were confident about the relatedness of floristically dissimilar associations, we occasionally initialized map classes using a different association at each park. We did this in the case of vegetated balds, which are characterized by a common structure and setting but whose constituent species vary significantly with geography.

We then used an agglomerative process to grow the map classes from their seeds.[42] At each step, we computed the level of floristic and modeling similarity (termed *joint similarity* hereafter) between each unassigned association and each nascent map class, by aggregating plot-level data.[43] We found the association–map class pair with the greatest pairwise joint similarity and joined them by assigning the association to that map class in the crosswalk. Association–map class similarities were

---

[41] We quantified model confusion as the out-of-bag error rate for a plot in a model attempting to discriminate between a pair of associations, built from the plots assigned to either of them. The R **randomForest** package (Liaw and Wiener 2002) provides this information as an optional 'votes' table output. See **Sections 2.5–7** for background in the modeling process.

[42] The interactive process described here was implemented in a spreadsheet.

[43] We computed similarity by aggregating plot-level data at each step because some plot-level QC was ongoing during this process and this prevented our needing to run random forests again with every change.

recalculated after each assignment, and the next most similar pair found. The process resulted in maximizing within-class similarity and minimizing between-class similarity, allowing more confident discrimination in the field and more reliable mapping.

Early in the process, assignments were easy because many associations clearly belonged together based on both floristics and modeling. The decisions became more difficult later. When we encountered associations whose floristic and modeling tendencies pointed to different map classes, we emphasized the floristics, unless some overriding structural or setting-based criterion was available to assist in field identification. When different patterns of similarity were observed at different parks, we made our decision based on the park where the majority of association plots occurred. Occasionally we went back to the plot data to unravel problems.

Associations that fit poorly to existing map classes were added as new classes if they represented a distinguishable concept and had enough plots to support modeling. It then occasionally become apparent that other associations that had already been assigned had a stronger affinity for the new class. The iterative process continued until all associations had been assigned.

Refining crosswalk

After the crosswalk was formed, we recalculated association-wide model similarity to each full map class, again from plot-level data. We re-examined associations that were a better fit to a map class other than the class with which they had been lumped. We often found that this mismatch arose from plots that were floristically distinct from most others assigned to the association. These outliers usually were easily recoded on an individual basis to an alternate association, but in several cases we found associations that contained a full subset of plots that were similar to each other but distinct from the rest. We formed new associations with these plot subsets and moved them to a different map class. Nielsen and Brunner (2021) includes several examples of these new associations.

In many cases, map class occurrences were confined to only one[44] or two[45] of the NCCN parks, which presented no challenge to the crosswalk since the constituent associations were also absent. However, occasionally a map class was present in a park, but with too few plots from which to construct a model of its distribution. In these cases, we lumped the constituent associations with the

[44] Examples included H63–ALPINE BUCKWHEAT PUMICE VEGETATION (at MORA only), C02–REDCEDAR, LABRADOR-TEA, SLOUGH SEDGE AND SPHAGNUM BOG (OLYM only), and C22–SUBALPINE LARCH WOODLAND (at NOCA only).

[45] Examples included C03–SITKA SPRUCE, WESTERN HEMLOCK AND WOOD SORREL FOREST (absent at NOCA), H57–GREEN FESCUE DRY MEADOW (absent at OLYM), and H52–COW PARSNIP MEADOW (absent at MORA, at least at mappable patch size).

most similar map class that was mappable at that park.[46] These are the only cases where the crosswalk between association and map class differs between parks.

In general, the outcomes of the crosswalking process confirmed our belief that unless aberrant vegetation structure is present, modeling tendencies and full floristic character tend to track each other extremely well. By maximizing the floristic distinctions between the map classes, we simultaneously created a highly mappable classification. The description for each map class in Nielsen et al. (2021c) contains a list of its component associations.

2.4.2. Development of other map classes

Natural sparse and abiotic map classes

Classification and mapping efforts were primarily focused on vegetated communities, but sparsely vegetated and abiotic areas occupy a large proportion of each NCCN park. To fill these areas of the map, we developed map classes that were simple for field crews to discriminate but would provide useful habitat context. We developed a "rock-dominated" set of map classes distinguished by the geomorphological origin of the mineral substrate, including R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE, R72–COLLUVIAL BARREN and R73–BEDROCK BARREN. We also developed an "H_2O-dominated" set of classes composed of W81–FRESH WATER and W82–EXPOSED SNOW AND ICE.

Disturbed and cultural map classes

Several other classes were created to handle areas of uncertain vegetation impacted by recent fires or anthropogenic disturbance (the latter most often in the mapped buffer around the park). In this category were M92–BURNED WITH UNCERTAIN VEGETATION, M93–TIMBERLAND WITH UNCERTAIN VEGETATION, M94–DEVELOPMENT, M95–ROADS IN PARK and M96–CLEARED CORRIDORS IN PARK. Details on discriminating these and the preceding types are contained in the map class descriptions (Nielsen et al. 2021c).

[46] Examples included B33–UPLAND RED ALDER, BIGLEAF MAPLE AND CONIFER FOREST, treated as S45–VINE MAPLE SHRUBLAND at MORA, and H56–SUBALPINE SUMMER-DRY GRASS-FORB MEADOW, treated as H57–GREEN FESCUE DRY MEADOW at NOCA.

2.4.3. Resulting map classification

The map classes present in the NOCA map are summarized in **Table 3**. See Nielsen et al. (2021c) for detailed map class descriptions and an explanation of the map class name coding system.

Table 3. Map classes present in the NOCA map, the other NCCN park maps in which they appear, and the number of training plots called to each at NOCA.

Map class code and full name	Other parks	Plot count
C04–Moist western hemlock, Douglas-fir and foamflower forest	MORA, OLYM	50
C05–Western hemlock, Douglas-fir and sword fern forest	MORA, OLYM	135
C06–Western hemlock, Douglas-fir and salal forest	MORA, OLYM	37
C07–North Cascades dry Douglas-fir forest	–	70
C09–Ponderosa pine and Douglas-fir woodland	–	61
C10–Moist silver fir, western hemlock and foamflower forest	MORA, OLYM	60
C11–Mesic silver fir and western hemlock forest	MORA, OLYM	96
C12–Silver fir, hemlock and Alaska blueberry forest	MORA, OLYM	29
C13–Mountain hemlock, silver fir and Cascade azalea forest	MORA, OLYM	105
C14–Silver fir, big huckleberry and beargrass forest	MORA, OLYM	11
C15–Lodgepole pine and Douglas-fir woodland	MORA, OLYM	78
C16–North Cascades Douglas-fir and subalpine fir woodland	–	62
C20–Subalpine fir and Sitka valerian forest and woodland	MORA, OLYM	87
C21–Mountain hemlock, subalpine fir and heather woodland	MORA, OLYM	91
C22–Subalpine larch woodland	–	68
C25–North Cascades subalpine fir and whitebark pine woodland	–	57
C26–Conifer krummholz and treed cliff	MORA, OLYM	79
B30–Successional gravel bar shrubland	MORA, OLYM	11
B31–Broadleaf riparian and swamp forest	MORA, OLYM	52
B33–Upland red alder, bigleaf maple and conifer forest	OLYM	26
B34–Bigleaf maple and Douglas-fir debris apron forest	–	36
B35–Upland paper birch and conifer forest	–	25
S40W–Low elevation shrub-dominated wetland	MORA, OLYM	30
S41W–Subalpine willow wetland	MORA, OLYM	14
S42–Sitka willow riparian shrubland	OLYM	13
S43–Sitka alder shrubland	MORA, OLYM	90
S44–Thimbleberry shrubland, tall forbs and bracken fern	OLYM	18
S45–Vine maple shrubland	MORA, OLYM	74
S46–Snowbrush and Scouler's willow shrubland	–	67
S47–Successional huckleberry shrubland	MORA, OLYM	64
S48–Subalpine heather shrubland	MORA, OLYM	78
S49–Alpine heather shrubland	MORA, OLYM	72

Table 3 (continued). Map classes present in the NOCA map, the other NCCN park maps in which they appear, and the number of training plots called to each at NOCA.

Map class code and full name	Other parks	Plot count
H50W–Lowland marsh and meadow	MORA, OLYM	22
H51W–Subalpine herbaceous wetland	MORA, OLYM	26
H52–Cow parsnip meadow	OLYM	9
H53–Showy sedge and Sitka valerian meadow	MORA, OLYM	56
H54–Moist talus vegetation	OLYM	35
H57–Green fescue dry meadow	MORA	75
H58–Bedrock balds and sparsely vegetated forest openings	MORA, OLYM	64
H60W–Black alpine sedge wetland	MORA, OLYM	20
H62–Alpine sparse herbaceous vegetation	MORA, OLYM	38
R71–Alluvial barren and debris-covered ice	MORA, OLYM	66
R72–Colluvial barren	MORA, OLYM	247
R73–Bedrock barren	MORA, OLYM	126
W81–Fresh water	MORA, OLYM	139
W82–Exposed snow and ice	MORA, OLYM	33
M92–Burned with uncertain vegetation	MORA, OLYM	–
M93–Timberland with uncertain vegetation	MORA, OLYM	–
M94–Development	MORA, OLYM	–
M95–Roads in park	MORA, OLYM	–
M96–Cleared corridors	–	–

Development of the NVC for the Pacific Northwest continued on a somewhat parallel track to ours, as we worked on finalizing the mapping associations and map classes presented here. We compared the relationship of our mapping associations and map classes with the hierarchical placement of the related associations in the most recent NVC update, USNVC (2019). At the NVCS group level, there is good correspondence, with our map classes mostly composed of associations that are members of a single group, or of an amalgam of associations from groups that are poorly represented in the project area and do not overlap with other map classes. There is less congruence at the alliance level, with one cause being that our map classes are generally less beholden to dominance and encompass a broader range of indicator species. Forests are somewhat more finely delineated in map classes than the current NVC alliances, but non-forests are a bit more coarsely lumped. Structural characteristics appear to be more important in distinguishing forested map classes than the corresponding alliances, but less important in distinguishing dwarf shrubland and herbaceous map classes.

2.4.4. Descriptions

Summary and setting narratives

The map class summary and setting paragraphs in Nielsen et al. (2021c) were compiled from plot-level floristics, vegetation structure data, summarized environmental variables and expert knowledge.

We edited the narratives to reflect park-specific characteristics and added observations based on the final map and plot data. Representative plot photos were selected for each class; these were generally obtained from the park in which it was most common.

Floristics tables

We generated constancy and cover information for each of the resulting map classes, based on the complete set of full ocular plots across the three parks. Because of the large number of records that were uncertain at the species level, we used the SCM taxa described above (and documented in INR 2021b) instead of species as the taxonomic units for the analysis. We used the tables generated to assign descriptive names to the map classes. Each map class description in Nielsen et al. (2021c) contains a condensed version of the constancy and cover results.

Indicator species analysis

Although we relied on the tools described above (SCM, pSCM and DIT) to reliably discriminate between map classes on the basis of plot floristics, those tools will not be available to users in the field, unless they are carrying a mobile device. Instead, we created a park-specific list of indicator species that are helpful for distinguishing each pair of map classes, which we have included in Nielsen et al. (2021c) for pairs that are likely to be occasionally confused.

We derived indicators from the constancy and cover data. *Presence indicators* are SCM taxa that are significantly more likely to be present in one of the map classes than in the other, based on the constancy tables. We rated the strength of presence indicators by the constancy ratio between the two classes and put those ratings on a comparable scale for both sides of each map class pair. *Cover indicators* are taxa that are likely to occur in significantly greater abundance in one of the classes than in the other, based on the cover tables. We prioritized listing taxa that occur reasonably often in the favored map class, but in some cases only less common taxa are good indicators. For this reason, we listed a significant number of indicators. Lack of presence of an indicator is not evidence against a map class; however, absence of taxa listed as occurring at high frequency in the floristics table for a map class can be construed that way.

2.4.5. Key

We have mentioned several times the difficulties we encountered with field plots that had been assigned to association based on a dichotomous key (see **Section 1.2.1** and elsewhere). However, although "the key is not the classification" (Crawford et al. 2009), it is where a typical user will start. In our map class key (Brunner et al. 2021), we aimed to provide as much help to a field user as possible without leading them astray by oversimplification. We strongly urge users who have keyed to a map class to carefully consult the map class description, including the indicators for closely-related alternate classes.

Much of the key was built using automated methods such as multivariate hierarchical clustering via **hclust** (R Core Team 2018) and classification trees via **rpart** (Therneau et al. 2015). Because setting and structural characteristics are easiest for a non-botanist to identify, we prioritized them in the key where possible, mostly at higher levels. At the lower levels (e.g., within conifer forests), the breaks were mostly determined by floristics. We transformed plot species composition information into

binary true/false characteristics based on presence, prominence, dominance of individual species and functional groups (e.g., broadleaf trees, total vascular cover), relative abundance (e.g., cover of *Acer circinatum* significantly greater than that of *Alnus viridis*), and quantifiable setting variables (e.g., south-facing). We used **rpart** to determine the optimal structure and best key break variables based on pools of samples drawn from the full-ocular plots, assuming these decisions would translate to new plots encountered in the field.

To help keep users from taking a wrong turn based on a single criterion, we added additional floristic and setting-based characteristics at most breaks to lend additional confidence. The additional criteria were pulled from surrogate variables in the classification tree, from a break-specific indicator species analysis using the R **indicspecies** package (De Cáceres and Legendre 2009), and from setting and structure notes. After each break was written, it was applied to the current plot pool and the resulting subdivided pool was fed into the next break. We minimized misclassification by only including criteria that correctly classified 95% or more of the plot pool entering the break; we tried to find a way to shepherd the misclassified plots home later in the key. The key was validated in the office with over 200 field plots per park and was also briefly tested in the field.

2.5. Independent data selection and pre-processing

We used an implementation of the random forests machine learning algorithm (Breiman 2001) to predict map class from field training data (discussed in **Sections 2.1–2.4**) and wall-to-wall independent predictor data. We used several broad categories of predictor data: (a) four separate years of aerial imagery from the National Agricultural Imagery Program (NAIP); (b) satellite imagery from the Landsat-8 and Sentinel-2 satellites, collected during multiple distinct seasons; (c) topographic and hydrologic metrics developed from standard digital elevation models; and (d) climate normals over the period 1981–2010. Several other types of potential predictor data—soils, surface geology, and geological landform information, infrastructure development locations, and maps of fire history—were considered for use and ultimately rejected. These layers had poor spatial registration or were incomplete or inconsistent over the project area. We felt their use would result in mapping artifacts and add little predictive power, since correlated information was available already in the other predictors. **Table 4** summarizes the data sources with their spatial resolutions and the dates to which they apply. The selection, acquisition and pre-processing of the data are described in the sections that follow.

Table 4. Sources of predictive modeling layers.

Data type	Spatial res (m)	Data source	Applicable timeframe(s)
4-band color-infrared aerial imagery	1	National Agriculture Imagery Program, State of Washington	2009, 2011, 2013, 2015
Historic mid-summer image, Landsat-5	30	USGS (2019a)	Aug 23, 1985
Current mid-summer image, Sentinel-2	10, 20	USGS (2019a)	Aug 11, 2017
Current minimum-snow image, Landsat-8	30	USGS (2019a)	Sep 11, 2015
Elevation	10	USGS (2019b)	–
Climate normals	~800	PRISM Climate Group (2019)	1981–2010

All data required pre-processing to ensure consistent spatial registration and reduce sources of noise. The process of selecting and obtaining the various datasets and the end products of the pre-processing are described here.

2.5.1. Aerial imagery

We acquired 4-band color infrared NAIP imagery as uncompressed quarter quads from four separate collections, in 2009, 2011, 2013, and 2015. The 2015 imagery was the main data source allowing mapping at 3-meter resolution. However, deep shadows which lowered mapping accuracy often occurred north of steep slopes (**Figure 6**). Because the shadow locations varied between image collections, we mitigated this problem by combining the best-illuminated portions of each into a mosaic.

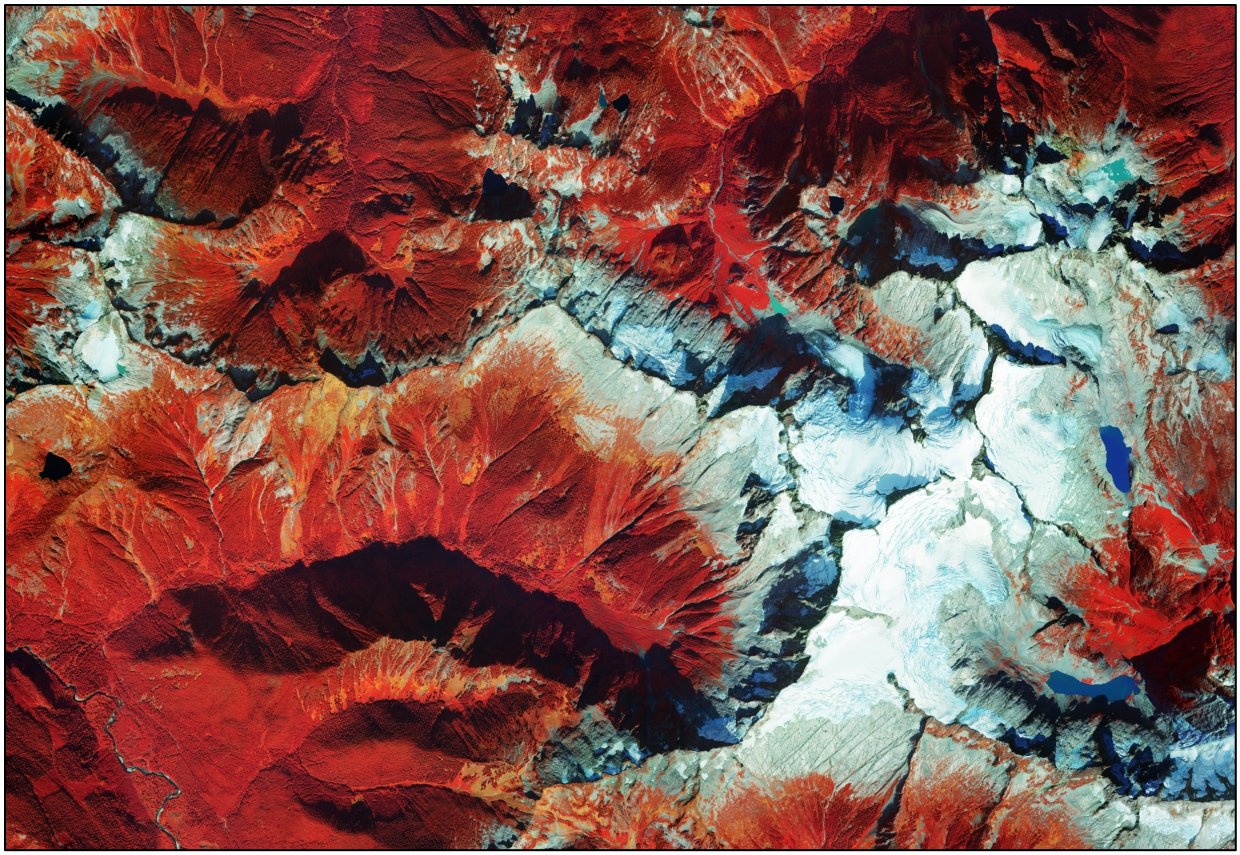


Figure 6. Aerial imagery with untreated shadows. The deep shadows seen in this 2015 NAIP imagery would interfere with accurate mapping unless treated.

Correction of 2015 imagery

We mosaicked the uncompressed quarter quads from each NAIP collection and generated aerial imagery metrics (**Section 2.6.1**). Making use of topographic information (**Section 2.6.3**), we then built a predictive model to identify shadows in the 2015 imagery by digitizing shadow and non-shadow training data, identifying shadows using a random forests model, and iteratively selecting additional training data to home in on problem areas. When satisfied with the results, we converted the shadow mask to a shapefile and buffered each feature by a variable distance, using a formula that yielded a buffer area roughly proportional to the size of the feature. Our hypothesis was that over a given region, the histogram of pixel values for each image band within corrected shadows should resemble that within the adjacent unshaded areas. We broke the project area into overlapping tiles, derived a crosswalk between shadow pixel values and corrected values based on matching the shadow and buffer histograms, and applied this to all shadow pixels. **Figure 7** illustrates a resulting corrected image.

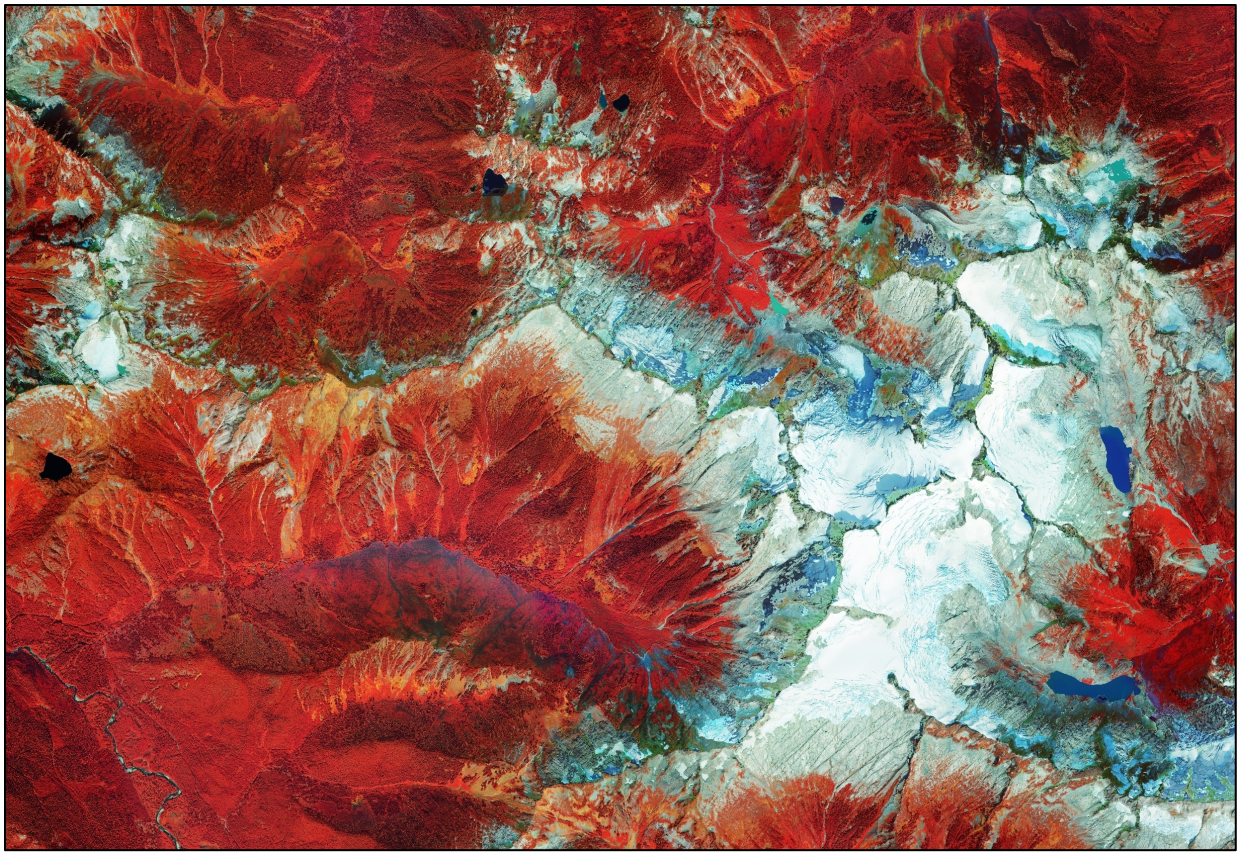


Figure 7. Aerial imagery with histogram-matched shadows. Here the shadowed areas have been matched to the surroundings.

Image merging and correction

Although shadow pixel values in the corrected 2015 image showed reasonable correspondence to the underlying land cover, lack of direct illumination resulted in a major reduction in local variance which could not be corrected. Because of the importance of high-resolution texture in accurate identification of land cover types (see **Section 2.6.1**), we incorporated an additional method of shadow treatment. We applied the model generated from 2015 imagery to the other imagery years, yielding shadow/non-shadow masks for each year. These masks were used to produce a merged image by selecting the first non-shadowed year from the sequence (2015, 2013, 2009, 2011) subject to the condition that if a given pixel was located within the digitized fire perimeters (**Section 2.1.2**), only imagery collected after the fire year could be selected. The year 2011 had lowest priority in the merge sequence because high snowpack that year obscured the ground and delayed vegetation development at high elevations through much of the summer. The merged image replaced many of the shadows in the 2015 image with illuminated data from other years. Although the spectral characteristics differed somewhat from year to year due to the lack of radiometric normalization in NAIP data, we felt that for modeling purposes the result was far superior to leaving the shadows untreated. The areas that were shadowed in all imagery years were corrected using the procedure applied to the 2015 imagery above. The resulting image, suitable for generating texture metrics (**Section 2.6.1**), is shown in **Figure 8**.

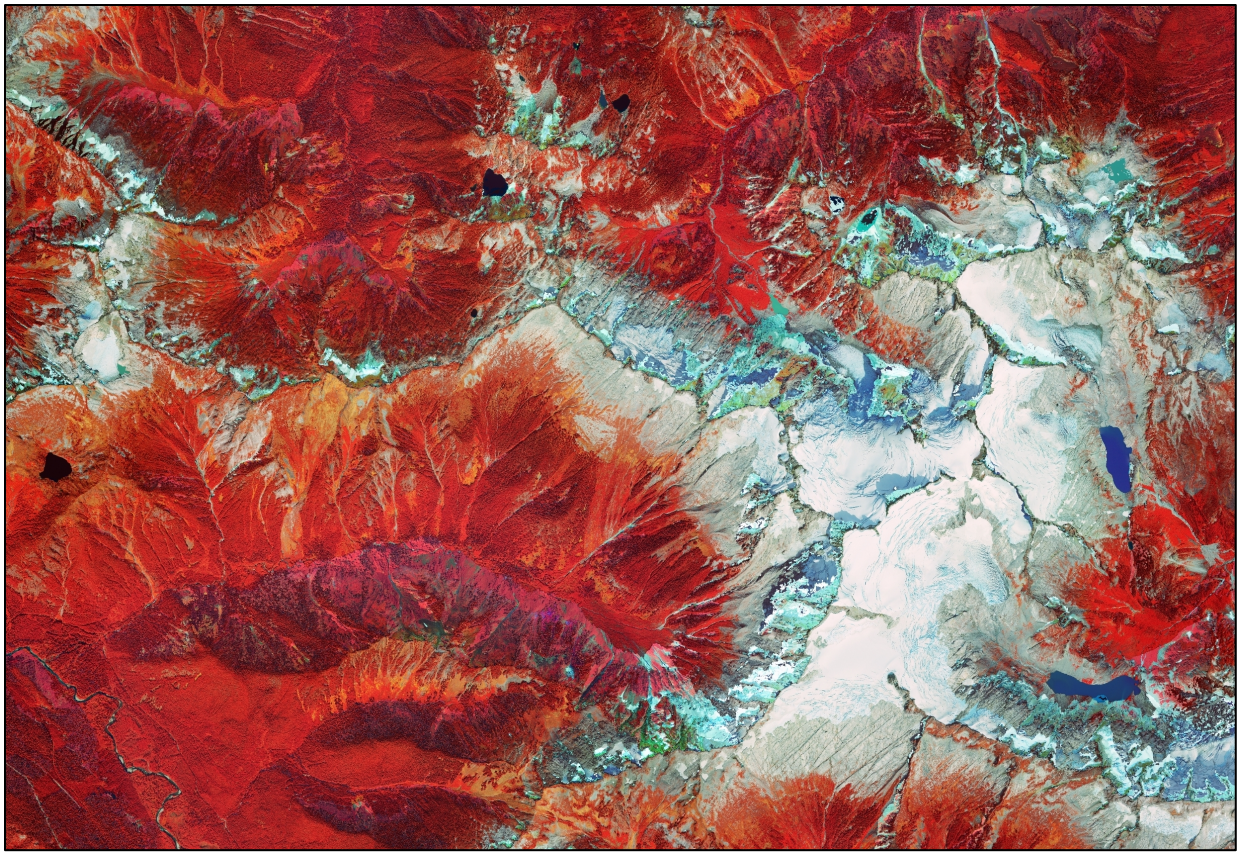


Figure 8. Aerial imagery merged across years. Here the shadowed areas have been filled with data from the 2009–13 images. Areas shadowed in all images were matched to the surroundings.

2.5.2. Satellite imagery

We searched the image archive at GLOVIS (USGS 2019a) for cloud-free Landsat-5, Landsat-8, and Sentinel-2 images collected between early June and late September in all years since 1982. Images from outside that seasonal window were mostly snow-covered or had very low sun angles and were not useful for vegetation mapping. In the map training fieldwork phase, a Landsat-5 image collected on July 16, 2006 was used to guide sampling and produce field map sheets. A midsummer Sentinel-2 image collected on August 11, 2017 (**Figure 9**) was used as the primary satellite image for modeling. A Landsat-8 image collected on September 11, 2015 was also used in modeling, since it had the minimum snow cover of any available images, allowing more effective mapping of higher elevation areas. Finally, a Landsat-5 image collected on August 23, 1985 served as the starting point for historic change detection over the intervening time period.

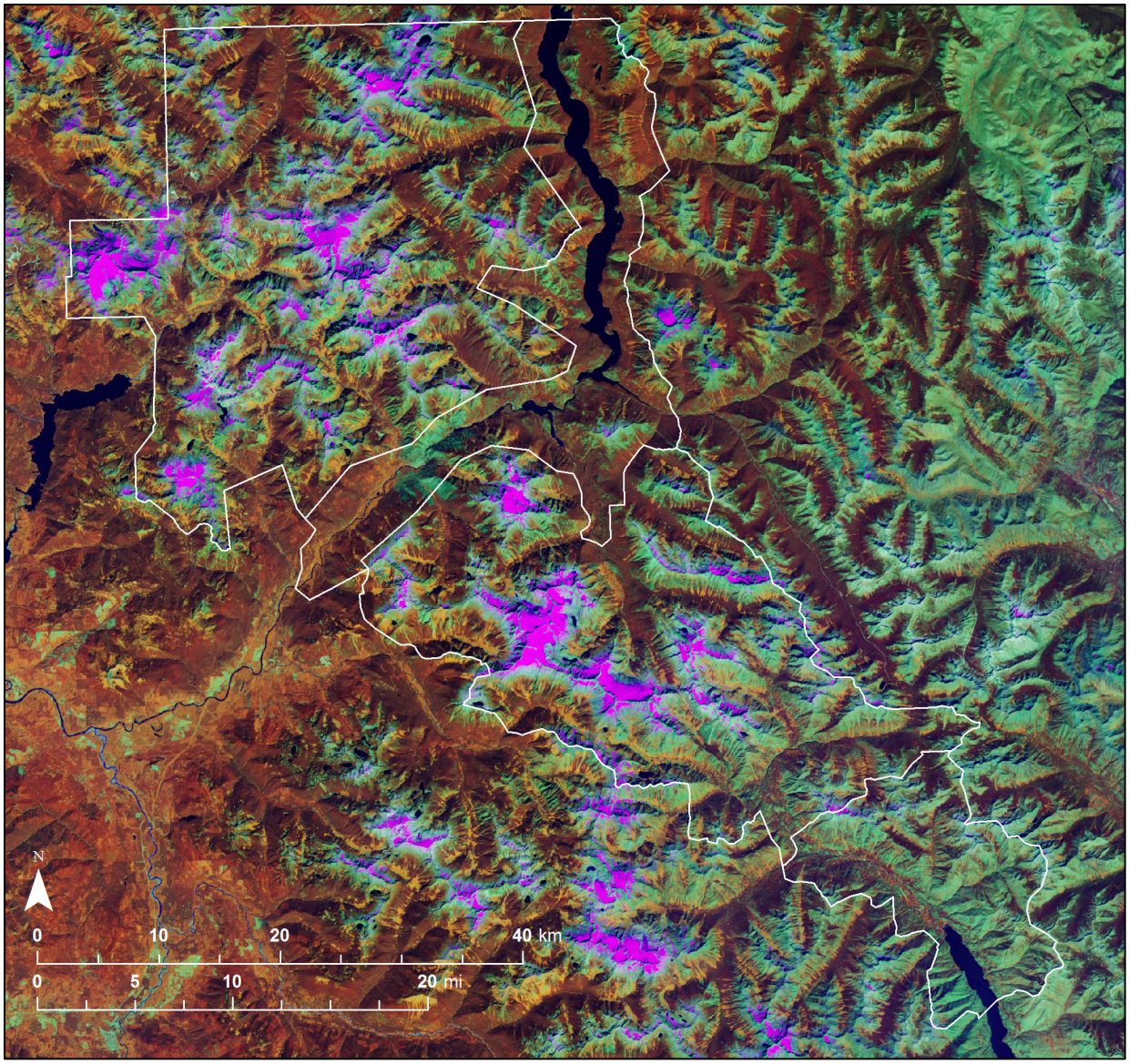


Figure 9. Base Sentinel-2 satellite image, collected August 11, 2017. The intensity of shades of red represents near-infrared reflectance from the land surface, green intensity represents mid-infrared reflectance, and blue intensity represents red reflectance.

