National Park Service
U.S. Department of the Interior

Natural Resource Stewardship and Science

Mount Rainier National Park Vegetation Classification and Mapping Project Report

Natural Resource Report NPS/NCCN/NRR—2021/2253

ON THIS PAGE
Looking north to Grand Park from Mount Fremont
Image credit Eric Nielsen

ON THE COVER
Looking south from subalpine meadows in Grand Park
Image credit NPS

Mount Rainier National Park Vegetation Classification and Mapping Project Report

Natural Resource Report NPS/NCCN/NRR—2021/2253

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. Mount Rainier National Park vegetation classification and mapping project report. Natural Resource Report NPS/NCCN/NRR—2021/2253. National Park Service, Fort Collins, Colorado. https://doi.org/10.36967/nrr-2286419.

NPS 105/175300, 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. Mount Rainier National Park vegetation map. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

Supplement 2: Map class descriptions

Nielsen, E. M., R. L. Brunner, C. Copass, and L. K. Wise. 2021. Mount Rainier National Park map class descriptions. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

Supplement 3: Map class key

Brunner, R. L., C. Copass, L. K. Wise, and E. M. Nielsen. 2021. Mount Rainier National Park map class key. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

Supplement 4: Accuracy assessment contingency tables

Institute for Natural Resources. 2021. Mount Rainier National Park accuracy assessment contingency tables. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

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 Mount Rainier National Park (MORA).

The map is based on a vegetation classification developed during the project and was created using an inductive modeling approach. Data used to construct the classification were collected between 2005 and 2015 and included plots from Olympic National Park and North Cascades National Park Complex. 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 MORA, collected during the same years. Independent field accuracy assessment data were collected in 2011, supplemented in 2014 and 2019, and applied to the final map generated later.

The map development process was organized around the random forests machine learning algorithm. The modeling used 1,900 plots representing 124 vegetation associations and 37 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 the 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, contains 33 natural vegetated classes, five mostly unvegetated natural classes, and four classes representing burned areas or anthropogenic disturbance. Coniferous forests and woodlands cover about three-fifths of the park. Upper montane forest codominated by silver fir (*Abies amabilis*), mountain hemlock (*Tsuga mertensiana*) and/or Alaska-cedar (*Callitropsis nootkatensis*) is the most abundant forest zone by far. Lower montane forest dominated by silver fir and western hemlock (*Tsuga heterophylla*), and subalpine forest and woodland dominated by subalpine fir (*Abies lasiocarpa*) and mountain hemlock are about equally abundant. Lowland forest dominated by Douglas-fir (*Pseudotsuga menziesii*) and western hemlock is more limited, covering less than ten percent of the park. Each of the forest zones are found throughout the park in appropriate habitat, but subalpine types are most abundant in the northeastern park quadrant and lowland forests are associated primarily with the lower Ohanapecosh/Cowlitz, Carbon and Nisqually river valleys. Broadleaf and mixed forests occupy less than two percent of the park, mainly near major rivers, and often in an early successional state

following disturbance by flooding. Shrublands cover nine percent, mostly as high-elevation mountain-heather, post-fire successional shrublands and tall shrubs in avalanche tracks. Herbaceous vegetation occupies just over five percent, mainly in lush subalpine and sparse alpine meadows. Sparsely vegetated and entirely bare rock, especially colluvial deposits, cover thirteen percent of the park, and exposed snow and ice occupy eight percent. Lake and river surfaces round out most of the remaining two percent.

The accuracy assessment (AA) was based on 761 independent field-collected plots representing all the vegetated classes, as well as alluvial, colluvial and bedrock barrens, which also often host vegetation communities. They were gathered from an inference area covering 6.9% of the park. The overall map accuracy based on this sample was 86.9%. After correcting for map class prevalence in the inference area, the overall accuracy was 83.3%. Six of the 35 classes evaluated in the AA failed to meet the 80% NPS standard for user's accuracy; seven fell short of the standard for producer's accuracy. The AA discussion in the report contains a review of all classes failing to meet either standard, considers possible remedies for each, and provides recommendations to NPS for possible modifications of the map 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

Katherine Braun, OLYM GIS specialist — production of map sheets for mapping and accuracy assessment (AA) fieldwork, production of final maps, metadata assistance
Bill Bennett — mapping crew member
John Boetsch, network data manager/ecologist — creation and maintenance of field plot database
Karl Brown, national VIP manager — programmatic guidance
Tammy Cook, biologist — programmatic guidance
Catharine Copass, network VIP projects manager — project oversight and coordination
Risako Dan — mapping crew member
Pete Del Zotto — classification crew leader, mapping crew leader
Holly Faulstich — mapping crew member
Beth Fallon MORA vegetation ecologist — project support, report peer review
Lise Grace, NCCN publication manager — report editing and publication assistance
Michelle Hansen — mapping crew member
Midori Hiramatsu — mapping crew member
Mark Huff, *NCCN Program Coordinator* — programmatic guidance, project support
Molly Immen — classification crew member
Sarah Koenig — classification crew member
Laurie Kurth, MORA vegetation ecologist — project support
Allen McCoy, regional GIS specialist — data procurement
Stacy McDonough — mapping crew leader
Allen Nabors — classification crew member, mapping crew member
Rebecca Peace — mapping crew member
Arnie Peterson — classification crew member
Evan Raimundo — mapping crew member
Tynan Ramm-Granberg — mapping data QC assistance, vegetation classification assistance
Judy Runge — classification crew member
Darin Swinney, MORA GIS specialist — project support
Jessica Waite — mapping crew member
Lou Whiteaker, MORA vegetation ecologist — project support

Institute for Natural Resources, Portland State University

Rachel Brunner, vegetation ecologist — field data QC lead, vegetation classification and image processing assistance, mapping crew member

Treg Christopher — AA crew member

John Christy, wetlands ecologist — vegetation classification assistance

Mike Conroy — AA crew member

Eleanor Gaines, INR-Portland director — project management assistance

Jimmy Kagan, ecologist — INR PI (2008–11), AA crew member, vegetation classification assistance

Adam Kotaich, shorebird monitor — mapping crew member

Matt Lee, field data manager — field data QC, vegetation classification assistance

Eric Nielsen, geospatial scientist — INR PI (2012–20), vegetation classification, image processing, modeling, statistical analysis, sample and field protocol designs, mapping and AA crew member

Matt Noone, remote sensing analyst — AA crew member

Claudine Tobalske — segmentation of aerial imagery for 2008 field sampling

Jason van Warmerdam — AA crew member

Lindsey Koepke Wise, biodiversity data manager — report editing, formatting and graphics production, vegetation taxonomy assistance, classification crew member

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 P06AC00021 (originally J8W07080007) 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)
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 metrics (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)
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 match tool (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 specifies the presence of a class where it is not actually 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 inductive 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. MORA 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 MORA project was the first of the parks to be sampled in each phase, with training data collection primarily in 2008 and accuracy assessment (AA) primarily in 2011. The training phase was done before the development of a consistent plot protocol allowing confident quality control of sample location and assigned class, and the AA effort was hobbled by a heavy snowpack that persisted through the summer of 2011. The comparative weakness of the MORA plot data presented major challenges to achieving acceptable map accuracy and AA sampling density. INR performed several subsequent field sampling efforts in a mostly successful effort to overcome these challenges.

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. Along these gradients, competition and historic factors may result in gradual changes in species prominence that make field assessment of breaks based on thresholds of species cover difficult; 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

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

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 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 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

[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.

associations were combined into map classes based on their joint floristic and modeling similarities. Despite the 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

Mount Rainier National Park (MORA) is located in the Cascade Range of western Washington State, lying about 80 km (50 miles) southeast of Seattle. It is the third largest of the North Coast and Cascades Network parks, with which it is shown in **Figure 1**, and lies just west of the Cascades crest, though Mount Rainier itself is much higher than the peaks and ridges that form the hydrologic divide. The park is surrounded by other protected lands, including wilderness areas in the Snoqualmie and Gifford Pinchot National Forests, although some portions of this land are managed for timber production or resort-based recreation.

[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.

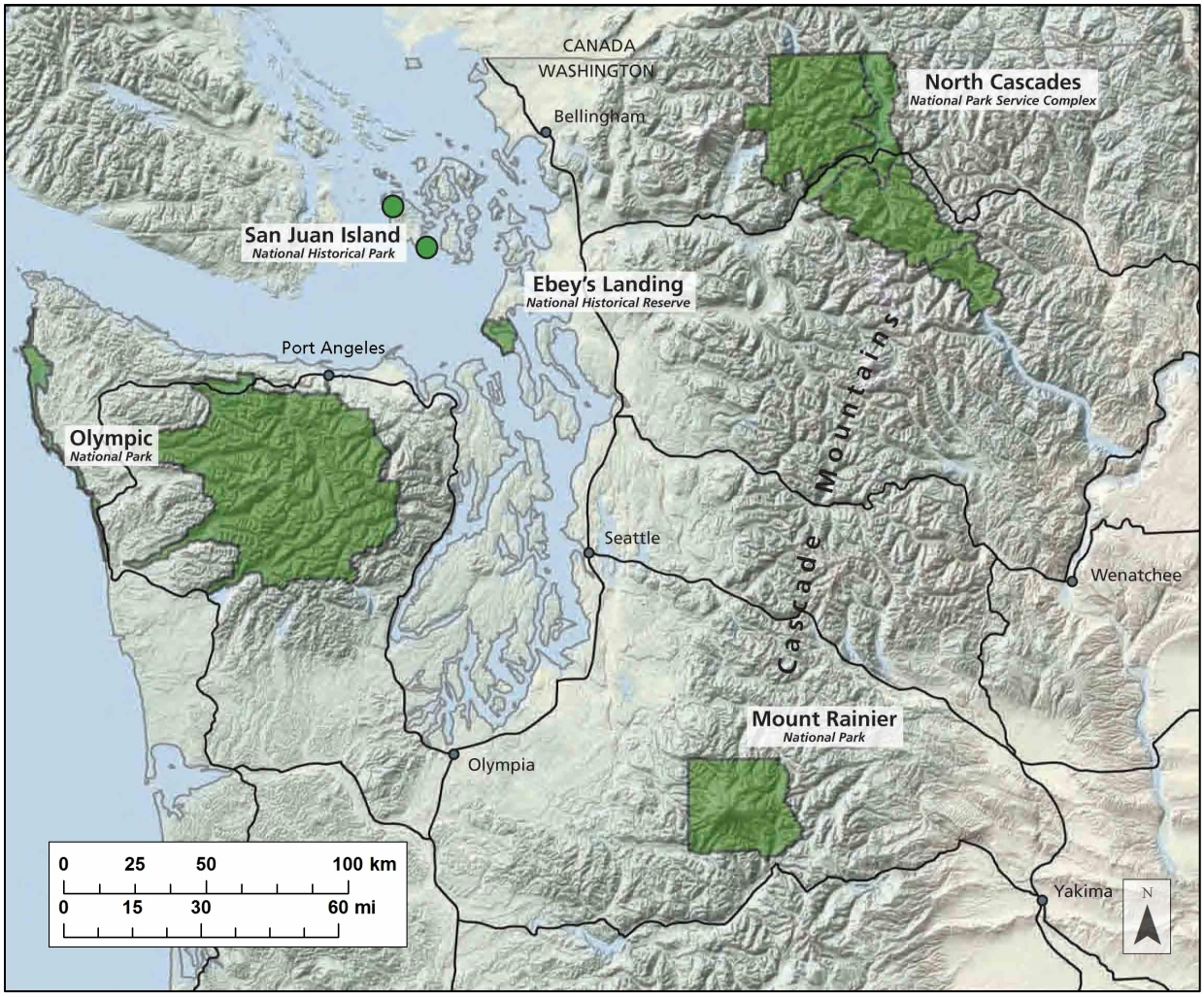


Figure 1. Map of North Coast and Cascades Network National Parks**, from Copass and Ramm-Granberg (2016b).** Fort Vancouver National Historic Site and Lewis & Clark National Historical Park lie farther south and are not shown.

The park (**Figure 2**) encompasses the active stratovolcano Mount Rainier, the highest mountain in the Cascade Range (4,392 meters; 14,410 feet). The mountain hosts some 90 km^2 of glaciers, including both the largest and the lowest terminal altitude glaciers in the contiguous United States (Riedel and Larrabee 2011). The park also includes extensive montane old growth conifer forests, luxuriant meadows and subalpine parklands and austere alpine plant communities developed on pumice deposits. Several major rivers originate from glaciers and emerge from the park: the Carbon River from the northwest, the White from the northeast, the Ohanapecosh and Cowlitz from the southeast, the Nisqually from the southwest, and the Puyallup from the west. The valley of the lower Carbon supports an inland temperate rainforest similar to forests on the Olympic Peninsula. The east side of the park, and especially the northeastern quadrant, lies in a rain shadow cast by the mountain itself. Reduced precipitation and geologically recent pumice deposits there (see Crandell 1969) result in distinct vegetation communities, especially in the alpine and subalpine zones. Elevations in the park range from 460 meters (1,500 feet) to the mountain's summit.

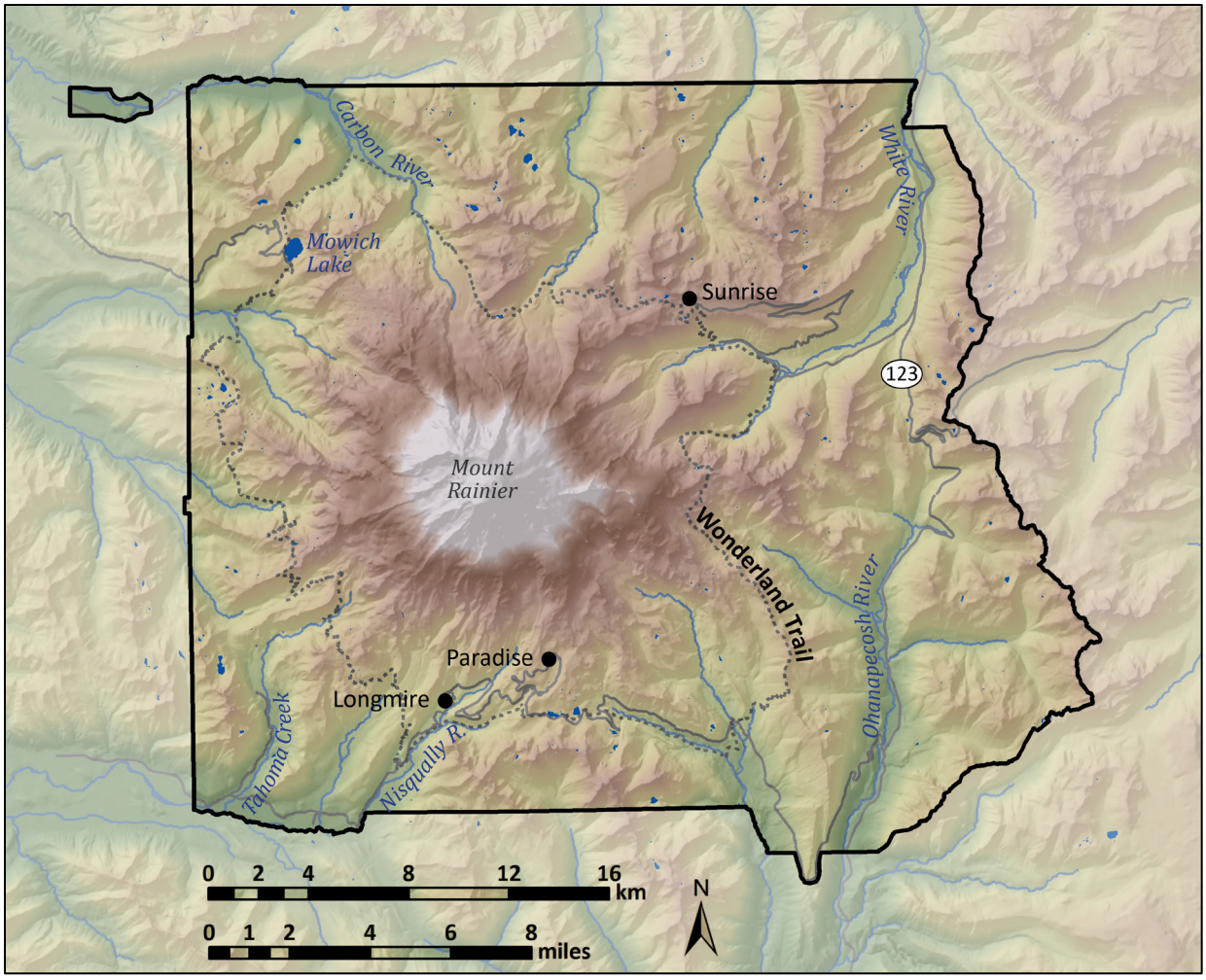


Figure 2. Map of Mount Rainier National Park, illustrating topography, rivers, roads and other features.

We defined the project area as the park (95,593 hectares; 236,215 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 half the total project extent of 190,008 hectares (469,520 acres).

1.3.2. Environmental setting, bioclimatic zones and major vegetation types

Continuing subduction under the North American plate over at least 40 million years has resulted in a series of volcanos here. Mount Rainier, whose building began just 500,000 years ago, is merely the current manifestation. During the height of Pleistocene continental glaciations, massive walls of ice blocked valleys and forced lava flows to form the distinctive ridgelines that radiate from the peak. The central cone and surrounding ridges have been alternately built and collapsed during eruptive periods over the last 11,000 years, resulting in the lahar surfaces and layers of tephra which underlie the vegetation found today. Although Mount Rainier is considered active, no magmatic eruptions have occurred in the last thousand years (USGS 2020).

The tremendous vertical relief (3,900 meters; 12,800 feet), a strong east-west precipitation gradient, and variations in topographic and hydrologic setting result in a wide range of distinct habitats. A total of 821 native vascular plant taxa are thought to occur within the park (Rochefort 2010).[6] Though this is lower than that documented at OLYM (see Buckingham et al. 1995) and NOCA (see Biven and Rochefort 2010), the number is compatible on a per-area basis. A variety of disturbance processes (e.g., lahars, mass movement, avalanches, fire) further increase landscape complexity. The major vegetation zones are discussed below.

Lowland forests

Lowland forests dominated by Douglas-fir (*Pseudotsuga menziesii*), western hemlock (*Tsuga heterophylla*) and western redcedar (*Thuja plicata*) occupy most of the landscape up to about 910 meters (3,000 feet) elevation. Sitka spruce (*Picea sitchensis*) is occasionally present below about 610 meters (2,000 feet) in the valley of the Carbon River. Understories vary depending on site moisture and stand history. Riparian forests include the broadleaved trees red alder (*Alnus rubra*) and black cottonwood (*Populus trichocarpa*), but mature examples are not abundant, presumably due to the frequency of disturbance along the glacier-fed rivers and the competitive advantage of conifers on drought-prone deep cobble surfaces. Upland broadleaf forests are even rarer; where they occur they occupy landslide-affected areas and are dominated by bigleaf maple (*Acer macrophyllum*) with substantial Douglas-fir.

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. Mesic forests characterized by silver fir and western hemlock are predominant up to about 1,250 meters (4,100 feet) elevation. Talus slopes and the lower portions of avalanche tracks are occupied by tall vine maple (*Acer circinatum*) shrublands; these extend downward into the lowland zone where suitable habitat exists.

Upper montane forests

The upper montane zone is characterized by closed forests with substantial cover of higher elevation species such as Alaska-cedar (*Callitropsis nootkatensis*), mountain hemlock (*Tsuga mertensiana*), subalpine fir (*Abies lasiocarpa*) and noble fir (*Abies procera*). These forests are predominant up to about 1,580 meters (5,200 feet) elevation. The zone is much more abundant at MORA than at the other NCCN parks, in part because of the use of noble fir as a defining feature of the zone but also because of the existence of much suitable territory in the appropriate elevation range along the major ridges radiating from the mountain. Tall shrublands of Sitka alder (*Alnus viridis*), often with Alaska-cedar, are frequent in avalanche tracks and other disturbed areas in the upper montane zone.

[6] Surveys performed for this project since Rochefort (2010) have documented additional species in the park. For example, INR made confident identification of quaking aspen (*Populus tremuloides*) in a bald adjacent to the Muddy Fork Cowlitz just below Box Canyon.

Subalpine environments

Above the closed forest zone, upper montane forests transition into subalpine woodlands and tree islands dominated by mountain hemlock and subalpine fir. Subalpine *parklands*, mosaics of wooded areas with dwarf shrublands and meadows, are predominant up to about 1,880 meters (6,200 feet). In the northeastern quadrant of the park—especially on Sunrise Ridge and in the Sourdough Mountains—the zone is considerably more extensive than elsewhere and dry subalpine fir woodlands can be found up to about 2,060 meters (6,800 feet). Whitebark pine (*Pinus albicaulis*) and lodgepole pine (*Pinus contorta*) are often present in these stands. Subalpine 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*), Cascade blueberry (*Vaccinium deliciosum*) and white mountain-heather (*Cassiope mertensiana*) with some taller shrubs such as big huckleberry (*Vaccinium membranaceum*) and Sitka mountain-ash (*Sorbus sitchensis*) in protected areas, especially near trees. Herbaceous subalpine vegetation is represented by several meadow types, with common species including subalpine lupine (*Lupinus latifolius*), fan-leaf cinquefoil (*Potentilla flabellifolia*), American bistort (*Bistorta bistortoides*), showy sedge (*Carex spectabilis*), green fescue (*Festuca viridula*), Gray's lovage (*Ligusticum grayi*), wandering daisy (*Erigeron glacialis*) and Sitka valerian (*Valeriana sitchensis*).

Alpine environments

In the alpine zone, which generally ranges upward from about 1,880 meters (6,200 feet), tree cover is reduced to stunted krummholz of mountain hemlock, subalpine fir and Alaska-cedar. Dwarf shrublands and meadows transition into sparser alpine variants with shorter vegetation and fewer species. Common species include pink and white mountain-heathers, showy sedge, black alpine sedge (*Carex nigricans*), prairie lupine (*Lupinus lepidus*), partridgefoot (*Luetkea pectinata*), spreading phlox (*Phlox diffusa*), Parry's rush (*Juncus parryi*), alpine aster (*Oreostemma alpigenum*), Davis' knotweed (*Aconogonon davisiae*) and Piper's woodrush (*Luzula piperi*). Eventually vegetation gives way almost completely to barren bedrock, talus, permanent snowfields and glaciers.

1.3.3. Human history

There is evidence of at least 8,500 years of indigenous use of the landscapes now included in the park. The traditional lands of the Cowlitz, Muckleshoot, Nisqually, Puyallup, Squaxin Island and Yakama people overlap with the present-day park boundary. Indigenous people hunted and gathered resources seasonally at all elevations of the park, and used fire to improve ungulate habitat (Burtchard 2007). Archeological sites, interpreted as seasonal residential base camps, have been found in subalpine areas all around the mountain. The density of these sites suggests that subalpine environments were of particular utility for hunting, wool gathering and collecting a wide variety of seasonally available foods including berries and perennial plants (Burtchard 2007).

Early European explorers in the region included Captain George Vancouver, who gave the mountain its English name in 1792. With the westward American expansion in the nineteenth century came explorers and climbers, and interest quickly developed in establishing a park to encourage tourism and preserve the natural resources. Mount Rainier National Park was established by Congress in 1899, the fifth national park in the country. In its early years, development activities focused on the

development of an extensive road network and lodges and other visitor facilities at Longmire, Paradise and Sunrise. Limited mining operations, especially in Glacier Basin, continued through the 1950s; the last claims in the park were not acquired by the federal government until 1984 (Burtchard et al. 2017). Mount Rainier's proximity to major cities in western Washington contribute to its popularity; in recent years upwards of two million people have visited annually (NPS 2020a).

1.3.4. Previous vegetation studies

Inventory, classification, and mapping of vegetation have been ongoing at Mount Rainier since the late nineteenth century (NPS 2012). The first comprehensive flora (Jones 1938) contained 729 species.[7] C. Frank Brockman conducted and published a wide variety of vegetation studies during his 1928–1941 tenure as "Information Ranger." These included studies in many major ecosystems (see NPS 2020b). The USGS published many reports in the mid-twentieth century regarding the history of glaciation and pyroclastic deposition in the park; many of these are relevant and useful references for understanding patterns of vegetation distribution and recovery processes (e.g., Sigafoos and Hendricks 1961, Mullineaux 1974).

Academic researchers, especially from the University of Washington and Oregon State University, have long been active in the park. Drawing partly on the work of Hamann (1972), Henderson (1974) presented a detailed classification of subalpine and alpine meadow and dwarf shrub communities, gave an overview of their distribution in the park, and proposed successional relationships between the types. The later work of Edwards (1980) and Rochefort and Peterson (1996) further contributed to the understanding of the structure and dynamics of subalpine and alpine meadows. The influence of climate change on vegetation has been a recurring topic of investigation since the late 1960s, when park managers initiated a study of tree establishment in subalpine meadows. Franklin et al. (1971) and Henderson (1974) noted increased establishment in the early twentieth century following the conclusion of the Little Ice Age. Other researchers have followed up on this work, investigating the impacts of climate change on conifer growth (Peterson and Peterson 2001), forest composition (e.g., Kroiss and Hille Ris Lambers 2015) and flowering phenology (Sethi et al. 2020).

Several previous vegetation mapping efforts have relied on interpretation of aerial photography to extrapolate findings from extensive field sampling to park-wide maps. Hemstrom (1979) and Hemstrom and Franklin (1982) used increment cores to estimate the dates of stand replacement events and extended the results via photo-interpretation to reconstruct a spatially-explicit fire chronology of the park. Franklin et al. (1988) created a classification of the forests at Mount Rainier which included fourteen mature forest types and five seral types; they illustrated their distributions with a park-wide map. The single previous vegetation map based on satellite imagery was produced by Pacific Meridian Resources (1997), who used a modified supervised classification method to map 14 vegetated and three unvegetated classes, in addition to some elements of vegetation structure,

[7] The most recent flora published (Biek 2000) includes 871 species. A recent inventory lists 973 vascular plant taxa, 821 of which are native (Rochefort 2010).

from 30-meter resolution Landsat TM imagery collected in 1988 or 1991.[8] The classification focused on forests; all herbaceous and shrub-dominated vegetation were lumped into a single class each.

Remote sensing techniques have also been used in some specialized mapping applications more recently in the park. Stueve et al. (2009) applied change detection techniques to declassified high-resolution satellite imagery from 1970 and aerial photography from 2003 to map recent tree establishment above continuous treeline. Moore et al. (2019) used the panchromatic and thermal bands from a 2014 Landsat 8 scene to estimate the extent of clean and debris-covered ice in the Carbon, Winthrop and Emmons Glaciers. While the latter study is not specifically vegetation-related, its results provide some useful context for interpretation of our mapped alluvial barren and exposed snow and ice classes.

1.4. Project timeline

The following timeline describes the primary activities during each year of the 15-year project. Only activities at MORA 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. Image segmentation and map training fieldwork. NAIP and satellite image collection and processing, development of association-level and higher-level NVC.
2009 — Development of fieldwork protocols for additional map training, limited map training fieldwork, development of higher level NVC, satellite image collection and processing, development of lidar processing and predictor metrics methodologies.
2010 — NAIP image collection and processing, development of predictor metrics methodologies.
2011 — Production of draft vegetation map for use in stratification of accuracy assessment sampling, AA sample design, AA fieldwork.
2012 — AA data quality control, NAIP and satellite image collection and processing.
2013 — AA data quality control.
2014 — Additional map training and AA fieldwork for poorly sampled regions. Training data quality control, AA data quality control, NAIP image collection and processing.
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.
2016 — Training data quality control, floristics quality control, mapping associations development, NAIP image collection and processing.
2017 — Training data quality control, floristics quality control, mapping associations development, map classification development, development of new topographic predictor metrics, refinement of nested texture metrics methodology.

[8] PMR (1997) specifies 1988 in one place and 1991 in another.

2018 — Training data quality control, mapping associations development, map classification development, satellite image collection and processing, draft map production.
2019 — NPS draft map review. Additional AA fieldwork for poorly sampled classes. 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 1,596 field plots, all collected at MORA, 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.[9] 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

[9] 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 spatial and floristic QC. Unfortunately, most MORA plots were collected before these improvements were made.

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 MORA 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 MORA 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 floristics plots			Model plots [A]
				MORA	OLYM	NOCA	MORA
VIP classification	2006–07	NPS	Full	186	228	79	213
VIP mapping protocol X	2008	NPS	Partial [D]	44	0	0	488
VIP mapping protocol M [B]	2009–11	NPS	Partial [D]	33	10	0	51
VIP mapping protocol Y	2009–11	NPS	Full	151	1,094	0	236
VIP mapping protocol Z [C]	2012–15	NPS	Full	0	233	1,612	0
VIP mapping protocol Q	2014, 2019	INR, NPS	Partial [D]	1	0	0	172
PMR accuracy assessment	2005–06	NPS	Full	61	18	10	132
Monitoring reconnaissance	2005–14	NPS	Full	91	46	48	135
UW forest community	2015	UW	Full	165	0	0	167
Total	**–**	**–**	**–**	**732**	**1,629**	**1,749**	**1,594**

[A] There is a large amount of overlap with full floristics plots collected at MORA. 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.[10] The protocol included collection of a comprehensive species list

[10] 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 (**Figure 3**). Crews stepped away from trails into an accessible point within each polygon and assessed the surroundings. They made estimates of crown cover 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.[11] The very large polygons often contained multiple vegetation types; crews noted this when they were aware of it, but provided no details on their spatial arrangement. Protocol X was the main source of map training data at MORA; a completed fieldsheet is shown in **Figure 4**.

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.

[11] 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.

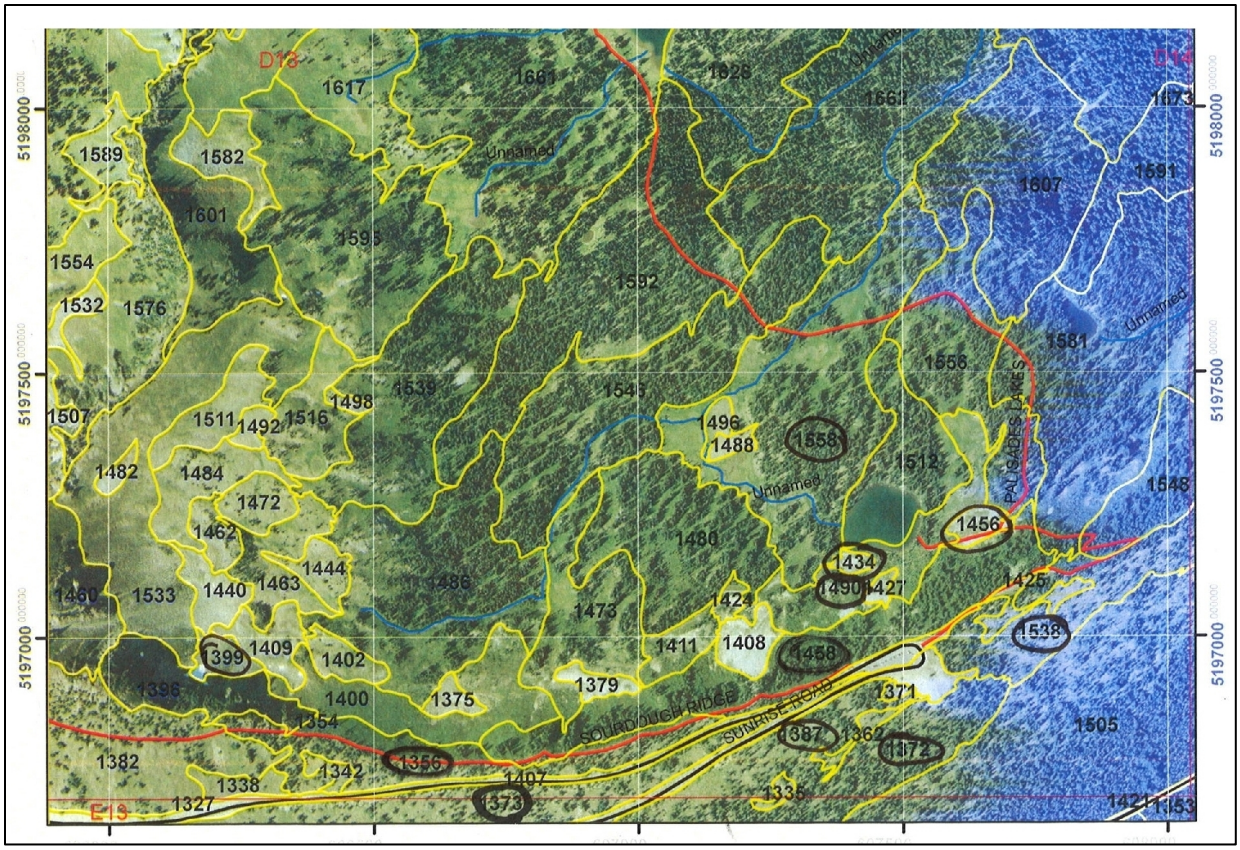


Figure 3. Completed mapping phase map sheet (protocol X, 2008); tick marks are at 500-meter intervals.