All satellite images were converted to at-sensor reflectance (e.g., Chander et al. 2009), and a simple dark object atmospheric correction (Chavez 1988) was applied to approximate surface reflectance. We developed a novel process for spatial coregistration of the satellite images with the elevation dataset. We began by coregistering the minimum-snow image—which showed the greatest illumination contrast due to its acquisition at a time of relatively low sun elevation angle—to the elevation data. A cosine(i) image of illumination intensity at the time of image acquisition was created based on local slope and aspect; it served as a reference for aligning the satellite near-infrared band using the ERDAS Imagine Autosync tool. We then coregistered the midsummer image to the

minimum-snow image using their respective near-infrared bands. The resulting coregistered images were resampled via cubic convolution to a common extent and pixel size.

The satellite images were then topographically normalized to reduce the effect of variable illumination on at-sensor reflectance. We did this via a modified version of the stratified c-correction method (Twele et al. 2006), using the normalized difference moisture index (NDMI; Wilson and Sader 2002) for stratification of pixels into distinct correction groups. The normalization process reduced the effects of shading, causing individual land cover types to exhibit more consistent reflectance across the image, regardless of slope and aspect (**Figure 10**).

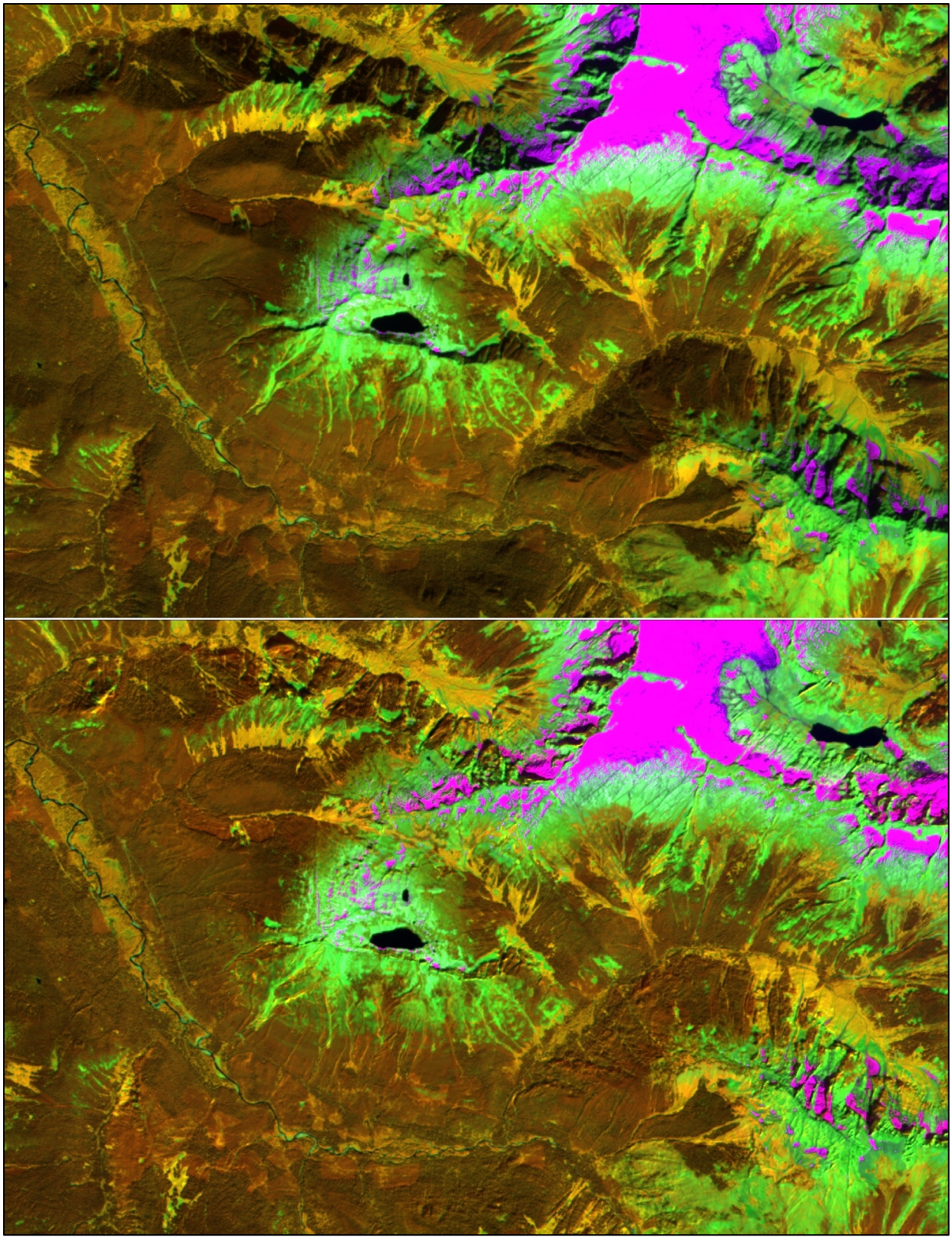


Figure 10. Sentinel-2 image before and after topographic normalization (upper and lower images respectively). Snow and ice appear pink, sparsely vegetated areas green, broadleaved trees and shrubs yellow, and conifers reddish-brown.

2.5.3. Elevation and climate data

We downloaded 10-meter resolution elevation data for the project area from the 3D Elevation Program (3DEP; USGS 2019b). We also downloaded a range of 30-year monthly climate normals at approximately 800-meter resolution from the PRISM Climate Group (2019), including January, April (see **Figure 11**), July and October precipitation, minimum and maximum temperature, mean dew point temperature, and maximum vapor pressure deficit. For processing efficiency, the elevation data were converted to integer format using a vertical unit of 0.25 feet. The climate data were clipped to the project area, reprojected and resampled to 30-meter resolution using bilinear interpolation.

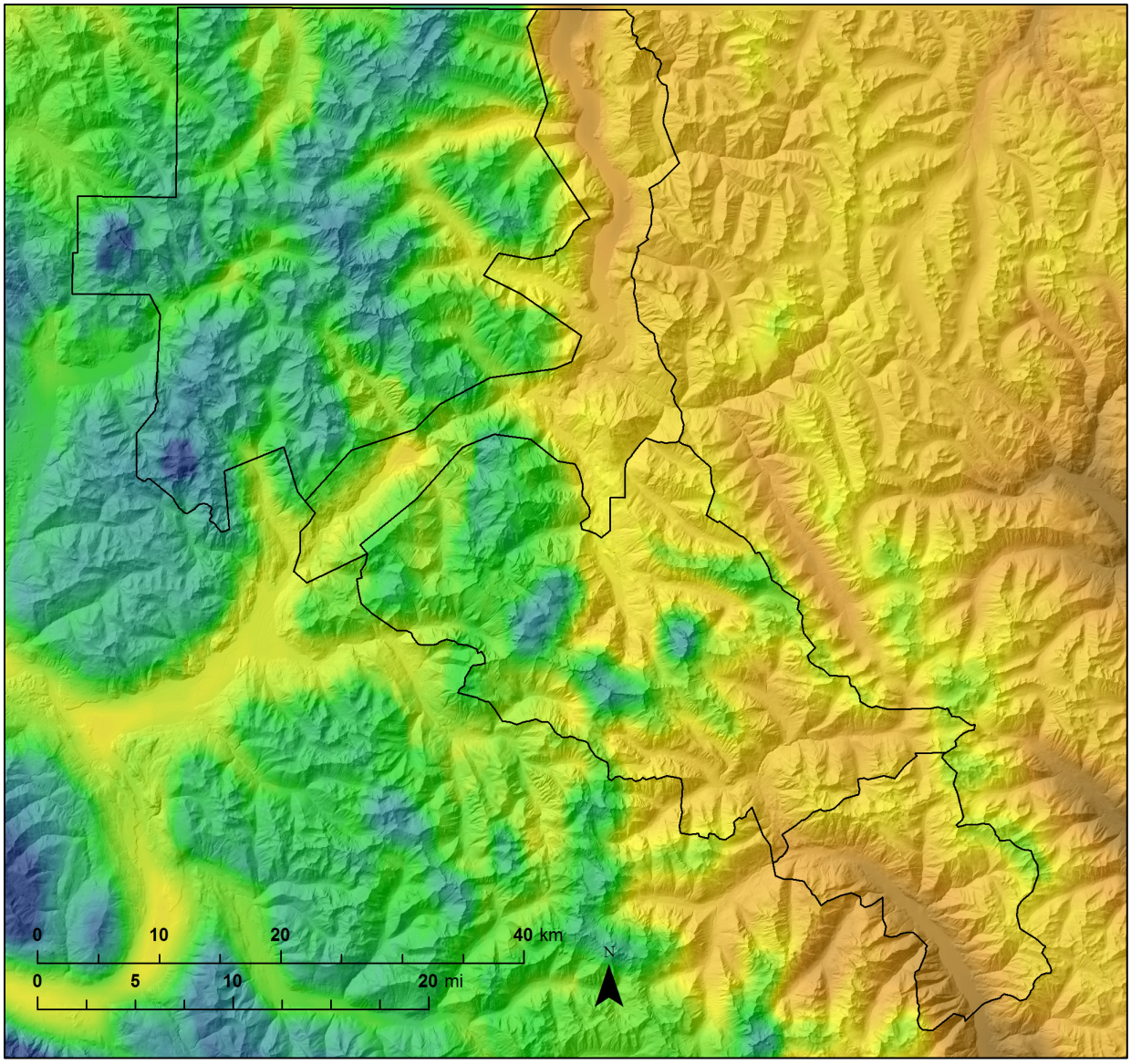


Figure 11. Average April precipitation from PRISM data, illustrating the strong gradient across the complex. Blue represents the wettest areas (about 30 cm of rain-equivalent precipitation) and burnt orange the driest (about 3 cm).

2.6. Predictive metrics

We used spatial contextual information, variable transformations, and noise minimization techniques to produce predictive metrics with stronger relationships to vegetation patterns than the raw independent data. The metrics fall into five main categories: metrics derived from aerial imagery, satellite imagery, topographic information, hydrologic information, and climate data. Each metric category is followed by a table detailing the predictive metrics produced from that data source. For each metric, we give an *effective resolution*. This combines characteristics of the data source as well as algorithmic factors to estimate the square dimensions surrounding any point over which land cover will influence the metric. It is used later in the predictor selection process to simultaneously optimize model error rate and effective spatial resolution (see **Section 2.7.4**).

2.6.1. Aerial imagery metrics

A variety of metrics representing spectral response and spatial patterning were calculated from the aerial imagery (**Table 5**). Two main types of metrics were produced. Reflectance metrics, produced from the shadow-corrected 2015 image, are based on responses in different spectral bands from a single imaged pixel. Texture metrics, produced from the shadow-corrected multi-year merged image, are based on local variability in spectral responses, measured across a moving window incorporating numerous pixels. The processing is described in greater detail below.

Table 5. Aerial imagery-based predictive metrics, the effective spatial resolution at which they respond, and a brief description or reference to a methodology.

Metric name(s)	Effective res (m)	Description
r1_md, r1_mx	3	Median and maximum red band value over source pixels
g1_md, g1_mx	3	Median and maximum green band value over source pixels
n1_md, n1_mx	3	Median and maximum near-IR band value over source pixels
u1_md, u1_mx	3	Median and maximum near-IR:green band contrast over source pixels
v1_md, v1_mx	3	Median and maximum near-IR:red band contrast (NDVI, Rouse et al. 1974, Tucker and Sellers 1986) over source pixels
w1_md, w1_mx	3	Median and maximum red:green band contrast over source pixels
x1_md, x1_mx	3	Median and maximum green:blue band contrast over source pixels
y1_md, y1_mx	3	Median and maximum red:blue band contrast over source pixels
r1a, r1b, r1c	3	Texture metric via filter ‘a’, ‘b’, ‘c’ at 1m resolution on red band
r2a, r2b, r2c	6	Texture metric via filter ‘a’, ‘b’, ‘c’ at 2m resolution on red band
r3a, r3b, r3c	9	Texture metric via filter ‘a’, ‘b’, ‘c’ at 3m resolution on red band
r4a, r4b, r4c	12	Texture metric via filter ‘a’, ‘b’, ‘c’ at 4m resolution on red band
r6a, r6b, r6c	18	Texture metric via filter ‘a’, ‘b’, ‘c’ at 6m resolution on red band
r9a, r9b, r9c	27	Texture metric via filter ‘a’, ‘b’, ‘c’ at 9m resolution on red band
rca, rcb, rcc	36	Texture metric via filter ‘a’, ‘b’, ‘c’ at 12m resolution on red band
rda, rdb, rdc	54	Texture metric via filter ‘a’, ‘b’, ‘c’ at 18m resolution on red band
rea, reb, rec	81	Texture metric via filter ‘a’, ‘b’, ‘c’ at 27m resolution on red band

Table 5 (continued). Aerial imagery-based predictive metrics, the effective spatial resolution at which they respond, and a brief description or reference to a methodology.

Metric name(s)	Effective res (m)	Description
rfa, rfb, rfc	108	Texture metric via filter 'a', 'b', 'c' at 36m resolution on red band
gRF, nRF, uRF, vRF, wRF	3–108	All the above combinations of resolution (R) and convolution filter (F) applied to green band, near-infrared band, near-IR:green contrast, near-IR:red contrast, and green:red contrast
ra_13, rb_13, rc_13	3	NDTI of *r1a* contrasted with *r3a*, *r1b* with *r3b*, and *r1c* with *r3c*
ra_26, rb_26, rc_26	6	NDTI of *r2a* contrasted with *r6a*, *r2b* with *r6b*, and *r2c* with *r6c*
ra_39, rb_39, rc_39	9	NDTI of *r3a* contrasted with *r9a*, *r3b* with *r9b*, and *r3c* with *r9c*
ra_4c, rb_4c, rc_4c	12	NDTI of *r4a* contrasted with *rca*, *r4b* with *rcb*, and *r4c* with *rcc*
ra_6d, rb_6d, rc_6d	18	NDTI of *r6a* contrasted with *rda*, *r6b* with *rdb*, and *r6c* with *rdc*
ra_9e, rb_9e, rc_9e	27	NDTI of *r9a* contrasted with *rea*, *r9b* with *reb*, and *r9c* with *rec*
ra_cf, rb_cf, rc_cf	36	NDTI of *rca* contrasted with *rfa*, *rcb* with *rfb*, and *rcc* with *rfc*
gF_RS, nF_RS, uF_RS, vF_RS, wF_RS	3–36	All the above combinations of convolution filter (F) and two resolutions (R,S) applied to green band, near-infrared band, near-IR:green contrast, near-IR:red contrast, and green:red contrast
d1c, d2c, d3c, d4c, d6c, d9c, dcc, ddc, dec, dfc	3–108	Cross-band contrast between *v1c & r1c, v2c & r2c, v3c & r3c, v4c & r4c, v6c & r6c, v9c & r9c, vcc & rcc, vdc & rdc, vec & rec, vfc & rfc*
e1c, e2c, e3c, e4c, e6c, e9c, ecc, edc, eec, efc	3–108	Cross-band contrast between *n1c & r1c, n2c & r2c, n3c & r3c, n4c & r4c, n6c & r6c, n9c & r9c, ncc & rcc, ndc & rdc, nec & rec, nfc & rfc*

Spectral metrics

Response metrics were produced for the red, green and near-infrared bands of the 1-meter resolution imagery. In addition, several vegetation indices were calculated from the raw band values: the normalized difference vegetation index (NDVI; Rouse et al. 1974, Tucker and Sellers 1986) and parallel contrast metrics between the near-infrared and green bands, and between the green and red bands. Each metric was summarized to the 3-meter mapping resolution by taking the median and the maximum 1-meter value within each 3-meter modeling pixel.

Nested texture metrics

Most information in high-resolution imagery is contextual and expressed in the spatial patterning of pixel neighborhoods; the eye's ability to identify many features based solely on the patterning and arrangement of gray-scale brightness values illustrates this point. We devised a method called *nested texture metrics* (NTM) to extract this information and provide it as predictor data to the modeling process. The texture metrics represent local variability at a range of pixel resolutions corresponding to distinct spatial scales at which various vegetation and landscape features occur.

Each of the spectral metrics described above was first median-aggregated[47] to a variety of coarser resolutions (2, 3, 4, 6, 9, 12, 18, 27, and 36 meters). We then used three different 3x3-cell[48] convolution filters to extract different aspects of patterning from each of the aggregated datasets as well as the original 1-meter dataset: (a) standard deviation of the center cell and the eight nearest neighbor (8NN) cells; (b) a 'speckle' filter, the absolute value difference between the center cell and the median of the 8NN, divided by the median of the 8NN and then smoothed by an additional 3x3-cell median filter; and (c) a non-trending variance filter accomplished via an alternating-cell convolution kernel $\begin{bmatrix} +4 & -5 & +4 \\ -5 & +4 & -5 \\ +4 & -5 & +4 \end{bmatrix}$. The results were converted to 3-meter resolution by a combination of median aggregation and cubic convolution resampling designed to maintain high-resolution detail.

Normalized difference texture index

We developed the normalized difference texture index (NDTI) to minimize the impact of variability in view angle and illumination characteristics between flight lines. The index works on the principle that because these artifacts affect textures similarly across a range of pixel resolutions, they can be partially canceled out by contrasting textures computed at two different pixel resolutions. Texture differences remaining after this cancellation result from image patterns at spatial scales intermediate between the two resolutions. NDTI metrics were produced by contrasting metrics computed at the following pairs of resolutions: 1m/3m, 2m/6m, 3m/9m, 4m/12m, 6m/18m, 9m/27m, and 12m/36m.

$$NDTI_{ab} = \frac{(\sigma_a - \sigma_b)}{(\sigma_a + \sigma_b)}$$

where a and b represent the two source texture resolutions and σ represents the source texture metric computed at the given resolution.

Cross-band contrast metrics

We produced another set of metrics to contrast corresponding metrics computed on NDVI against the red band, and on the near-infrared band against the red band. A formula like that used for NDTI was used, based only on the results from the 'c' convolution filter.

2.6.2. Satellite imagery metrics

We calculated a variety of image transformations from the satellite imagery (**Table 6**). The metrics differed somewhat depending on whether the source imagery was obtained by Sentinel-2, Landsat-8 or Landsat-5. All applicable metrics were produced for the current midsummer and minimum-snow

---

[47] GIS data is typically aggregated to a coarser resolution by taking the mean value of the finer resolution input pixels across each of the output pixels. Summarizing by the median value instead reduces smoothing near land cover transitions and increases the isolation of scale-dependent texture signals.

[48] The term *cell* is generally synonymous with *pixel*, but we mean it in a more abstract sense—generally in the context of a data-processing algorithm—than we do *pixel*, which is usually associated with the local contribution to some larger "picture."

images and for the historic midsummer image, but only those from current images were used in modeling.

Table 6. Satellite imagery-based predictive metrics, the effective spatial resolution at which they respond, and a brief description or methodology reference.[A] Where two resolutions are shown, the first is for Sentinel imagery, the second for Landsat.

Metric name(s)	Effective res (m)	Description
grn	10, 30	Green reflectance: 543–577 nm (Sentinel-2), 530–590 nm (Landsat-8) or 520–600 nm (Landsat-5)
red	10, 30	Red reflectance: 650–680 nm (Sentinel-2), 640–670 nm (Landsat-8) or 630–690 nm (Landsat-5)
re1	20	Red edge reflectance: 698–712 nm (Sentinel-2 only)
re2	20	Red edge reflectance: 733–747 nm (Sentinel-2 only)
re3	20	Red edge reflectance: 773–793 nm (Sentinel-2 only)
nir	10, 30	Near-infrared reflectance: 785–899 nm (Sentinel-2), 850–880 nm (Landsat-8) or 760–900 nm (Landsat-5)
sw1	20, 30	Shortwave reflectance: 1565–1655 nm (Sentinel-2), 1570–1650 nm (Landsat-8) or 1550–1750 nm (Landsat-5)
sw2	20, 30	Shortwave reflectance: 2100–2280 nm (Sentinel-2), 2110–2290 nm (Landsat-8) or 2080–2350 nm (Landsat-5)
temp	100	Thermal band response: 10.60–11.19 µm (Landsat-8 only)
ndvi, ndvip	10, 30	Normalized difference vegetation index (Tucker and Sellers 1986)
ndmi, ndmip	20, 30	Normalized difference moisture index (Wilson and Sader 2002)
ndfi, ndfip	20, 30	Normalized difference forest index = *ndvi* + *ndmi*
nbr, nbrp	20, 30	Normalized burn ratio (Key and Benson) 2002
ndsi, ndsip	20, 30	Normalized difference snow index (Hall et al. 1995)
ndgr, ndgrp	10, 30	Normalized contrast between *grn* and *red*
ndng, ndngp	10, 30	Normalized contrast between *nir* and *grn*
ndsw, ndswp	20, 30	Normalized contrast between *sw1* and *sw2*
tcb	20, 30	Tasseled cap brightness (Kauth and Thomas 1986, Huang et al. 2002)
tcg	20, 30	Tasseled cap greenness (Kauth and Thomas 1986, Huang et al. 2002)
tcw	20, 30	Tasseled cap wetness (Kauth and Thomas 1986, Huang et al. 2002)
di	20, 30	Disturbance index (Healey et al. 2005)
ndre, ndrep	20	Normalized difference red edge index (Barnes et al. 2000, Sentinel-2 only)
ccci	20	Canopy chlorophyll content index (Barnes et al. 2000, Sentinel-2 only)
mcari	20	Modified chlorophyll absorption ratio index (Daughtry et al. 2000, Sentinel-2 only)
resav	20	Red edge soil-adjusted vegetation index (Cao et al. 2013, Sentinel-2 only)

[A] The indices ending in 'p' were developed during this work. They were calculated by adding 2 to the denominator of the standard formula for the metric, to compensate for index overestimation on dark surfaces such as water and deep shadow.

2.6.3. Topographic metrics

A variety of metrics describing the influence of local topography on vegetation composition were calculated from the 10-meter resolution 3DEP data (**Table 7**). The more complex novel metrics created during this project are briefly described here.

Table 7. Topographic predictive metrics, the effective spatial resolution at which they respond, and a brief description or reference to a methodology.

Metric name(s)	Effective res (m)	Description
elev	10	Bare earth elevation.
slope	10	Slope in degrees (Esri 2013).
east, south	10	"Eastiness" = sin(aspect) and "southiness" = sin(aspect-90°).
cur30, cur150, cur750	30, 150, 750	3x3-cell total curvature (Esri 2013) from elevation aggregated to 30m, 150m and 750m resolution.
cpl30, cpl150, cpl750	30, 150, 750	3x3-cell planimetric curvature (Esri 2013) from elevation aggregated to 30m, 150m, and 750m resolution.
cpr30, cpr150, cpr750	30, 150, 750	3x3-cell profile curvature (Esri 2013) from elevation aggregated to 30m, 150m and 750m resolution.
heat	10	Relative heat load (McCune and Keon 2002).
raddir, raddur	30	Direct solar radiation and duration of direct illumination across full year (Esri 2013); distinct from heat load in that cast topographic shadows are modeled.
topodry	30	Elevation-scaled heat index = *raddir* * (1 – (*elev* / highest elev in WA)).
mp630, mp3150	30, 150	Morphometric protection (SAGA-GIS, Conrad et al. 2015) from elevation aggregated to 30m over 630m radius, and to 150m over 3150m radius.
tpp300, tpp1500, tpp7500	30, 150, 750	Topographic position percentile, the percentile rank of cell elevation relative to surrounding elevations within a 300m, 1500m and 7500m radius.
tpmi300, tpmi1500, tpmi7500	30, 150, 750	Minimum elevation differential within 300m, 1500m and 7500m. See text for methodology.
tpma300, tpma1500, tpma7500	30, 150, 750	Maximum elevation differential within 300m, 1500m and 7500m. See text for methodology.
cold300, cold1500, cold7500	30, 150, 750	Cold air accumulation calculated over surrounding 300m, 1500m and 7500m. See text for methodology.
rough30, rough90, rough270	10, 30, 90	Surface roughness at 30m, 90m and 270m scales. See text for methodology.

Minimum and maximum elevation differentials

We devised two multi-resolution metrics to quantitatively represent landform position. Four bisecting lines of length 21 times the cell resolution were constructed for each cell, oriented in the N-S, NE-SW, E-W, and SE-NW directions. The mean elevation along each line was determined, and differences between the central cell's elevation and each of the four means were calculated. The minimum of these four differences (*minimum elevation differential* or *tpmi*) and the maximum (*maximum elevation differential* or *tpma*) are relevant with respect to landform position. For example, a peak would have high values of both *tpma* and *tpmi*, while a level ridgeline would have a high *tpma*

and a *tpmi* near zero. A gap in a ridgeline would have a high *tpma* and a fairly large negative *tpmi*. The metrics were calculated at a variety of cell sizes to represent terrain morphology at a variety of spatial scales.

Cold air accumulation

Cold air accumulation in basins is a major driver of vegetation patterns in mountainous terrain. We developed an original approach for simulating this process, using the four elevation differentials created above. Locations at which the sum of the elevation differentials across perpendicular axes is a negative number have some tendency to accumulate cold air draining from above. The greater the magnitude of this negative number, the greater will be the tendency for cold air to enter from above and become trapped, and the colder that air is likely to be.

For the four elevation differentials ed_{NS}, ed_{EW}, ed_{NESW} and ed_{SENW} we found the minimum sum of each of the perpendicular pairs:

$$ed_{\perp,min} = \min(ed_{NS} + ed_{EW} \ , \ ed_{NESW} + ed_{SENW})$$

By analogy with the compound topographic index (Moore 1991)—a hydrologic metric that similarly integrates the influence of a size-varying contribution area with the local tendency to disperse that input—we represented cold air accumulation at a cell using:

$$cold = \ln\left(\frac{ed_{\perp,min}}{s}\right)$$

where s is the slope in the downward direction from the cell at the same spatial scale over which the elevation differentials were calculated. Cold air accumulation was determined at each of the cell sizes for which elevation differentials were produced.

Surface roughness

We defined surface roughness as local variability in aspect that is non-trending across an analysis window, scaled up by the local slope. The non-trending criterion is important—for example, a window centered on a north-south oriented ridgeline would show a strong change in aspect from west-facing to east-facing, but this would not indicate surface roughness. To accomplish this, we again used the alternating-cell convolution kernel $\begin{bmatrix} +4 & -5 & +4 \\ -5 & +4 & -5 \\ +4 & -5 & +4 \end{bmatrix}$, this time applied to four transformations of aspect: $\sin(asp)$, $\sin(asp - 45°)$, $\sin(asp - 90°)$ and $\sin(asp - 135°)$, summed these four directional measures of aspect variability, and multiplied by the mean slope across the analysis window. Roughness was computed at a range of spatial scales.

2.6.4. Hydrologic metrics

The hydrologic metrics were derived from processing within a landscape context rather than from a simple pixel-based perspective, since they depend on upstream areas in addition to the immediate surroundings. We first created a hydrologic flow accumulation layer based on the bare earth elevation, correcting for poorly modeled flow due to lack of information on road culvert locations. We used the flow accumulation layer to create a channel network, calibrating it using an NPS

streams data layer. The channel network was used as an input to a variety of distance metrics describing proximity to channels exceeding various flow thresholds. The predictive hydrologic metrics are shown in **Table 8**.

Table 8. Hydrologic predictive metrics, the effective spatial resolution at which they respond, and a brief description or reference to a methodology.

Metric name(s)	Effective res (m)	Description
vd_perm, vd_major	10	Vertical distance above permanent channel and major river networks (Conrad et al. 2015)
hd_perm, hd_major	10	Horizontal distance to permanent channel and major river networks
dtw	10	Cartographic depth to water index (White et al. 2012)
wetness	10	SAGA wetness index (Conrad et al. 2015), closely related to Compound Topographic Index (Moore 1991)
upland	10	Log-scaled cost distance to channel network, see text

Flow accumulation and channel networks

Hydrologic *flow accumulation* is a spatial representation of the catchment area contributing to flow at each gridded location in a drainage network. Its computation was important both as a step in the channel delineation process and also as a key input needed to generate several predictive metrics. The flow accumulation algorithm in SAGA-GIS (Conrad et al. 2015), when used to delineate channel networks, produced anastomosing effects in flat areas and appeared to realistically represent hydrologic processes for incorporation into predictive metrics.

Hydrologic modeling was performed at 10-meter resolution. In order to represent the impact of spatial precipitation patterns on channel development, we created a weighted grid by rescaling PRISM annual precipitation to a fraction of the maximum value in the study area. We then filled sinks in the elevation grid, using the Wang & Liu (2006) method with *minslope* = 0.01, and modeled flow accumulation based on the weighted precipitation grid, using SAGA's **Catchment Area (Top-Down)** method with multiple flow directions and *convergence* = 1.1.

We used the flow accumulation results to delineate channel networks, also in SAGA-GIS. Two alternate channel networks were created from the flow accumulation result. One was calibrated to represent all permanent channels, the other to represent only unconstrained rivers in major valleys. Various minimum thresholds of flow necessary to result in a channel were tested; the resulting networks were visually compared to stream representations in USGS 1:24,000 quad sheets. The best match to the represented permanent streams was found using a flow accumulation threshold of 20,000, corresponding to an average catchment area of approximately 200 hectares. A threshold of 8,000,000 (corresponding to about 80,000 ha) was used for major rivers; this resulted in delineation of channels downstream of the approximate location where their floodplains begin to widen substantially.

Riparian influence and metrics generation

We devised a metric to express the degree of floristic riparian influence at any location. The first step was to determine the total flow quantity associated with each section of the channel network. The channel network was broken into discrete channel reaches defined by network intersections. Many channel segments were composed of anastomosing flow pathways, in which flow was modeled in several adjacent parallel paths; it was therefore necessary to consider the several paths as all contributing to a single total flow value. We accomplished this by associating each flow accumulation cell with the nearest delineated reach[49] and averaging across reach length.

We classified channel reaches into five categories based on average reach flow, with thresholds between the categories spaced in a regular geometric progression ranging from the minimum to the maximum channel reach flow in the study area. We then created a cost function to describe the degree of riparian influence in the perpendicular direction away from the channel. The cost function was proportional to the square of slope, which emphasized slope breaks and was able to represent physiographic features such as fluvial terraces and natural levees. We calculated the least cost distance from each cell to each of the five channel size categories using this function.

The riparian influence metric was fit to its practical impact on species composition by examining the cost function values at the locations of training plots assigned to riparian vs. non-riparian associations. This resulted in an estimate of a cost function cutoff for each of the five flow categories that most accurately separated the plots with riparian floristics from those with upland floristics. A logarithmic relationship was found to best fit the relationship between the five cost function cutoffs and the mean flow quantity across all reaches in each of the five flow categories. We then iteratively modified the initial cutoffs until they exactly fit the logarithmic model. For each of the flow categories, we assumed that no further floristic riparian influence would be exerted beyond the cost distance cutoff. Finally, an "uplandness" index was created using:

$$upland = \log_{10}\left[1 + \min\left(\frac{C_A}{T_A}, \frac{C_B}{T_B}, \frac{C_C}{T_C}, \frac{C_D}{T_D}, \frac{C_E}{T_E}\right)\right]$$

where $C_{A..E}$ represent the slope-based cost distances to each of the five flow categories and $T_{A..E}$ represent the cost cutoffs used to define the extent of riparian influence for each category.

2.6.5. Climate metrics

The climate data required no additional processing to form predictive metrics. The predictors (**Table 9**) were simply the 1981–2010 normals provided by the PRISM Climate Group (2019).