POLY-ID NE 1490 MAP NO. D-13 Date: 8/5/08

Surveyors (last name) Nabors, Hansen

GPS Taken: (Y) GPS NOTES:

GPS UNIT Q UTM E 607582 UTM N 5197407

Point NOTES:

Assoc. at point: CARNIG/CALLEP/CARSPE

Assoc. notes:

Other Associations in polygon:

1)

2)

3)

FUELTYPE: GR1 Fuels notes:

Photos taken? (Y)

1) 2255 ✓

2) 2256 ✓

3) 2257 ✓

4) 2258

	SPECIES	CODE	COVER
T1			
T2			
T3			
T4			
T5			

Polygon NOTES: Balloon 607560 5197247

☑ GPS Downloaded ☑ Pictures Downloaded ☒ Data Entered ☐ Has ocular?

Figure 4. Completed mapping phase data sheet (protocol X, 2008).

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.[12] 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

[12] 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 occurrences of vegetation types that had been poorly represented in the targeted sampling.

maps were used in all subsequent field efforts. 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.[13] 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).

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.[14] 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.

Mapping protocol Z (2012–14)

Field sampling was guided by a revised stratified sample design.[15] 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**).

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

---

[13] 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.

[14] 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.

[15] 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.

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.

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.[16] 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 and abiotic map classes[17] by assigning PI locations where it was possible to do so confidently. Generally, we approached this

[16] 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.

[17] The map classes that received PI plots at MORA were C26–CONIFER KRUMMHOLZ AND TREED CLIFF, B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND, B31–BROADLEAF RIPARIAN AND SWAMP FOREST, S40W–LOW ELEVATION

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 specified clearly incorrect classes.[18]

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

SHRUB-DOMINATED WETLAND, H50W–LOWLAND MARSH AND MEADOW, R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE, R72–COLLUVIAL BARREN, R73–BEDROCK BARREN, W81–FRESH WATER and W82–EXPOSED SNOW AND ICE.

[18] The very limited number of such plots at MORA focused on confusion between R73–BEDROCK BARREN and other abiotic types.

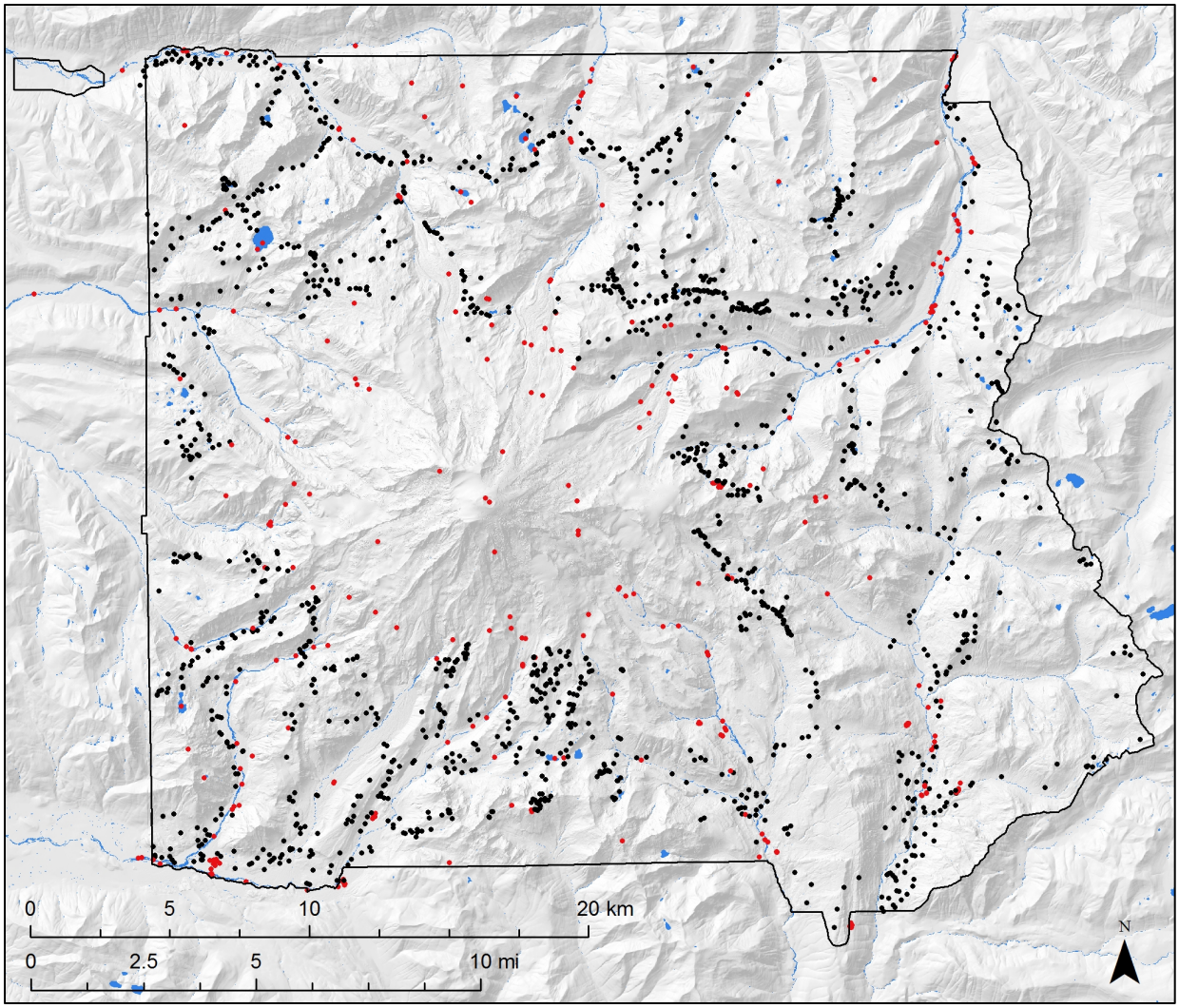


Figure 5. Model training plot locations**.** Of the 1,900 plots, 1,594 were collected in the field (shown in black), and 306 were photo-interpreted (shown in red).

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, 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 and lidar imagery,

which were the finest-resolution predictor datasets 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 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, lidar imagery, 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 16, 2016.[19] Fires were the main source of disturbance. We digitized the perimeters and entered the dates of all documented fires in the park since 1984. 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 assess the extent of disturbance. If conditions no longer resembled those present at the time of sampling, the plots were not used in modeling.[20]

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

[19] 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.

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

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 OLYM was probably really *E. paucicapitatus*, 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 classification of these office-created ocular records[21] 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

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

prominent species and contained the most significant species in each vegetation layer,[22] 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.[23] 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 observations 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, distributional data and community affinities.[24] 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*.

---

[22] 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.

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

[24] 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.

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.[25] 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.[26] 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.

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.[27]

---

[25] 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.

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

[27] 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

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.[28] 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[29] multiplier, applied to the transformed and standardized cover values.[30] To approximate the NVC's structural emphasis, we calculated total cover for each lifeform and for all vascular vegetation,[31] and incorporated those in the analysis as if they were additional taxa.

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

---

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.

[28] 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.

[29] 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).

[30] 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.

[31] 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.

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;[32] (b) associations distinguished from others based solely on total vegetative cover, either cumulative or in a single layer;[33] (c) associations named and defined based on the presence of a single common species (often a dwarf shrub such as *Vaccinium deliciosum* or *Juniperus communis*), regardless of the other vegetation present;[34] and (d) associations that were excessively heterogeneous in species composition (as represented in the floristic calibration plots), occurring in a variety of settings.[35]

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

---

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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).[36]

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.[37] 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 covered. For instance, we found that a trace amount of *Orthilia secunda* was a consistent indicator of the most common successional mid-slope 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.

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.[38] 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

---

[36] 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.

[37] 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).

[38] 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.

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 floristic 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.

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.[39] 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

[39] 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.

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. 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.[40] 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 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

[40] 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.

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[41] or two[42] 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 most similar map class that was mappable at that park.[43] 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 was present, modeling tendencies and full floristic character tended 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 map 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 and M95–ROADS IN PARK. Details on discriminating these and the preceding types are contained in the map class descriptions (Nielsen et al. 2021c).

---

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

[42] 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).

[43] 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 MORA map are summarized in **Table 3**. See Nielsen et al. (2021c) for detailed map class descriptions, an explanation of the map class name coding system, and crosswalks to two other prominent classifications of vegetation at Mount Rainier, Franklin et al. (1988) and Henderson (1974).

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

Map class code and full name	Other parks	Plot count
C03–Sitka spruce, western hemlock and wood-sorrel forest	OLYM	11
C04–Moist western hemlock, Douglas-fir and foamflower forest	NOCA, OLYM	62
C05–Western hemlock, Douglas-fir and sword fern forest	NOCA, OLYM	100
C06–Western hemlock, Douglas-fir and salal forest	NOCA, OLYM	67
C10–Moist silver fir, western hemlock and foamflower forest	NOCA, OLYM	68
C11–Mesic silver fir and western hemlock forest	NOCA, OLYM	117
C12–Silver fir, hemlock and Alaska blueberry forest	NOCA, OLYM	57
C13–Mountain hemlock, silver fir and Cascade azalea forest	NOCA, OLYM	85
C14–Silver fir, big huckleberry and beargrass forest	NOCA, OLYM	85
C15–Lodgepole pine and Douglas-fir woodland	NOCA, OLYM	0
C20–Subalpine fir and Sitka valerian forest and woodland	NOCA, OLYM	133
C21–Mountain hemlock, subalpine fir and heather woodland	NOCA, OLYM	42
C23–Mount Rainier subalpine fir and whitebark pine woodland	–	46
C26–Conifer krummholz and treed cliff	NOCA, OLYM	34
B30–Successional gravel bar shrubland	NOCA, OLYM	43
B31–Broadleaf riparian and swamp forest	NOCA, OLYM	40
S40W–Low elevation shrub-dominated wetland	NOCA, OLYM	24
S41W–Subalpine willow wetland	NOCA, OLYM	9
S43–Sitka alder shrubland	NOCA, OLYM	32
S45–Vine maple shrubland	NOCA, OLYM	35
S47–Successional huckleberry shrubland	NOCA, OLYM	62
S48–Subalpine heather shrubland	NOCA, OLYM	74
S49–Alpine heather shrubland	NOCA, OLYM	50
H50W–Lowland marsh and meadow	NOCA, OLYM	32
H51W–Subalpine herbaceous wetland	NOCA, OLYM	51
H53–Showy sedge and Sitka valerian meadow	NOCA, OLYM	77
H56–Subalpine summer-dry grass-forb meadow	OLYM	52
H57–Green fescue dry meadow	NOCA	60
H58–Bedrock balds and sparsely vegetated forest openings	NOCA, OLYM	9
H60W–Black alpine sedge wetland	NOCA, OLYM	21
H62–Alpine sparse herbaceous vegetation	NOCA, OLYM	34

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

Map class code and full name	Other parks	Plot count
H63–Alpine buckwheat pumice vegetation	–	11
H64–Alpine lupine pumice vegetation	–	14
R71–Alluvial barren and debris-covered ice	NOCA, OLYM	55
R72–Colluvial barren	NOCA, OLYM	71
R73–Bedrock barren	NOCA, OLYM	29
W81–Fresh water	NOCA, OLYM	81
W82–Exposed snow and ice	NOCA, OLYM	22
M92–Burned with uncertain vegetation	NOCA, OLYM	0
M93–Timberland with uncertain vegetation	NOCA, OLYM	0
M94–Development	NOCA, OLYM	0
M95–Roads in park	NOCA, OLYM	0

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 field workers unless they are carrying a mobile device. As an alternative, 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**–**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 lidar data and standard digital elevation models; (d) climate normals over the period 1981–2010; and (e) vegetation canopy information from lidar data. 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 16, 2016
Current minimum-snow image, Landsat-8	30	USGS (2019a)	Sep 27, 2015
Elevation	10	USGS (2019b)	–
Climate normals	~800	PRISM Climate Group (2019)	1981–2010
Lidar bare earth and highest hit elevation	1	Watershed Sciences, Inc.	2008

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 recent data source allowing mapping at 3-meter resolution. However, deep shadows which lowered mapping accuracy often occurred north of steep slopes (**Figure 6** illustrates this with an excerpt from NOCA). Because

the shadow locations varied between image collections, we mitigated this problem by combining the best-illuminated portions of each into a mosaic.

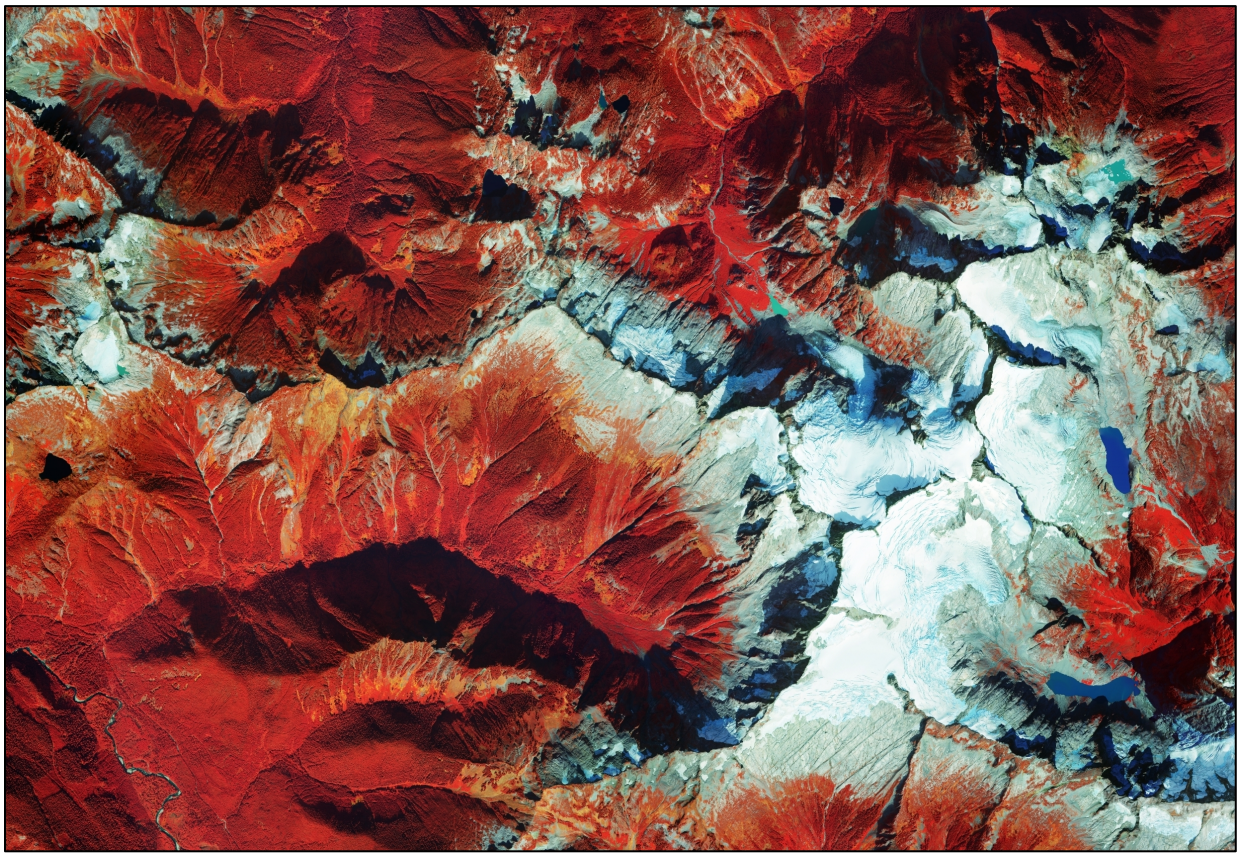


Figure 6. Aerial imagery with untreated shadows**.** The deep shadows seen in this 2015 NAIP imagery would interfere with accurate mapping unless treated.

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.

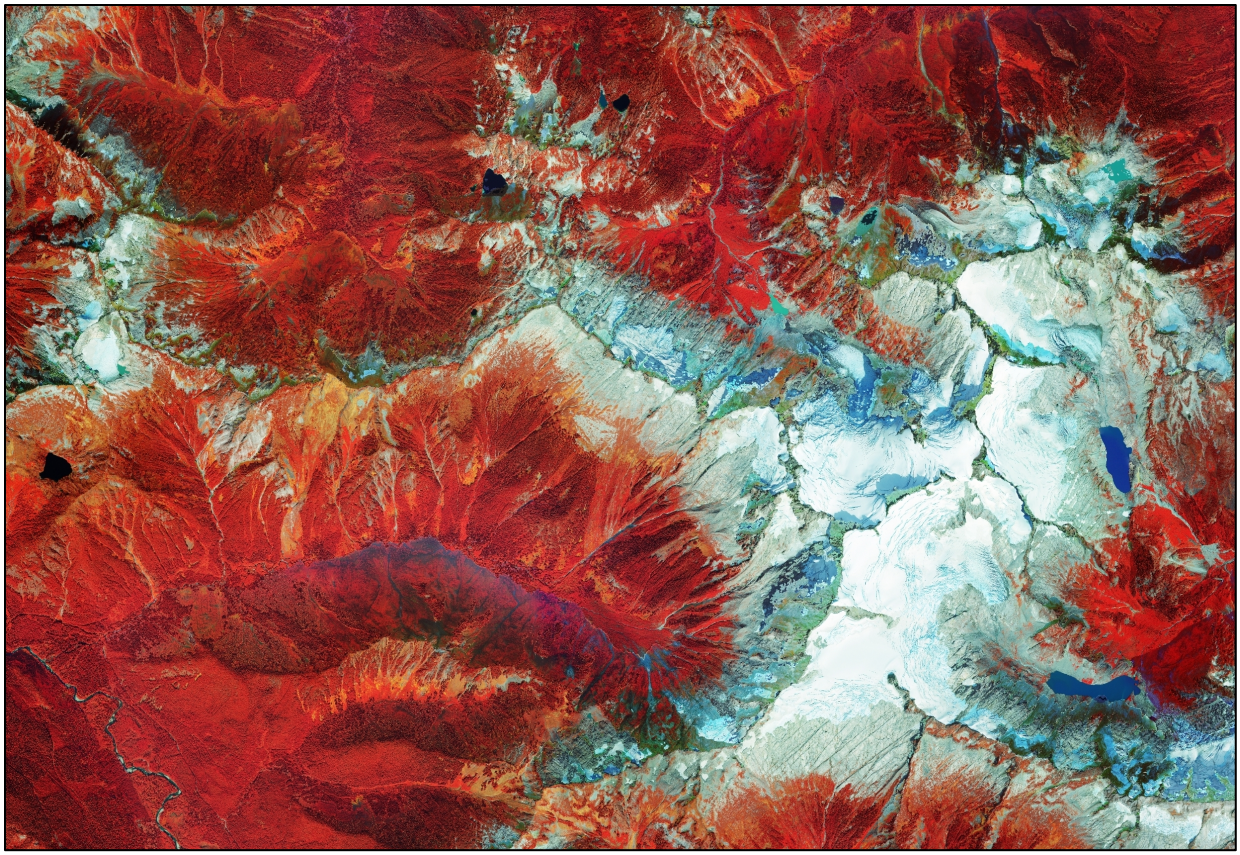


Figure 7. Aerial imagery with histogram-matched shadows**.** Here the shadowed areas have been matched to the surroundings.

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**.

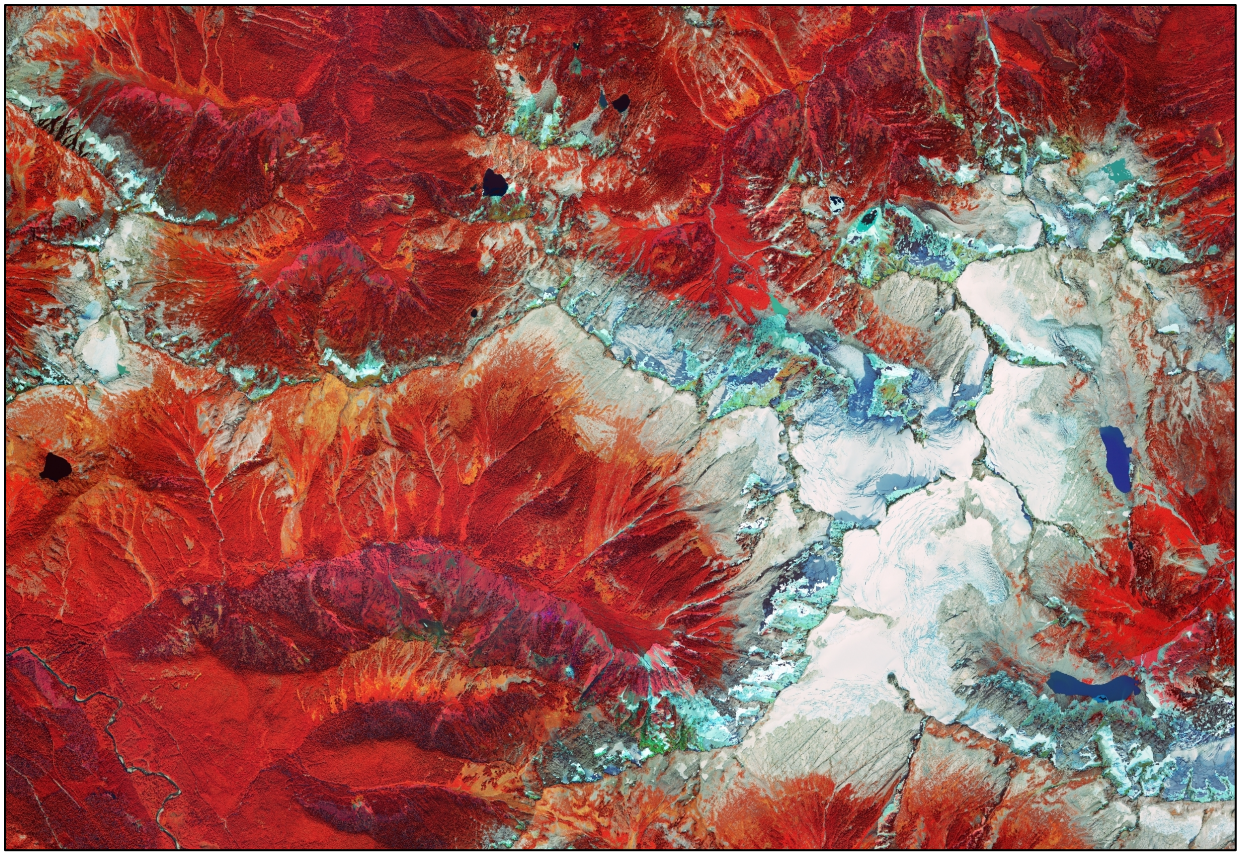


Figure 8. Aerial imagery merged across years**.** Here the shadowed areas have been filled with data from the 2009–13 images. Areas shadowed in all images were matched to the surroundings.

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 26, 2004 was used to guide sampling and produce field map sheets. A midsummer Sentinel-2 image collected on August 16, 2016 was used as the primary satellite image for modeling. A Landsat-8 image collected on September 27, 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.

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 9**).

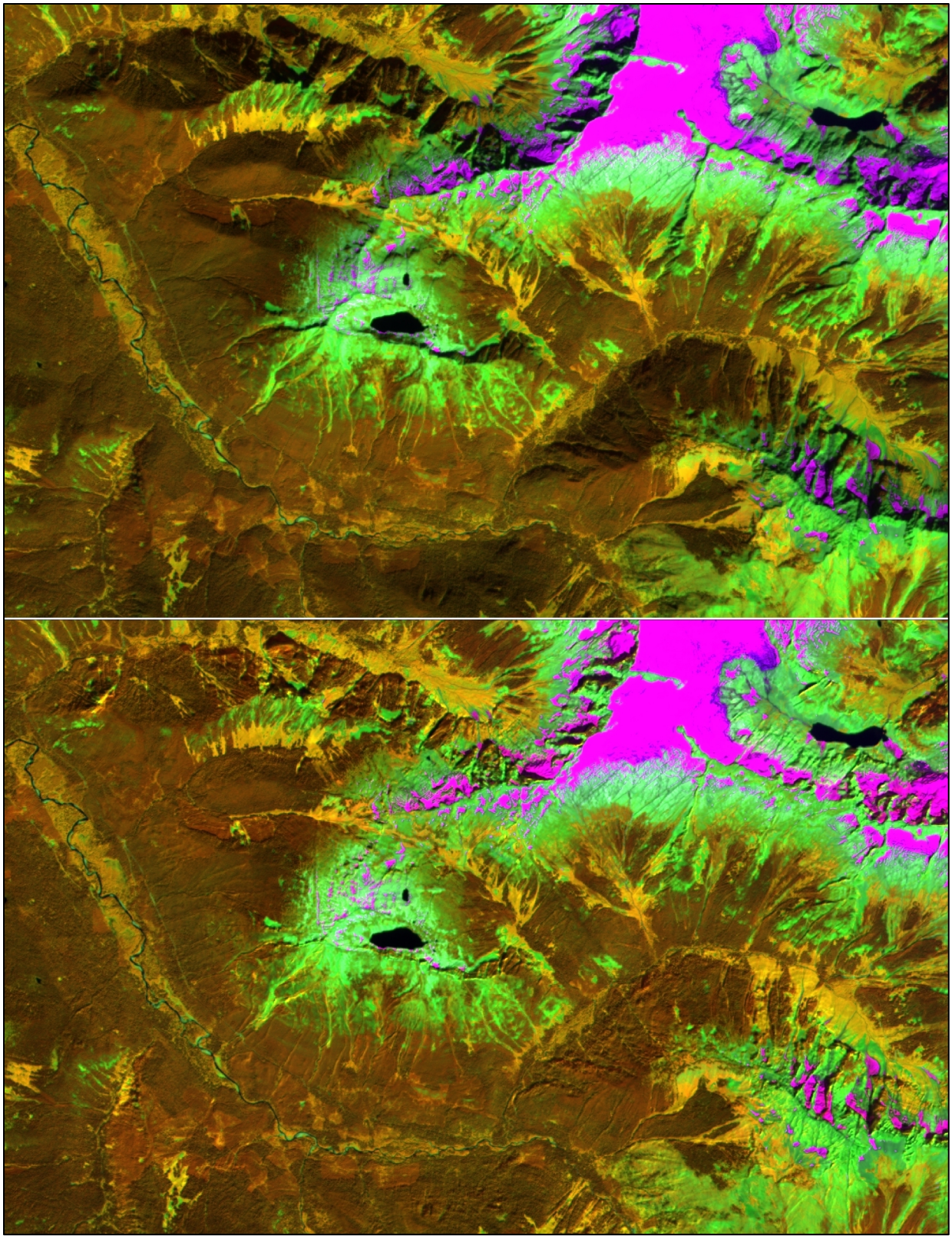


Figure 9. Sentinel-2 image before and after topographic normalization **(upper and lower images respectively).** Snow and ice appear pink, sparsely vegetated areas green, broadleaved trees and shrubs yellow, and conifers reddish-brown.

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, 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.

2.5.4. Lidar data

Lidar data were collected and processed by Watershed Sciences (now Quantum Spatial) of Portland, Oregon in the summer and fall of 2008. The collection covered nearly the entire park; only a few small areas totaling several hectares along the southeastern park boundary were omitted. Very little area outside the park was collected. We used the 1-meter gridded return intensity, bare earth elevation and highest hit elevation datasets produced by Watershed Sciences. We did not evaluate the point cloud data as the gridded products appeared sufficient for our needs. We mosaicked the tiled data into single rasters and found the height above ground of vegetation and other elements by subtracting the bare earth elevation from the highest hit (see **Figure 10**). We converted both elevation and height from floating point to integer format for data storage and processing efficiency, using vertical units of 0.25 feet for elevation and 0.01 feet for height.

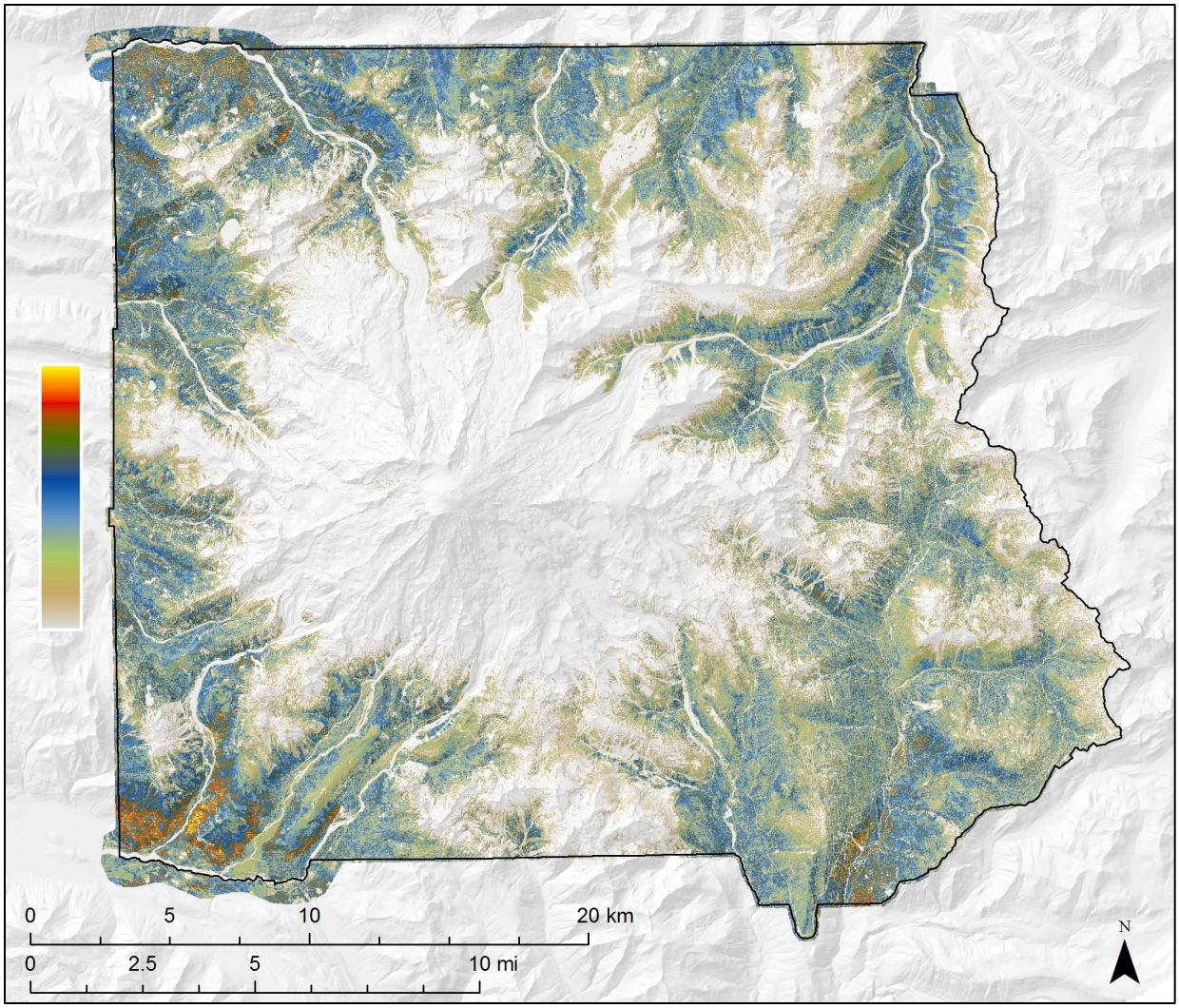


Figure 10. Lidar-derived vegetation height, ranging from zero to 200 feet along the color ramp shown.

An artifact in the vegetation height layer was observed in steep areas, where height values proportional to ground slope were found, even in areas completely barren of vegetation. Some high and steep cliffs resulted in height artifacts of 100 meters or more. Even on gentler slopes of 30 degrees or less, invalid height data on the order of several meters occurred frequently. Because of the importance of the vegetation height layer for high-resolution mapping, these artifacts presented a serious challenge.

We modeled locations at which artifacts could be expected by comparing local values of both slope and height to the median values in a moving window of twice the dimension of the typical maximum artifact width. Several related metrics were produced and visually evaluated. The most promising were combined and threshold values above which the artifact was generally present were established through visual inspection. Flagged pixels were then clumped and sieved, keeping only clumps larger than 36 m^2, since most artifacts seemed to be at least that size. Height values of the detected artifacts were recoded to missing data.

Two versions of the filtered height dataset were created, using different values of the artifact detection threshold. The first, with a low detection threshold, was used in producing canopy height predictor metrics for random forests modeling. This resulted in an aggressive process which erred toward over-filtering of the height data, with the intention of assigning missing data areas to adjacent vegetation types during the post-processing mapping phase. The second filtered height variant, with a higher detection threshold, was used in the polygon segmentation process, which required as few areas of missing data as feasible. This resulted in erring toward over-segmentation, as some artifact areas were not filtered. However, the more aggressive filter used in the modeling process prevented the erroneous height data from being used to assign a vegetation type. In general, the artifact detection model appeared to work well at high elevations in open areas but was less successful in predominantly forested areas. Canopy metrics were therefore occasionally impacted in forested areas on very steep terrain.