[49] "Distance" to the reach was evaluated via a cumulative slope cost function.

Table 9. Climate predictive metrics, the effective spatial resolution at which they respond, and a brief description or reference to a methodology.

Metric name(s)	Effective res (m)	Description
ppt_jan, ppt_apr, ppt_jul, ppt_oct	~800	Average precipitation for month.
tmax_jan, tmax_apr, tmax_jul, tmax_oct	~800	Average daily maximum temperature for month.
tmin_jan, tmin_apr, tmin_jul, tmin_oct	~800	Average daily minimum temperature for month.
tdew_jan, tdew_apr, tdew_jul, tdew_oct	~800	Average daily mean dew point temperature for month.
vmax_jan, vmax_apr, vmax_jul, vmax_oct	~800	Average daily maximum vapor pressure deficit for month.

2.7. Modeling

We used a machine learning algorithm, *random forests* (RF; Breiman 2001, Liaw and Wiener 2002), to build models for predicting map class presence using the quality-controlled training plots resulting from the work in **Section 2.3**. We used RF because of its tendency to avoid overfitting to training data and its ability to isolate signals in noisy datasets (Cutler et al. 2007). The large number of map classes, with widely varying quantities of available training data, presented a modeling challenge: how to simultaneously produce models that are good at both "easy" prediction tasks (e.g., discriminating between low and high elevation types) and "hard" tasks (e.g., discriminating between two tall shrubland types occurring in similar settings), while avoiding bias against the rarest classes and also making maximum use of all available training data. To address this, we wrapped the RF algorithm in a factorial binary process in which each map class was modeled against every other. This allowed each model to specialize in distinguishing a single pair of map classes, choosing appropriate predictors for that task. During the prediction phase each class "competed" with each other class; the class with the lowest cumulative loss margin across all contests at a pixel was considered the best answer there. The predictor selection, model creation, and model prediction phases discussed below all ran on binary models.

2.7.1. Model predictor data

All the metrics discussed in **Section 2.6** were resampled to a fixed 3-meter resolution grid over the coincident extent of all metrics. The resampling method used depended on the data source. We used nearest neighbor resampling to maintain the finest resolution possible for all metrics derived from NAIP; the predictor sampling grid was taken from these rasters to prevent any spatial shifting. Satellite imagery was resampled using cubic convolution, which results in less smoothing than bilinear interpolation and maintains crisper boundaries. The non-imagery layers were resampled using bilinear interpolation.

2.7.2. Model training data

Following the quality control process, the training plots represented relatively continuous patches of the assigned map class, spanning the full area defined by the plot center location and radius. Any patches of alternate types within the plot were assumed to be less than nine meters on a side. Training data were created from the predictor metrics by extracting the 3-meter pixel values at 13 points distributed across each training circle, with the most distant four points lying on the circumference

(**Figure 12**).[50] The primary reason for extracting data from multiple locations at each plot was the necessity of training models at the same spatial scale at which they were predicted.[51] In addition, this scheme allowed better representation of the range of predictor variation within each plot, including providing training data near transitions to adjacent vegetation types. The assigned association and map class calls and the extracted predictor values were then imported into R using functions provided by the **rgdal** package (Bivand et al. 2014). The training data represented 48 distinct map class calls; flowing water and impounded water were modeled separately but were later merged into the single map class W81–FRESH WATER, and B31–BROADLEAF RIPARIAN & SWAMP FOREST was modeled as distinct east- and west-side variants before merging.

[50] Because the 13 component samples from each plot are not statistically independent, we used only one of the 13 in any given random forests tree during the predictor selection and model creation steps. This avoided introducing pseudo-replication of training data and preserved the independence of the out-of-bag samples, while making use of the predictor variability within each plot.

[51] A commonly used alternative, summarizing predictor data over the plot area, would have introduced artificial smoothing into the training data that would not be applicable for predicting onto finer resolution pixels.

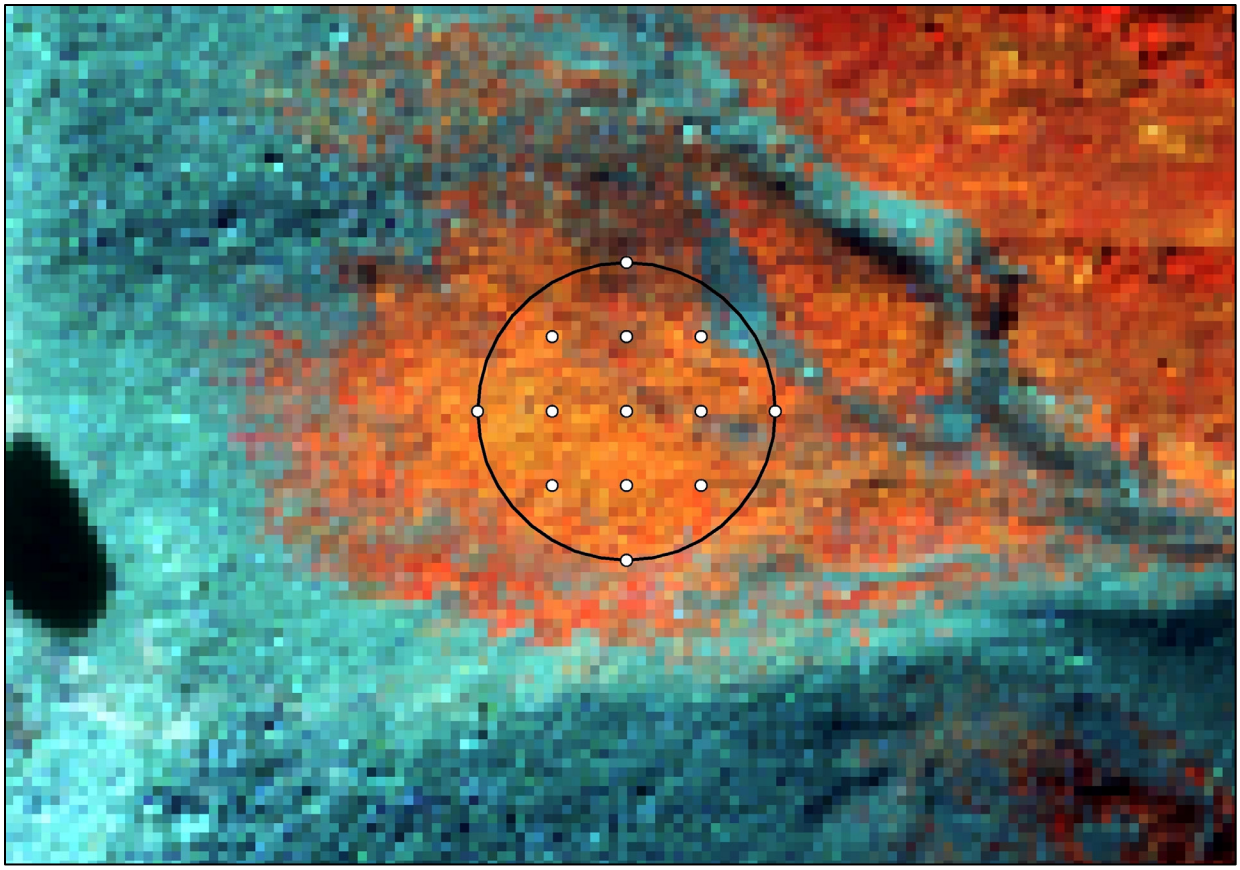


Figure 12. Training data extraction. Predictors were extracted from the 3-meter resolution metrics at 13 points distributed across each training circle. The vegetation patch represented was assigned to H53–SHOWY SEDGE AND SITKA VALERIAN MEADOW; the imagery is color-infrared 2015 NAIP at 1-meter resolution.

2.7.3. Model binarization

Each two-class combination of the 48 modeled map classes was treated separately, resulting in 1,128 distinct binary models.[52] This allowed each model to specialize in a single task—distinguishing two classes from one another—and gave us the freedom to treat issues of predictor collinearity more sensitively. For example, over the geography defined by all training samples in NOCA, there is a very strong negative correlation between elevation and maximum January temperature. But within the environmental subspace defined by the training plots assigned to C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST and C06–WESTERN HEMLOCK, DOUGLAS-FIR AND SALAL FOREST, those variables are only weakly correlated, with maximum January temperature being a very strong predictor and elevation comparatively weak. Using both in a model based on training data throughout NOCA would violate standards against excessively correlated predictors. But there is no

[52] We handled the computationally intensive process of predictor selection, model creation and model prediction at 3-meter resolution using three standard desktop computers, each running between three and five instances of R or Python simultaneously. They were connected to a network-attached storage device that hosted the training data, predictor data, and a shared status file that allowed the processes to distribute the tasks amongst themselves.

such violation for the single model C05 versus C06 and to exclude either predictor on this basis would unnecessarily reduce the model accuracy.

2.7.4. Predictor selection

We developed a novel predictor selection method to use with our multi-resolution predictor datasets, which reduced predictor collinearity[53] while also optimizing model accuracy, model effective spatial resolution, and the efficiency of the prediction process.

Initial selection

We used a stepwise variable selection process coded in R, which was based on maximizing RF cross-validated model accuracy at each step.[54] We organized the predictors into ten tiers based on the effective spatial resolution at which they were calculated,[55] with the finest scale predictors—the 3-meter resolution NAIP band responses—in the first tier.

At each tier, the process cycled through all available predictors, building 100 forests of 501 trees each, with each forest built from a single randomly selected point of the 13 for each plot. For each of the two map classes in the model, the *out-of-bag error rate*[56] for each plot, e_p, was compiled over each of the forests and converted to an estimate of the probability of plot misclassification by a single forest.[57] This probability estimate was then averaged across all plots to produce an overall error rate estimate for the model including the newly introduced predictor.[58] The predictor in the tier that resulted in the greatest decrease in model error rate was selected; any predictors (in that tier or others) with an absolute-valued Spearman rank correlation of 0.8 or greater to the selected predictor were eliminated from further consideration. If no predictors within the tier resulted in a decrease in

[53] Inclusion of substantially correlated predictors causes RF to overfit to those predictors, which is a major concern because our training data were gathered from such a small fraction of the project area.

[54] We considered using a process guided by an importance measure returned by RF, as in Evans and Cushman (2009). However, we found that a predictor's contribution to model accuracy is strongly dependent on which other predictors are included, and that an importance measure returned from a model based on all predictors was not indicative of its potential utility in a model based on a small subset.

[55] However, despite the availability of high-resolution topography from lidar across much of the park, we considered topographic and hydrologic predictors together with mid-resolution satellite imagery—after Sentinel imagery but before Landsat—to keep the emphasis on existing conditions as opposed to environmental setting.

[56] RF generates this by testing each tree of the model against the samples that were withheld from creating it.

[57] The expected misclassification by a single forest was of interest because the map was made based on a single forest. This step assumed normal distribution of e_p across forests.

[58] The model error rate here is defined as the higher error rate of the two modeled classes. By optimizing this quantity, rather than the overall (average) model error rate, we kept the error rate of the two classes balanced, which was an important assumption made by our prediction method.

model error rate, consideration moved to the following tier. After a predictor was selected, consideration always moved to the first tier again.[59]

Climate variables can act as proxies for geographic location, as they are generally arranged along broad spatial gradients. Their use as predictors can present a severe risk of overfitting to training data whose collection has been determined more by convenience than by a random sample. Because our climate predictors were derived from approximately 800-meter resolution data, they were in the final selection tier. We additionally limited models to only one climate predictor, to reduce the likelihood of overfitting to our often spatially constricted training data.

Often there was an inherent tradeoff between accuracy and spatial resolution. If satellite imagery or coarser scale texture metrics provide key information that is lacking in finer scale data, their use will increase accuracy but will also coarsen the model's resolution (**Figure 13**).

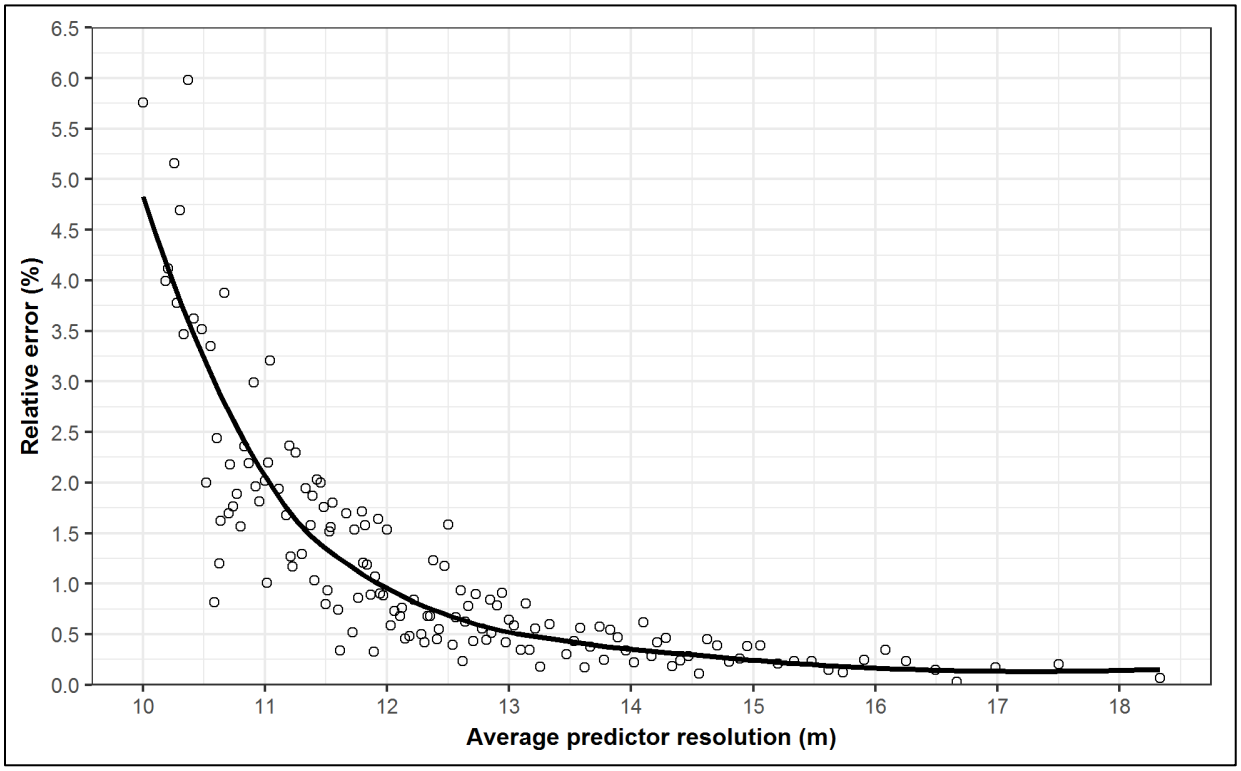


Figure 13. Average predictor resolution vs. median relative model error (the error increase attributable to leaving out coarser predictors). Predictors finer than 10-m resolution were treated as 10 m here; greater averages indicate increasing incorporation of coarser predictors in the model. The data were derived from only models with final error of 1% or greater, with a median final error of 3.7%; this would have been 8.5% using only predictors of 10-m or finer resolution. The best fit line is from a loess smoothing function.

[59] Because a predictor's value may not be recognized until a compatible predictor has been included.

Predictor switching

As seen in **Figure 13**, there is an optimal resolution at which to produce a model, which takes advantage of some of the predictive power of coarser resolution predictors, while maintaining responsiveness to fine scale vegetation transitions.[60] The predictor selection routine described above simply tried to minimize model error, but some of this may come at the unnecessary expense of coarser resolution. To address the tradeoff between the two, we created a new metric *errxres* that combined both model error and average predictor resolution:

$$errxres = avgres(err + 1\%)$$

where *err* is the model error (in percent) and *avgres* is the average predictor resolution. Starting with the predictors selected in the previous phase, we used another R script to drop the last selected predictors until the value of *errxres* was minimized. We then tested each of the remaining predictors, finding the model error rate that resulted from substituting any highly correlated predictors for them. Any substitutions that resulted in lowering *errxres* were accepted. The resulting predictor list was saved as an alternative set.

Choosing best set

To choose between the two sets of predictors produced, we used different decision-making criteria depending on whether we wanted to prioritize error rate or mapping resolution for the model. If the two map classes in the model were both larger patch size types (e.g., most conifer-dominated map classes), or if their environmental envelopes were so distinct that they wouldn't be found in close proximity to one another, we concluded that high spatial resolution in the resulting map was not as important as model error rate. In this case, we kept the set of predictors that resulted in the lowest error rate.

For pairs of map classes in which fine grain mapping was a high priority, we kept the predictor set that minimized the product $avgres(err')$, where *err'* was a transformed version of *err* that prioritized the reduction of model error to 5%, but only gave partial credit for reducing error lower than that.[61] We did this because the training samples were not perfectly pure[62] and we wanted to prioritize predictor resolution once a low error rate had been achieved. For example, a model to distinguish between a meadow and a woodland may have been trained with meadow samples that had occasional scattered trees. The best model in this case may have been one with a non-zero error rate against the training data. **Table 10** lists the most frequently included predictors across all binary models.

---

[60] The assumption is being made that the average spatial resolution of the predictors included in model is related to the effective resolution at which it "maps." Since RF is an inherently non-linear process, this is not necessarily true, though it is intuitively appealing.

[61] $err' = \begin{cases} \max\left(4, \frac{(err+15)}{4}\right) & if\ err \leq 5 \\ err & if\ err \geq 5 \end{cases}$

[62] See **Figure 12**. Small patches where an alternative map class might be preferable are present in many training plots.

Table 10. Most frequently used predictors in each selection tier. Up to ten predictors are shown for each tier, provided they were used in at least 2% of the models.[A] The tables in **Section 2.6** provide descriptions for each predictor.

Tier	Resolution or type	Predictor names[B] and number of models in which used (in parentheses)
1	3 meters	*n_d1c* (614), *n_w1c* (560), *r_n1_mx* (512), *n_g1b* (501), *n_w1a* (497), *n_e1c* (466), *n_n1a* (426), *r_w1_md* (424), *r_n1_md* (397), *n_n1c* (380)
2	6 meters	*n_nb_13* (456), *n_wa_13* (383), *n_wc_13* (383), *n_nc_13* (381), *n_wb_13* (381), *n_uc_13* (370), *n_gc_13* (369), *n_rc_13* (368), *n_vc_13* (366), *n_vb_13* (364)
3	9–10 meters	*s_nir* (431), *s_grn* (304), *s_ndgr* (251), *s_ndng* (220), *s_red* (220), *s_ndvi* (202), *n_d3c* (185), *n_e3c* (143), *n_g3b* (128), *n_g3c* (127)
4	12 meters	*n_wa_26* (111), *n_wb_26* (98), *n_d4c* (96), *n_ra_26* (93), *n_va_26* (93), *n_vb_26* (92), *n_ub_26* (91), *n_nb_26* (90), *n_wc_26* (90), *n_na_26* (89)
5	18–20 meters	*s_ndsi* (178), *s_sw1* (99), *s_tcw* (99), *s_ndsw* (87), *s_ndmi* (80), *s_tcb* (75), *n_wa_39* (54), *n_ra_39* (47), *s_re1* (47), *s_sw2* (47)
6A	topographic	*t_elev* (445), *t_topodry* (220), *t_slope* (144), *t_cpr750* (119), *t_tpma7500* (106), *t_rough30* (103), *t_cold7500* (101), *t_tpp300* (101), *t_tpp7500* (96), *t_raddur* (94)
6B	hydrologic	*h_hd_major* (209), *h_wetness* (142), *h_dtw* (118), *h_vd_major* (97), *h_upland* (95), *h_hd_perm* (61), *h_vd_perm* (42)
7	27–30 meters	*e_ndgrp* (32)
10	climate	*p_ppt_apr* (65), *p_ppt_oct* (46), *p_clidry* (34), *p_vmax_jul* (34), *p_ppt_jan* (30), *p_tmax_jan* (30), *p_vmax_apr* (26), *p_ppt_jul* (25)

[A] The 'p' variants of the summer satellite imagery metrics ("s") were omitted from modeling due to a programming error.

[B] Predictor names are preceded by a letter indicating to which source group they belong: "e" indicates late summer minimum-snow satellite imagery, "h" is hydrologic, "n" is aerial imagery NTM, "p" is climate, "r" is aerial imagery reflectance, "s" is summer satellite imagery, and "t" is topographic.

Additional predictors for abiotic map classes

Many models for abiotic map classes had a small number of predictors selected, since the most obvious difference between these and vegetated types is their lack of vegetation, which is easily ascertained from NAIP imagery. While these very simple models worked well under normal circumstances, we found that in deep shadows these models often performed poorly. The abiotic classes are generally restricted to environments that are easily described in terms of topographic and hydrologic metrics. For example, impounded water is found in areas with low slope and high topographic wetness, and barren colluvial deposits are found in concave areas with positive curvature. We added appropriate predictors to models involving these map classes to make sure they remained restricted to reasonable locations.

2.7.5. Model creation

We built a random forest of 507 trees for each map class pair,[63] using the predictors selected above and specifying a sample size for each class equal to the minimum number of training plots available

[63] 39 trees were generated for each of the 13 sample points at each training plot.

for either class.[64] The resulting model was saved for use later in the prediction phase. We then estimated model error rates using 1000 bootstrap samples. Each was constructed by holding out one plot from the least common class and a proportional number from the most common class, again randomly selecting from the 13 sample points available at each plot for both training and test sets. **Figure 14** illustrates the cumulative probability across error rate for all binary models. Of the 1,128 models, 27% had an error rate of zero, a considerably smaller proportion than at the other NCCN parks. Substantial error is concentrated in a fairly small number of models; like the other parks, 95% of the models showed less than 10% error.

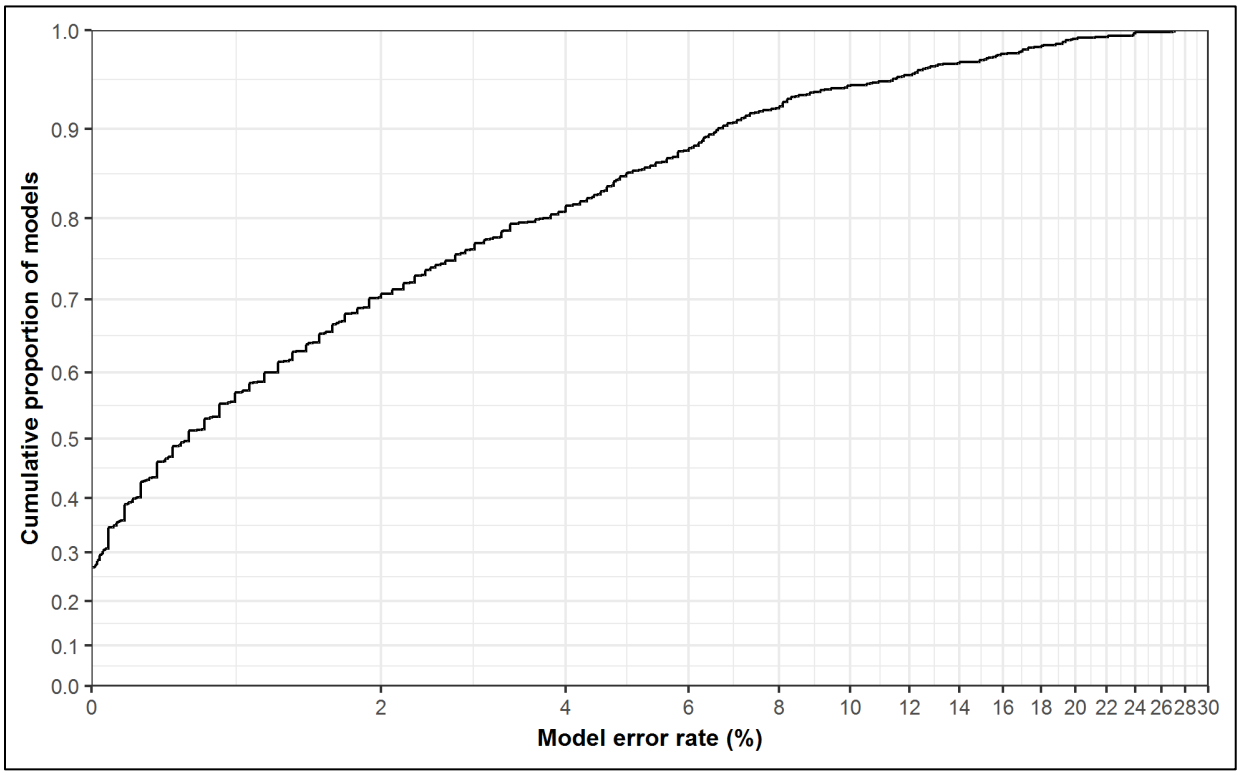


Figure 14. Error rate across all binary models. Recall that due to heterogeneity within training samples (i.e., "inclusions"), some model error is to be expected.

Significant model error is highly concentrated in a fairly small number of binary one-versus-one models; these are map class distinctions that are more likely to map poorly. Although the accuracy assessment (**Section 3**) provides more definitive metrics of map accuracy, some map classes were poorly sampled in the accuracy assessment; for those, model error may be useful supplementary information. **Table 11** lists the 40 models with the highest error rates.

[64] When the classes to be predicted are not represented evenly in the training data, the more common class has a tendency to be modeled with greater accuracy than the other. This effect can be alleviated by downsampling the more common class (see Evans and Cushman 2009). The same technique was used earlier during the predictor selection phase.

Table 11. Binary models with highest cross-validated error rates.

Map class 1 / Map class 2 (codes and abbreviated names)	Error rate (%)
H60W–Black alpine sedge / S49–Alpine heather	31.4
H60W–Black alpine sedge / S48–Subalpine heather	27.1
C05–W hemlock & sword fern / C07–N Casc dry Doug-fir	26.9
H51W–Subalpine herbaceous wetland / H60W–Black alpine sedge	24.0
C05–W hemlock & sword fern / C11–Mesic silver fir & w hemlock	23.9
H50W–Lowland marsh & meadow / S40W–Low elevation shrub wetland	23.9
R72–Colluvial barren / R73–Bedrock barren	23.8
H51W–Subalpine herbaceous wetland / S41W–Subalpine willow wetland	22.1
R71–Alluvial barren / R72–Colluvial barren	21.2
H53–Showy sedge & Sitka valerian / S47–Successional huckleberry	19.8
C04–Moist w hemlock & foamflower / C10–Moist silver fir & foamflower	19.7
B35–Upland paper birch & conifer forest / C07–N Casc dry Doug-fir	19.4
H57–Green fescue dry meadow / S47–Successional huckleberry	19.4
S44–Thimbleberry, forbs & bracken fern / S45–Vine maple	19.2
B33–Upland bigleaf maple & conifer / B35–Upland paper birch & conifer forest	19.2
S48–Subalpine heather / S49–Alpine heather	18.8
B30–Successional gravel bar / R71–Alluvial barren	18.1
C11–Mesic silver fir & w hemlock / C14–Silver fir & big huckleberry	18.0
H60W–Black alpine sedge / H62–Alpine sparse herbaceous	17.6
S42–Sitka willow riparian / S44–Thimbleberry, forbs & bracken fern	17.3
C13–Mtn hemlock & Cascade azalea / C14–Silver fir & big huckleberry	17.3
B31–Broadleaf riparian & swamp forest / S40W–Low elevation shrub wetland	17.0
R71–Alluvial barren / W81–Fresh water	17.0
C13–Mtn hemlock & Cascade azalea / C20–Subalp fir & Sitka valerian	16.9
H52–Cow parsnip / S44–Thimbleberry, forbs & bracken fern	16.8
C10–Moist silver fir & foamflower / C11–Mesic silver fir & w hemlock	16.2
C21–Mtn hemlock & heather / S48–Subalpine heather	15.9
H54–Moist talus vegetation / R72–Colluvial barren	15.7
S42–Sitka willow riparian / S43–Sitka alder	15.7
B35–Upland paper birch & conifer forest / C05–W hemlock & sword fern	15.5
C14–Silver fir & big huckleberry / C16–N Casc Doug-fir & subalp fir	15.3
C11–Mesic silver fir & w hemlock / C12–Silver fir & Alaska blueberry	15.2
B34–Bigleaf maple & Doug-fir debris apron / S45–Vine maple	15.1
C15–Lodgepole pine & Doug-fir / C16–N Casc Doug-fir & subalp fir	14.9
H62–Alpine sparse herbaceous / R72–Colluvial barren	14.9
H62–Alpine sparse herbaceous / R73–Bedrock barren	14.0
C05–W hemlock & sword fern / C06–W hemlock & salal	13.9

Table 11 (continued). Binary models with highest cross-validated error rates.

Map class 1 / Map class 2 (codes and abbreviated names)	Error rate (%)
S42–Sitka willow riparian / S45–Vine maple	13.3
S44–Thimbleberry, forbs & bracken fern / S46–Snowbrush & Scouler's willow	13.1
C04–Moist w hemlock & foamflower / C05–W hemlock & sword fern	13.0

2.7.6. Model prediction

The map class prediction at each 3-meter pixel was made by evaluating the results of each one-versus-one model and determining which class had the best overall performance. The "winner" of each model was determined using a simple threshold of 50% of the 507 trees. **Figure 15** shows the outcome of a single binary model in one small area. We accomplished this by creating a round-robin schedule of "contests" using the circle method of Reverend Kirkman (1847). Not all models needed to be evaluated at each pixel; after a map class had "lost" five contests, it was eliminated from contention and any subsequent models including it were skipped. The selection of the "winning" map class was made by comparing the total probability loss margin across all models, rather than by the number of contests lost, which removed the possibility of tie outcomes.

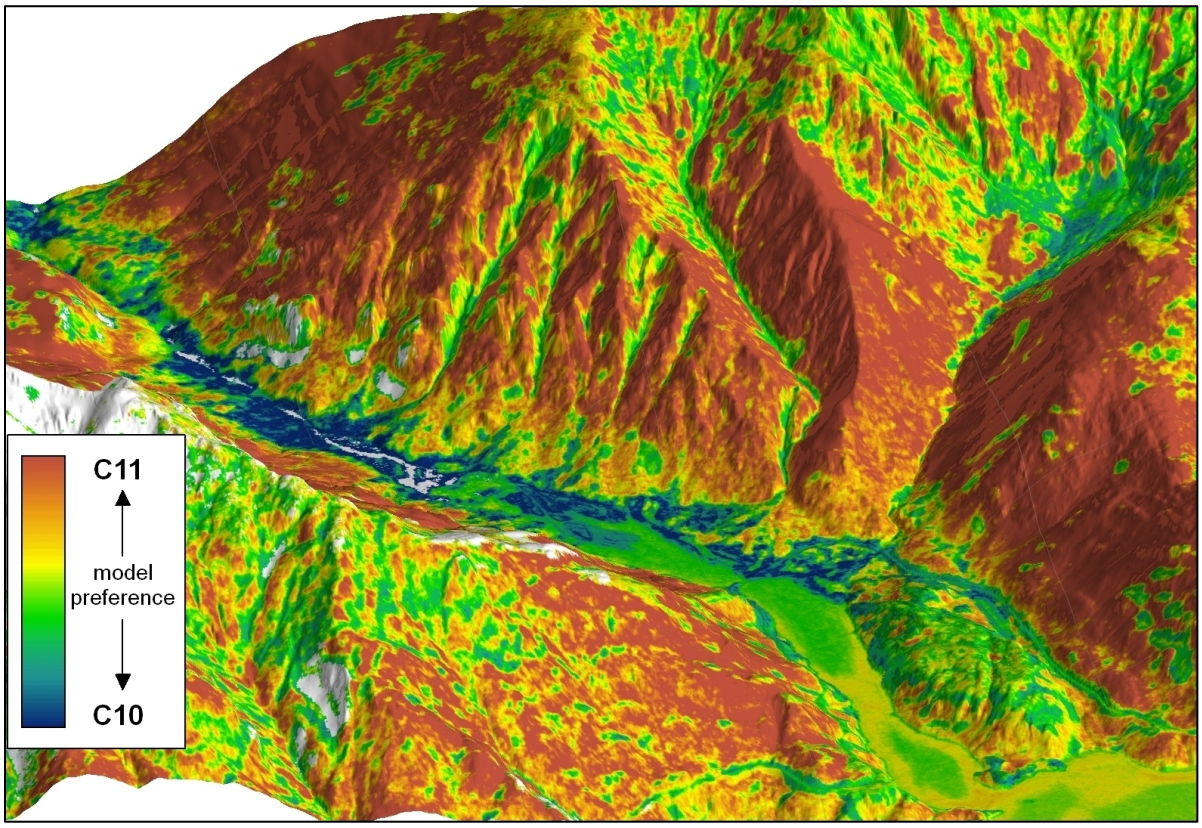


Figure 15. Binary model prediction example, for the model C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST versus C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST. C10 was favored mostly on toe slopes and valley bottoms, while C11 was preferred on middle and higher slopes. White areas were not predicted because one or both map classes had already been eliminated from contention.