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 six main categories: metrics derived from aerial imagery, satellite imagery, topographic information, hydrologic information, climate data, and vegetation canopy characteristics. 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
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[44] to a variety of coarser resolutions (2, 3, 4, 6, 9, 12, 18, 27, and 36 meters). We then used three different 3x3-cell[45] 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

[44] 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.

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

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 metrics based on mathematical transformations of 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 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)

[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.

Table 6 (continued). 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
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 (**Table 7**). These were derived primarily from the lidar bare earth elevation dataset, except in areas beyond that dataset's extent, where 10-meter resolution 3DEP data were substituted. 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	3	Bare earth elevation.
slope	3	Slope in degrees (Esri 2013).
east, south	3	"Eastiness" = sin(aspect) and "southiness" = sin(aspect-90°).
cur6, cur30, cur150, cur750	6, 30, 150, 750	3x3-cell total curvature (Esri 2013) from elevation aggregated to 6m, 30m, 150m and 750m resolution.
cpl6, cpl30, cpl150, cpl750	6, 30, 150, 750	3x3-cell planimetric curvature (Esri 2013) from elevation aggregated to 6m, 30m, 150m, and 750m resolution.
cpr6, cpr30, cpr150, cpr750	6, 30, 150, 750	3x3-cell profile curvature (Esri 2013) from elevation aggregated to 6m, 30m, 150m and 750m resolution.
heat	3	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* / max elev in WA)).
mp126, mp630, mp3150	6, 30, 150	Morphometric protection (SAGA-GIS, Conrad et al. 2015) from elevation aggregated to 6m over 126m radius, to 30m over 630m radius, and to 150m over 3150m radius.
tpp60, tpp300, tpp1500, tpp7500	6, 30, 150, 750	Topographic position percentile, the percentile rank of cell elevation relative to surrounding elevations within a 60m, 300m, 1500m and 7500m radius.
tpmi60, tpmi300, tpmi1500, tpmi7500	6, 30, 150, 750	Minimum elevation differential within 60m, 300m, 1500m and 7500m. See text for methodology.
tpma60, tpma300, tpma1500, tpma7500	6, 30, 150, 750	Maximum elevation differential within 60m, 300m, 1500m and 7500m. See text for methodology.
cold60, cold300, cold1500, cold7500	6, 30, 150, 750	Cold air accumulation calculated over surrounding 60m, 300m, 1500m and 7500m. See text for methodology.
rough3, rough9, rough30, rough90, rough270	1, 3, 10, 30, 90	Surface roughness at 3m, 9m, 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:

$$ed_{\perp,min} = \min(ed_{N-S} + ed_{E-W}\ ,\ ed_{NE-SW} + ed_{SE-NW})$$

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_drain, vd_perm, vd_major	6	Vertical distance above slope concavities, permanent channels and major river networks (Conrad et al. 2015)
hd_drain, hd_perm, hd_major	6	Horizontal distance to slope concavities, permanent channels and major river networks
dtw	6	Cartographic depth to water index (White et al. 2012)
wetness	6	SAGA wetness index (Conrad et al. 2015), closely related to Compound Topographic Index (Moore 1991)
upland	6	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.

Although lidar data allows flow modeling with much greater spatial detail than a 3DEP DEM would permit, this detail can be a source of difficulties as well. For example, where culverts allow water to pass under roads and remain in its natural path, modeled flow paths may be blocked and diverted by road prisms. We reduced such problems and also eased computation by performing hydrological modeling at 6-meter resolution rather than at the full lidar resolution. We reduced the resolution by aggregating based on the minimum, setting each 6-meter cell equal to the lowest value of the 36 constituent 1-meter cells. This resulted in a greatly reduced number of obstructed flow paths in subsequent modeling, as compared to aggregating based on the mean. 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.

Despite the minimum aggregation step above, flow paths along the upslope sides of roads continued to be a problem. Mismodeled flow eventually will find its way across the road, resulting in a stream in an incorrect location. A roads layer with very high spatial accuracy was needed in order to allow a targeted approach for flow correction near roads without compromising the quality of flow paths in

other areas. No existing roads layers were sufficient, so we hand-digitized park roads based on the lidar slope image. The centers of all existing roads were digitized; old roads were also digitized if the prism was still apparent in the slope image. For all areas adjacent to roads, the 6-meter resolution elevation was set to the minimum elevation in the surrounding 3x3-cell window, in an attempt to route flow downslope across the roads. The sink filling and flow accumulation procedures above were then repeated.

This approach improved flow delineation substantially but did not resolve all road-related flow path issues. To solve the remaining problems, we manually digitized short line segments across roads in areas where substantial amounts of flow were still being incorrectly routed along roadsides. We buffered each connecting segment by six meters and set all overlapping elevation cells to the minimum elevation within the buffer. After running the sink filling and flow accumulation procedures yet again, most flow path problems appeared to be resolved.

We used the flow accumulation results to delineate channel networks, also in SAGA-GIS. Three alternate channel networks were created from the flow accumulation result. One was calibrated to represent visible slope concavities, another to represent permanent channels, and the third 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 40,000, corresponding to an average catchment area of about 140 hectares. A threshold of 8,000,000 (corresponding to about 29,000 ha) was used for major rivers; this resulted in delineation of channels downstream of the approximate location where their floodplains begin to widen substantially. We chose a threshold of 1,000 (corresponding to about 3.6 ha) for slope concavities, because it produced results that generally matched the representation of discernible hollows on the quad sheets.

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[46] 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

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

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).

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.

2.6.6. Vegetation canopy metrics

In addition to the 3-meter resolution artifact-filtered vegetation height *ht3* and return intensity *intens*, a variety of predictors were formed by summarizing canopy height information over moving windows. Other predictors were produced by summarizing over segmented polygons. **Table 10** provides an overview of the metrics representing the vegetation canopy; they are described in greater detail below.

Table 10. Vegetation canopy 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
ht3	3	Local vegetation height
intens	3	Local return intensity
ht50	16, 32, 64, polygon	Median height within 9m, 18m and 36m radius circular moving windows and within segmented polygons
dif50	3	Difference between local vegetation height and median height within 9m, 18m and 36m radius circular moving windows
htmx	16, 32, 64	Maximum height within 9m, 18m and 36m radius circular moving windows
ht75, ht88	polygon	75th and 88th percentile height within segmented polygons
htsd	16, 32, 64, polygon	Standard deviation of height within 9m, 18m and 36m radius circular moving windows and within segmented polygons
htt50, htw50	16, 32, 64, polygon	Median height value of heights of tree-sized (over 5m) and woody-sized (over 80cm) vegetation, within 9m, 18m and 36m radius circular moving windows and within segmented polygons
cct, ccw	16, 32, 64, polygon	Fractional cover of tree-sized and woody-sized vegetation, within 9m, 18m and 36m radius circular moving windows and within segmented polygons
errmn	16, 32, 64, polygon	Canopy elevation relief ratio (see text) within 9m, 18m and 36m radius circular moving windows and within segmented polygons
err88	polygon	Canopy elevation relief ratio for 88th percentile height (see text) within segmented polygons
h1b3b	16, polygon	Height-based NDTI of 1m ‘b’ texture metric contrasted with 3m ‘b’ texture metric, summarized over 9m radius circular moving window and within segmented polygons
h2b6b	16, polygon	Height-based NDTI of 2m ‘b’ texture metric contrasted with 6m ‘b’ texture metric, summarized over 9m radius circular moving window and within segmented polygons
h3b9b	16, polygon	Height-based NDTI of 3m ‘b’ texture metric contrasted with 9m ‘b’ texture metric, summarized over 9m radius circular moving window and within segmented polygons
h4bcb	16, polygon	Height-based NDTI of 4m ‘b’ texture metric contrasted with 12m ‘b’ texture metric, summarized over 9m radius circular moving window and within segmented polygons
h6bdb	18, polygon	Height-based NDTI of 6m ‘b’ texture metric contrasted with 18m ‘b’ texture metric, summarized over 9m radius circular moving window and within segmented polygons
h9beb	27, polygon	Height-based NDTI of 9m ‘b’ texture metric contrasted with 27m ‘b’ texture metric, summarized over 9m radius circular moving window and within segmented polygons
hcbfb	36, polygon	Height-based NDTI of 12m ‘b’ texture metric contrasted with 36m ‘b’ texture metric, summarized over 9m radius circular moving window and within segmented polygons

Basic moving window metrics

Several summary metrics were generated over circular moving windows of 9-meter, 18-meter and 36-meter radius. To minimize artifacts near vegetation transitions, most were based on the median rather than the mean.

The predictors *ht50*, *htmx*, and *htsd* were calculated respectively as the median value, maximum value, and standard deviation of the heights within the circular window. The height difference between the center 3-meter cell and the median height in the surrounding variable-sized window was calculated as *dif50*.

We created several predictors to describe the height and density of the upper vegetation layers, if those layers were present within the circular window. The median tree height *htt50* was calculated as the median value of the height values that exceeded the tree height threshold of five meters. Rather than indicating the height of the dominant vegetation, *htt50* describes the height of any tree layer present, no matter how sparse it is. The median woody vegetation height *htw50* was defined similarly, but using a threshold of 80 centimeters, assuming that most height values over that correspond to woody vegetation of some kind. Tree and woody vegetation canopy cover (*cct* and *ccw*) were defined as the fraction of the window with height values over five meters and 80 centimeters, respectively.

Another descriptor of vertical canopy structure was derived based on the elevation-relief ratio of Pike and Wilson (1971). The mean canopy elevation relief ratio *errmn* specifies the fractional distance of the mean canopy height *htmn* between the minimum *htmin* (usually zero) and maximum height *htmx* in the focal window. It was computed by:

$$errmn = \frac{htmn\ - htmin}{htmx - htmin}$$

Most of these metrics were also summarized over polygons, described below.

Texture metrics

The NDTI was computed based on the 'b' texture metric from the height raster, at the same range of spatial scales as described in **Section 2.6.1**. These metrics represent measures of vertical canopy roughness at different horizontal scales. Finer scale NDTI metrics are sensitive to individual trees and small canopy openings, while coarser scale metrics are more responsive to variability in overall canopy height and larger openings. Several other roughness metrics were investigated, such as aerodynamic roughness (e.g., Menenti and Ritchie 1994) and fractal dimension (e.g., Isaaks and Srivastava 1988), but the gain in information expected from these metrics appeared to not justify the cost and complexity of producing them across the project area. The texture metrics were summarized over 9-meter circular moving windows and also over polygons, described below.

Polygon summaries

Hypothesizing that forest stands discernible on the basis of canopy structure would provide useful mapping units, we applied an automated segmentation process to the canopy height data to break forested areas into distinct homogeneous-appearing stands. These polygons were then used as stand-scale summary units for canopy metrics. Because the polygon-based summaries only incorporated

metric values from within each delineated stand, they were immune to the edge effects to which the moving window summaries were susceptible.[47]

Canopy segmentation descriptors

Local canopy height provides an obvious primary basis for delineating distinct forest stands, but there are two additional needs to produce a useful forest segmentation in Pacific Northwest forests. First, the segmentation must be sensitive to horizontal structural characteristics such as variations in tree spacing and canopy cover. Second, the process must be nonlinear with respect to height: it should be more sensitive to a given shift in canopy height for a stand at the lower end of the height spectrum (e.g., hardwoods, young conifer stands) than for tall, mature conifers. We met the first need by using a metric describing local variance in height, and the second by creating additional versions of the height and variance metrics based on log-transformed height.

We began with the lightly artifact-filtered vegetation height dataset created during the lidar pre-processing phase (**Section 2.5.4**), referred to as *vh* here. We first created a log-transformed version of vegetation height (*log(vh)*), and then computed a moving window 3x3-cell standard deviation of height (σ_{vh}) and log-transformed height ($\sigma_{log(vh)}$). We then used a 14-meter radius circle as a moving window to produce median-smoothed versions of *vh*, *log(vh)*, σ_{vh}, and $\sigma_{log(vh)}$. The focal operation aggregated the descriptors over a small forested area and reduced artifacts remaining in the filtered vegetation height dataset. The window size was chosen so that multiple tree crowns would be encompassed even in old forests, causing the smoothed metrics to vary only gradually except when encountering a significant transition in forest structure. The median-based filter preserved sharp boundaries at major transitions, where the use of a focal mean would have resulted in smearing of the metrics and consequent inability of the subsequent segmentation to locate boundaries precisely.

We rescaled the four resulting descriptors to standardized 8-bit values to ensure that all layers would receive equal weighting during the segmentation process. Cells with missing data were assigned to the maximum 8-bit value of 255. The rescaling was designed so that the highest valid values in each descriptor equaled approximately 240. The resulting value gap between valid and invalid data ensured that the segmentation process would distinguish areas of missing data from the valid data around them.

Canopy segmentation

To determine the optimal segmentation parameters, we created a subset of the segmentation metrics image for a small but diverse test area, and imported it into eCognition Developer. Its **multiresolution segmentation** method allowed specification of four main parameters: *weightings* (one for each of the input descriptors), *scale*, *shape*, and *compactness*. Through experimentation we determined that *shape* = 0.1 and *compactness* = 0.9 produced the most realistic and aesthetically

[47] In contrast to typical segmentation-based workflows, we did not use polygons as prediction units; we used them only to produce additional 3-meter resolution pixel-based predictors using an alternate method for summarizing canopy height information.

pleasing outputs across the range of the other parameters. Therefore, our main effort was concentrated on determining the optimal settings for *weighting* and *scale*.

We ran segmentations using 15 different test sets of weightings for the four descriptors, using a *scale* of 50 for all runs. The test sets were designed to sum to the same value, resulting in similar numbers of polygons being created in each run. Examining the resulting polygon sets, we selected ten locations with distinct types of transitions in tree-dominated vegetation that were often delineated incorrectly, and assessed the number of these transitions correctly flagged in each run. Three sets of weightings performed equally well, so we averaged their weightings and performed the test again. The averaged combination, with *weightings*[*vh*, *log(vh)*, σ_{vh}, $\sigma_{log(vh)}$] = [8, 6, 6, 4], performed as well or better than those it was derived from. We next ran the averaged weightings over a larger test area for a range of *scale* values, gradually reducing it until the size of the resulting polygons had generally fallen below the 30-meter resolution of the Landsat imagery that we anticipated would be important in later predicting the map class for each polygon. We took this as an indication of an overly fine segmentation, and selected the next coarsest value for *scale*, 40. We then used these parameters to segment the entire project area, producing a total (n) of about 280,000 polygons with an average area ($\bar{A}$) of approximately 3500 m^2.

In areas where the descriptor values changed rapidly, eCognition frequently produced multiple narrow parallel polygon sections with boundaries perpendicular to the transition direction. These generally did not correspond to any actual land cover, and frequently inappropriately connected two more distant patches into a single polygon. To address this issue, we filtered the raw polygons to eliminate their narrowest sections and merge them with adjacent larger polygons. The width of the filtering buffer was chosen to eliminate polygon constrictions, while preserving most linear tree-dominated patches (e.g., subalpine ribbon forests, riparian woody vegetation). All areas within four meters of polygon boundaries were dissolved into the adjacent polygon whose interior portion was nearest. This resulted in the elimination of many polygon sections narrower than about eight meters. Although portions of some linear features were eliminated in this process, most would not have been accurately classified due to their small size in relation to most predictor datasets. The number of polygons increased after filtering ($n \approx 385{,}000$; $\bar{A} \approx 2550$ m^2), because many polygons were broken into multiple patches after removal of the connecting constrictions.

A sample illustration of the final polygons in a forest-dominated area is shown in **Figure 11**, superimposed on lidar-derived canopy height and aerial imagery. The polygons capture the main variations apparent in imagery, and successfully discriminate between a range of conifer stands of different ages and densities, in addition to more recently disturbed avalanche chutes dominated by *Alnus viridis* with varying amounts of conifer regeneration, small meadow openings, and sparsely vegetated areas in a variety of settings.