Processing made use of the R **randomForest** (Liaw and Wiener 2002), **raster** (Hijmans 2018), and **rgdal** (Bivand et al. 2014) packages, and was made more efficient by dividing the project area into tiles of approximately 2000 by 2000 pixels each. Each concurrent prediction process loaded the full set of predictor rasters for a single tile into memory and evaluated all needed models, tracking the number of losses and total loss margin by map class. The results for each binary model and the tracking information were copied to the network-attached storage device.[65] The total loss margin, seen in **Figure 16**, can be interpreted as a map of model uncertainty.

[65] Although multithreaded prediction (using all available CPU cores in a single process) is possible in R, we encountered reliability issues with this approach, and also found it was significantly more efficient to use multiple single-threaded processes.

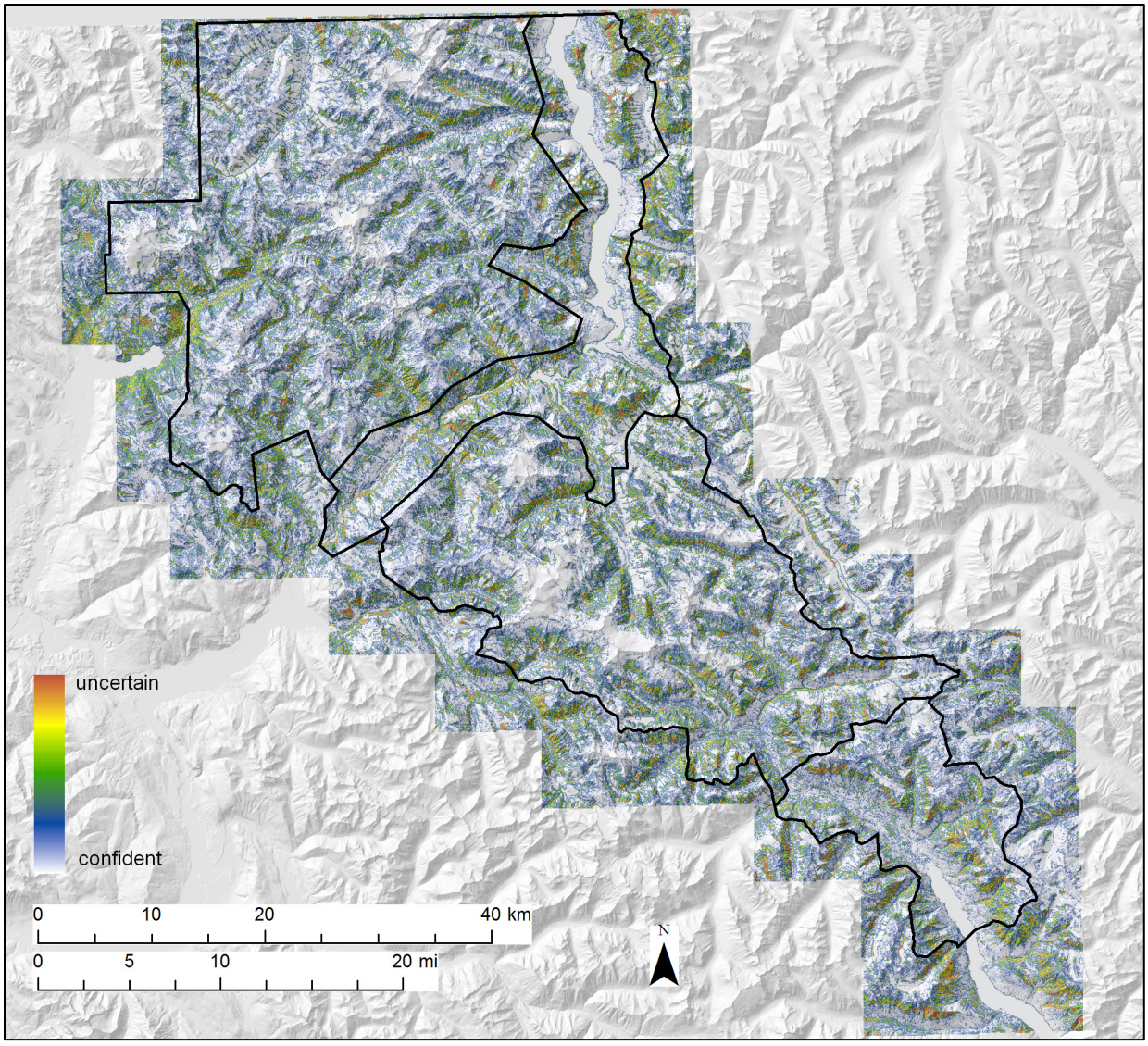


Figure 16. Prediction uncertainty. The predicted class won all contests in areas displayed as white. Colors from blue to red indicate that the best class lost at least one contest, by increasing amounts. Certainty is lowest where training data were inadequate, especially in difficult-to-access areas.

2.8. Post-processing

The random forests pixel-based predictions were converted into a final map by means of a sequence of post-processing steps. We used various approaches, described below, to add additional map classes for land cover types within the project area but not represented in training data. A filtering process was then used to convert the pixel-based predictions into a polygon-based map. After that, we did a final phase of manual map editing to address observed problems in a few areas.

2.8.1. Additional map classes

We defined several additional map classes—M92–BURNED WITH UNCERTAIN VEGETATION, M93–TIMBERLAND WITH UNCERTAIN VEGETATION, M94–DEVELOPMENT, M95–ROADS IN PARK and M96–

CLEARED CORRIDORS—to represent land cover types that were present in the project area but not in the training data.

Burned areas

The tasseled cap wetness index is particularly effective at estimating structural attributes in conifer-dominated forests (Cohen and Spies 1992). We created a mask of recently burned areas by subtracting the tasseled cap wetness calculated from the historic midsummer satellite imagery from that calculated from the current midsummer imagery (see **Sections 2.5.2** and **2.6.2**). We empirically determined a change threshold that was effective in flagging areas that had experienced severe burns between those dates, restricting the results to pixels within the digitized recent fires mask (**Section 2.1.2**). For areas identified as burned, if the model prediction was not a map class associated with early recovery from fire[66] (e.g., B33–UPLAND RED ALDER, BIGLEAF MAPLE AND CONIFER FOREST, S46–SNOWBRUSH AND SCOULER'S WILLOW SHRUBLAND, S47–SUCCESSIONAL HUCKLEBERRY SHRUBLAND), we recoded it to M92–BURNED WITH UNCERTAIN VEGETATION. The burned class included post-fire recovery areas as well as recent burns. We assumed that the model prediction would generally be an acceptable result for areas recovering from burns earlier than the mid-1980s. Because east-side post-fire communities are better represented in the classification and training data, burns in west-side forests are more likely to be mapped as M92–BURNED WITH UNCERTAIN VEGETATION.

Development

A variety of land cover types on both sides of the park boundary are actively maintained by human land-use practices (e.g., buildings for residential and commercial purposes, agriculture, roads). We designated three map classes to encompass these: M95–ROADS IN PARK (representing roads within the park boundaries only), M96–CLEARED CORRIDORS (power line corridors within the park, and the US-Canada boundary slash zone), and M94–DEVELOPMENT (representing everything else, including other development within the park).

We began by digitizing the roads, corridors and developed areas inside the park (using 2015 NAIP imagery), as well as major areas of development and agriculture outside the park boundary. Hand-mapping was done at 1:4,000 scale; road centerlines were buffered by either seven or 14 meters depending on their size. We created a managed areas mask by excluding the park and adjacent USFS wilderness from the project area, for use in later steps.

We added roads outside the park to M94–DEVELOPMENT by resampling the developed land cover classes in the 2016 National Land Cover Database (Yang et al. 2018) to our mapping resolution and removing areas that lay either outside the managed areas mask or within our digitized development and agriculture layer. Areas within the digitized development and agriculture layer were then added to M94–DEVELOPMENT unless they had modeled as a forest or woodland, a dense tall shrubland, or ocean, all of which can be reasonable map results in those places. Although ponds are a frequent feature in developed areas, we excluded them as there was a tendency to erroneously map fresh water in flat developed areas with cast shadows from adjacent trees. Digitized developed areas within the

[66] We referred to Franklin and Dyrness (1973) in compiling this list.

park were also included in M94–DEVELOPMENT. Digitized roads within the park and cleared corridors were included as separate map classes, M95–ROADS IN PARK and M96–CLEARED CORRIDORS.

Logging

We began by flagging disturbed forests, treating the impact of logging similarly to that of fires, by thresholding the historic change in tasseled cap wetness to detect areas that had experienced major canopy loss since the mid-1980s. We then applied a multi-stage majority filter and excluded areas that were smaller than a half-hectare, were within the park or adjacent wilderness areas, or had already been assigned to M94–DEVELOPMENT. Because flooding along major rivers was another significant cause of forest disturbance, areas that modeled as a typically riparian map class, were within five vertical feet of a major river and in a location with high hydrologic wetness (see **Section 2.6.4**) were also excluded.

The remaining disturbed areas were identified as potentially logged and were examined manually to remove those that did not appear to be within timber harvest boundaries. The rest were recoded to M93–TIMBERLAND WITH UNCERTAIN VEGETATION unless they had modeled as a forest, woodland or tall shrubland map class. The timberland class included early seral and planted forests, as well as recent regeneration harvests. We assumed that the model prediction would generally be an acceptable result for areas recovering from logging earlier than the mid-1980s.

2.8.2. Filtering

We converted the 3-meter pixel predictions to a polygon map via a sequence of filtering steps. Because lifeform can be predicted at very high accuracy but map classes are less easily distinguished, we began with a lifeform-specific majority filter that reassigned each pixel to the most common map class of the same lifeform among the neighboring pixels. No pixels were changed to a different lifeform than that to which they were predicted during this step. The analysis window ranged from 3-by-3 to 7-by-7 pixels depending on lifeform We next addressed fine scale speckle by applying two successive 3-by-3 pixel majority filters across all map classes with no lifeform specificity.

We then moved to filtering based on patch size and shape, beginning by removing very small patches of fewer than nine contiguous 3-meter pixels, reassigning pixels in those patches to the nearest persisting patches. The shortest distance from each pixel to any neighboring patch was determined and the mean depth of each patch (d) was found by summarizing over its constituent pixels. Through experimentation, we defined an additional parameter g to describe patch shape:

$$g = d^{3/2}\, a^{-1/4}$$

where a is the patch area. While d describes the average width of a patch, g is a shape parameter describing the width of a patch compared to its overall size. We then empirically determined map class-specific thresholds for d and g; patches for which either parameter exceeded its threshold were kept, while the others were eliminated, assigning the constituent pixels to the nearest adjacent patch. This allowed us to filter map classes that often occur in long slender strips (e.g. C26–CONIFER KRUMMHOLZ AND TREED CLIFF) differently than those that typically occur over more extensive areas.

We followed this with a final additional patch size filter, with a map class-specific size requirement ranging from nine to 49 pixels (81–441 m^2).

2.8.3. Map editing

We scanned the resulting polygon map for obvious errors that could be fixed easily by hand, finding few such issues. We then converted back to raster format and ran a final 3x3-pixel majority filter to eliminate any stray missing pixels from the map before converting back to polygons for the final time. The final map is available in Nielsen et al. (2021d).

3. Accuracy Assessment

3.1. Background

A map accuracy assessment (AA) determines the degree to which a map correctly represents on-the-ground conditions (see Lea and Curtis 2010, Congalton and Green 1999). A *confusion matrix* or *contingency table* tabulates the misassignments found between each possible pair of map classes. The information from this matrix is used to draw conclusions about the quality of mapping for each map class; the results allow an evaluation of potential map applications and applicable caveats. *User's accuracy* (UA) and *producer's accuracy* (PA) describe two relevant aspects of map accuracy.

UA is a reliability measure to estimate the probability that the map is correct where a particular class is mapped. It is inversely related to the false-positive or *commission error* (CE) rate (the probability of mapping the class where it is not present). Low UA may indicate that a class is *over-mapped* (mapped more often than it actually occurs). It also can be evidence that classes that are particularly confused with it are themselves *under-mapped* (mapped less often than they actually occur).

PA is a mappability measure to estimate the probability that the map is correct where a particular class is found on the ground. It is inversely related to the false-negative or *omission error* (OE) rate (the probability of omitting the class where it is present). Low PA may indicate that a class is under-mapped. Because PA is relative to the true land cover, rather than the mapped land cover, its calculation is dependent on an estimate of the true quantity of the class present in the study area. Thus, two distinct estimates of PA can be made. The first, relative to the number of plots found in the field, is calculated from a confusion matrix drawn directly from the sampled plots, the *sample contingency table* (SCT). The second, a more meaningful quantity, is scaled to an estimate of the true area occupied by each class, the *population contingency table* (PCT).

We followed the procedures and formulas provided by Lea and Curtis (2010) for sample design, sample protocol, and analysis, to the extent possible.[67] NPS standards specify an 80% accuracy goal for each individual map class hosting native vegetation communities. In addition to assessing the class-specific UA and PA against this standard, we used UA and PA in combination to produce an

[67] The logistics involved in mounting the AA field campaign with an experienced field crew—in addition to other project management considerations—required the AA fieldwork to occur earlier than would have been ideal. At the time of the fieldwork, the mapping associations and map classification were incomplete. The draft map relied on for the sample design had significant differences to the final map. Because we recognized the challenges that lay ahead in making the AA data compatible with an as-yet uncompleted classification and map, we had no choice but to violate some important guidance from Lea and Curtis (2010), collecting substantial field data during the fieldwork, and using it later in the office to arrive at a best call using the final classification (though we did lean heavily on the field crew's call whenever possible). Cognizant of the concern that the expertise possessed by team members would not be available to later users, we based our assessment solely on field materials that will be available—namely, the map class indicator species lists available in Nielsen et al. (2021c). Furthermore, all map production steps undertaken subsequent to the fieldwork—the creation of mapping associations, their crosswalk to map classes, changes in the modeling scheme, and (most importantly) post-processing and map editing—were done without reference to the AA field data, which were not fully processed until the final year of the project.

estimate of the true area occupied by each class in the park. These estimates were in turn used to adjust the overall map accuracy by area-weighting the per-class accuracies; this step was necessary to compensate for the stratified random sampling design that guided sample selection.

3.2. Sample design

We used a random sampling approach, stratified by mapped class, to select sample locations. We targeted all natural vegetated classes mapped within the park, including M92–BURNED WITH UNCERTAIN VEGETATION. We also targeted three nominally abiotic classes—R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE, R72–COLLUVIAL BARREN and R73–BEDROCK BARREN—because they often host a small amount of vegetation and have potential for confusion with several vegetated map classes, such as B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND, H62–ALPINE SPARSE HERBACEOUS VEGETATION and H58–BEDROCK BALDS AND SPARSELY VEGETATED FOREST OPENINGS. Accurate mapping of these abiotic classes is necessary in order to accurately map similar vegetated classes. All map classes were mapped on greater than 50 hectares, corresponding to a sampling goal of 30 plots each (Lea and Curtis 2010).

3.2.1. Inference area

The AA inference area was defined based on accessibility requirements. A primary sampling mask was created by buffering roads, trails, and common mountaineering routes by a fixed distance of 250 meters[68] and excluding barriers such as private property, impassible cliff bands, major streams, and slopes greater than 45 degrees. A secondary mask, to be used only for map classes that could not be adequately sampled within the primary mask, was created using a 500-meter buffer. A significant number of plots were collected outside the primary mask, so we treat the inference area as the region within the secondary mask. The resulting region (**Figure 17**) spanned 39,149 hectares, 14.2% of the total area of the complex.[69] Because crews were not able to visit the full area in which sample locations were generated, we will refer to the targeted sampling region as the *attempted inference area* (AIA).

[68] Restricting the sampling area to these corridors was necessary for safety and efficiency. Some segments of the trail network were not included. The single season available for sampling required that backpacking trips be planned for maximum efficiency, and some trail segments would have required too much effort to reach or did not fit well with other established sampling objectives.

[69] Although this falls well short of the 30% standard given in Lea and Curtis (2010), it was impossible to achieve sampling productivity goals with a larger inference area (see **Table 12** for a summary of the per-class inference area proportion). Given limited resources, we chose to prioritize the number of samples collected over expanding the inference area.

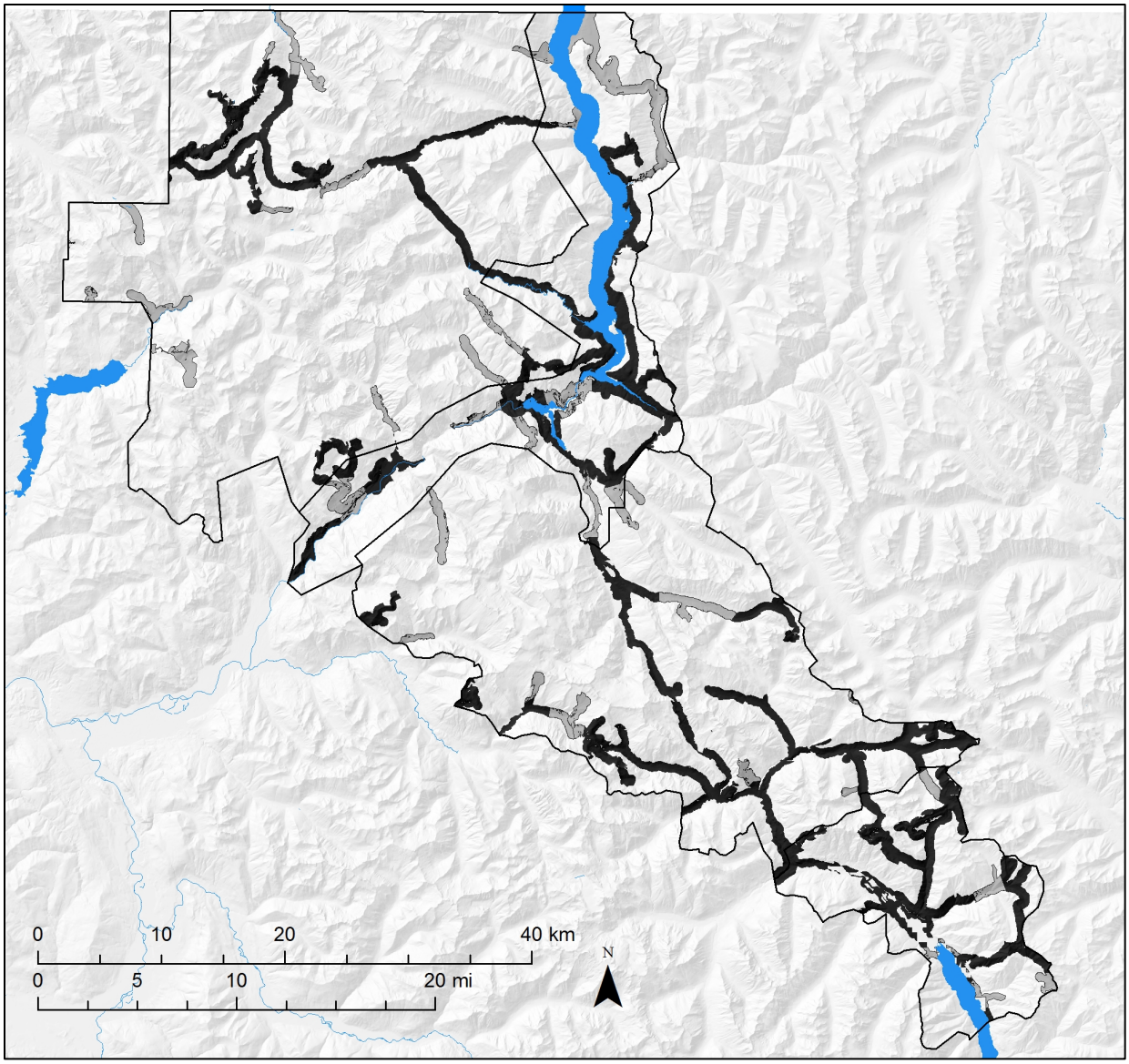


Figure 17. Accuracy assessment inference area. Field samples were acquired within the reached inference area, shown in black. The attempted inference area also includes gray-shaded areas which were not accessed by field crews.

3.2.2. Sample selection

We attempted to minimize map class membership ambiguity in the samples by targeting only points for which at least 75% of the 3-meter pixels in the surrounding 30 meters were identified as the target map class in the draft map. Stricter requirements (e.g., to allow for field and map positional error, see Lea and Curtis 2010) would have made achieving sampling goals for many of the map classes difficult or impossible. Because crew size limitations prevented independent observations for overlapping plots, and to minimize clumping of targets, we specified a minimum separation of 250 meters between targets, irrespective of mapped class.

We randomly selected 45 points within the AIA for each mapped class. 30 were designated as the initial sampling targets, and the other 15 points served as a reservoir for later target selection to replace rejected or missed plots. This allowed us to obtain a spatially representative collection of plots from each class—spreading samples across different field trips—without lowering the chances of obtaining an adequate number of samples for any.

3.3. Field data

3.3.1. Field logistics

NPS field crews collected field data from May through September 2016. Eight seasonal NPS employees worked in teams of two. As crews moved along roads and trails, they sampled every plot that the day's logistics and safety factors would allow. This increased the chance of reaching target map class totals without having to send crews back to areas of the park that can take multiple days to reach. After each seven day tour, a day was spent in the office entering a brief summary of each plot along the route: whether it was attempted, whether it was reached, and what the field map class call had been. This summary was used to assess the updated prognosis for achieving the sampling goal for each map class, based on the opportunities remaining in the target pool. Replacement plots for skipped targets were selected and high priority map classes were designated as those most in danger of falling short of sampling goals. Crew leaders planned their itineraries accordingly,[70] and missions planned for subsequent weeks were adjusted as some trips contained a greater proportion of high priority map classes.

3.3.2. Field protocol

Crews navigated to each target location and assessed the surroundings. If obvious structural vegetation transitions were within 30 meters, they moved the plot center to a more homogeneous point in the same vegetation type and updated the location using a GPS. If plot centers were not safely accessible and high confident assessments of plot location and vegetation call could be made, crews were permitted to make their observations from a distance. In this case, they noted the distance and bearing at which the plot lay from their observation point. Plots that could not be reached or assessed were discarded.

A 30-meter radius around the sample point was considered. If multiple distinct vegetation patches were present in that area, each was described separately. A full species list was compiled for the main patch around the center point. If other patches were present, the ten most abundant species in each were noted. The cover of each species was estimated to the nearest 1% cover. Species with cover of less than 1%, or present in the immediate vicinity and plausibly present within the plot, were documented as having trace cover. Overstory and understory tree species data were documented separately with a height of five meters used to discriminate them. The cover of abiotic types (e.g., water, colluvial rock, bedrock) was recorded if 10% or greater.

[70] Field crews were not knowledgeable of the class that each plot had been mapped to, to avoid influencing their judgment.

The map class with the best fit to each patch was noted, in addition to any plausible alternative calls.[71] Each call was assigned a confidence of high, medium or low. All patches were documented on a plot diagram, along with any nearby reference points, to provide additional confidence in assessing plot location. Slope, aspect and several other topographic variables were recorded. Photos were taken from the plot center in the four cardinal directions. Finally, a brief description of the plot and pros and cons for each of the map class calls were written. **Figure 18** contains a completed AA field form.

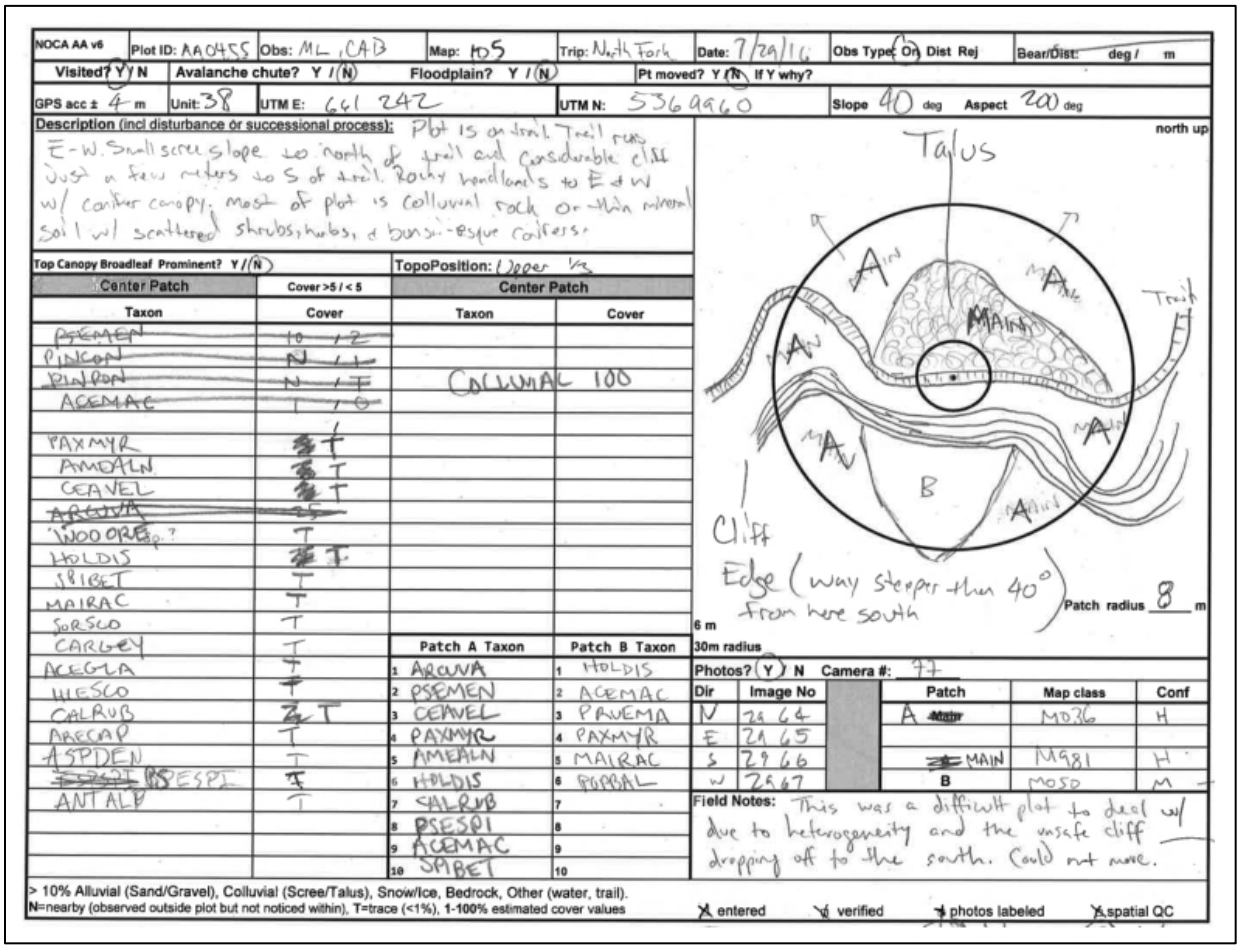


Figure 18. Completed accuracy assessment data sheet, collected in 2016.

[71] Descriptions of the nascent map classes, which later evolved into Nielsen et al. (2021c), were provided for this purpose.

3.3.3. Quality control

Accuracy assessment plot data went through a quality control process similar to that of map training data. The NPS phase of the spatial accuracy QC started after the field tour. At the conclusion of each tour, locations representing moved points were uploaded and reviewed for accuracy by a team member who had visited the plot. Subsequent INR QC verified and if necessary adjusted the location by comparing NAIP imagery and the field diagram.

Using the species cover data, plot description and diagram, and with reference to imagery and map class indicator species (see **Section 2.5.3**), each plot was checked for the correctness of the map class calls. Every effort was made to label the main patch for each plot with a single best map class call. However, there were two situations in which we chose to allow some flexibility. Sometimes, we were unable to make a single best call because the floristics field data were perfectly intermediate between two classes.[72] Both calls were treated as legitimate possible answers at the 79 plots where this occurred. In another 33 plots, the identified homogeneous patches were so small that we anticipated the possibility of a label mismatch when in fact the vegetation was correctly predicted. This might result from spatial offset between the GPS location, the layers by which the analyst assessed the plot, and the model-based prediction, as well as from filtering the predictions to MMU scale. If the boundary with another map class was within ten meters or less of the assessed point, we entered that map class as an additional correct answer. During the course of individual plot QC, if field-documented alternate patches were both clearly separable from the main patch and clearly assignable to a map class, we added them as additional AA data.

Throughout this process, the analyst did not have access to the final map polygons or labels. If a plot could not be confidently located or its correct map class could not be narrowed down to at most two possibilities, it was discarded. Twenty-one plots were rejected on these grounds, most of them because of uncertain location, extreme heterogeneity, or ambiguity due to the impact of disturbance.

3.3.4. Field plot totals and reached inference area

A total of 916 field plot samples were collected for the accuracy assessment; 895 of them passed the QC process. Fifty-six additional AA locations were added in alternate patches, for a total of 951 field plots. We determined the *reached inference area* (RIA) by removing from the AIA any route portions at least 1 km long and lacking sampled plots. The RIA totaled 28,560 hectares, equivalent to 73.0% of the attempted inference area or 10.4% of the full park. **Table 12** gives the overall effectiveness at reaching the accessible portions of each field-targeted map class and the class-specific inference area fraction. We fell well short of the standard of Lea and Curtis (2010), who specify that "a minimum of at least the most accessible 30th percentile of abundant classes should be included in even the most difficult of access situations." That was achieved for only seven of the 45 targeted map classes.

[72] Burned areas were also given latitude; either the best fit vegetated map class or M92–BURNED WITH UNCERTAIN VEGETATION were considered correct.

Table 12. Map class-specific accuracy assessment inference areas. For each class, the mapped area in (a) the park, (b) the attempted inference area (AIA) and (c) the reached inference area (RIA), followed by (d) the fraction reached of the area mapped in the AIA (a measure of field effectiveness), and (e) the fraction of the total mapped area represented in the RIA (a measure of the representativeness of the inference area).

Class code and abbreviated name	Mapped in park (ha)	Mapped in AIA (ha)	Mapped in RIA (ha)	% of AIA reached	% Rep in RIA
R73–Bedrock barren	34,899	530	439	82.9	1.3
C26–Conifer krummholz & treed cliff	7,167	225	167	74.5	2.3
R71–Alluvial barren	3,705	329	175	53.1	4.7
C21–Mtn hemlock & heather	18,832	1,304	953	73.1	5.1
R72–Colluvial barren	14,140	970	751	77.4	5.3
S43–Sitka alder	14,551	1,044	794	76.0	5.5
C12–Silver fir & Alaska blueberry	11,169	823	662	80.4	5.9
C13–Mtn hemlock & Cascade azalea	17,255	1,672	1,199	71.7	6.9
H60W–Black alpine sedge	295	29	21	71.0	7.0
H62–Alpine sparse herbaceous	4,850	525	347	66.1	7.2
S45–Vine maple	8,087	964	635	65.9	7.9
S48–Subalpine heather	5,495	633	458	72.5	8.3
S49–Alpine heather	4,837	566	404	71.4	8.4
H54–Moist talus vegetation	2,349	231	198	85.5	8.4
C11–Mesic silver fir & w hemlock	12,320	1,829	1,211	66.2	9.8
H58–Bedrock balds & forest openings	3,262	394	342	86.8	10.5
M92–Burned	8,788	1,210	935	77.3	10.6
H53–Showy sedge & Sitka valerian	1,521	193	168	87.1	11.0
C05–W hemlock & sword fern	15,358	3,404	1,762	51.8	11.5
C10–Moist silver fir & foamflower	7,306	1,761	1,050	59.6	14.4
C25–N Casc subalp fir & whitebark pine	3,891	648	590	91.0	15.2
S47–Successional huckleberry	2,692	481	410	85.2	15.2
C20–Subalp fir & Sitka valerian	6,507	1,312	1,091	83.1	16.8
H51W–Subalpine herbaceous wetland	298	58	54	92.2	18.1
C06–W hemlock & salal	3,783	1,461	691	47.3	18.3
C15–Lodgepole pine & Doug-fir	3,856	954	719	75.4	18.6
C16–N Casc Doug-fir & subalp fir	7,268	1,497	1,398	93.4	19.2
S42–Sitka willow riparian	1,223	323	238	73.8	19.5
B35–Upland paper birch & conifer forest	2,843	902	556	61.6	19.5
H57–Green fescue dry meadow	1,898	427	387	90.6	20.4
C22–Subalpine larch	3,548	826	741	89.7	20.9
S46–Snowbrush & Scouler's willow	3,771	892	790	88.6	20.9
S44–Thimbleberry, forbs & bracken fern	254	71	57	80.6	22.6

Table 12 (continued). Map class-specific accuracy assessment inference areas. For each class, the mapped area in (a) the park, (b) the attempted inference area (AIA) and (c) the reached inference area (RIA), followed by (d) the fraction reached of the area mapped in the AIA (a measure of field effectiveness), and (e) the fraction of the total mapped area represented in the RIA (a measure of the representativeness of the inference area).

Class code and abbreviated name	Mapped in park (ha)	Mapped in AIA (ha)	Mapped in RIA (ha)	% of AIA reached	% Rep in RIA
S41W–Subalpine willow wetland	243	63	59	93.1	24.2
B30–Successional gravel bar	314	124	84	67.2	26.6
C09–Ponderosa pine & Doug-fir	3,647	1,151	1,033	89.7	28.3
B31–Broadleaf riparian & swamp forest	2,025	872	601	68.9	29.7
C14–Silver fir & big huckleberry	1,754	701	525	74.9	29.9
C07–N Casc dry Doug-fir	5,055	2,229	1,681	75.4	33.3
S40W–Low elevation shrub wetland	282	117	97	82.7	34.3
B33–Upland bigleaf maple & conifer	1,275	642	443	69.1	34.8
H50W–Lowland marsh & meadow	158	74	57	77.6	36.1
C04–Moist w hemlock & foamflower	3,159	1,608	1,160	72.1	36.7
B34–Bigleaf maple & Doug-fir debris apron	2,094	870	845	97.1	40.4
H52–Cow parsnip	92	54	47	86.2	50.5

3.4. Photo-interpretation

An additional 36 randomly generated plots were assessed within the AIA for mapped abiotic classes that were not targeted for field sampling (W81–FRESH WATER and W82–EXPOSED SNOW AND ICE). These points were photo-interpreted to map class. Although not included in the AA contingency tables, these classes occupy substantial portions of the park and we felt it was important to have an estimate of their accuracy.

3.5. Sampling outcomes

A total of 987 sample plots passed the QC process or were photo-interpreted (**Figure 19**).

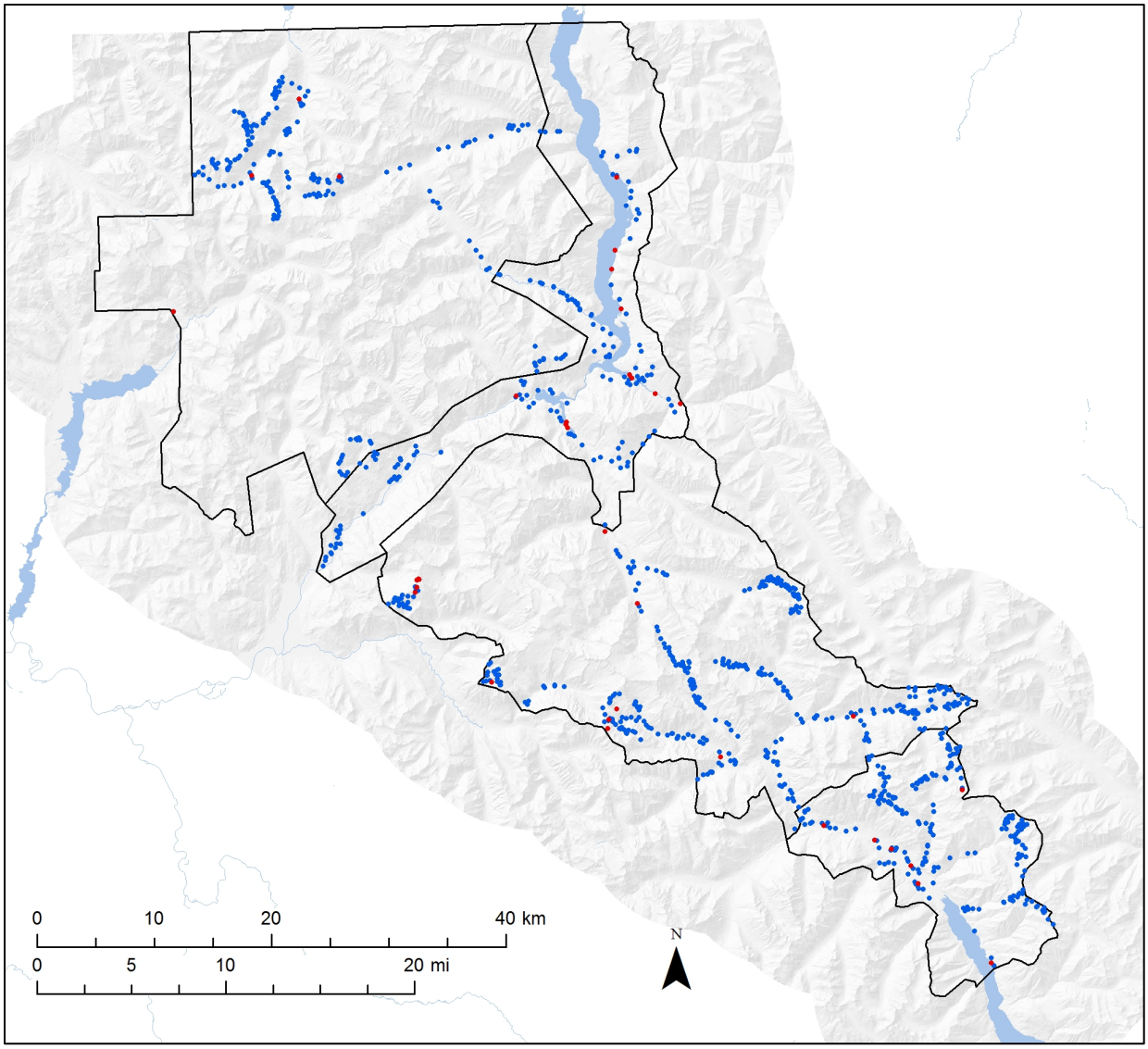


Figure 19. Accuracy assessment (AA) plot locations. 951 quality-controlled field-collected AA plots (blue dots) and 36 photo-interpreted points (red dots) are shown.

The final coordinates for each plot were used to extract the predicted map class label from the final map. **Table 13** gives, for each targeted map class, the mapped area (in hectares), the number of plots mapped as and identified as the class, and the fraction of the sampling goal that was achieved. Small numbers of mapped plots result in uncertain estimates of user's accuracy, whereas small numbers of identified plots result in uncertain estimates of producer's accuracy (and uncertain map area corrections for any classes confused with it).

Table 13. Accuracy assessment (AA) plot totals. For each class, the mapped area in the park, the number of AA plots mapped as and identified as the class, and the fraction of the sampling goal achieved (minimum based on mapped and identified plots). The sampling goal was 30 plots per class. Poorly sampled classes are listed first.

Class code and abbreviated name	Mapped area in park (ha)	# of plots mapped as	# of plots identified as	% of goal achieved
H52–Cow parsnip	92	2	3	7
H60W–Black alpine sedge	295	3	3	10
B30–Successional gravel bar	314	6	7	20
C06–W hemlock & salal	3,783	8	9	27
S40W–Low elevation shrub wetland	282	8	9	27
R71–Alluvial barren	3,705	10	8	27
C11–Mesic silver fir & w hemlock	12,320	11	13	37
B35–Upland paper birch & conifer forest	2,843	14	11	37
H50W–Lowland marsh & meadow	158	12	11	37
C05–W hemlock & sword fern	15,358	13	12	40
C12–Silver fir & Alaska blueberry	11,169	16	12	40
C14–Silver fir & big huckleberry	1,754	17	12	40
S44–Thimbleberry, forbs & bracken fern	254	12	14	40
C10–Moist silver fir & foamflower	7,306	13	18	43
H51W–Subalpine herbaceous wetland	298	18	13	43
R73–Bedrock barren	34,899	16	13	43
S41W–Subalpine willow wetland	243	16	15	50
S42–Sitka willow riparian	1,223	15	17	50
H58–Bedrock balds & forest openings	3,262	16	18	53
C07–N Casc dry Doug-fir	5,055	16	17	53
C25–N Casc subalp fir & whitebark pine	3,891	26	16	53
C26–Conifer krummholz & treed cliff	7,167	17	16	53
H54–Moist talus vegetation	2,349	27	16	53
B33–Upland bigleaf maple & conifer	1,275	19	20	63
M92–Burned	8,788	25	20	67
C04–Moist w hemlock & foamflower	3,159	23	21	70
S45–Vine maple	8,087	27	21	70
B31–Broadleaf riparian & swamp forest	2,025	25	22	73
C09–Ponderosa pine & Doug-fir	3,647	24	25	80
C15–Lodgepole pine & Doug-fir	3,856	24	24	80
C21–Mtn hemlock & heather	18,832	24	28	80
S47–Successional huckleberry	2,692	24	27	80
H62–Alpine sparse herbaceous	4,850	25	24	80
C13–Mtn hemlock & Cascade azalea	17,255	26	30	87

Table 13 (continued). Accuracy assessment (AA) plot totals. For each class, the mapped area in the park, the number of AA plots mapped as and identified as the class, and the fraction of the sampling goal achieved (minimum based on mapped and identified plots). The sampling goal was 30 plots per class. Poorly sampled classes are listed first.

Class code and abbreviated name	Mapped area in park (ha)	# of plots mapped as	# of plots identified as	% of goal achieved
B34–Bigleaf maple & Doug-fir debris apron	2,094	26	26	87
H57–Green fescue dry meadow	1,898	26	27	87
H53–Showy sedge & Sitka valerian	1,521	31	27	90
S48–Subalpine heather	5,495	28	43	93
S46–Snowbrush & Scouler's willow	3,771	29	33	97
C16–N Casc Doug-fir & subalp fir	7,268	30	36	100
S49–Alpine heather	4,837	32	31	103
C20–Subalp fir & Sitka valerian	6,507	38	32	107
C22–Subalpine larch	3,548	33	45	110
S43–Sitka alder	14,551	41	41	137
R72–Colluvial barren	14,140	49	55	163

The sampling goal of 30 samples per mapped class, needed for a confident assessment of UA, was achieved for only seven of the 45 classes included in the AA. Two classes (H52–COW PARSNIP MEADOW and H60W–BLACK ALPINE SEDGE WETLAND) had less than five mapped occurrences sampled. The success rate for identified plots, needed for a confident assessment of PA, was similar. Only nine classes were identified at least 30 times. The same two classes that were poorly sampled from the UA perspective were also poorly sampled for PA.

On the other hand, several classes were oversampled, although for only one was this drastic (R72–COLLUVIAL BARREN). Although the plan had been to remove plots corresponding to types that had been adequately sampled from the crew's target lists, there were difficulties in implementing the plan given limited office time between field tours. Crews themselves were not able to steer away from oversampled types because they were unaware of the map class label attached to each point.

3.6. Analysis

A total of 45 classes were included in the AA analysis, including all classes hosting natural vegetation communities. Three nominally abiotic classes (R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE, R72–COLLUVIAL BARREN and R73–BEDROCK BARREN) were also included; field sampling permitted a confident assessment of their accuracy and they often provide important habitat both for unmappably sparse plant communities and for wildlife populations. The two classes for which most AA samples were photo-interpreted[73] were excluded from further analysis. We noted that the 42 samples mapped as these classes indicated a user's accuracy of 97.6%, but felt that because

[73] W81–FRESH WATER and W82–EXPOSED SNOW AND ICE.

they were not field-sampled and generally do not host natural vegetation communities, their inclusion would artificially inflate the overall accuracy. Four classes mapped via PI or deductive modeling and unlikely to represent areas of high conservation significance[74] were also excluded from further analysis. Three field-sampled plots fell into one of these classes; all were correctly mapped. A total of 941 plots, all field-collected, remained in the analysis.

The predicted map class was extracted from the 3-meter pixel at each plot center point and compared to the field-sourced map class calls. For plots with two best calls, we allowed either as a correct answer; 29 of these 79 plots were counted as correct because the map matched the second of the two best calls (11 were mapped as M92–BURNED WITH UNCERTAIN VEGETATION). For plots with a secondary patch call within ten meters of the assessed center point, we also allowed that as a correct answer.[75] Of these 33 plots, 11 were called correct based on matching the secondary patch call. If the plot was counted as incorrect, it was labeled as the first best map class call from the primary patch.

The sample contingency table was created by indexing the observed map class against the predicted map class for each plot and summing across all plots. User's accuracy was calculated for each map class by dividing the number of correct samples by the total number of samples mapped as that class. Overall sample-based map accuracy was calculated by dividing the total number of correct calls by the total number of samples; however, this measure is misleading as it is biased by the use of the stratified random sampling design, which does not sample map classes in proportion to their prevalence in the project area (sample-based PA are similarly biased). The resulting SCT is shown in INR (2021a). The overall accuracy based on the raw samples is 84.5%.

To address the bias introduced by the stratified random sample design, a population contingency table (PCT) was created by reweighting the proportions represented by the cells in each row of the SCT by the fraction of the reached inference area mapped to that class. Each cell of the PCT, rather than containing raw sample counts, represents the estimated proportion of the RIA that is mapped as the class defined by the cell's row and identified as the class defined by the cell's column. We recalculated PA and overall accuracy based on the PCT; the revised measures represent the best estimates of the results that would have been obtained had the AA sample design been based on a simple random sample. The resulting PCT is shown in INR (2021a). The overall accuracy, after correcting for map class prevalence in the inference area, is 82.0%. Note that the mapped areas in the table sum to 27,022 hectares rather than the 28,560 hectares contained in the RIA. The six classes excluded from the analysis were mapped on the remaining portion of the RIA.

The kappa coefficient, which provides an accuracy measure that accounts for the probability of map class agreement by chance alone, was calculated. Kappa is a proportion ranging from 0–100%, where a value of zero indicates a map that is no more accurate than would be expected by chance alone. 90% confidence intervals were calculated for all accuracy estimates. Finally, a corrected area

[74] M93–TIMBERLAND WITH UNCERTAIN VEGETATION, M94–DEVELOPMENT, M95–ROADS IN PARK and M96–CLEARED CORRIDORS.

[75] This was intended to partially address nonthematic errors resulting from spatial misregistration (Foody 2002).

estimate was created for each map class by multiplying the sum of the proportions in each column by the total mapped area of all the map classes in the PCT.

The class-specific user's accuracies are summarized in **Table 14**. The accuracy estimate met the 80% standard for 34 of the 45 map classes. However, for only 15 was the 90% confidence interval entirely above the 80% mark. Many of the 11 that failed to meet the standard were poorly sampled. Only two classes, C25–NORTH CASCADES SUBALPINE FIR AND WHITEBARK PINE WOODLAND and H54–MOIST TALUS VEGETATION, were conclusively demonstrated to fail to meet the standard via the 90% confidence interval.

Table 14. Map class-specific user's accuracy (UA) for each assessed map class, with poorly mapped classes first. Asterisks indicate true values that are at least 90% confident to lie either fully above or below the 80% accuracy target.

Class code and full name	# of plots mapped as	UA estimate	UA 90% conf interval
C05–Western hemlock, Douglas-fir and sword fern forest	13	54%*	27–80%
C25–North Cascades subalpine fir and whitebark pine woodland	26	54%*	36–72%
H54–Moist talus vegetation	27	56%*	38–73%
C12–Silver fir, hemlock and Alaska blueberry forest	16	63%	39–86%
C14–Silver fir, big huckleberry and beargrass forest	17	65%	43–87%
C07–North Cascades dry Douglas-fir forest	16	69%	47–91%
B35–Upland paper birch and conifer forest	14	71%	48–95%
H51W–Subalpine herbaceous wetland	18	72%	52–92%
C20–Subalpine fir and Sitka valerian forest and woodland	38	74%	61–87%
H53–Showy sedge and Sitka valerian meadow	31	74%	60–89%
R71–Alluvial barren and debris-covered ice	10	80%	54–100%
S45–Vine maple shrubland	27	78%	63–93%
C11–Mesic silver fir and western hemlock forest	11	82%	58–100%
M92–Burned with uncertain vegetation	25	80%	65–95%
S44–Thimbleberry shrubland, tall forbs and bracken fern	12	83%	61–100%
C06–Western hemlock, Douglas-fir and salal forest	8	88%	62–100%
S49–Alpine heather shrubland	32	81%	68–94%
R73–Bedrock barren	16	81%	62–100%
C10–Moist silver fir, western hemlock and foamflower forest	13	85%	64–100%
C04–Moist western hemlock, Douglas-fir and foamflower forest	23	83%	67–98%
C16–North Cascades Douglas-fir and subalpine fir woodland	30	83%	70–96%
H62–Alpine sparse herbaceous vegetation	25	84%	70–98%

Table 14 (continued). Map class-specific user's accuracy (UA) for each assessed map class, with poorly mapped classes first. Asterisks indicate true values that are at least 90% confident to lie either fully above or below the 80% accuracy target.

Class code and full name	# of plots mapped as	UA estimate	UA 90% conf interval
B34–Bigleaf maple and Douglas-fir debris apron forest	26	85%	71–98%
H58–Bedrock balds and sparsely vegetated forest openings	16	88%	71–100%
S41W–Subalpine willow wetland	16	88%	71–100%
C15–Lodgepole pine and Douglas-fir woodland	24	88%	74–100%
H50W–Lowland marsh and meadow	12	92%	74–100%
H52–Cow parsnip meadow	2	100%	75–100%
B31–Broadleaf riparian and swamp forest	25	88%	75–100%
C13–Mountain hemlock, silver fir and Cascade azalea forest	26	88%	76–100%
S47–Successional huckleberry shrubland	24	92%*	80–100%
H57–Green fescue dry meadow	26	92%*	82–100%
C26–Conifer krummholz and treed cliff	17	94%*	82–100%
H60W–Black alpine sedge wetland	3	100%*	83–100%
B33–Upland red alder, bigleaf maple and conifer forest	19	95%*	84–100%
C22–Subalpine larch woodland	33	94%*	86–100%
C09–Ponderosa pine and Douglas-fir woodland	24	96%*	87–100%
S43–Sitka alder shrubland	41	95%*	88–100%
S48–Subalpine heather shrubland	28	96%*	89–100%
B30–Successional gravel bar shrubland	6	100%*	92–100%
S40W–Low elevation shrub-dominated wetland	8	100%*	94–100%
S42–Sitka willow riparian shrubland	15	100%*	97–100%
S46–Snowbrush and Scouler's willow shrubland	29	100%*	98–100%
C21–Mountain hemlock, subalpine fir and heather woodland	24	100%*	98–100%
R72–Colluvial barren	49	100%*	99–100%

The class-specific producer's accuracies, obtained from the PCT, are summarized in **Table 15**. The accuracy estimate met the 80% standard for 32 of the 45 map classes. However, for only 17 of them was the 90% confidence interval entirely above the 80% mark. Most of the 13 classes that failed to meet the standard were poorly sampled. Two classes, S44–THIMBLEBERRY SHRUBLAND, TALL FORBS AND BRACKEN FERN and C22–SUBALPINE LARCH WOODLAND, were conclusively demonstrated to fail to meet the standard.

Table 15. Map class-specific producer's accuracy (PA) for each assessed map class, with poorly mapped classes first. Figures are taken from population contingency table. Asterisks indicate true values that are at least 90% confident to lie either fully above or below the 80% accuracy target.

Class code and full name	# of plots identified as	PA estimate	PA 90% conf interval
S44–Thimbleberry shrubland, tall forbs and bracken fern	14	39%*	15–63%
C05–Western hemlock, Douglas-fir and sword fern forest	12	65%	44–87%
H58–Bedrock balds and sparsely vegetated forest openings	18	67%	47–87%
C22–Subalpine larch woodland	45	70%*	61–78%
C06–Western hemlock, Douglas-fir and salal forest	9	72%	43–100%
C11–Mesic silver fir and western hemlock forest	13	72%	50–93%
C15–Lodgepole pine and Douglas-fir woodland	24	72%	49–95%
C12–Silver fir, hemlock and Alaska blueberry forest	12	73%	44–100%
S48–Subalpine heather shrubland	43	73%	64–82%
C10–Moist silver fir, western hemlock and foamflower forest	18	74%	59–89%
C07–North Cascades dry Douglas-fir forest	17	74%	56–93%
S40W–Low elevation shrub-dominated wetland	9	80%	49–100%
C16–North Cascades Douglas-fir and subalpine fir woodland	36	76%	66–87%
H53–Showy sedge and Sitka valerian meadow	27	78%	59–96%
H52–Cow parsnip meadow	3	90%	58–100%
B34–Bigleaf maple and Douglas-fir debris apron forest	26	82%	67–96%
H54–Moist talus vegetation	16	86%	64–100%
R72–Colluvial barren	55	83%	72–93%
C13–Mountain hemlock, silver fir and Cascade azalea forest	30	83%	72–93%
S42–Sitka willow riparian shrubland	17	83%	65–100%
C14–Silver fir, big huckleberry and beargrass forest	12	88%	66–100%
S49–Alpine heather shrubland	31	83%	72–95%
C20–Subalpine fir and Sitka valerian forest and woodland	32	85%	72–97%
C04–Moist western hemlock, Douglas-fir and foamflower forest	21	86%	70–100%
B33–Upland red alder, bigleaf maple and conifer forest	20	87%	71–100%
C25–North Cascades subalpine fir and whitebark pine woodland	16	88%	72–100%
S47–Successional huckleberry shrubland	27	88%	76–100%
B30–Successional gravel bar shrubland	7	95%	79–100%
B35–Upland paper birch and conifer forest	11	94%*	81–100%
S46–Snowbrush and Scouler's willow shrubland	33	91%*	82–100%
C21–Mountain hemlock, subalpine fir and heather woodland	28	92%*	83–100%
S41W–Subalpine willow wetland	15	95%*	83–100%
H60W–Black alpine sedge wetland	3	100%*	83–100%
H62–Alpine sparse herbaceous vegetation	24	93%*	85–100%
C09–Ponderosa pine and Douglas-fir woodland	25	94%*	86–100%

Table 15 (continued). Map class-specific producer's accuracy (PA) for each assessed map class, with poorly mapped classes first. Figures are taken from population contingency table. Asterisks indicate true values that are at least 90% confident to lie either fully above or below the 80% accuracy target.

Class code and full name	# of plots identified as	PA estimate	PA 90% conf interval
H57–Green fescue dry meadow	27	94%*	87–100%
R71–Alluvial barren and debris-covered ice	8	100%*	94–100%
H50W–Lowland marsh and meadow	11	100%*	95–100%
S43–Sitka alder shrubland	41	99%*	96–100%
H51W–Subalpine herbaceous wetland	13	100%*	96–100%
R73–Bedrock barren	13	100%*	96–100%
C26–Conifer krummholz and treed cliff	16	100%*	97–100%
S45–Vine maple shrubland	21	100%*	98–100%
M92–Burned with uncertain vegetation	20	100%*	98–100%
B31–Broadleaf riparian and swamp forest	22	100%*	98–100%

Classes with accuracies less than the 80% target should be considered as candidates for merging with other classes (Lea and Curtis 2010). To assist with this task, **Table 16** lists these classes as well as the classes with which each is confused.

Table 16. Significantly confused map classes and the classes with which they are confused. For classes with user's or producer's accuracy less than 80%, all classes with which confusion exists are given with the proportion of the reached inference area (RIA) affected by confusion in either direction between the pair.

Class code and abbreviated name	Minimum (UA, PA)	Confused with (proportion of RIA affected)
S44–Thimbleberry, forbs & bracken fern	39%	S45–Vine maple (0.26%) S46–Snowbrush & Scouler's willow (0.04%) H53–Showy sedge & Sitka valerian (0.02%)
C05–W hemlock & sword fern	54%	C07–N Casc dry Doug-fir (2.56%) C11–Mesic silver fir & w hemlock (1.00%) C15–Lodgepole pine & Doug-fir (0.50%) C06–W hemlock & salal (0.50%) C10–Moist silver fir & foamflower (0.30%)
C25–N Casc subalp fir & whitebark pine	54%	C22–Subalpine larch (1.17%)
H54–Moist talus vegetation	56%	S48–Subalpine heather (0.16%) H62–Alpine sparse herbaceous (0.08%) R71–Alluvial barren (0.06%) S47–Successional huckleberry (0.03%) S49–Alpine heather (0.03%) C21–Mtn hemlock & heather (0.03%)

Table 16 (continued). Significantly confused map classes and the classes with which they are confused. For classes with user's or producer's accuracy less than 80%, all classes with which confusion exists are given with the proportion of the reached inference area (RIA) affected by confusion in either direction between the pair.

Class code and abbreviated name	Minimum (UA, PA)	Confused with (proportion of RIA affected)
C12–Silver fir & Alaska blueberry	63%	C11–Mesic silver fir & w hemlock (0.56%) C13–Mtn hemlock & Cascade azalea (0.48%) C21–Mtn hemlock & heather (0.15%) C16–N Casc Doug-fir & subalp fir (0.15%) C10–Moist silver fir & foamflower (0.15%)
C14–Silver fir & big huckleberry	65%	C16–N Casc Doug-fir & subalp fir (0.52%) C13–Mtn hemlock & Cascade azalea (0.11%) C20–Subalp fir & Sitka valerian (0.11%) C10–Moist silver fir & foamflower (0.11%)
H58–Bedrock balds & forest openings	67%	M92–Burned (0.55%) R72–Colluvial barren (0.08%) C15–Lodgepole pine & Doug-fir (0.08%)
C07–N Casc dry Doug-fir	69%	C05–W hemlock & sword fern (2.56%) C06–W hemlock & salal (0.39%) B35–Upland paper birch & conifer forest (0.29%) C15–Lodgepole pine & Doug-fir (0.11%) S45–Vine maple (0.09%)
C22–Subalpine larch	70%	C25–N Casc subalp fir & whitebark pine (1.17%) C20–Subalp fir & Sitka valerian (0.11%) S41W–Subalpine willow wetland (0.01%)
B35–Upland paper birch & conifer forest	71%	C07–N Casc dry Doug-fir (0.29%) B33–Upland bigleaf maple & conifer (0.23%) C04–Moist w hemlock & foamflower (0.15%)
C06–W hemlock & salal	72%	C05–W hemlock & sword fern (0.50%) C07–N Casc dry Doug-fir (0.39%) C15–Lodgepole pine & Doug-fir (0.32%)
C11–Mesic silver fir & w hemlock	72%	C05–W hemlock & sword fern (1.00%) C12–Silver fir & Alaska blueberry (0.56%) C04–Moist w hemlock & foamflower (0.41%) C10–Moist silver fir & foamflower (0.30%)
C15–Lodgepole pine & Doug-fir	72%	C05–W hemlock & sword fern (0.50%) C06–W hemlock & salal (0.32%) C09–Ponderosa pine & Doug-fir (0.22%) C07–N Casc dry Doug-fir (0.11%) H58–Bedrock balds & forest openings (0.08%)
H51W–Subalpine herbaceous wetland	72%	H53–Showy sedge & Sitka valerian (0.02%) H57–Green fescue dry meadow (0.01%) S48–Subalpine heather (0.01%) S41W–Subalpine willow wetland (0.01%)

Table 16 (continued). Significantly confused map classes and the classes with which they are confused. For classes with user's or producer's accuracy less than 80%, all classes with which confusion exists are given with the proportion of the reached inference area (RIA) affected by confusion in either direction between the pair.

Class code and abbreviated name	Minimum (UA, PA)	Confused with (proportion of RIA affected)
S48–Subalpine heather	73%	S49–Alpine heather (0.29%) H54–Moist talus vegetation (0.16%) S47–Successional huckleberry (0.13%) H62–Alpine sparse herbaceous (0.05%) H53–Showy sedge & Sitka valerian (0.02%) H51W–Subalpine herbaceous wetland (0.01%)
C20–Subalp fir & Sitka valerian	74%	C13–Mtn hemlock & Cascade azalea (0.49%) C16–N Casc Doug-fir & subalp fir (0.42%) C04–Moist w hemlock & foamflower (0.19%) C14–Silver fir & big huckleberry (0.11%) S47–Successional huckleberry (0.11%) C22–Subalpine larch (0.11%) C21–Mtn hemlock & heather (0.11%) S43–Sitka alder (0.07%)
C10–Moist silver fir & foamflower	74%	C04–Moist w hemlock & foamflower (0.37%) C16–N Casc Doug-fir & subalp fir (0.34%) C05–W hemlock & sword fern (0.30%) C11–Mesic silver fir & w hemlock (0.30%) C13–Mtn hemlock & Cascade azalea (0.17%) C12–Silver fir & Alaska blueberry (0.15%) C14–Silver fir & big huckleberry (0.11%)
H53–Showy sedge & Sitka valerian	74%	H57–Green fescue dry meadow (0.13%) S43–Sitka alder (0.04%) S47–Successional huckleberry (0.04%) H51W–Subalpine herbaceous wetland (0.02%) S44–Thimbleberry, forbs & bracken fern (0.02%) H52–Cow parsnip (0.02%) S48–Subalpine heather (0.02%)
C16–N Casc Doug-fir & subalp fir	76%	B34–Bigleaf maple & Doug-fir debris apron (0.59%) C14–Silver fir & big huckleberry (0.52%) C20–Subalp fir & Sitka valerian (0.42%) C10–Moist silver fir & foamflower (0.34%) C04–Moist w hemlock & foamflower (0.19%) C12–Silver fir & Alaska blueberry (0.15%)
S45–Vine maple	78%	S44–Thimbleberry, forbs & bracken fern (0.26%) S42–Sitka willow riparian (0.09%) B33–Upland bigleaf maple & conifer (0.09%) C07–N Casc dry Doug-fir (0.09%)

Sample and population contingency tables were also constructed at the lifeform/land-use level by lumping each map class into a category based on its dominant vegetation; the results are in INR (2021a) and are summarized below in **Table 17**. All AA plots that successfully passed through the QC process were used, including those of the nine map classes excluded from the map class level analysis (these fell into the last five categories in the table).

Table 17. Accuracy of map aggregated to lifeform/land-use level. Figures taken from population contingency table. Asterisks indicate true values that are at least 90% confident to lie either fully above or below the 80% accuracy target.

Lifeform	# of plots mapped as	UA estimate	UA 90% conf interval	# of plots identified as	PA estimate	PA 90% conf interval
Broadleaf tree	90	90%*	84–96%	86	93%*	87–99%
Conifer	359	99%*	98–100%	367	98%*	97–99%
Tall shrub	112	94%*	90–98%	112	96%*	91–99%
Shrubland	120	97%*	94–100%	139	83%*	75–88%
Herbaceous	160	85%*	80–90%	142	88%*	81–96%
Water	36	97%*	91–100%	35	100%*	99–100%
Rock	76	97%*	94–100%	76	97%*	92–100%
Snow & ice	6	100%*	92–100%	7	73%	34–100%
Developed	3	100%*	83–100%	3	100%*	83–100%
Other disturbed	25	80%	65–95%	20	100%*	98–100%

The user's and producer's accuracy estimates exceeded the 80% standard at 90% confidence for all natural land-use categories other than the **SHRUBLAND** producer's accuracy, which did not achieve that level of confidence. The most significant lifeform mapping errors among natural vegetation were **SHRUBLAND** mapping as **HERBACEOUS** (S48–SUBALPINE HEATHER SHRUBLAND mapping as H54–MOIST TALUS VEGETATION and S49–ALPINE HEATHER SHRUBLAND mapping as H62–ALPINE SPARSE HERBACEOUS VEGETATION), and **BROADLEAF TREE** mapping as **CONIFER** (B34–BIGLEAF MAPLE AND DOUGLAS-FIR DEBRIS APRON FOREST mapping as C16–NORTH CASCADES DOUGLAS-FIR AND SUBALPINE FIR WOODLAND and C09–PONDEROSA PINE AND DOUGLAS-FIR WOODLAND). The user's accuracy estimate for **OTHER DISTURBED**—which included M92–BURNED WITH UNCERTAIN VEGETATION and M93–TIMBERLAND WITH UNCERTAIN VEGETATION—was borderline at 80%, primarily due to H58–BEDROCK BALDS AND SPARSELY VEGETATED FOREST OPENINGS mapping as M92–BURNED WITH UNCERTAIN VEGETATION. The producer's accuracy estimate for **SNOW & ICE** was lower than 80%, though the sample size was small and it was not conclusively demonstrated to fail to meet the standard via the 90% confidence interval. The poor result was due to mapping of one W82–EXPOSED SNOW AND ICE plot as R72–COLLUVIAL BARREN.