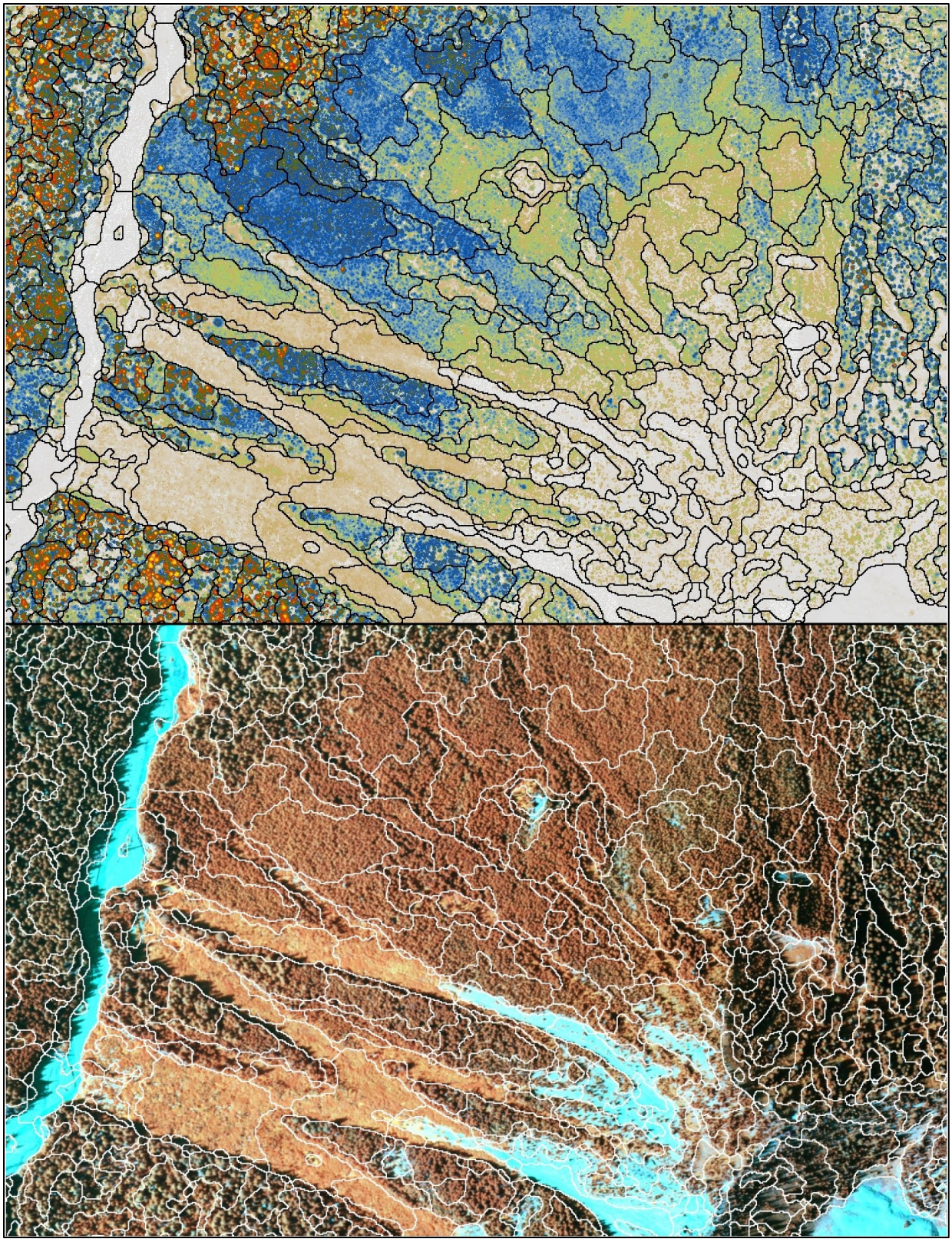


Figure 11. Polygons resulting from segmentation of lidar metrics, superimposed on vegetation height (top; color ramp as shown in Figure 10) and false-color NAIP imagery (bottom).

Metrics

Many of the summary metrics discussed above—*ht50*, *htsd*, *htt50*, *htw50*, *cct*, *ccw*, *errmn*, and the texture metrics—were created over the polygons in addition to the moving windows. Some additional metrics were created specifically for use with the polygons. The 75th and 88th percentile height values within each polygon were determined by repeatedly computing the median of the height pixels over the polygons. After finding the first median, *ht50* (discussed earlier), all cells with heights less than *ht50* for the polygon in which they occurred were masked. Taking the median of the remaining cells resulted in the 75th percentile height (*ht75*). The process was repeated once more, resulting in the 88th percentile height (*ht88*). The height of the dominant vegetation layer at most sites corresponds well to *ht88*, so it is useful for discriminating between distinct vegetation physiognomies at the polygon scale. Finally, a correlate to *errmn* was created based on *ht88*. The 88th percentile height elevation relief ratio *err88* was computed by:

$$err88 = \frac{ht88 - htmin}{htmx - htmin}$$

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.

The vegetation canopy predictors described in **Section 2.6.6** could only be produced where lidar data were available. This forced us to produce a separate set of models in order to map the sections of the project area that lacked lidar. The steps described in **Sections 2.7.3–6** were performed twice: the *lidar run* included the lidar predictors, and was used wherever lidar data were available and appeared reliable; the *no-lidar run* left them out, and was used only where lidar data were unavailable or unreliable.

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 and lidar; 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**).[48] 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.[49] 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 38 distinct map class calls; flowing water and impounded water were modeled separately, but were later merged into the single map class W81–FRESH WATER.

[48] 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.

[49] 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.

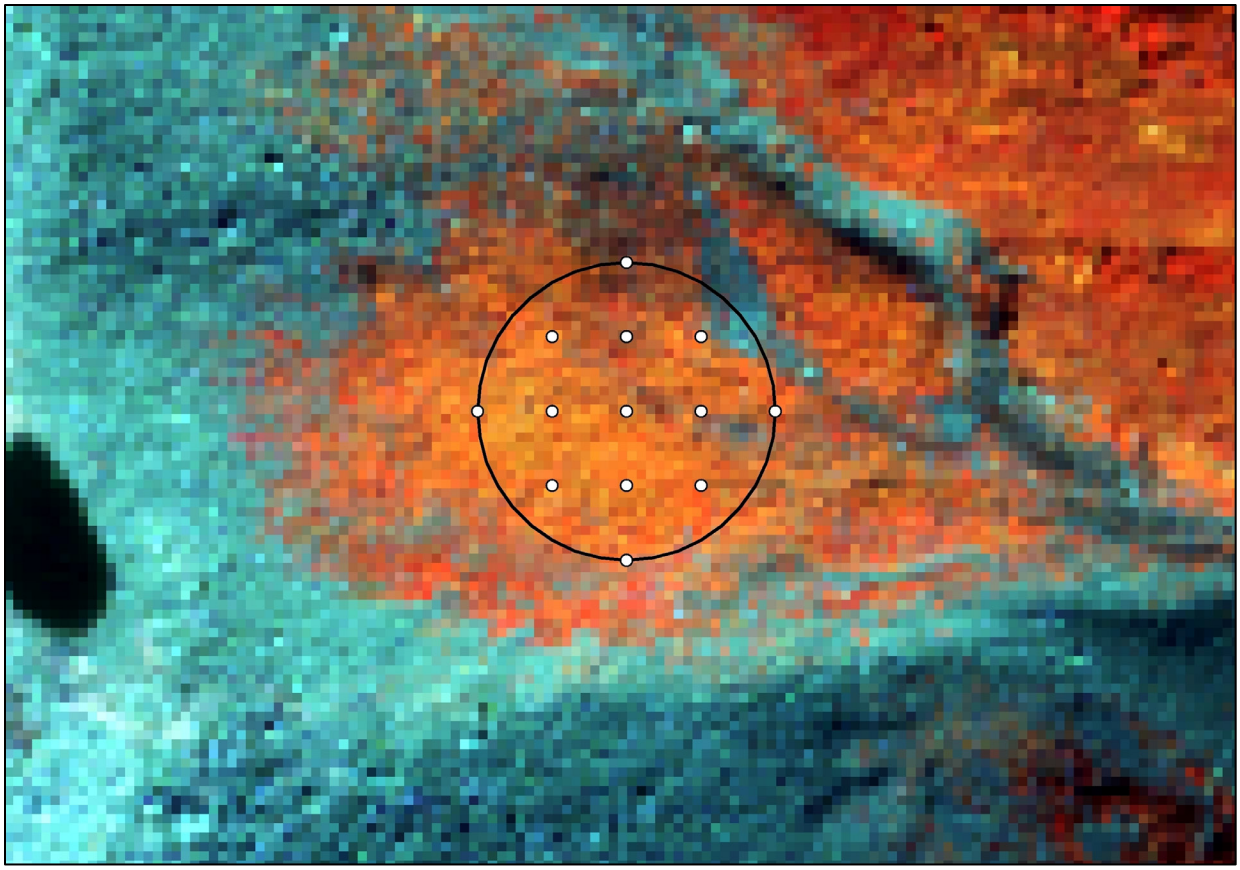


Figure 12. Training data extraction. Predictors were extracted from the 3-meter resolution metrics at 13 points distributed across each training circle. The vegetation patch represented was assigned to H53–SHOWY SEDGE AND SITKA VALERIAN MEADOW; the imagery is color-infrared 2015 NAIP at 1-meter resolution.

2.7.3. Model binarization

Each two-class combination of the 38 modeled map classes was treated separately, resulting in 703 distinct binary models.[50] 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

[50] 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[51] 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.[52] We organized the predictors into ten tiers based on the effective spatial resolution at which they were calculated,[53] with the finest scale predictors—the 3-meter resolution NAIP band responses and lidar-derived canopy height—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*[54] 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.[55] This probability estimate was then averaged across all plots to produce an overall error rate estimate for the model including the newly introduced predictor.[56] 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

[51] 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.

[52] 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.

[53] 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.

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

[55] 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.

[56] 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.[57]

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**; from lidar run).

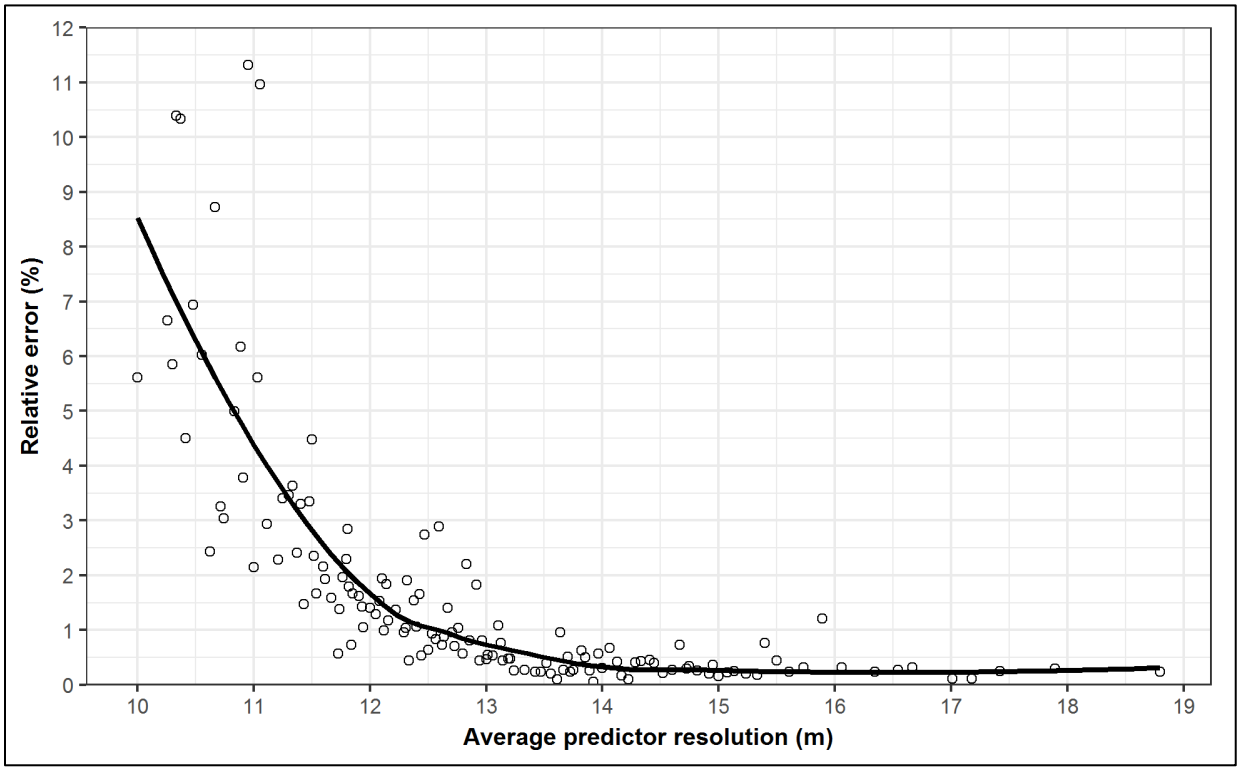


Figure 13. Average predictor resolution vs. median relative model error (the error increase attributable to leaving out coarser predictors). Predictors finer than 10-m resolution were treated as 10 m here; greater averages indicate increasing incorporation of coarser predictors in the model. The data were derived from only models with final error of 1% or greater, with a median final error of 3.5%; this would have been 12.0% using only predictors of 10-m or finer resolution. The best fit line is from a loess smoothing function.

[57] 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.[58] 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.[59] We did this because the training samples were not perfectly pure[60] 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 11** lists the most frequently included predictors across all binary models.

---

[58] 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.

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

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

Table 11. Most frequently used predictors in each selection tier (lidar run). 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[A] and number of models in which used (in parentheses)
1	3 meters	*c_ht3* (518), *c_intens* (425), *n_d1c* (411), *n_e1c* (308), *r_v1_mx* (237), *r_n1_mx* (208), *r_x1_mx* (198), *r_u1_md* (195), *n_v1b* (194), *r_y1_md* (175)
2	6 meters	*n_wa_13* (218), *n_nc_13* (216), *n_nb_13* (213), *n_gc_13* (208), *n_wb_13* (203), *n_wc_13* (203), *n_uc_13* (200), *n_rb_13* (199), *n_vc_13* (199), *n_rc_13* (192)
3	9–10 meters	*s_ndgb* (210), *s_nir* (184), *n_d3c* (107), *n_e3c* (92), *n_g3c* (75), *n_r3c* (65), *n_g3a* (62), *n_r3b* (61), *n_g3b* (60), *n_n3c* (57)
4	12 meters	*n_wa_26* (68), *n_ua_26* (58), *n_vb_26* (52), *n_wb_26* (50), *n_na_26* (48), *n_nb_26* (48), *n_d4c* (46), *n_gc_26* (45), *n_ra_26* (44), *n_uc_26* (43)
5	18–20 meters	*s_tcw* (98), *c_dif50_9* (81), *c_htsd_9* (75), *c_ccw_9* (71), *s_ndmi* (67), *c_h1b3b_9* (59), *s_sw1* (55), *c_errmn_9* (44), *c_h3b9b_9* (41), *c_cct_9* (37)
6A	topographic	*t_elev* (290), *t_tpp7500* (90), *t_raddir* (68), *t_tpma7500* (68), *t_topodry* (65), *t_cpr750* (56), *t_rough30* (56), *t_mp3150* (48), *t_tpp300* (48), *t_rough3* (41)
6B	hydrologic	*h_wetness* (84), *h_vd_major* (80), *h_hd_major* (71), *h_dtw* (61), *h_upland* (51), *h_vd_perm* (48), *h_vd_drain* (36), *h_hd_perm* (18), *h_hd_drain* (17)
10	climate	*p_ppt_jul* (19), *p_vmax_jul* (19), *p_tdew_jul* (17), *p_vmax_jan* (15), *p_vmax_oct* (14)

[A] Predictor names are preceded by a letter indicating to which source group they belong: "c" indicates vegetation canopy, "e" is 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,[61] using the predictors selected above and specifying a sample size for each class equal to the minimum number of training plots available for either class.[62] The resulting model was saved for use later in the prediction phase. We then

[61] 39 trees were generated for each of the 13 sample points at each training plot.

[62] 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

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 in the no-lidar run. Of the 703 models, 36% had an error rate of zero. Substantial error is concentrated in a fairly small number of models; 94% of the models showed less than 10% error.