3.7. Discussion

The accuracy assessment phase field observations are the reference by which the map's accuracy is measured, but these observations are not infallible. Key decisions in the field regarding the extent of the vegetation types perceived, the locations of boundaries between them, and the cover occupied by the species present within each can vary between observers. The vegetation classification itself (Nielsen et al. 2021c, Nielsen and Brunner 2021) is a somewhat subjective entity, with few hard rules for discriminating classes other than the weight of statistical evidence from ocular data, which are incomplete for many AA plots. Observers may also disagree about the degree to which a text-based map class description matches a vegetation patch in the field. To borrow a term from taxonomy, the circumscriptions of the map classes and mapping associations may not be consistently understood and applied. In some cases, the accuracy assessment plot quality control process will have corrected for these inconsistencies; in other cases, not.

The failure to meet inference area goals for most classes and the geographic bias toward areas of the park which we were able to reach limit the confidence we can attach to many of our conclusions. In these cases, Lea and Curtis (2010) warn that "extending the results of the thematic accuracy assessment from the inference area to the rest of the project must be justified by assumptions, rather than by statistical inference." In the following discussion, we have supplemented the AA analysis with photo-interpretation and consideration of context, in an attempt to provide additional evidence and to make these assumptions as transparent as possible.

3.7.1. Undersampled map classes

Based on their mapped area, the assessed classes each require 30 samples of mapped occurrences. As documented in **Table 13**, this was only achieved for seven of the 45 classes. Several causes for this failure are described below.

Eight of the 38 undersampled classes were rare (totaling less than 100 mapped hectares) in the reached inference area. These classes were typically concentrated in one or two parts of the RIA and even if the sampling goals had been achieved, autocorrelation amongst these plots in both floristics and mapping tendencies would have likely made their application to the full project area questionable.

Five other classes (C26–CONIFER KRUMMHOLZ AND TREED CLIFF, S46–SNOWBRUSH & SCOULER'S WILLOW, H54–MOIST TALUS VEGETATION, H62–ALPINE SPARSE HERBACEOUS VEGETATION and R73–BEDROCK BARREN) are very often located in inaccessible areas, either due to high elevations and steep slopes or to impenetrable vegetation. It is likely that field efficiency or safety concerns often resulted in crews skipping plots from these classes in order to maximize overall plot collection or stay on their itinerary.

Of the remaining 25 undersampled classes, six had fewer than 15 mapped occurrences in the AA dataset. One of these, B35–UPLAND PAPER BIRCH & CONIFER FOREST, had not been targeted, due to changes in the map classification made between the time of AA sampling in 2016 and production of the final maps in 2019–20. Two pairs of map classes—C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST and C06–WESTERN HEMLOCK, DOUGLAS-FIR AND SALAL FOREST, and C10–

MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST and C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST—had been split from one another between the time of AA sampling and production of the final map. The 30 samples targeted in each original map class were divided between the pairs of final classes. In addition, four plots targeted for the first pair wound up mapping as C07–NORTH CASCADES DRY DOUGLAS-FIR FOREST instead, and four targeted for the second pair mapped as C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST in the final map. The impacts of these classification changes, as well as five unreached plots in each of the original classes, resulted in the four final classes falling substantially short of sampling goals. Finally, an oversight resulted in the creation of only 15 targets for R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE, and several of these were impacted by channel migration between the draft and final map image dates, resulting in shifts to B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND or W81–FRESH WATER.

The above classification changes—and many others—were made in order to address challenges of mapping and field identification. While the goals of these updates were achieved, they had the by-product of reducing the number of AA plots available for assessing the final classes. An ideally executed project would have deferred AA fieldwork until completion of the final map—or at least map classification—but project management concerns took precedence here. For many map classes, the sample sizes are too small to confidently assess whether the 80% accuracy standard was achieved, as reflected in the wide confidence intervals seen in **Table 14** and **Table 15**. The small sample sizes should be kept in mind when considering the mapping error rates discussed below.

3.7.2. Map classes failing to meet accuracy standards

A list of the map classes failing to meet accuracy standards, the classes they are most confused with, and a possible corrective action that could be taken (if any) are shown in **Table 18**. Since every area of the map must be labeled, the only corrective action we consider is that of merging confused classes. This is likely to result in overall improvements only if the classes to be merged are confused primarily with each other. Otherwise, any poorly mapping area will simply get attributed into a different bin, perhaps bringing a different class below the accuracy target. We first review the apparent mapping errors for which merging classes does not appear to be an option.

Table 18. Map classes failing to meet accuracy standards or confused with those classes. '+' indicates accuracy estimates of 80% or higher; asterisks indicate accuracy less than 80% at 90% confidence. The classes accounting for the most mismapped area are listed under "confusion with," along with the fraction contributed to the total mismapped area in parentheses. A possible corrective action is noted for each.

Class code and abbreviated name	UA %	PA %	Confusion with	Corrective action
C04–Moist w hemlock & foamflower	+	+	C11 (31%), C10 (29%)	–
C05–W hemlock & sword fern	54	65	C07 (53%), C11 (21%)	consider merge with C07
C06–W hemlock & salal	+	72	C05 (41%), C07 (32%)	none, no reciprocity
C07–N Casc dry Doug-fir	69	74	C05 (74%)	consider merge with C05
C10–Moist silver fir & foamflower	+	74	C04 (21%), C16 (20%)	none, no reciprocity
C11–Mesic silver fir & w hemlock	+	72	C05 (44%), C12 (25%)	none, no reciprocity
C12–Silver fir & Alaska blueberry	63	73	C11 (37%), C13 (32%)	none, no reciprocity

Table 18 (continued). Map classes failing to meet accuracy standards or confused with those classes. '+' indicates accuracy estimates of 80% or higher; asterisks indicate accuracy less than 80% at 90% confidence. The classes accounting for the most mismapped area are listed under "confusion with," along with the fraction contributed to the total mismapped area in parentheses. A possible corrective action is noted for each.

Class code and abbreviated name	UA %	PA %	Confusion with	Corrective action
C13–Mtn hemlock & Cascade azalea	+	+	C20 (37%), C12 (36%)	–
C14–Silver fir & big huckleberry	65	+	C16 (60%)	none, no reciprocity
C15–Lodgepole pine & Doug-fir	+	72	C05 (41%), C06 (26%)	none, no reciprocity
C16–N Casc Doug-fir & subalp fir	+	76	B34 (26%), C14 (23%)	consider merge with B34
C20–Subalp fir & Sitka valerian	74	+	C13 (30%), C16 (26%)	none, no reciprocity
C22–Subalpine larch	+	70*	C25 (91%)	consider merge with C25
C25–N Casc subalp fir & whitebark pine	54*	+	C22 (100%)	consider merge with C22
B34–Bigleaf maple & Doug-fir debris apron	+	+	C16 (54%), S46 (22%)	consider merge with C16
B35–Upland paper birch & conifer forest	71	+	C07 (44%), B33 (35%)	none, no reciprocity
S44–Thimbleberry, forbs & bracken fern	+	39*	S45 (82%)	consider merge with S45
S45–Vine maple	78	+	S44 (50%)	consider merge with S44
S48–Subalpine heather	+	73	S49 (44%), H54 (24%)	consider merge with S49
S49–Alpine heather	+	+	S48 (56%), H62 (30%)	consider merge with S48
H51W–Subalpine herbaceous wetland	72	+	H53 (40%), H57 (20%)	none, no reciprocity
H53–Showy sedge & Sitka valerian	74	78	H57 (45%)	consider merge with H57
H54–Moist talus vegetation	56*	+	S48 (42%), H62 (21%)	none, no reciprocity
H57–Green fescue dry meadow	+	+	H53 (69%), S49 (25%)	consider merge with H53
H58–Bedrock balds & forest openings	+	67	M92 (78%)	consider merge with M92
M92–Burned	+	+	H58 (80%)	consider merge with H58

Map classes for which merging is not a viable option

No remedy is possible for the following apparent mapping errors, due to non-reciprocity of errors within the confused classes. The classes are considered in order of decreasing severity. Asterisks indicate estimates that are 90% confident to lie either above or below the 80% accuracy target; all other estimates given are not statistically significant with respect to the target. Recommendations to NPS are given in boldface.

H54–MOIST TALUS VEGETATION (UA 56%*, PA 86%) at least occasionally *maps over*[76] several other alpine and subalpine map classes—especially S48–SUBALPINE HEATHER SHRUBLAND and H62–ALPINE SPARSE HERBACEOUS VEGETATION—but the sample is geographically biased. Although H54 is mapped throughout the park, the AA plots all lay within a narrow range, and nearly all the incorrect

[76] In this terminology, "X *maps over* Y" means that class X was repeatedly mapped in locations where class Y was discovered on the ground. "Y *is mapped over by* X" would be an equivalent formulation.

samples were located in Fisher Creek Basin. The user's accuracy in the four other areas where the class was sampled was 87%. It seems likely that geographic bias in training data led to poor mapping in Fisher Creek Basin. ***We recommend NPS make additional effort to verify the mapping of moist, protected talus vegetation in other areas of the park.***[77]

C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST (UA 63%, PA 73%) has some *two-way confusion*[78] with C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST and C13–MOUNTAIN HEMLOCK, SILVER FIR AND CASCADE AZALEA FOREST, and also sometimes maps over several other montane and upper montane conifer forest types. There appears to be no geographic pattern to the error. This class occurs in intermediate settings, with the best call often influenced by spotty distributions of indicator tree species and understory variability in response to fine scale patterns of soil moisture. ***In floristically intermediate areas, we recommend that NPS regard mosaics of small mapped forest patches as indicating the proportions in which several vegetation types may be represented locally, rather than always seeking a fine-scale spatial correspondence.***

C14–SILVER FIR, BIG HUCKLEBERRY AND BEARGRASS FOREST (UA 65%, PA 88%) maps over C16–NORTH CASCADES DOUGLAS-FIR AND SUBALPINE FIR WOODLAND and several other somewhat moister upper montane classes. However, most of the errors are confined to a small area along the McAlester trail below Stiletto Peak, and in other places C16 maps over C14 instead. C14 is spottily distributed in the park, and the AA plots were confined to a very narrow geographic portion of its distribution. The C14 plots in the complex are good fits to the dry montane forest association *Abies amabilis-Pseudotsuga menziesii/Vaccinium membranaceum* and are floristically distinct from C16. However, it is possible that not enough plots are available to model the class well.

B35–UPLAND PAPER BIRCH AND CONIFER FOREST (UA 71%, PA 94%*) maps over C07–NORTH CASCADES DRY DOUGLAS-FIR FOREST on two plots separated by just 300 meters above Ruby Creek. B35 often occurs as broadleaf tree openings in C07 forest, and appears to have been mapped here where *Acer circinatum*, rather than *Betula papyrifera*, is the main broadleaf component in gaps between conifers. B35 also has some two-way confusion with B33–UPLAND RED ALDER, BIGLEAF MAPLE AND CONIFER FOREST above the lower Skagit River. While east-side occurrences of B35 are clearly distinct from B33, they intergrade on the west side. Although B33 is strongly associated with areas recovering from logging on the pre-dam Skagit floodplain, the toe slopes just above were not well sampled in our training data and there may be areas intermediate between B33 and B35 there.

C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST (UA 82%, PA 72%) is mapped over by C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST in two different locations. One of the plots may indicate a significant under-mapping of C11 near Little Beaver Creek as it is well lower than the current mapped range. These can be difficult classes as they are both frequently depauperate

[77] If the problem is found elsewhere, a revised map using additional training plots—using these AA plots and perhaps PI plots—would likely resolve the problem. INR could produce this easily.

[78] i.e., each of the types occasionally maps over the other.

and must often be distinguished mainly based on prominence of *Abies amabilis*, which can vary over fine spatial scales. C11 also has some general confusion with various moist montane map classes.

C06–WESTERN HEMLOCK, DOUGLAS-FIR AND SALAL FOREST (UA 88%, PA 72%) is mapped over by C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST and C07–NORTH CASCADES DRY DOUGLAS-FIR FOREST at one plot each, on the eastern edge of its mapped range. It may be generally under-mapped on the west side of Ross Lake. It is not possible to assess the western part of the mapped range of C06 due to lack of plots in that area.

C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND (UA 88%, PA 72%) may be slightly under-mapped, with one plot each mapped as C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST, C06–WESTERN HEMLOCK, DOUGLAS-FIR AND SALAL FOREST and H58–BEDROCK BALDS AND SPARSELY VEGETATED FOREST OPENINGS. The two former plots both had very low cover of *Pinus contorta*, and the latter was a very sparse woodland. All had C15 mapped in adjacent areas. C15 also mapped at two plots identified as C09–PONDEROSA PINE AND DOUGLAS-FIR WOODLAND on the east side of Ross Lake, where very little C09 was mapped. Only a small amount of *Pinus ponderosa* was present. All of these plots were intermediate in species composition.

H51W–SUBALPINE HERBACEOUS WETLAND (UA 72%, PA 100%*) appears to be over-mapped, based on errors in three distinct regions. Two H53–SHOWY SEDGE AND SITKA VALERIAN MEADOW plots, one km apart in the same drainage, were mapped as H51W, and a single plot each of two other upland subalpine map classes were also. One plot was floristically intermediate to H51W, and all the mismapped plots were near streams and/or wetland areas. It appears that inaccuracies in the available topographic data may have resulted in models that project H51W to occasionally extend somewhat further above drainages than it should.

C20–SUBALPINE FIR AND SITKA VALERIAN FOREST AND WOODLAND (UA 74%, PA 85%) appears to be over-mapped, particularly into difficult floristic areas that lack *Abies lasiocarpa*, but are not a clear fit to any other type. It maps over four plots called to C16–NORTH CASCADES DOUGLAS-FIR AND SUBALPINE FIR WOODLAND, three of them near lower McAlester Creek above Bridge Creek, at lower elevations than typical for C20. All these plots contain significant *Picea engelmannii* and no *Abies lasiocarpa*, and are generally not a good fit to any map class. C20 also maps over three plots called to C13–MOUNTAIN HEMLOCK, SILVER FIR AND CASCADE AZALEA FOREST. Two of these were on the western side of the C20 distribution around Cascade Pass and Park Creek Pass. Despite their lack of *Abies lasiocarpa*, their understories were intermediate to C20. Another C13 plot, south of McAlester Pass, also had significant *Picea engelmannii*. To a lesser extent, C20 is mapped over by several other map classes, in various areas. For the most part, the poor mapping results for C20 appear to reflect difficulties with the classification, particularly in areas with *Picea engelmannii* but lacking *Abies lasiocarpa*.

C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST (UA 85%, PA 74%) appears to be under-mapped, at least in NPSS. Of the eight plots where it was found but not mapped, six were in NPSS, including two in the upper Thunder Creek drainage, one on the North Fork Cascade River, two neighboring plots on North Fork Bridge Creek, and one on the main stem of Bridge Creek. The

under-mapping appears to be most notable in the Stehekin watershed, where C10 is mapped only in widely scattered small patches. These three plots mapped as C16–NORTH CASCADES DOUGLAS-FIR AND SUBALPINE FIR WOODLAND, C04–MOIST WESTERN HEMLOCK, DOUGLAS-FIR AND FOAMFLOWER FOREST, and C14–SILVER FIR, BIG HUCKLEBERRY AND BEARGRASS FOREST. The plots were all intermediate and atypical in character; nothing like them had been encountered during training data collection despite fairly dense sampling in the region. Although C10 is under-mapped here, the errors likely are restricted to very particular locales and don't affect a very large area.

Map classes for which merging may be a viable option

There are seven groups of map classes that could conceivably be aggregated for improved accuracy, based on this analysis. They are considered below in order of decreasing severity of the apparent mapping issue they would address. Asterisks indicate estimates that are 90% confident to lie either above or below the 80% accuracy target; all other estimates given are not statistically significant with respect to the target. Recommendations to NPS are given in boldface.

S44–THIMBLEBERRY SHRUBLAND, TALL FORBS AND BRACKEN FERN (UA 83%, PA 39%*) and S45–VINE MAPLE SHRUBLAND (UA 78%, PA 100%*) could be merged; the combined class S44+S45 would have UA 87% and PA 99%*. S44's low PA[79] is primarily due to confusion with S45 in the north part of the park. In that region, S44 typically occurs on toe slopes below S45, with an intermediate zone where they intermingle. Very few of these areas were accessible for sampling, so these conclusions were drawn entirely from plots along Little Beaver Creek and Brush Creek. Three heterogeneous and/or floristically ambiguous plots there were called to S44 but mapped as S45. In the southeast part of the park, areas upslope of S44 are usually S43–SITKA ALDER SHRUBLAND, S46–SNOWBRUSH AND SCOULER'S WILLOW SHRUBLAND or S47–SUCCESSIONAL HUCKLEBERRY SHRUBLAND. There, S44 is confused primarily with S46, not with S45. Although S44 does appear to be under-mapped in the north, the plots are fairly ambiguous and the very distinct set of neighboring map classes in the southern part of its distribution argue for keeping S44 and S45 distinct. ***We recommend the classes be kept separate, with the caveat that S44 appears to be under-mapped in NPSN, where it may be erroneously labeled as S45. We also recommend that NPS additionally investigate the mapping of toe-slope shrublands in the northwestern part of the park, many of which are difficult to access.***

C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST (UA 54%, PA 65%) and C07–NORTH CASCADES DRY DOUGLAS-FIR FOREST (UA 69%, PA 74%) could be merged; the combined class C05+C07 would have UA 83% and PA 93%*. Six plots have two-way confusion between these classes, although both classes are confused with other classes on five additional plots (C05 with C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST and others, C07 with B35–UPLAND PAPER BIRCH AND CONIFER FOREST and others). Of the six mislabeled plots, three are within a few meters of a mapped transition to the other type. Of the other three, two are intermediate in floristics. C07 is generally restricted to within 5–7 km of Ross Lake, while C05 is mapped across the park outside the

[79] Note that S44's PA from the sample contingency table is a much more reasonable 71%. The low population contingency table result is due to the much greater extent over which S45 is mapped.

Stehekin watershed, though it is also most abundant around Ross Lake. Near Ross Lake, they are generally mapped in large patches (500 meters or more across) segregated by aspect. ***We recommend the classes be kept separate, because the mislabeled plots don't provide strong evidence, and C05 is for the most part a west-side forest type, while C07 is a distinctly east-side type. The area around Ross Lake is transitional between the two, and some ambiguity is inevitable. In broad floristic transitional areas, we recommend that NPS regard alternating map patches of the same vegetation structure as a sign of floristics generally intermediate between the types.***

C25–NORTH CASCADES SUBALPINE FIR AND WHITEBARK PINE WOODLAND (UA 54%*, PA 88%) and C22–SUBALPINE LARCH WOODLAND (UA 94%*, PA 70%*) could be merged; the combined class C22+C25 would have UA 100%* and PA 98%*. Both classes are primarily confused with the other, and the AA plots are not ambiguous. While the general habitat of the two classes is very similar and they frequently occur in close proximity, C25 is mapped on warmer and drier aspects, on steeper slopes, and at slightly higher elevations than C22. In several areas—south of Stiletto Peak, on Rainbow Ridge and around Rainbow Lake, around Rennie Peak—C25 maps over C22, perhaps extending farther downslope, into lusher vegetation on gentler terrain, than it should. Considering the fairly strong model which discriminates the classes (well-balanced, with 11.9% average error and 42 predictors), the poor AA result is strange. The training plots for C22 tended to have lusher understories than many of the AA plots where it was found but mapped as C25. It may also be that the environmental niche varies with geography, as most of the incorrectly mapped AA plots were located a half-kilometer or farther from training plots. It is possible that the model simply wasn't trained sufficiently. ***NPS may want to merge these classes, given the large improvement in accuracy it would allow. However, we have left them as is, because*** *Larix lyallii* ***may be of management interest, the classes have clear floristic differences, and we believe the AA overestimates the amount of confusion and that the mapping quality varies in different areas.***[80]

H58–BEDROCK BALDS AND SPARSELY VEGETATED FOREST OPENINGS (UA 88%, PA 67%) and M92–BURNED WITH UNCERTAIN VEGETATION (UA 80%, PA 100%*) could be merged; the combined class H58+M92 would have UA 93%* and PA 100%*. Both classes are primarily confused with the other. All four plots are at balds that have burned forests surrounding them. In all cases H58 was modeled, but the area was recoded to M92 during post-processing. The errors resulted in the comparatively low spatial resolution of the satellite-based change detection process that was used to detect burned areas; the change detected in the surrounding burned forests bled over into the balds where little actual change had occurred. This is a particularly challenging error because many of the balds at the park occur in a matrix with forests that are subject to frequent fire. ***We recommend the classes be kept separate. Although unburned balds are certainly frequently mapped as burned forests within recent fire perimeters, that is not a reason to give up on mapping balds throughout the park.***

S48–SUBALPINE HEATHER SHRUBLAND (UA 96%*, PA 73%) and S49–ALPINE HEATHER SHRUBLAND (UA 81%, PA 83%) could be merged; the combined class S48+S49 would have UA 98%* and PA

[80] Because the poor results in several areas reflect an inadequacy of training data in those areas, a revised map using these AA plots as additional training would likely result in substantial improvements. INR could produce this easily.

85%. S49 maps over S48 on five plots, and the reverse happens on one plot. While this represents a slight majority of the problem plots for S49, S48 is equally confused with H54–MOIST TALUS VEGETATION, and is occasionally mapped over by other classes as well. The confused S48/S49 plots are not floristically ambiguous, but all but one are within 20 meters of a mapped transition to the correct type. ***Given the significantly different types that each class is confused with, and their usefulness in delineating the subalpine/alpine transition, we recommend the classes be kept separate.***

H53–SHOWY SEDGE AND SITKA VALERIAN MEADOW (UA 74%, PA 78%) and H57–GREEN FESCUE DRY MEADOW (UA 92%*, PA 94%*) could be merged; the combined class H53+H57 would have UA 88%* and PA 96%*. Three plots are mislabeled between this pair; this is only a quarter of the problem plots for H53. One of them is floristically ambiguous, and two are within 20 meters of a mapped transition to the correct type. These classes can be intermeshed at fine scale depending on soil texture and slight changes in drainage pattern, though modal occurrences are quite distinct. ***We recommend the classes be kept separate.***

C16–NORTH CASCADES DOUGLAS-FIR AND SUBALPINE FIR WOODLAND (UA 83%, PA 76%) and B34–BIGLEAF MAPLE AND DOUGLAS-FIR DEBRIS APRON FOREST (UA 85%, PA 82%) could be merged; the combined class C16+B34 would have UA 91%* and PA 85%. Four plots are mislabeled between this pair; this is only a quarter of the problem plots for C16. Of the four, two are floristically ambiguous, and one of the others is 25 meters from a mapped transition to the correct type. There does not appear to be much to be gained by merging these classes, whose modal occurrences are quite distinct. ***We recommend the classes be kept separate.***

3.7.3. Other known mapping issues

We observed several other mapping issues that will likely be noticed from time to time. ***Further investigation of these issues by NPS may be warranted.***

H62–ALPINE SPARSE HERBACEOUS VEGETATION was mapped at three plots that were field-identified as S49–ALPINE HEATHER SHRUBLAND. Each of the plots was within ten meters of denser S49 that was mapped correctly. It appears that this error is fairly common in sparser occurrences of S49 (which varies widely in density and appearance). This is an understandable mistake as *Phyllodoce empetriformis*, *Cassiope mertensiana* and *Luetkea pectinata* are among the most dominant plants in both S49 and H62. Apart from a few indicators that are often absent, H62 is generally distinguished from S49 by the sparseness of its vegetation. It does appear that the map may begin to show H62 at a higher level of vegetation cover than would be ideal.

R73–BEDROCK BARREN was mapped at three plots that were field-identified as R72–COLLUVIAL BARREN. Many areas of talus at NOCA have essentially weathered in place and occur in geomorphological settings more similar to bedrock than typical talus slopes below cliffs. This may be the primary source of this error, in addition to possible inconsistency of discriminating the classes in training data.

4. Vegetation of North Cascades National Park Complex

4.1. Vegetation map

The vegetation map (**Figure 20**, and a higher-resolution version at Nielsen et al. 2021d) illustrates the distribution of the 51 map classes across the park complex and surrounding buffer. The map contains nearly 470 million pixels aggregated into patches of no less than nine 3-meter pixels (81 m^2). The estimated area of each class, based on its mapped area modified by the correction factor from the AA population contingency table, is shown in **Table 19**. The map classes vary widely in abundance, with most of them limited in extent. The most common eight classes collectively occupy half the complex, while 21 classes cover less than 1% each. M93–TIMBERLAND WITH UNCERTAIN VEGETATION is absent in the park, though it is present immediately adjacent in several areas; it is not shown in the table.

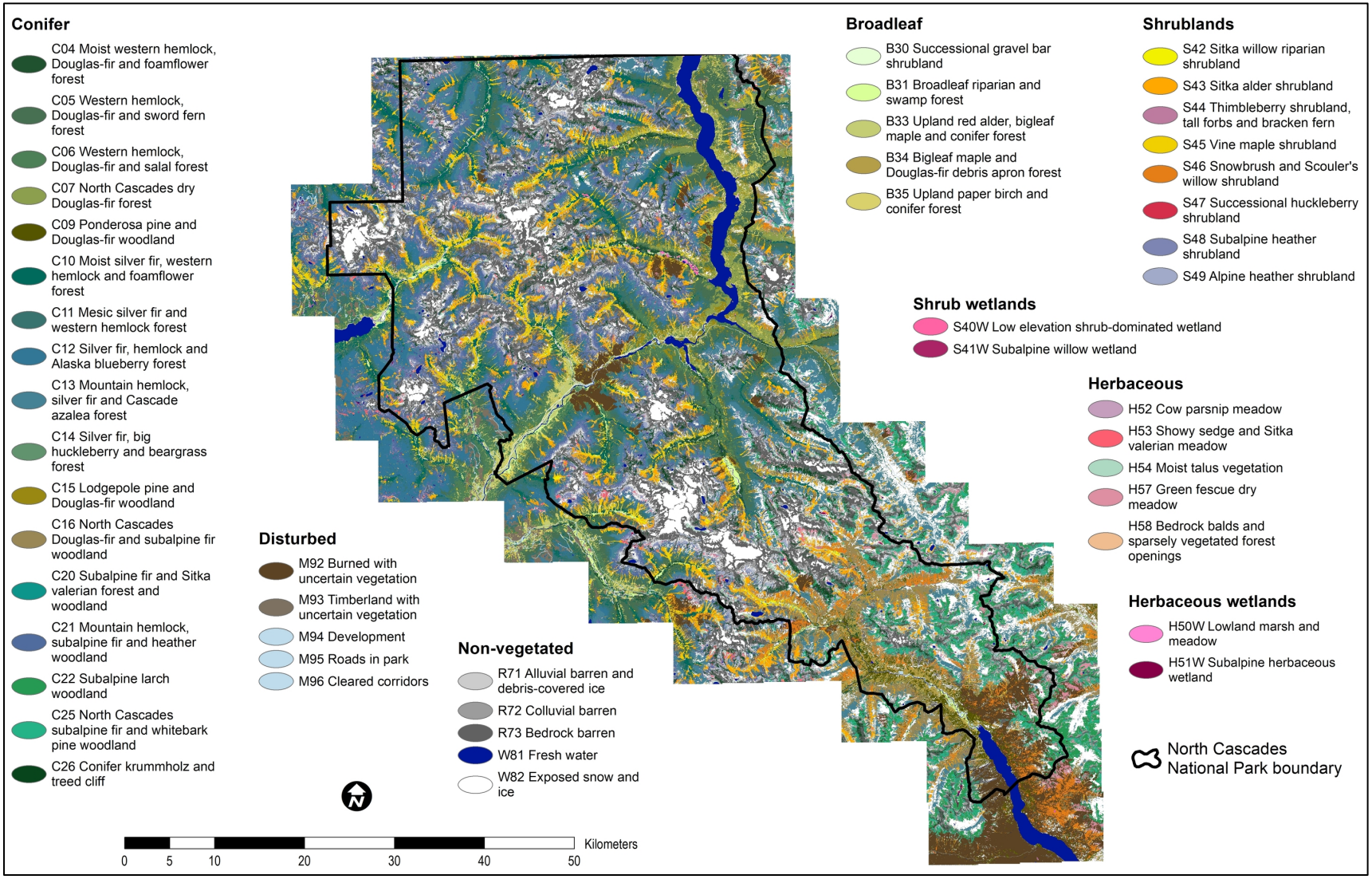


Figure 20. Vegetation map of North Cascades National Park Complex.

Table 19. Map class estimated area and proportion of park complex, listed by area.

Class code and full name	Est area in complex (ha)	Est area in complex (ac)	Proportion of complex (%)
R73–Bedrock barren	31,627	78,152	11.46
C21–Mountain hemlock, subalpine fir and heather woodland	20,210	49,940	7.32
R72–Colluvial barren	18,337	45,312	6.65
C13–Mountain hemlock, silver fir and Cascade azalea forest	17,795	43,972	6.45
S43–Sitka alder shrubland	13,940	34,446	5.05
C11–Mesic silver fir and western hemlock forest	13,633	33,688	4.94
C05–Western hemlock, Douglas-fir and sword fern forest	11,868	29,326	4.30
C12–Silver fir, hemlock and Alaska blueberry forest	9,114	22,521	3.30
W82–Exposed snow and ice	9,007	22,257	3.26
C10–Moist silver fir, western hemlock and foamflower forest	8,687	21,466	3.15
W81–Fresh water	8,128	20,085	2.95
C16–North Cascades Douglas-fir and subalpine fir woodland	7,876	19,462	2.85
M92–Burned with uncertain vegetation	7,030	17,371	2.55
S48–Subalpine heather shrubland	6,888	17,021	2.50
C26–Conifer krummholz and treed cliff	6,746	16,670	2.44
S45–Vine maple shrubland	6,443	15,921	2.33
C20–Subalpine fir and Sitka valerian forest and woodland	6,063	14,982	2.20
C07–North Cascades dry Douglas-fir forest	5,523	13,648	2.00
C22–Subalpine larch woodland	5,315	13,134	1.93
C06–Western hemlock, Douglas-fir and salal forest	4,808	11,881	1.74
C15–Lodgepole pine and Douglas-fir woodland	4,641	11,468	1.68
S49–Alpine heather shrubland	4,603	11,374	1.67
H62–Alpine sparse herbaceous vegetation	4,527	11,186	1.64
H58–Bedrock balds and sparsely vegetated forest openings	4,260	10,527	1.54
S46–Snowbrush and Scouler's willow shrubland	3,974	9,820	1.44
C09–Ponderosa pine and Douglas-fir woodland	3,816	9,430	1.38
C04–Moist western hemlock, Douglas-fir and foamflower forest	3,373	8,335	1.22
R71–Alluvial barren and debris-covered ice	2,964	7,324	1.07
S47–Successional huckleberry shrubland	2,839	7,015	1.03
B34–Bigleaf maple and Douglas-fir debris apron forest	2,489	6,150	0.90
C25–North Cascades subalpine fir and whitebark pine woodland	2,310	5,708	0.84
B35–Upland paper birch and conifer forest	2,098	5,184	0.76
H57–Green fescue dry meadow	1,969	4,865	0.71
H54–Moist talus vegetation	1,848	4,567	0.67
B31–Broadleaf riparian and swamp forest	1,782	4,403	0.65
B33–Upland red alder, bigleaf maple and conifer forest	1,711	4,228	0.62

Table 19 (continued). Map class estimated area and proportion of park complex, listed by area.

Class code and full name	Est area in complex (ha)	Est area in complex (ac)	Proportion of complex (%)
S42–Sitka willow riparian shrubland	1,603	3,961	0.58
C14–Silver fir, big huckleberry and beargrass forest	1,377	3,403	0.50
H53–Showy sedge and Sitka valerian meadow	1,308	3,232	0.47
S44–Thimbleberry shrubland, tall forbs and bracken fern	1,006	2,486	0.36
S40W–Low elevation shrub-dominated wetland	363	897	0.13
M95–Roads in park	332	820	0.12
B30–Successional gravel bar shrubland	328	811	0.12
H60W–Black alpine sedge wetland	295	729	0.11
M96–Cleared corridors	238	588	0.09
S41W–Subalpine willow wetland	230	568	0.08
H51W–Subalpine herbaceous wetland	215	531	0.08
H50W–Lowland marsh and meadow	145	358	0.05
H52–Cow parsnip meadow	141	348	0.05
M94–Development	119	294	0.04

4.2. Vegetation overview

The map classes can be broadly broken into ten groups based on their dominant lifeform and land-use characteristics: (a) conifer-dominated, (b) broadleaf tree-dominated, (c) tall shrublands, (d) shrublands and dwarf-shrublands, (e) herbaceous vegetation, (f) rock barrens, (g) exposed snow and ice, (h) water, (i) natural and semi-natural disturbed landscapes (including burned and logged areas), and (j) development. A map made by merging map classes into these groups is shown in **Figure 21**, and the relative abundance of map classes within each group is illustrated by **Figure 22**.

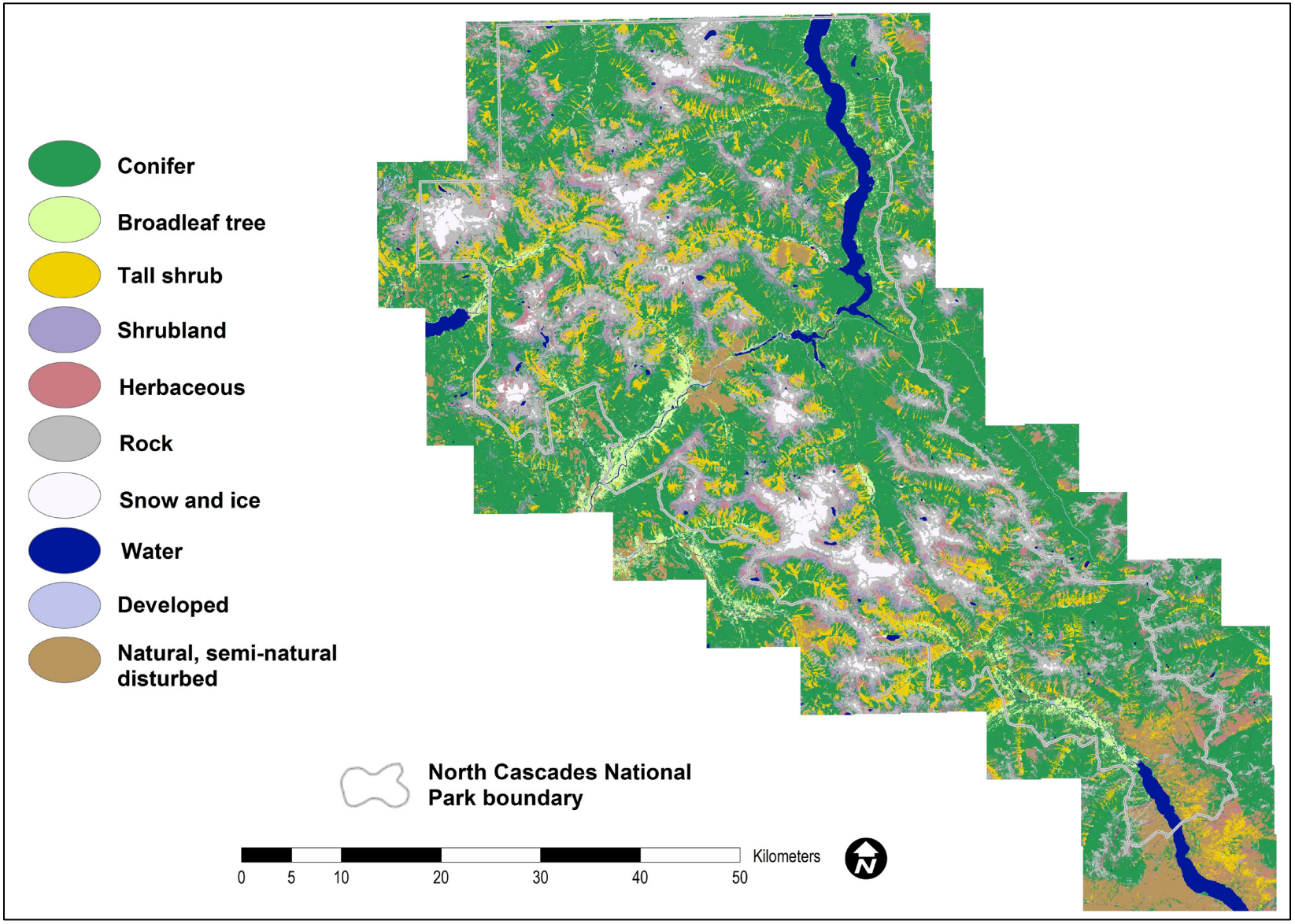


Figure 21. Lifeform map of North Cascades National Park Complex.

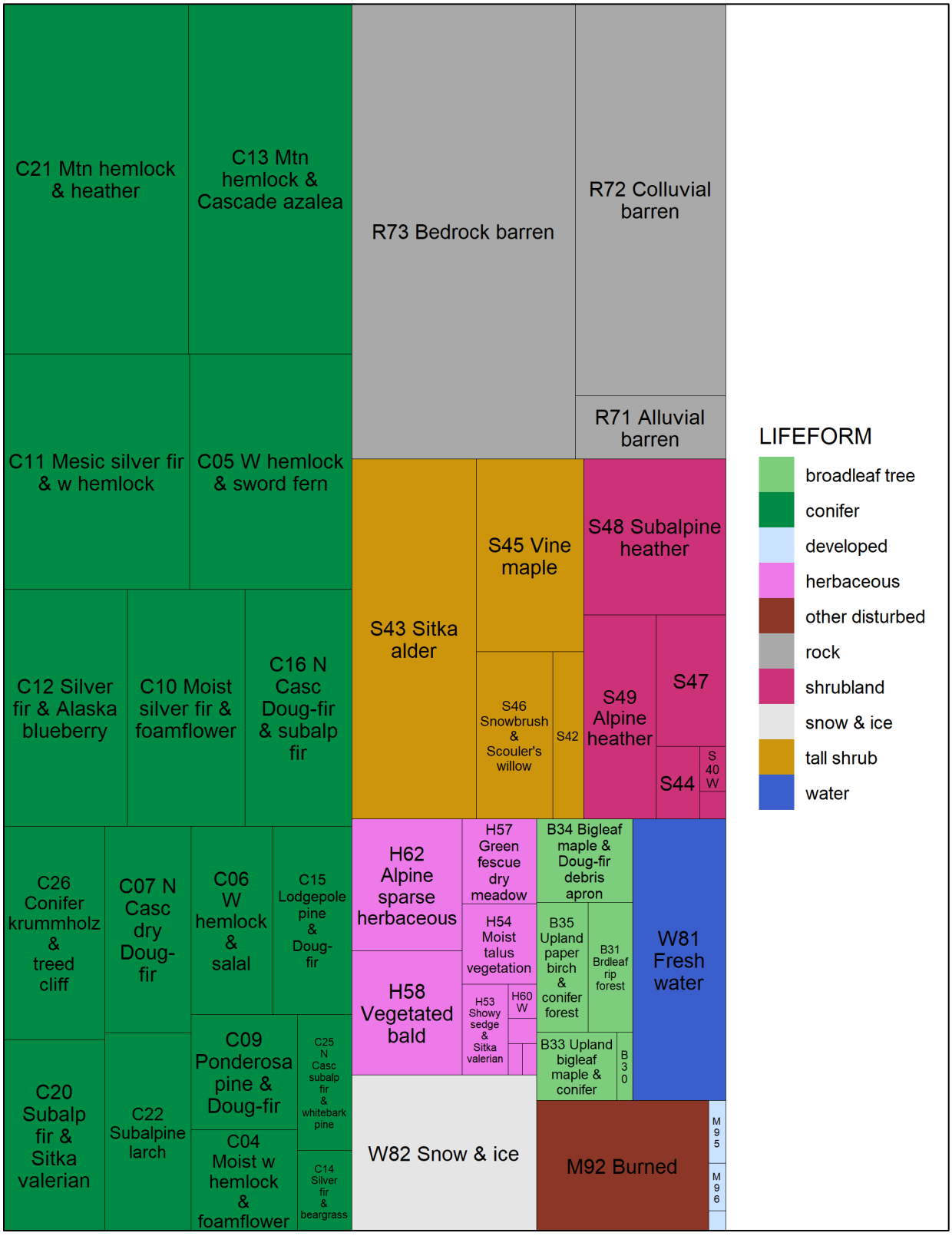


Figure 22. Relative abundance of map classes, grouped by lifeform/land-use category.

Analysis of the distribution patterns of plant species provides a more fundamental (but also more detailed and complex) way of understanding vegetation patterns in the park. **Table 20** lists the most common species documented in the training plots, as well as other species that are key components of the mapped vegetation classes. Complete floristics tables are provided in INR (2021b).

Table 20. Common species in plots and other important species discussed in the text. Frequency in full-ocular training plots is given and ranked relative to all other species. Elevation zones (L=lowland, LM=lower montane, UM=upper montane, S=subalpine, A=alpine) with which species are most associated are marked with an 'X.' Bullets '•' indicate zones of less common but still notable occurrence. Zones where the species is not appreciably present are indicated by '–.' Scientific names follow Hitchcock and Cronquist (2018); species are listed alphabetically. INR (2021b) has a complete list.

Scientific name	Common name	Frequency	Rank	L	LM	UM	S	A
Abies amabilis	silver fir	33.0%	4	•	X	X	•	–
Abies grandis	grand fir	4.7%	121	X	–	–	–	–
Abies lasiocarpa	subalpine fir	32.5%	5	–	•	X	X	X
Acer circinatum	vine maple	24.9%	8	X	•	–	–	–
Acer glabrum	Rocky Mountain maple	14.6%	27	X	X	•	–	–
Acer macrophyllum	bigleaf maple	11.1%	49	X	–	–	–	–
Alnus rubra	red alder	5.8%	100	X	–	–	–	–
Alnus viridis	Sitka alder	9.9%	76	•	X	X	–	–
Amelanchier alnifolia	western serviceberry	21.2%	13	X	X	X	–	–
Arctostaphylos nevadensis	pinemat manzanita	6.8%	88	X	X	X	•	–
Arnica latifolia	broad-leaved arnica	14.5%	28	–	–	X	X	–
Athyrium distentifolium	alpine lady fern	4.9%	118	–	–	X	–	X
Athyrium filix-femina	lady fern	14.4%	30	X	X	–	–	–
Berberis nervosa	dwarf Oregon-grape	17.0%	20	X	•	–	–	–
Betula papyrifera	paper birch	3.3%	160	X	–	–	–	–
Calamagrostis rubescens	pinegrass	14.4%	31	X	X	X	–	–
Callitropsis nootkatensis	Alaska-cedar	15.5%	22	–	•	X	X	•
Carex aquatilis	water sedge	0.7%	310	X	–	–	X	–
Carex nigricans	black alpine sedge	6.1%	97	–	–	–	X	X
Carex spectabilis	showy sedge	9.5%	68	–	–	–	X	X
Cassiope mertensiana	white mountain-heather	13.8%	35	–	–	–	•	X
Ceanothus velutinus	snowbrush	7.4%	87	X	X	–	–	–
Chamaenerion angustifolium	fireweed	16.6%	21	•	X	X	X	–
Chimaphila umbellata	pipsissewa	18.6%	17	X	X	X	–	–
Clintonia uniflora	queen's cup	15.0%	25	•	X	X	–	–
Cornus stolonifera, C. occidentalis	red-osier dogwood	6.2%	96	X	•	–	–	–
Cryptogramma acrostichoides	American parsley fern	9.6%	67	X	•	X	•	X
Eremogone capillaris	mountain sandwort	11.2%	48	–	–	•	X	X

Table 20 (continued). Common species in plots and other important species discussed in the text. Frequency in full-ocular training plots is given and ranked relative to all other species. Elevation zones (L=lowland, LM=lower montane, UM=upper montane, S=subalpine, A=alpine) with which species are most associated are marked with an 'X.' Bullets '•' indicate zones of less common but still notable occurrence. Zones where the species is not appreciably present are indicated by '–.' Scientific names follow Hitchcock and Cronquist (2018); species are listed alphabetically. INR (2021b) has a complete list.

Scientific name	Common name	Frequency	Rank	L	LM	UM	S	A
Festuca viridula	green fescue	10.9%	51	–	–	•	X	–
Gaultheria shallon	salal	5.1%	115	X	•	–	–	–
Goodyera oblongifolia	rattlesnake-plantain	22.3%	12	X	X	X	–	–
Heracleum maximum	cow parsnip	4.1%	132	–	X	–	•	–
Larix lyallii	subalpine larch	6.3%	92	–	–	–	X	–
Linnaea borealis	twinflower	15.4%	24	X	X	•	–	–
Lomatium brandegeei	Brandegee's biscuitroot	7.6%	81	–	–	X	X	–
Luetkea pectinata	partridgefoot	18.8%	16	–	–	•	X	X
Lupinus latifolius	subalpine lupine	20.3%	14	–	–	•	X	•
Luzula piperi	Piper's woodrush	4.8%	120	–	–	–	–	X
Micranthes tolmiei	Tolmie's saxifrage	4.2%	131	–	–	–	–	X
Orthilia secunda	one-sided wintergreen	20.3%	15	•	X	X	–	–
Paxistima myrsinites	Oregon-box	40.0%	3	X	X	X	X	•
Phlox diffusa	spreading phlox	10.1%	59	–	–	•	X	X
Phyllodoce empetriformis	pink mountain-heather	23.2%	9	–	–	•	X	X
Picea engelmannii	Engelmann spruce	11.6%	43	–	•	X	X	–
Pinus albicaulis	whitebark pine	9.2%	70	–	–	•	X	X
Pinus contorta	lodgepole pine	7.5%	84	•	X	•	–	–
Pinus monticola	western white pine	8.1%	76	X	X	X	–	–
Pinus ponderosa	ponderosa pine	5.6%	104	X	–	–	–	–
Populus trichocarpa	black cottonwood	5.2%	112	X	–	–	–	–
Potentilla flabellifolia	fan-leaf cinquefoil	6.2%	94	–	–	–	X	•
Prunus emarginata	bitter cherry	5.3%	110	X	X	•	–	–
Pseudotsuga menziesii	Douglas-fir	41.9%	2	X	X	•	–	–
Rubus leucodermis	whitebark raspberry	3.8%	139	X	•	–	–	–
Rubus nutkanus	thimbleberry	14.9%	26	X	X	•	–	–
Rubus spectabilis	salmonberry	11.5%	45	X	X	–	–	–
Salix commutata	undergreen willow	1.7%	218	–	–	–	X	–
Salix scouleriana	fire willow	12.2%	41	•	X	X	–	–
Salix sitchensis	Sitka willow	3.8%	139	X	X	–	–	–
Sambucus racemosa	red elderberry	10.9%	50	X	X	–	–	–

Table 20 (continued). Common species in plots and other important species discussed in the text. Frequency in full-ocular training plots is given and ranked relative to all other species. Elevation zones (L=lowland, LM=lower montane, UM=upper montane, S=subalpine, A=alpine) with which species are most associated are marked with an 'X.' Bullets '•' indicate zones of less common but still notable occurrence. Zones where the species is not appreciably present are indicated by '–.' Scientific names follow Hitchcock and Cronquist (2018); species are listed alphabetically. INR (2021b) has a complete list.

Scientific name	Common name	Frequency	Rank	L	LM	UM	S	A
Senecio integerrimus	tall western groundsel	6.3%	93	–	–	•	X	–
Sorbus sitchensis	Sitka mountain-ash	18.5%	18	–	–	X	X	–
Spiraea douglasii	rose spirea	1.2%	181	X	X	–	–	–
Spiraea lucida	white spirea	14.5%	29	X	X	X	–	–
Spiraea splendens	mountain spirea	2.8%	169	–	–	X	X	–
Struthiopteris spicant	deer fern	3.8%	137	X	X	•	–	–
Taxus brevifolia	Pacific yew	8.6%	73	X	X	–	–	–
Thuja plicata	western redcedar	26.5%	7	X	X	–	–	–
Tiarella trifoliata	foamflower	13.9%	34	X	X	X	–	–
Tsuga heterophylla	western hemlock	28.8%	6	X	X	•	–	–
Tsuga mertensiana	mountain hemlock	22.7%	10	–	–	X	X	•
Vaccinium deliciosum	Cascade blueberry	22.4%	11	–	–	•	X	X
Vaccinium membranaceum	big huckleberry	42.1%	1	•	•	X	X	–
Vaccinium ovalifolium	Alaska blueberry	13.8%	36	•	X	X	–	–
Vaccinium parvifolium	red huckleberry	15.4%	23	X	X	–	–	–
Valeriana sitchensis	Sitka valerian	17.5%	19	–	–	X	X	–

The distribution patterns of the most important species and map classes within each of the lifeform/land-use categories are discussed below.[81] For purposes of discussion, we treat natural abiotic areas (including rock barrens, snow and ice, and water) as a single unit, and we treat wetlands as a separate unit, despite lumping them by their dominant lifeform in the lifeform map. As above, area estimates are based on the mapped area modified by the correction factor from the AA population contingency table. Species occurrence frequencies are relative to the full-ocular training plot dataset.

4.2.1. Conifers

The dry summers and relatively warm winters of the Pacific Northwest favor the development of coniferous forest as the climax lifeform where local conditions permit. Conifer-dominated vegetation is the most abundant lifeform in the complex, covering nearly half the landscape, accounting for one-third of the map classes, and ranging from stately low elevation forests to snow-sculpted krummholz clinging to rocky ridges. The classes are evenly distributed across elevation zones: lowland forests

[81] The map class descriptions (Nielsen et al. 2021c) contain greater detail about the species composition, habitat and distribution associated with each map class.

cover 11% of the complex, lower montane forests occupy 10%, upper montane forests 13%, and subalpine woodlands 12%. Krummholz and treed cliffs occupy only 2.5%. The seventeen species of conifers are supplanted by other vegetation only where disturbance, snowpack, saturated soils or lack of soil development prevent their establishment and persistence.

West side

Lowland forests

The pioneer species Douglas-fir (*Pseudotsuga menziesii*) and the shade-tolerant western hemlock (*Tsuga heterophylla*) dominate west-side lowland conifer forests. Douglas-fir is found in most parts of the park; it is the most abundant tree in the plot dataset and the second most abundant plant overall. Given enough light availability for establishment, it can thrive at all but the highest elevations and is absent only at the wettest sites. Western hemlock is the major successional tree in west-side lowland forests and extends up through the montane zone. These species, with lesser amounts of western redcedar (*Thuja plicata*), contribute most of the overstory to the map classes C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST, C06–WESTERN HEMLOCK, DOUGLAS-FIR AND SALAL FOREST and, in moister settings, C04–MOIST WESTERN HEMLOCK, DOUGLAS-FIR AND FOAMFLOWER FOREST. Other lowland conifer forest associates, none of which are normally more than prominent, include western white pine (*Pinus monticola*), grand fir (*Abies grandis*) and the small understory tree Pacific yew (*Taxus brevifolia*). Dwarf Oregon-grape (*Berberis nervosa*) and red huckleberry (*Vaccinium parvifolium*) are prominent in the understory of most lowland conifer stands except in the moistest settings, where foamflower (*Tiarella trifoliata*) is omnipresent in a lush mixed understory. Dense successional forests often have very sparse understories, sometimes limited to scattered rattlesnake plantain (*Goodyera oblongifolia*) and pipsissewa (*Chimaphila umbellata*). Vancouverian species such as salal (*Gaultheria shallon*) and deer fern (*Struthiopteris spicant*) are significantly less common at NOCA than in the other NCCN parks.

Montane forests

The lower montane zone is characterized by codominance of silver fir (*Abies amabilis*) with western hemlock. Entry to this zone is marked by increased canopy prominence of silver fir, the primary successional species throughout all but the driest of mid-elevation forests. C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST occurs on valley bottoms and moist toe slopes, while higher and drier slopes are occupied by C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST. Alaska blueberry (*Vaccinium ovalifolium*) is a dominant understory species in many lower montane stands and is especially abundant in wet-mesic settings where it is usually joined by foamflower and queen's cup (*Clintonia uniflora*). Mesic mid-slope stands are often composed of dense silver fir with sparse understories of scattered one-sided wintergreen (*Orthilia secunda*).

Upper montane forests are transitional to the subalpine zone above. Here, the high elevation species mountain hemlock (*Tsuga mertensiana*), subalpine fir (*Abies lasiocarpa*) and Alaska-cedar (*Callitropsis nootkatensis*) become prominent in closed forests dominated by silver fir. The lush C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST is transitional from the lower montane, with high elevation tree species usually subordinate to silver fir, western hemlock and/or Douglas-fir. At higher elevations, it gives way to the very abundant C13–MOUNTAIN HEMLOCK, SILVER FIR AND

CASCADE AZALEA FOREST. Alaska-cedar is regularly prominent and occasionally codominant in the west-side upper montane, showing a preference for wetter forests, especially those regenerated after avalanche disturbance. The most common species in the park, big huckleberry (*Vaccinium membranaceum*) is the dominant understory species in the upper montane zone, and is equally prominent on the west and east sides.

Subalpine forests and woodlands

The most abundant conifer class, C21–MOUNTAIN HEMLOCK, SUBALPINE FIR AND HEATHER WOODLAND, signals the entry to the subalpine zone upslope of C13–MOUNTAIN HEMLOCK, SILVER FIR AND CASCADE AZALEA FOREST. It is extensive on upper slopes and shoulders of west-side mountain ridges. At yet higher elevations, heavy snow accumulation limits the growth of conifers other than C26–CONIFER KRUMMHOLZ AND TREED CLIFF in exposed locations where wind sweeps it clear.

East side

Lowland forests

Patterns differ east of the Cascade crest, though one constant is Douglas-fir, which is perhaps even more prominent on the east side. As a drought-tolerant pioneer species, it has a particular advantage here where fire return intervals are shorter. C07–NORTH CASCADES DRY DOUGLAS-FIR FOREST is abundant around Ross Lake, especially on south-facing lower slopes. Ponderosa pine (*Pinus ponderosa*) is often codominant with Douglas-fir at the low elevations in the Stehekin Valley, where C09–PONDEROSA PINE AND DOUGLAS-FIR WOODLAND is the most abundant conifer type. Understories on the east side include many components of the more continental East Cascades and Rocky Mountain floras that are rare or absent on the west side and at other NCCN parks. Widespread examples in lowland forests include white spirea (*Spiraea lucida*) and pinegrass (*Calamagrostis rubescens*).

Montane forests

The lower montane zone is essentially eliminated in the southeast, squeezed between lowland forests and the upper montane zone. However, dry bedrock benches and ridgelines in this elevation zone above the middle Skagit and around Ross Lake often host lodgepole pine (*Pinus contorta*) in C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND. Otherwise, lower montane forests around Ross Lake are generally similar to west-side types.

The upper montane zone is reached quickly on the east side, as shade-intolerant subalpine fir extends to lower elevations here. Although C16–NORTH CASCADES DOUGLAS-FIR AND SUBALPINE FIR WOODLAND shares importance with the west-side mountain hemlock map classes around Ross Lake, it becomes the primary upper montane type in the Stehekin, where mountain hemlock becomes limited to particularly moist sites. The third most common species in the park, Oregon-box (*Paxistima myrsinites*), is a dominant understory species across many east-side conifer forest types, and also is prominent in a variety of higher elevation shrublands and meadows.

Subalpine forests and woodlands

Subalpine fir is by far the dominant species in east-side subalpine forests and woodlands, although it is very prominent in drier west-side subalpine woodlands as well. The most widespread east-side

subalpine class is C20–SUBALPINE FIR AND SITKA VALERIAN FOREST AND WOODLAND, which is found on middle to upper slopes at all but the most sheltered north-facing sites. Dry woodlands above that—especially on steep south- and west-facing slopes in the southeast—are dominated by whitebark pine (*Pinus albicaulis*) and subalpine larch (*Larix lyallii*). Subalpine larch is found on somewhat moister sites in C22–SUBALPINE LARCH WOODLAND, while whitebark pine is found in the driest and most exposed sites where C25–NORTH CASCADES SUBALPINE FIR AND WHITEBARK PINE WOODLAND prevails amongst rock outcrops. Engelmann spruce (*Picea engelmannii*) is a frequent but rarely dominant associate found throughout the east-side upper montane and subalpine zones. Continental flora elements in the understories of east-side subalpine forests and woodlands include Brandegee's biscuitroot (*Lomatium brandegeei*) and tall western groundsel (*Senecio integerrimus*).

4.2.2. Broadleaf trees

Communities dominated by deciduous broadleaf trees are a distinguishing feature of disturbed areas and occupy three percent of the complex. They occur on valley bottom floodplains, in sites recovering from fire or logging, and on toe slopes habitually impacted by mass movement.

Floodplains

The colonization phase of floodplain successional dynamics is represented by B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND. The dominant woody plants in these communities are red alder (*Alnus rubra*), black cottonwood (*Populus trichocarpa*), and Sitka willow (*Salix sitchensis*). Red alder is increasingly rare on the east side; riparian broadleaf stands along the Stehekin are usually dominated by black cottonwood. Without repeated disturbance, these successional shrublands mature into B31–BROADLEAF RIPARIAN AND SWAMP FOREST. Salmonberry (*Rubus spectabilis*), red elderberry (*Sambucus racemosa*) and lady fern (*Athyrium filix-femina*) are common and abundant understory components in these forests.

Uplands

Bigleaf maple (*Acer macrophyllum*) is the most common deciduous species in the complex and is present in lower elevation upland areas throughout. It is a dominant in two broadleaf forest classes. B33–UPLAND RED ALDER, BIGLEAF MAPLE AND CONIFER FOREST often represents recovery from logging disturbance and is also common on west-side debris aprons. It is present on lower slopes, toe slopes, and in the logged and hydrologically modified floodplain of the Skagit River below the Gorge Dam. B34–BIGLEAF MAPLE AND DOUGLAS-FIR DEBRIS APRON FOREST is the east-side analogue and is especially abundant on toe slopes along the Stehekin River and its tributaries. Paper birch (*Betula papyrifera*) is found on sites that have been disturbed by fire or mass movement, and is especially abundant around Ross Lake. B35–UPLAND PAPER BIRCH AND CONIFER FOREST captures these mixed broadleaf-conifer patches, which are often embedded in C07–NORTH CASCADES DRY DOUGLAS-FIR FOREST.

4.2.3. Upland tall shrubs

Tall shrublands in uplands cover nine percent of the complex, occupying avalanche tracks, montane talus slopes, riparian benches, toe-slope debris aprons, and east-side slopes recovering from fire.

Avalanches and mass movement

Avalanches tend to impact the same slopes year after year, carving out chutes through montane and subalpine forests. The regular disturbances favor resilient and rapidly resprouting shrubs rather than tall and brittle trees. Sitka alder (*Alnus viridis*) bends rather than breaks when walloped by snow, and thrives in avalanche zones and moist talus. Although the species is not extremely common in the dataset, it is generally dominant where it occurs, and S43–SITKA ALDER SHRUBLAND is the most abundant non-forest vegetation type in the complex. Red elderberry, salmonberry and lady fern are common associates.

Other tall shrubland classes occur on avalanche toe slopes, where Sitka alder gives way to a variety of other shrub species. Often found just above large streams and rivers, S42–SITKA WILLOW RIPARIAN SHRUBLAND forms dense thickets that are codominated by Sitka willow, Sitka alder and red-osier dogwood (*Cornus sericea*) with a mix of moisture-loving forbs from lowlands and middle elevations. Drier talus and avalanche toe slopes host S45–VINE MAPLE SHRUBLAND, in which vine maple (*Acer circinatum*) is often joined by young bigleaf maple and a variety of smaller shrubs and herbaceous plants. These sites have strong floristic similarities to B33–UPLAND RED ALDER, BIGLEAF MAPLE AND CONIFER FOREST. Vine maple is also very common in the understory of lowland conifer forests.

Fire recovery

Fire willow (*Salix scouleriana*), snowbrush (*Ceanothus velutinus*) and bitter cherry (*Prunus emarginata*) form considerable patches where they have resprouted after fire. S46–SNOWBRUSH AND SCOULER'S WILLOW SHRUBLAND covers east-side midslopes for years following intense fire, gradually transitioning into forested patches which reflect the spatial distribution of disturbance. Western serviceberry (*Amelanchier alnifolia*) frequently joins these species and also ranges widely across other dry woodlands and shrublands, usually at low cover.

4.2.4. Upland shrublands

Shorter shrublands in uplands cover nearly six percent of the complex, exploiting disturbance in the montane zone and exposed areas with poor soil development at higher elevations. The lowest elevation shrublands are on lower montane toe slopes with thimbleberry (*Rubus nutkanus*) and an assortment of tall forbs; these are mapped as S44–THIMBLEBERRY SHRUBLAND, TALL FORBS AND BRACKEN FERN. At higher elevations, above the elevation range of S46–SNOWBRUSH AND SCOULER'S WILLOW SHRUBLAND, areas dominated by a mix of big huckleberry, Cascade blueberry (*Vaccinium deliciosum*), mountain spirea (*Spiraea splendens*) and Sitka mountain-ash (*Sorbus sitchensis*) are scattered through the upper montane and subalpine zones. These successional shrublands, mapped as S47–SUCCESSIONAL HUCKLEBERRY SHRUBLAND, typically occur between forests and subalpine meadows, and are often associated with recovery from fire.

The iconic mountain-heather shrublands of the subalpine and alpine zones are dominated by pink mountain-heather (*Phyllodoce empetriformis*), white mountain-heather (*Cassiope mertensiana*) and Cascade blueberry. These are particularly prominent on the shoulders of west-side ridges. S48–SUBALPINE HEATHER SHRUBLAND is found just above C21–MOUNTAIN HEMLOCK, SUBALPINE FIR AND HEATHER WOODLAND, but there is a substantial transition zone in which the two classes mosaic at a scale of 10–20 meters, resulting in an extensive landscape of tree islands and heather openings.

Succession toward mountain hemlock is usually evident in these areas. A gradual transition to S49–ALPINE HEATHER SHRUBLAND is evident with increasing elevation and exposure. This class is characterized by more compact vegetation with considerably fewer forbs, increased abundance of white mountain-heather, and reduced abundance of Cascade blueberry.

4.2.5. Upland herbaceous vegetation

Lingering snowpack, low temperatures, desiccating winds and repeated disturbance create conditions under which only herbaceous plants can survive. A diverse assortment of upland herbaceous plant communities—including lush forb meadows, rocky graminoid meadows and sparse alpine cushion plants—share five percent of the complex. Herbaceous communities often transition across short distances, responding to finer-scale changes in topography, substrate and soil moisture than adjacent forests.

Lower elevations

The lowest elevation herbaceous communities in the complex occur in forest openings on bedrock-limited soils and are mapped as H58–BEDROCK BALDS AND SPARSELY VEGETATED FOREST OPENINGS. They are found throughout the complex but are especially abundant in the Stehekin Valley and around Ross Lake. The abundant east-side balds are distinct from those elsewhere in the NCCN; although we have generally treated the map class as herbaceous, here they are usually dominated by dwarf shrubs, especially pinemat manzanita (*Arctostaphylos nevadensis*). Openings in lower montane forests may also occur on toe slopes, where herbaceous communities with cow parsnip (*Heracleum maximum*) and other tall forbs persist due to habitual disturbance from landslides and avalanche runouts. These meadows, H52–COW PARSNIP MEADOW, are limited in extent and usually occur in a mosaic with S44–THIMBLEBERRY SHRUBLAND, TALL FORBS AND BRACKEN FERN, discussed above.