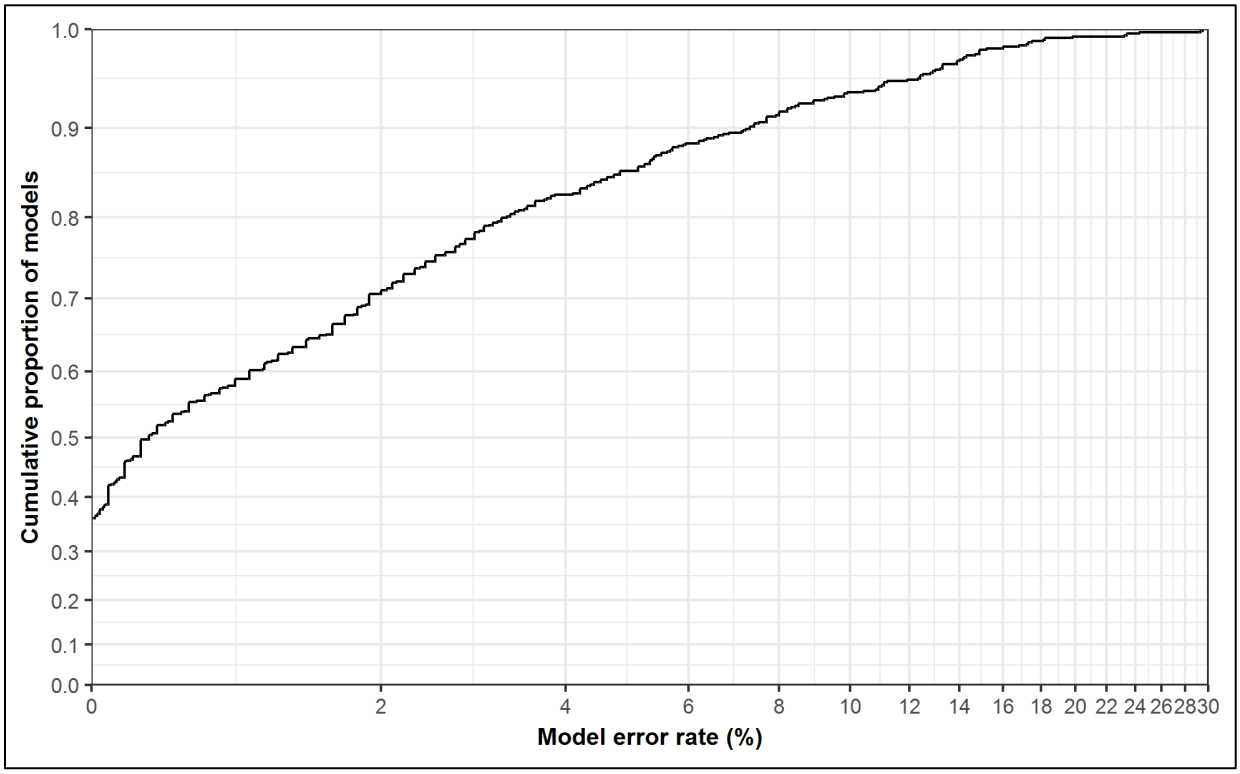


Figure 14. Error rate across all binary models (no-lidar run). Recall that due to heterogeneity within training samples (i.e., "inclusions"), some model error is to be expected.

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 accuracy assessment; for those, model error may be useful supplementary information. **Table 12** lists the 40 models with the highest error rates in the lidar run, used to provide map predictions over most of the park.

more common class (see Evans and Cushman 2009). The same technique was used earlier during the predictor selection phase.

Table 12. Binary models with highest cross-validated error rates (lidar run).

Map class 1 / Map class 2 (codes and abbreviated names)	Error rate (%)
H60W–Black alpine sedge / S49–Alpine heather	29.1
C10–Moist silver fir & foamflower / C11–Mesic silver fir & w hemlock	25.8
C05–W hemlock & sword fern / C06–W hemlock & salal	23.9
H56–Subalpine summer-dry grass-forb / H57–Green fescue dry meadow	17.6
H62–Alpine sparse herbaceous / R72–Colluvial barren	17.4
C12–Silver fir & Alaska blueberry / C14–Silver fir & big huckleberry	16.9
H60W–Black alpine sedge / H62–Alpine sparse herbaceous	16.9
C05–W hemlock & sword fern / C11–Mesic silver fir & w hemlock	16.7
H51W–Subalpine herbaceous wetland / S41W–Subalpine willow wetland	16.6
H53–Showy sedge & Sitka valerian / S48–Subalpine heather	16.0
C20–Subalp fir & Sitka valerian / C23–Mt Rainier subalp fir & whitebark pine	15.2
S48–Subalpine heather / S49–Alpine heather	15.1
H60W–Black alpine sedge / S48–Subalpine heather	15.0
R72–Colluvial barren / R73–Bedrock barren	14.7
C03–Sitka spruce & wood-sorrel / C04–Moist w hemlock & foamflower	14.5
H62–Alpine sparse herbaceous / S49–Alpine heather	14.3
C12–Silver fir & Alaska blueberry / C13–Mtn hemlock & Cascade azalea	14.2
H56–Subalpine summer-dry grass-forb / S49–Alpine heather	14.1
H51W–Subalpine herbaceous wetland / H53–Showy sedge & Sitka valerian	14.0
C04–Moist w hemlock & foamflower / C10–Moist silver fir & foamflower	13.3
H56–Subalpine summer-dry grass-forb / S48–Subalpine heather	13.2
H57–Green fescue dry meadow / S48–Subalpine heather	12.7
C13–Mtn hemlock & Cascade azalea / C20–Subalp fir & Sitka valerian	12.6
C04–Moist w hemlock & foamflower / C11–Mesic silver fir & w hemlock	12.6
H51W–Subalpine herbaceous wetland / H56–Subalpine summer-dry grass-forb	12.6
C20–Subalp fir & Sitka valerian / C21–Mtn hemlock & heather	12.5
S47–Successional huckleberry / S48–Subalpine heather	12.4
S41W–Subalpine willow wetland / S48–Subalpine heather	11.8
H53–Showy sedge & Sitka valerian / H57–Green fescue dry meadow	11.7
C11–Mesic silver fir & w hemlock / C12–Silver fir & Alaska blueberry	11.5
H51W–Subalpine herbaceous wetland / H60W–Black alpine sedge	11.4
H53–Showy sedge & Sitka valerian / H56–Subalpine summer-dry grass-forb	11.2
H50W–Lowland marsh & meadow / S40W–Low elevation shrub wetland	11.1
R71–Alluvial barren / R72–Colluvial barren	11.0
R71–Alluvial barren / W81–Fresh water	10.7
C04–Moist w hemlock & foamflower / C05–W hemlock & sword fern	10.3
H57–Green fescue dry meadow / S47–Successional huckleberry	10.2

Table 12 (continued). Binary models with highest cross-validated error rates (lidar run).

Map class 1 / Map class 2 (codes and abbreviated names)	Error rate (%)
H53–Showy sedge & Sitka valerian / S47–Successional huckleberry	10.2
H56–Subalpine summer-dry grass-forb / H60W–Black alpine sedge	9.9
C10–Moist silver fir & foamflower / C12–Silver fir & Alaska blueberry	9.7

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.

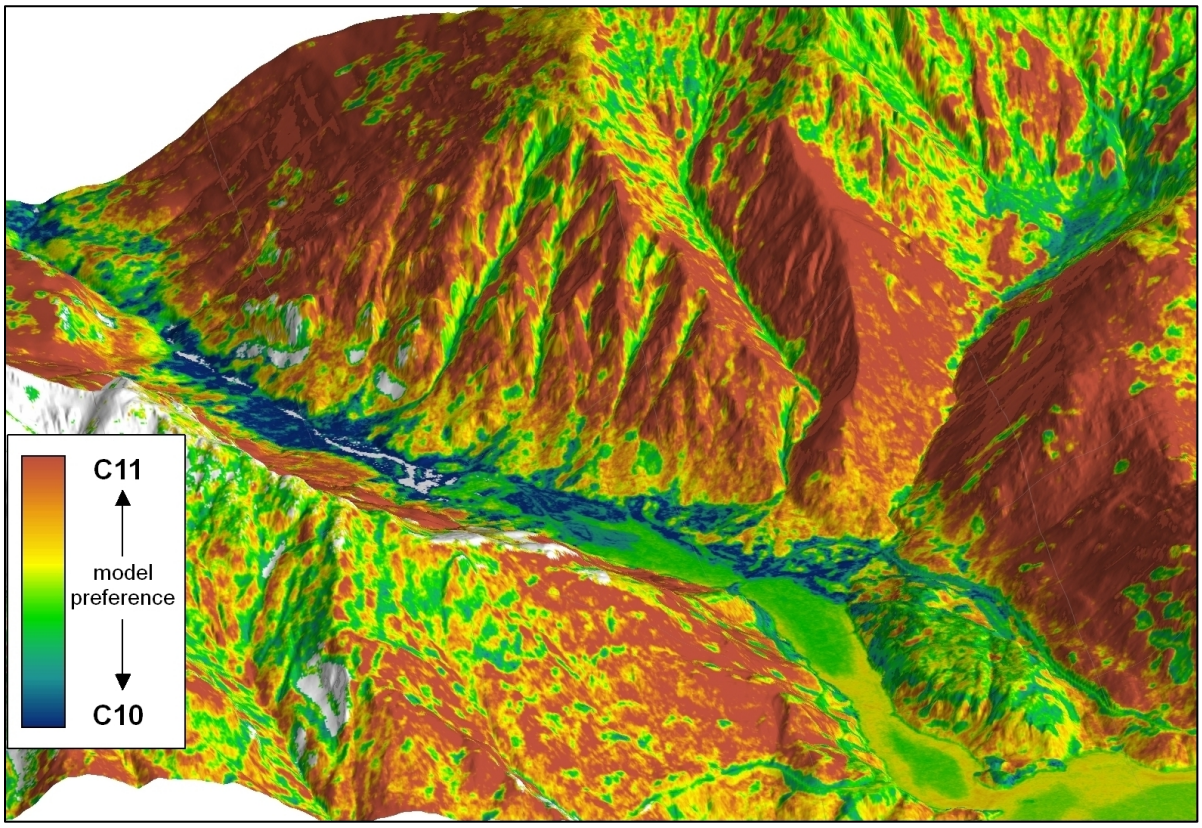


Figure 15. Binary model prediction example, for the model C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST versus C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST. C10 was favored mostly on toe slopes and valley bottoms, while C11 was preferred on middle and higher slopes. White areas were 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.[63] The total loss margin, seen in **Figure 16**, can be interpreted as a map of model uncertainty.

[63] 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.

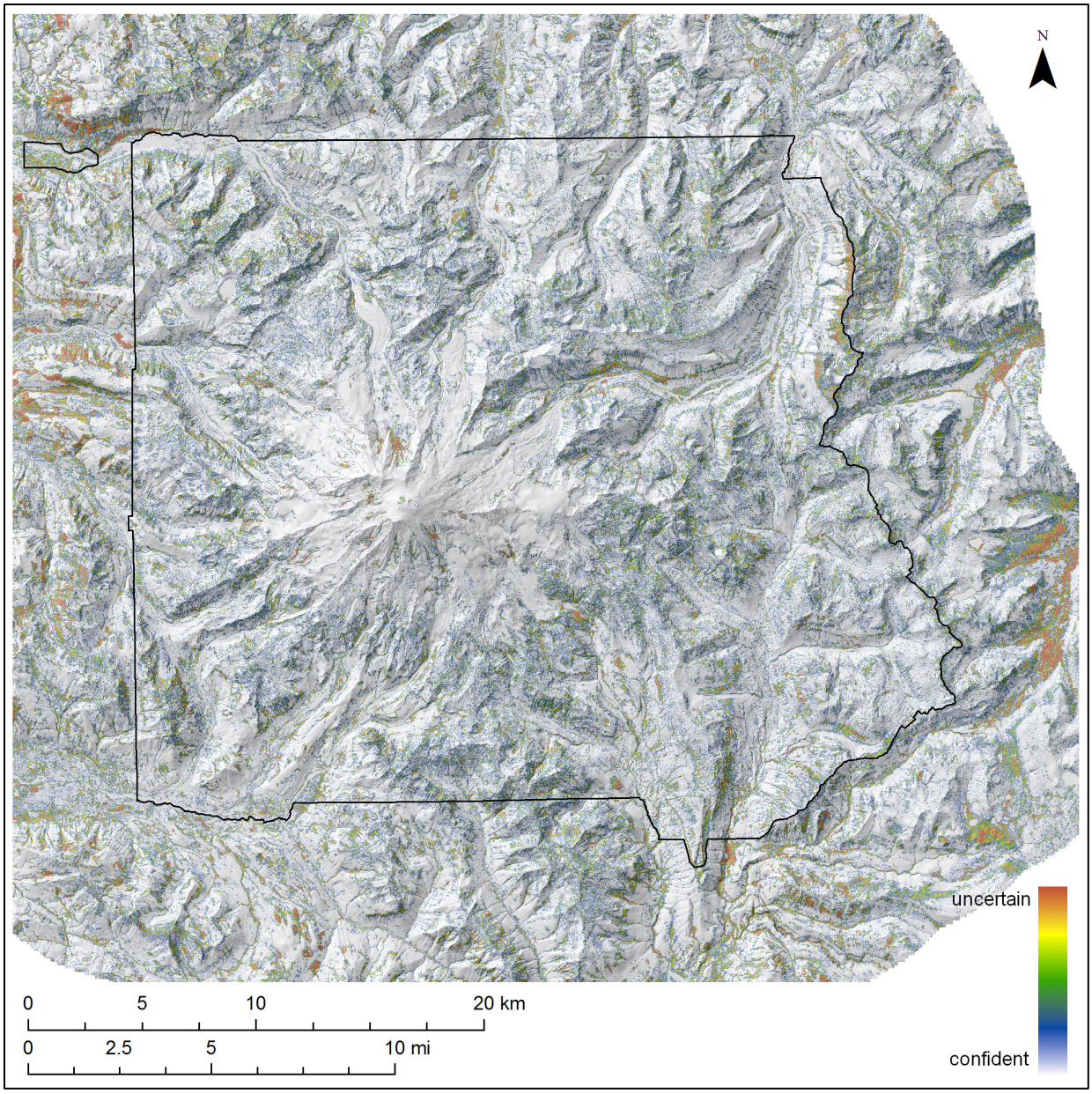


Figure 16. Prediction uncertainty in the no-lidar run. The predicted class won all contests in areas displayed as white. Colors from blue to red indicate that the best class lost at least one contest, by increasing amounts. Certainty is lowest where training data were inadequate, especially outside the park.

After predicting both the lidar and no-lidar runs, we created the final modeled map by using the lidar run result wherever it was available and the no-lidar result elsewhere.

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 and M95–ROADS IN PARK—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[64] (e.g., S45–VINE MAPLE 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 two map classes to encompass these: M95–ROADS IN PARK (representing roads within the park boundaries only) and M94–DEVELOPMENT (representing everything else, including other development within the park).

We began by digitizing the roads and developed areas inside the park (using lidar data where available and 2015 NAIP otherwise), 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

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

to M94–DEVELOPMENT unless they had modeled as a forest, woodland, or dense tall shrubland, 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 park were also included in M94–DEVELOPMENT. Digitized roads within the park were included as a separate map class, M95–ROADS IN PARK.

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 mapped occurrences of C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND by scouring the park, armed with field data, aerial imagery, lidar, and INR and NPS ecologists' local knowledge, and digitizing them at 1:4,000 scale. We scanned the remainder of the polygon map for obvious errors, correcting them by hand by either recoding entire polygons or splitting them into pieces first. Issues that frequently required correction included areas that had changed substantially over the decade since lidar acquisition—especially along flood-ravaged portions of the Carbon and Nisqually Rivers—and poor mapping along the edges of roads in the park. We then converted back to pixels and ran a final 3-by-3 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.[65] NPS standards specify an 80% accuracy goal for each individual map class hosting native vegetation communities. In addition to assessing the

[65] Project management considerations required doing the field portion of the AA many years earlier than would have been ideal. At the time of the AA fieldwork we had yet to create the map classification, and were using a draft version of the alliances (NatureServe 2012) and the provisional NVC associations (Crawford et al. 2009), which were later discovered to have a number of issues (necessitating our creation of the mapping associations). The early draft map we relied on for our sample design in some areas bears little resemblance 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), relying substantially on later work in the office to arrive at a best call using the final classification. 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 INR (2021a). 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.

class-specific UA and PA against this standard, we used UA and PA in combination to produce an 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. The map classes had not yet been determined at the time of field sampling, so we targeted areas mapped to our understanding at the time of the alliances that were later described in NatureServe (2012). In general, our targets represented all natural vegetated classes present in the final map except for C03–SITKA SPRUCE, WESTERN HEMLOCK AND WOOD-SORREL FOREST, B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND, S40W–LOW ELEVATION SHRUB-DOMINATED WETLAND, S41W–SUBALPINE WILLOW WETLAND, H50W–LOWLAND MARSH AND MEADOW and M92–BURNED WITH UNCERTAIN VEGETATION. However, because several of the targeted alliances were subsequently split or otherwise reorganized into multiple map classes, some map classes received significantly less sampling effort than others. We also targeted two nominally abiotic classes, R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE and R72–COLLUVIAL 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 and H62–ALPINE SPARSE HERBACEOUS VEGETATION. Accurate mapping of these abiotic classes is necessary in order to accurately map similar vegetated classes. While R73–BEDROCK BARREN also has the potential for confusion with vegetated map classes, it is difficult and often unsafe to access, and we deemed it not a good use of field time to attempt it. All the targeted 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 by buffering park roads and trails by 100 meters.[66] This very narrow buffer was necessitated by spatial inaccuracies in the park trails data, which often exceeded 200 meters,[67] and the recognition that the true distance from the trail might be as large as the summed buffer distance and the local trail spatial inaccuracy. We chose not to use a slope-based cost distance model because of concerns about the relative spatial accuracy of the trails layer and the topographic data. The resulting region (see **Figure 17**) spanned 11,759 hectares, 12.3% of the total area of the park.[68] 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).