Higher elevations

Lush meadows of subalpine lupine (*Lupinus latifolius*), Sitka valerian (*Valeriana sitchensis*), showy sedge (*Carex spectabilis*) and other herbaceous species are mapped as H53–SHOWY SEDGE & SITKA VALERIAN MEADOW. Found throughout the park but more abundant on the moist west side, they are often found in and around C20–SUBALPINE FIR AND SITKA VALERIAN FOREST AND WOODLAND. In rocky areas, this community frequently grades into H54–MOIST TALUS VEGETATION, which typically occurs as patches of alpine lady fern (*Athyrium distentifolium*) and diverse forbs within protected and often sub-irrigated portions of talus slopes. Exclusively alpine plants such as Piper's woodrush (*Luzula piperi*) and Tolmie's saxifrage (*Micranthes tolmiei*) are joined by partridgefoot (*Luetkea pectinata*), which ranges through a variety of subalpine and alpine communities, to form H62–ALPINE SPARSE HERBACEOUS VEGETATION. This is the most abundant herbaceous map class in the complex—though again more abundant on the west side—and is found on rocky upper slopes where snow lingers late into the growing season.

Smaller amounts of snow accumulation on the east side limit the extent of the preceding classes, though H54–MOIST TALUS VEGETATION still takes advantage of moist pockets within talus. The dominant high-elevation herbaceous class here is H57–GREEN FESCUE DRY MEADOW, which occurs in large expanses on south-facing slopes and is also found in fine mosaics with C22–SUBALPINE LARCH WOODLAND and C25–NORTH CASCADES SUBALPINE FIR AND WHITEBARK PINE WOODLAND. These dry

meadows are characterized by green fescue (*Festuca viridula*), mountain sandwort (*Eremogone capillaris*), and spreading phlox (*Phlox diffusa*).

4.2.6. Wetlands

One-half of one percent of the complex is mapped as one of the five wetland map classes. Low-gradient areas conducive to wetland formation are mostly limited to lowland valleys and high-elevation headwaters basins, often where glaciers have carved out flat-bottomed cirques. There are riparian communities and occasional mid-slope seeps in the intermediate montane zones, but few large wetlands.

Lowlands

Lowland herbaceous wetlands, mapped as H50W–LOW ELEVATION HERBACEOUS WETLAND, are usually dominated by water sedge (*Carex aquatilis*) and other graminoids and often occur along pond and lake margins. In larger valleys such as the Big Beaver, Little Beaver, and Thunder Creek Valleys, they form dynamic complexes with S40W–LOW ELEVATION SHRUB SWAMP wetlands dominated by Sitka willow and rose spirea (*Spiraea douglasii*), and with swampy B31–BROADLEAF RIPARIAN AND SWAMP FOREST. Herbaceous wetlands around ponds and lakes may be ringed by a shrub wetland and limited to a narrow strip between that and the water. Beaver activity is important in creating and maintaining many of these wetlands, which are more abundant farther north in the park, but are scattered in the Stehekin Valley as well.

Subalpine and alpine

Herbaceous subalpine wetlands, mapped as H51W–SUBALPINE HERBACEOUS WETLAND, usually have prominent fan-leaf cinquefoil (*Potentilla flabellifolia*) and black alpine sedge (*Carex nigricans*) with a range of other forbs and sedges. These wetlands are found throughout the high elevations on cirque floors and in other headwaters basins. Wetlands in similar environments but dominated by dense patches of undergreen willow (*Salix commutata*) are especially abundant in the rain shadow and are mapped as S41W–SUBALPINE WILLOW WETLAND. The two types often form mosaics, with undergreen willow wetlands adjacent to low-gradient streams and surrounded by herbaceous wetlands. Dense turfy patches of black alpine sedge and other sedges, usually with prominent partridgefoot, occur throughout the alpine zone in depressions holding snow beds or collecting melt from above. These perched wetlands are mapped as H60W–BLACK ALPINE SEDGE WETLAND; patches are often inclusions within subalpine or alpine heather shrublands, with rock barrens usually nearby.

4.2.7. Natural abiotic areas

Unvegetated natural areas are extensive in the complex, cumulatively occupying over a quarter of the landscape. R73–BEDROCK BARREN and R72–COLLUVIAL BARREN are the first and third most abundant single classes, covering eleven and seven percent of the complex respectively, while R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE occupies only one percent. Although most of these barrens may entirely lack vascular plants, many include diverse sparse vegetation. The most common vascular plant species in barrens as a group are Tolmie's saxifrage, American parsley fern (*Cryptogramma acrostichoides*) and whitebark raspberry (*Rubus leucodermis*). W82–EXPOSED SNOW AND ICE and W81–FRESH WATER each cover about three percent, the latter consisting primarily of

Ross Lake (the dammed upper Skagit Valley) and the northern portion of Lake Chelan (a glacially carved basin).

4.2.8. Natural and semi-natural disturbed landscapes

Nearly three percent of the complex is mapped as M92–BURNED WITH UNCERTAIN VEGETATION. These are areas that have experienced severe fire in the last 35 years but do not resemble any of the fire-adapted map classes discussed above. They were not well-sampled in our fieldwork; vegetation cover is likely low and probably consists of a variety of early successional plants, including fireweed (*Chamaenerion angustifolium*).

4.2.9. Development

Roads, cleared corridors and developed areas cumulatively occupy 0.3% of the complex, almost entirely in the Highway 20 corridor.

4.3. Influence of disturbance

The USNVC and the mapping classification are best developed for stable climax and late seral vegetation types, but disturbances are a major driver of vegetation composition in the complex. Post-disturbance trajectories may follow consistent patterns of vegetation colonization or recovery represented in the classification, but also may result in unique combinations of species that do not fit the classification well. Both scenarios are discussed below.

Fire is the predominant cause of vegetation disturbance in the complex, affecting tens of thousands of acres each decade on average. Recent notable burns include the upper Skagit fires (2016), the Goodell and Wolverine fires (2015), the Lone Mountain fire (2014), the Big Beaver fires (2013) and the Flick Creek fire near Stehekin (2006). Sites burned within the last 35 years that do not resemble typical post-fire vegetation classes are mapped as M92–BURNED WITH UNCERTAIN VEGETATION. Variability in pre-fire vegetation, fire severity, and propagule availability result in diverse recovery pathways, limiting our capability to map these areas with more specificity.

Avalanches represent another important agent of change. They periodically shatter tree trunks, favoring shorter and more flexible tall shrub plant communities along their established paths. These communities, which repeat regularly across the upper and lower montane zones, are mapped as S43–SITKA ALDER SHRUBLAND and S45–VINE MAPLE SHRUBLAND. The less severe impacts in adjacent areas result in less consistent outcomes that are difficult to predict or map. The outer flanks of avalanche tracks often contain battered conifer forests with variable understories, and concentrated debris deposition zones in avalanche runouts often feature a haphazard and opportunistic mix of subalpine plants displaced from above with montane plants from nearby. Species composition in these areas varies from site to site and would not be easily placed in any classification.

Periodic flooding occurs along the wild rivers and streams in the park. Rivers in the lower valleys regularly change course, washing away established forests and over time resulting in a patchwork of even-aged broadleaf forests and older conifer-dominated forests. Less destructive flooding can kill standing trees or bury the understory in cobbles, leading to atypical plant communities that are not captured in the classification. In the absence of continued flooding, abandoned channels, banks, and

bars are eventually colonized and follow trajectories toward climax conifer forests. Colonizing woody plants and their herbaceous associates often show consistent floristics, which we map as B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND, but a wide variety of species are possible depending on propagule availability, substrate and water table depth. West-side drainages are especially susceptible to flooding when warm spring storms rapidly melt deep snowpacks in the montane zone. The mainstem of the Skagit River is an exception, due to the construction of hydroelectric dams which have curtailed flooding in the lower valley and resulted in a shift toward older riparian forests.

Glaciers are found perched on benches carved into the sides of high rocky peaks. Though imposing, they have receded significantly, both over the Holocene (Osborn et al. 2012, Beget 1984) and much more recently (Pelto 2017). They have left lakes in their wakes, scoured bedrock benches, and piled unsorted glacial till in moraines of various ages. The vegetation (or lack thereof) on these landforms depends on age, climate, water table and propagule availability. Older moraines host subalpine plant communities such as S48–SUBALPINE HEATHER SHRUBLAND or even conifer forests. Younger moraines are less likely to host a cohesive plant community and are instead dominated by a smattering of whatever nearby plant species happen to get a toehold. These early seral assemblages are often dissimilar to all map classes, but are mapped as their best match: for alpine moraines, this is often S49–ALPINE HEATHER SHRUBLAND or H62–ALPINE SPARSE HERBACEOUS VEGETATION.

We created several map classes to account for land cover types not treated in the associations: areas significantly disturbed by fire or logging, and developed and agricultural land within the park and adjacent mapped areas. These land cover types were not inventoried either due to access issues or because they do not contain native vegetation represent significant conservation value. Although they cannot be described floristically, they were mapped to general land-cover/land-use categories to prevent gaps in the map coverage. The most abundant of these types within the complex are recently burned areas; these are discussed above.

Protective legislation and inaccessibility prevented clearcutting in most areas prior to the establishment of the National Park Complex. Nonetheless, some forest stands within the current boundary had been harvested, most notably in the lower Stehekin Valley and in the Skagit Valley below the current Gorge Dam. This included logging up the walls of the Skagit and selective removal of western redcedar in lower Skagit tributaries. The upper Skagit was significantly logged prior to damming (Luxenberg 1986), but that area is now submerged beneath Ross Lake and those logged patches are not apparent in the vegetation map. Regenerated cuts within the park are handled well by the association-level classification. More recently logged areas outside the park boundary that could not be confidently assigned to one of the other map classes were coded as M93–TIMBERLAND WITH UNCERTAIN VEGETATION. Atypical conditions are found in these areas due to forest management practices such as replanting and the use of herbicides to suppress growth of deciduous trees and shrubs (Washington DNR 2018).

M94–DEVELOPMENT includes developed sites within the park including visitor centers, housing, and maintenance facilities. It also includes the community of Newhalem, hydroelectric infrastructure, as well as various farmed, residential, and industrial lands outside the park. M96–CLEARED CORRIDORS

represents vegetation management and removal under power lines and along the international border. Park roads are mapped as M95–ROADS IN PARK.

4.4. Guidelines for map use

Before using the map products, users should thoroughly review both the map class descriptions and the accuracy assessment. The map represents existing vegetation as of summer 2015, although it may reflect the impacts of disturbance occurring before August 11, 2017. Vegetation patches smaller than 500 m^2 may not appear in the map; patches smaller than 90 m^2 are definitely not captured. Narrow ribbon-like artifacts may be present near transitions between distinct lifeforms. In order to capture real vegetation that occurs in elongated slender patches, we did not aggressively filter these artifacts.

For some map uses, the fine floristic distinctions between our map classes will likely be unnecessary. We've provided some guidance on merging map classes into dominant lifeform groups, but likely other combinations will be useful. When combining map classes into broader categories (e.g. silver fir forests, mountain-heather dwarf-shrublands), consider floristic similarity, spatial proximity (e.g., "are the classes typically found adjacent on the ground?"), and confusion (e.g., "how confused are the classes in the accuracy assessment?").

Planning of management or monitoring activities based on the vegetation map should always incorporate a consideration of the assessed accuracy of the map classes involved. Whether user's or producer's accuracy is a more appropriate metric depends on the issue. If a monitoring study requires field sampling within a map class, the class user's accuracy should be considered before sending crews to randomly generated locations. For map classes with lower user's accuracies, additional steps should be taken—at a minimum, examining recent aerial imagery—to ensure the sample locations are indeed occupied by the target class. On the other hand, the practicality of delineating the spatial bounds of a vegetation type is a function of the class producer's accuracy. Map classes with low producer's accuracy are not mapped in many places where they are present, so their distribution will be less clear. In some cases, an application might require consideration of the full population contingency table. For instance, an assessment of the impacts of a mapped disturbance event on habitat availability would need to estimate the fractional composition of map classes in the disturbed area. Although a simple summary based on the mapped classes would be easy, a better approach might be to apply area estimate corrections based on the population contingency table, as was done to estimate map class extents for **Table 19**.

Literature Cited

Agee, J. K. 1993. Fire ecology of Pacific Northwest forests. Island Press, Washington, District of Columbia.

Agee, J. K., and J. Kertis. 1987. Forest types of the North Cascades National Park Service Complex. Canadian Journal of Botany 65:1520–1530.

Agee, J. K., and S. G. Pickford. 1985. Vegetation and fuel mapping of the North Cascades National Park Service Complex. National Park Service, Cooperative Park Studies Unit, University of Washington, Seattle, Washington.

Almack, J. A., W. L. Gaines, R. H. Naney, P. H. Morrison, J. R. Eby, G. F. Wooten, M. C. Snyder, S. H. Fitkin, and E. R. Garcia. 1993. North Cascades grizzly bear ecosystem evaluation. Report to the Interagency Grizzly Bear Committee in fulfillment of requirements identified in the 1982 Grizzly Bear Recovery Plan.

Alverson, E., and J. Arnett. 1986. From the steppe to the alpine: a botanical reconnaissance of the Lake Chelan-Sawtooth Ridge Area, Washington. Douglasia Occasional Papers 2:1–63.

Antonova, N., C. Copass, and S. Clary. 2013. Landsat-based monitoring of landscape dynamics in the North Cascades National Park Service Complex: 1985–2009. Natural Resource Data Series NPS/NCCN/NRDS—2013/532. National Park Service, Fort Collins, Colorado.

Barnes, E., T. Clarke, S. Richards, P. Colaizzi, J. Haberland, M. Kostrzewski, P. Waller, C. Choi, E. Riley, and T. Thompson. 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. Pages 16–19 *in* University of Minnesota. Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, Minnesota.

Beget, J. E. 1984. Tephrochronology of Late Wisconsin deglaciation and Holocene glacier fluctuations near Glacier Peak, North Cascade Range, Washington. Quaternary Research 21:304–316.

Bishop, C. M. 2006. Pattern recognition and machine learning. Springer, New York.

Biven, M. M., and R. M. Rochefort. 2010. Vascular plant inventory of North Cascades National Park Service Complex. Natural Resource Technical Report. NPS/NCCN/NRTR—2010/369. Natural Resource Program Center, Fort Collins, Colorado.

Bivand, R., T. Keitt and B. Rowlingson. 2014. R package 'rgdal' version 0.8-16. Available at: https://cran.r-project.org/web/packages/rgdal/index.html (accessed 01 July 2020).

Boxberger, D. L. 1996. An ethnographic overview and assessment of North Cascade National Park Service Complex. Cultural Resources Division, Pacific West Region. National Park Service, Seattle, Washington.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

Brunner, R. L., E. M. Nielsen, and C. Copass. 2017. A data-driven method for assembling map classes from vegetation associations in Washington National Parks. Organized poster session, Ecological Society of America Conference, Portland, Oregon. Abstract available at: https://eco.confex.com/eco/2017/webprogram/Paper62429.html.

Brunner, R. L., C. Copass, L. K. Wise, and E. M. Nielsen. 2021. North Cascades National Park Complex map class key. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Burke Herbarium. 2020. Burke herbarium image collection. Available at: https://biology.burke.washington.edu/herbarium/imagecollection.php (accessed 30 December 2020).

Cao, Q., Y. Miao, H. Wang, S. Huang, S. Cheng, R. Khosla, and R. Jiang. 2013. Non-destructive estimation of rice plant nitrogen status with Crop Circle multispectral active canopy sensor. Field Crops Research 154:133–144.

Chander, G., B. L. Markham, and D. L. Helder. 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+ and EO-1 ALI sensors. Remote Sensing of Environment 113:893–903.

Chavez, P. S. 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sensing of Environment 24:459–479.

Cohen, W. B., and T. A. Spies. 1992. Estimating structural attributes of Douglas-fir–western hemlock forest stands from Landsat and SPOT imagery. Remote Sensing of Environment 41:1–17.

Comulada, A. B. 1981. A botanical reconnaissance of the Chilliwack River Valley in the North Cascades National Park, Washington. MS thesis, Western Washington University, Bellingham, Washington.

Congalton, R. G., and K. Green. 1999. Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, Boca Raton, Louisiana.

Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, and J. Böhner. 2015. System for Automated Geoscientific Analyses (SAGA) version 2.1.4. Geoscientific Model Development 8:1991–2007.

Copass, C., and T. Ramm-Granberg. 2016a. Ebey's Landing National Historic Reserve vegetation inventory and mapping project. Natural Resource Report NPS/NCCN/NRR—2016/1127. National Park Service, Fort Collins, Colorado.

Copass, C., and T. Ramm-Granberg. 2016b. Vancouver National Historic Reserve vegetation inventory and mapping project. Natural Resource Report NPS/NCCN/NRR—2016/1128. National Park Service, Fort Collins, Colorado.

Crawford, R. C., C. B. Chappell, C. C. Thompson and F. J. Rocchio. 2009. Vegetation classification of Mount Rainier, North Cascades and Olympic National Parks. Natural Resource Technical Report NPS/NCCN/NRTR—2009/D-586. National Park Service. Fort Collins, Colorado. Available at: https://irma.nps.gov/DataStore/Reference/Profile/661669 (accessed 01 July 2020).

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

Cwynar, L. C. 1987. Fire and the Forest History of the North Cascade Range. Ecology 68:791–802.

Daughtry, C., C. Walthall, M. Kim, E. B. De Colstoun, and J. McMurtrey. 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sensing of Environment 74:229–239.

De Cáceres, M., and P. Legendre. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90:3566–3574. Code available at: https://cran.r-project.org/web/packages/indicspecies/index.html (accessed 01 July 2020).

De Cáceres, M., X. Font, and F. Oliva. 2010. The management of vegetation classifications with fuzzy clustering. Journal of Vegetation Science 21:1138–1151. Code available at: https://cran.r-project.org/web/packages/vegclust/index.html (accessed 01 July 2020).

Douglas, G. 1972. Subalpine plant communities of the western North Cascades, Washington. Arctic and Alpine Research 4:147–166.

Douglas, G., and L. C. Bliss. 1977. Alpine and high subalpine plant communities of the North Cascades range, Washington and British Columbia. Ecological Monographs 47:113–150.

Esri. 2013. ArcGIS Desktop 10.2. Environmental Systems Research Institute, Redlands, California.

Evans, J. S., and S. A. Cushman. 2009. Gradient modeling of conifer species using random forests. Landscape Ecology 24:673–683.

Federal Geographic Data Committee (FGDC). 2008. National vegetation classification standard, Version 2, FGDC-STD-005-2008. Available at: https://www.fgdc.gov/standards/projects/FGDC-standards-projects/vegetation/NVCS_V2_FINAL_2008-02.pdf (accessed 01 July 2020).

Fleischner, T. L., and S. Weisberg. 1992. Tidewater to timberline, forest to steppe: natural history of the Greater North Cascades Ecosystem. Pages 4–21 *in* M. Friedman and P. Lindholt, editors. Cascadia Wild: protecting an international ecosystem. Greater Ecosystem Alliance, Bellingham, Washington.

Foody, G. M. 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80:185–201.

Franklin, J. F., and C. T. Dyrness. 1988. Natural vegetation of Oregon and Washington. Oregon State University Press, Corvallis, Oregon.

Franklin, J. F., W. H. Moir, M. A. Hemstrom, S. E. Greene, and B. G. Smith. 1988. The forest communities of Mount Rainier National Park. Scientific Monograph Series No. 19. U. S. Department of Interior, National Park Service, Washington, District of Columbia.

Franklin, J. F., and J. M. Trapper. 1963. Plant communities of the northern Cascade Range: A reconnaissance. Northwest Science 37:163–164.

Grossman, D. H., D. Faber-Langendoen, A. S. Weakley, M. Anderson, P. Bourgeron, R. Crawford, K. Goodin, S. Landaal, K. Metzler, K. D. Patterson, M. Pyne, M. Reid, and L. Sneddon. 1998. International classification of ecological communities: terrestrial vegetation of the United States. Volume I. The National Vegetation Classification System: development, status, and applications. The Nature Conservancy, Arlington, Virginia.

Hall, D. K., G. A. Riggs, and V. V. Salomonson. 1995. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. Remote Sensing of Environment 54:127–140.

Healey, S. P., W. B. Cohen, Z. Yang, and O. N. Krankina. 2005. Comparison of tasseled cap-based Landsat data structures for use in forest disturbance detection. Remote Sensing of Environment 97:301–310.

Henderson, J. A., R. D. Lesher, D. H. Peter, and D. C. Shaw. 1992. Field Guide to the Forested Plant Associations of the Mt. Baker- Snoqualmie National Forest. USDA-FS PNW Research Station. R6 ECOL TP 028-91, Portland, Oregon.

Hijmans, R. J. 2018. R package 'raster': geographic data analysis and modeling, version 2.8-4. Available at: https://cran.r-project.org/web/packages/raster/index.html (accessed 01 July 2020).

Hitchcock, C. L., and A. Cronquist. 1973. Flora of the Pacific Northwest, first edition. University of Washington Press, Seattle, Washington.

Hitchcock, C. L., and A. Cronquist. 2018. Flora of the Pacific Northwest, second edition. Burke Museum Herbarium & University of Washington Press, Seattle, Washington.

Hoffman, R. L., A. Woodward, P. Haggerty, K. Jenkins, P. Griffin, M. J. Adams, J. Hagar, T. Cummings, D. Duriscoe, K. Kopper, J. Riedel, L. Marin, G. S. Mauger, K. Bumbaco, and J. S. Littell. 2015. North Cascades National Park Service Complex: natural resource condition assessment. Natural Resource Report NPS/NCCN/NRR—2015/901. National Park Service, Fort Collins, Colorado.

Huang, C., B. Wylie, L. Yang, C. Homer, and G. Zylstra. 2002. Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. International Journal of Remote Sensing 23:1741–1748.

Institute for Natural Resources (INR). 2021a. North Cascades National Park Complex accuracy assessment contingency tables. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Institute for Natural Resources (INR). 2021b. Taxonomic tables for Mount Rainier, Olympic and North Cascades National Parks. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2283943.

Kagan, J. S., E. M. Nielsen, M. D. Noone, J. C. van Warmerdam, L. K. Wise, G. Kittel, and C. Copass. 2012. Lewis and Clark National Historic Park vegetation classification and mapping project report. Natural Resource Report NPS/LEWI/NRR—2012/597. National Park Service, Fort Collins, Colorado.

Kauth, R. J., and G. S. Thomas. 1976. The tasseled cap: a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data: 4B41–4B51. Purdue University, West Lafayette, Indiana.

Key, C. H., and N. C. Benson. 2002. Measuring and remote sensing of burn severity. U. S. Geological Survey Wildland Fire Workshop, October 31–November 3, 2000. USGS Open File Report 02-11. USGS, Los Alamos, New Mexico.

Kirkman, T. P. 1847. On a problem in combinatorics. Cambridge and Dublin Mathematical Journal 2:191–204.

Kopper, K. Personal communication. Notes on fire history of North Cascades. Fire Ecologist, National Park Service, Washington.

Lea, C. 2011. Vegetation classification guidelines: National Park Service Vegetation Inventory, version 2.0. Natural Resource Report NPS/NRPC/NRR—2011/374. National Park Service, Fort Collins, Colorado. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2170603 (accessed 09 December 2020).

Lea, C., and A. C. Curtis. 2010. Thematic accuracy assessment procedures: National Park Service Vegetation Inventory, version 2.0. Natural Resource Report NPS/2010/NRR—2010/204. National Park Service, Fort Collins, Colorado. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2124829.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2002:18–22. Available at: https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf. Code available at: https://cran.r-project.org/web/packages/randomForest/index.html (accessed 01 July 2020).

Lillybridge, T., B. Kovalchik, C. Williams, and B. Smith. 1995. Field guide for forested plant associations of the Wenatchee National Forest. USDA Forest Service Report PNW-GTR-359. U. S. Forest Service PNW Research Station, Portland, Oregon.

Louter, D. 1998. Contested terrain: North Cascades National Park Service Complex, Washington, an administrative history. National Park Service, Seattle, Washington. Available at: https://www.nps.gov/parkhistory/online_books/noca/adhi/intro.htm (accessed 01 July 2020).

Luxenberg, G. A. 1986. Historic resource study: North Cascades National Park Service Complex, Washington. National Park Service, Seattle, Washington. Available at: https://www.nps.gov/parkhistory/online_books/noca/hrs/index.htm (accessed 01 July 2020)

McCune, B., and D. Keon. 2002. Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science 13:603–606.

McCune, B. 2007. Improved estimates of incident radiation and heat load using non-parametric regression against topographic variables. Journal of Vegetation Science 18:751–754.

Moore, I. D. 1991. Digital terrain modeling: a review of hydrological, geomorphological and biological applications. Hydrological Processes 5:3–30.

National Park Service (NPS). 2018. Vegetation mapping inventory. Available at: https://www.nps.gov/im/vegetation-inventory.htm (accessed 01 July 2020).

NatureServe. 2012. NCCN alliance descriptions: Forested and a subset of non-forested alliances from Mount Rainier, North Cascades & Olympic National Parks. Interim report. NatureServe, Arlington, Virginia.

Nielsen, E. M., and R. L. Brunner. 2021. Vegetation associations for mapping Pacific Northwest national parks. Institute for Natural Resources, Portland State University, Portland, Oregon. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2283945 (available June 2021).

Nielsen, E. M., C. Copass, R. L. Brunner, and L. K. Wise. 2021a. Mount Rainier National Park vegetation classification and mapping project report. Natural Resource Report NPS/NCCN/NRR—2021/2253. National Park Service, Fort Collins, Colorado. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2286419.

Nielsen, E. M., C. Copass, R. L. Brunner, and L. K. Wise. 2021b. Olympic National Park vegetation classification and mapping project report. Natural Resource Report NPS/NCCN/NRR—2021/2255. National Park Service, Fort Collins, Colorado. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2286420.

Nielsen, E. M., R. L. Brunner, C. Copass, and L. K. Wise. 2021c. North Cascades National Park Complex map class descriptions. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Nielsen, E. M., C. Copass, R. L. Brunner, and K. Braun. 2021d. North Cascades National Park Complex vegetation map. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285192.

Oksanen, J., F. Guillaume Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2019. R package 'vegan': community ecology, version 2.5-6. Available at: https://cran.r-project.org/web/packages/vegan/index.html (accessed 09 December 2020).

Oliver, C. D., A. B. Adams, and R. J. Zasoski. 1985. Disturbance patterns and forest development in a recently deglaciated valley in the northwestern Cascade Range of Washington, U. S. A. Canadian Journal of Forest Research 15:221–232.

Osborn, G., B. Menounos, C. Ryane, J. Riedel, J. J. Clague, J. Koch, D. Clark, K. Scott, and P. T. Davis. 2012. Latest Pleistocene and Holocene glacier fluctuations on Mount Baker, Washington. Quaternary Science Reviews. 49:33–51.

Pacific Meridian Resources (PMR). 1997. Vegetation and landform database development. Report to the National Park Service. Pacific Meridian Resources, Portland, Oregon.

Pelto, M. 2017. North Cascade Range, Washington USA. Pages 101–129 *in* M. Pelto. Recent climate change impacts on mountain glaciers. John Wiley & Sons Ltd., West Sussex, UK.

Pojar, J., A. Mackinnon, and P. B. Alaback. 2004. Plants of the Pacific Northwest coast: Washington, Oregon, British Columbia and Alaska, second edition. Lone Pine Publishing, Vancouver, Canada.

Prichard, S. J. 2003. Spatial and temporal dynamics of fire and vegetation change in Thunder Creek watershed, North Cascades National Park, Washington. PhD Dissertation. College of Forest Resources, University of Washington, Seattle, Washington.

PRISM Climate Group. 2019. PRISM climate data. Available at: https://prism.oregonstate.edu/ (accessed 01 July 2020).

R Development Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/ (accessed 01 July 2020).

Ramm-Granberg, T., F. J. Rocchio, C. Copass, R. Brunner, and E. Nielsen. 2021. Revised vegetation classification for Mount Rainier, North Cascades, and Olympic National Parks: Descriptions and identification keys for plant associations and wetland alliances. North Coast and Cascades Network. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2279820.

Risvold, A. M., and R. W. Fonda. 2001. Community composition and floristic relationships in montane wetlands in the North Cascades, Washington. Northwest Science 75:157–167.

Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering. 1974. Monitoring vegetation systems in the Great Plains with ERTS. Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, volume 1A:309–317.

Tabor, R. W., and R. A. Haugerud. 1999. Geology of the North Cascades: a mountain mosaic. The Mountaineers, Seattle, Washington.

Taylor, R. J., and G. W. Douglas. 1977. Plant ecology and natural history of Chowder Ridge, Mt. Baker: a potential alpine Research Natural Area in the western north Cascades. Northwest Science 52:35–50.

Therneau, T., and B. Atkinson. 2018. R package 'rpart': recursive partitioning and regression trees, version 4.1–13. Available at: https://cran.r-project.org/web/packages/rpart/index.html (accessed 01 July 2020).

Tucker, C. J., and P. J. Sellers. 1986. Satellite remote sensing of primary production. International Journal of Remote Sensing 7:1395–1416.

Twele, A. 2006. The effect of stratified topographic correction on land cover classification in tropical mountainous regions. ISPRS Commission VII Mid-term Symposium: remote sensing from pixels to processes. Enschede, The Netherlands.

United Nations Educational Scientific, and Cultural Organization (UNESCO). 1973. International classification and mapping of vegetation. Series 6: Ecology and Conservation. UNESCO, Paris.

United States Geological Survey (USGS). 2012. National Hydrography Dataset. Available at: https://www.usgs.gov/core-science-systems/ngp/national-hydrography (accessed 01 July 2020).

United States Geological Survey (USGS). 2019a. GLOVIS: global visualization viewer. Available at: https://glovis.usgs.gov/ (accessed 01 July 2020).

United States Geological Survey (USGS). 2019b. 3D elevation program. Available at: https://www.usgs.gov/core-science-systems/ngp/3dep (accessed 01 July 2020).

United States National Vegetation Classification (USNVC). 2019. United States National Vegetation Classification Database, V2.03. Federal Geographic Data Committee, Vegetation Subcommittee, Washington District of Columbia. Available at: http://usnvc.org/ (accessed 01 July 2020).

Vanbianchi, R., and S. J. Wagstaff. 1988. A floristic survey of Big Beaver Valley. Douglasia Occasional Papers 3:1–61.

Wagstaff, S., and R. J. Taylor. 1980. Botanical reconnaissance in the Stetattle Creek Research Natural Area, North Cascades National Park, Washington. Western Washington University, Bellingham, Washington.

Wang, L., and H. Liu. 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science 20:193–213.

Washington Natural Heritage Program (WNHP). 2011. Ecosystems of Washington. Available at: https://www.dnr.wa.gov/NHPecosystems (accessed 01 July 2020).

White, B., J. Ogilvie, D. M. H. Campbell, D. Hiltz, B. Gauthier, H. K. Chisholm, H. K. Wen, P. N. C. Murphy, and P. A. Arp. 2012. Using the cartographic depth-to-water index to locate small streams and associated wet areas across landscapes. Canadian Water Resources Journal 37:333–347.

Williams, C. K., and T. R. Lillybridge. 1983. Forested Plant Associations of the Okanogan National Forest. USDA Forest Service Report R6-ECOL-132b-1983. U. S. Forest Service PNW Research Station, Portland, Oregon.

Wilma, D. 2003. Upper Skagit River Hydroelectric Project. Available at: https://www.historylink.org/File/5347 (accessed 01 July 2020).

Wilson, E. H., and S. A. Sader. 2002. Detection of forest harvest type using multiple dates of Landsat TM imagery. Remote Sensing of Environment 80:385–396.

Wooten G., and P. Morrison. 2006. Classification of vascular plant communities of the North Cascades using discrete space boundary analysis. Available at: https://www.semanticscholar.org/paper/CLASSIFICATION-OF-VASCULAR-PLANT-COMMUNITIES-OF-THE-Wooten-Morrison/ee8118cb46322483a9c53b3f1b1474529b9ef494 (accessed 01 July 2020).

Yang, L., S. Jin, P. Danielson, C. Homer, L. Gass, S. M. Bender, A. Case, C. Costello, J. Dewitz, J. Fry, M. Funk, B. Granneman, G. C. Liknes, M. Rigge, and G. Xian. 2018. A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. ISPRS Journal of Photogrammetry and Remote Sensing 146:108–123. Available at: https://www.mrlc.gov/data/nlcd-2016-land-cover-conus (accessed 01 July 2020).

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS 168/175301, May 2021

Natural Resource Stewardship and Science
1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525