---

[66] 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 field 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.

[67] After the 2011 AA fieldwork, but before returning to the field for additional training data collection in 2014, we substantially improved the trails data by digitizing problem areas using the lidar bare earth data.

[68] Although this falls well short of the Lea and Curtis (2010) standard, it was impossible to achieve sampling productivity goals with a larger inference area (see **Table 13** for a summary of the per-class inference area

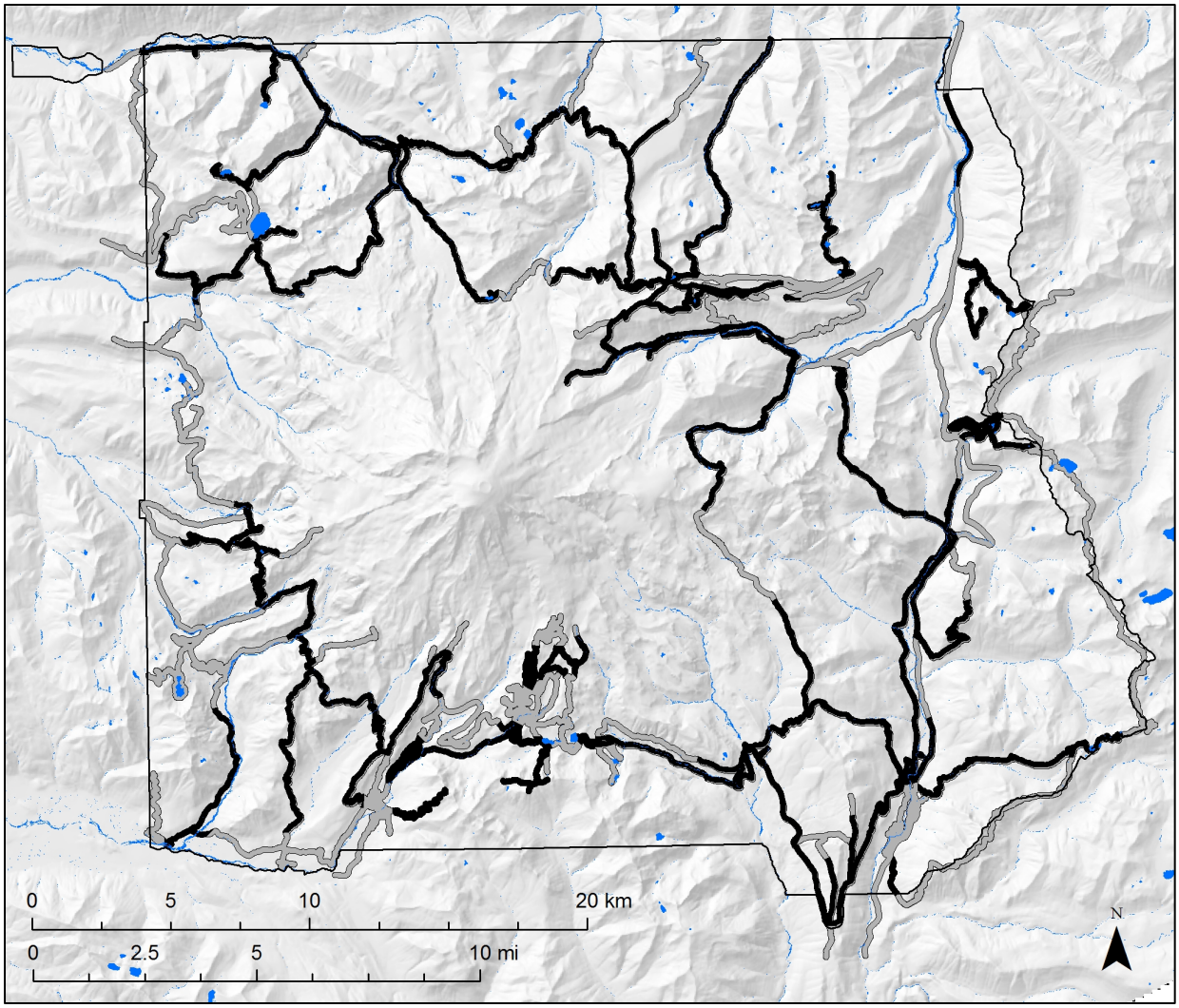


Figure 17. Accuracy assessment inference area. Field samples were acquired within the reached inference area, shown in black. The attempted inference area also includes gray-shaded areas which were not accessed by field crews.

3.2.2. Sample selection

A La Niña weather pattern resulted in exceptionally heavy snowpack that persisted in much of the park through the entire summer of 2011, the main year designated for AA fieldwork. These conditions made it extremely difficult to obtain an adequate, geographically representative sample even for the more common vegetation types. Access to much of the inference area—even many lower elevation portions—required traversing higher elevation areas on park trails that were often snow-covered. In addition, delayed phenology made identification of some herbaceous vegetation communities difficult even into the fall.

---

proportion). Given limited resources, we chose to prioritize the number of samples collected over expanding the inference area.

The draft map used for selecting samples was modeled at the polygon level rather than the pixel-based method we later adopted. We attempted to minimize map class membership ambiguity in the samples by locating them some minimum distance from polygons mapped as different alliances. Initially, draft map polygons were internally buffered by 44 meters, assuming a 15-meter expected maximum field positioning error and a 12-meter allowable map positional error, in addition to the half-hectare MMU (see Lea and Curtis 2010). In addition, because crew size limitations prevented independent observations for overlapping plots, we began with a minimum separation of 80 meters between samples, irrespective of mapped class.

To make obtaining sufficient samples for each stratum more likely, we generated many more target points for each mapped class than the stratum-specific sampling goal. This was necessary because of the difficulties in accessing or properly assessing many locations, in addition to the rough terrain which made many randomly generated points unsafe or cost-prohibitive to reach. Inaccuracies in the GIS trail layer added to the challenge, often requiring substantially longer off-trail ventures than we had planned.

In the first sample generation phase, we attempted to create 100 random sample locations for each class subject to the accessibility criteria and the 44-meter internal buffer. For most classes, less than 100 suitable locations could be found, but if at least twice the sample goal for the class were located, we deemed that sufficient. If fewer than twice the sample goal were located for any class using the 44-meter internal buffer, the minimum buffer to polygon edge was reduced by 11 meters, the minimum sample separation was reduced by ten meters, and another sample generation phase was performed. This procedure was repeated, with the goal of achieving a bare minimum of 50 targeted plots per class. No buffer distances less than 11 meters were used (or 22 meters for tree or tall shrub-dominated classes), nor were sample locations of any vegetation type placed nearer than 40 meters apart. Reduction of the buffers was essential for several forested map classes and for nearly all non-forest classes. A total of 1,556 accuracy assessment sample locations were generated.

We assigned a random number to each of the targets generated for each class and coded the targets with the 30 lowest numbers within each class as required plots, with the remainder being optional. Crews were to plan their field logistics around collecting the required plots for each map class, with additional accessible optional plots encountered being sampled if time permitted. 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. If required plots were found to be inaccessible, additional required plots were designated from the remaining optional plots to maintain the goal of 30 plots per class.

3.3. Field data

3.3.1. Field logistics

Fieldwork planning and execution was greatly complicated by the snowpack. As a planning tool, we built a predictive snow-cover model based on elevation and aspect, parameterizing it using trail segment conditions posted on the park website. The map helped both in determining the optimal time window for the work, and in assessing what portions of the park could be prioritized for earlier

access. To assist in organizing the tours and tracking the geographic spread of collected samples, we divided park trail and road segments into 44 sampling regions with common access points. The median elevation of each region was used to sequence the tours to maximize sampling efficiency with respect to snow cover.

INR field crews collected field data from August 21 to September 23, 2011. The short sampling window began when snowmelt allowed access to a reasonable portion of the park and lasted until incipient winter weather made additional fieldwork unproductive or unsafe. Crews worked in staggered shifts of approximately a week each; fieldwork took place continuously except for two poor-weather periods. In general, crews worked in teams of one or two individuals. As they moved along roads and trails, they sampled every required plot that the day's logistics and safety factors would allow and spent any additional time available sampling optional plots along the route. Occasionally we updated the prognosis for achieving the sampling goal for each class based on the opportunities remaining in the target pool and reassigned some optional plots to high priority status for sampling.

3.3.2. Field protocol

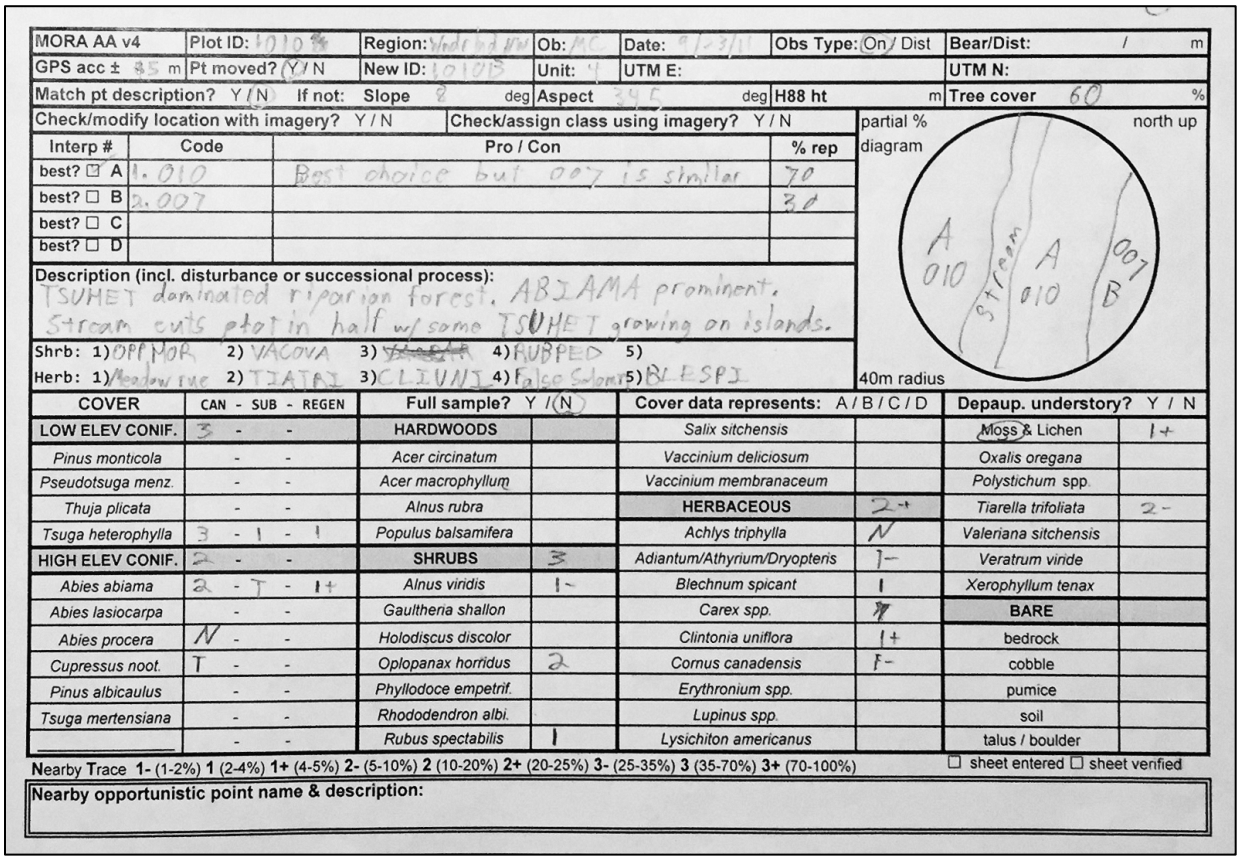
Crews navigated to each target location and assessed the surroundings. If transitions to alternate alliances were nearby, they moved the plot center to a more homogeneous point in the same vegetation type and updated the location using a GPS. If plot centers were not safely accessible and high confident 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.

We used a half-hectare assessment area for all vegetation types, corresponding to a 40-meter assessment radius around plot center. Crews identified the plant associations (if their level of experience permitted) or alliances (otherwise) covering at least 20% of the plot area.[69] Multiple calls occurred frequently in small patch vegetation types dominated by herbaceous plants and dwarf shrubs. Although spatial heterogeneity was less of an issue for most forest plots, ambiguity in association and even alliance were occasionally encountered, usually due to intermediate combinations of understory species or borderline composition situations (e.g., *Abies amabilis* at or near 10% canopy cover). Up to four associations or alliances with a plausible fit to the vegetation present were recorded. The crew's confidence in each call was recorded as high, medium or low.

A variety of data were collected to assist with subsequent quality control. Cover estimation for a fixed set of important indicator species and unvegetated ground components, including all tree species and a variety of shrubs, forbs, and graminoids, allowed for subsequent checks on the vegetation calls. These estimates, which were collected for plots assessed on-site only, was documented using a class-based system, with tree species broken into canopy, sub-canopy, and regeneration layer components. The identities of the five most prevalent species in both the shrub and

[69] The plant association descriptions and keys (Crawford et al. 2009) and the alliance descriptions that were available at the time were provided for this purpose.

herbaceous layers were also recorded. Crews noted topographic setting and sketched a diagram showing nearby reference points to assist in quality checking plot locations. Finally, they provided a brief description of the plot and pros and cons for each of their association or alliance calls. **Figure 18** contains a completed AA field form.



MORA AA v4 | Plot ID: 1010 | Region: Wonderland NW | Ob: MC | Date: 9/23/11 | Obs Type: On / Dist | Bear/Dist: / m
GPS acc ± 35 m | Pt moved? Y / N | New ID: 101013 | Unit: 4 | UTM E: | UTM N:
Match pt description? Y / N If not: Slope 8 deg | Aspect 345 deg | H88 ht m | Tree cover 60 %
Check/modify location with imagery? Y / N | Check/assign class using imagery? Y / N

Interp #	Code	Pro / Con	% rep
best? ☑ A	1. 010	Best choice but 007 is similar	70
best? ☐ B	2. 007		30
best? ☐ C			
best? ☐ D			

Description (incl. disturbance or successional process): TSUHET dominated riparian forest. ABIAMA prominent. Stream cuts plot in half w/ some TSUHET growing on islands.

Shrb: 1) OPEMOR 2) VACOVA 3) ~~VACCAE~~ 4) RUBPED 5)
Herb: 1) Meadow rue 2) TIAJAP 3) CLIUNI 4) False Solomon 5) BLESPI

partial % diagram — north up — 40m radius

COVER	CAN - SUB - REGEN	Full sample? Y / N		Cover data represents: A / B / C / D		Depaup. understory? Y / N	
LOW ELEV CONIF.	3 - -	HARDWOODS		*Salix sitchensis*		Moss & Lichen	1+
Pinus monticola	- -	*Acer circinatum*		*Vaccinium deliciosum*		*Oxalis oregana*	
Pseudotsuga menz.	- -	*Acer macrophyllum*		*Vaccinium membranaceum*		*Polystichum* spp.	
Thuja plicata	- -	*Alnus rubra*		HERBACEOUS	2+	*Tiarella trifoliata*	2-
Tsuga heterophylla	3 - 1 - 1	*Populus balsamifera*		*Achlys triphylla*	N	*Valeriana sitchensis*	
HIGH ELEV CONIF.	2 - -	SHRUBS	3	*Adiantum/Athyrium/Dryopteris*	1-	*Veratrum viride*	
Abies abiama	2 - T - 1+	*Alnus viridis*	1-	*Blechnum spicant*	1	*Xerophyllum tenax*	
Abies lasiocarpa	- -	*Gaultheria shallon*		*Carex* spp.	1	BARE	
Abies procera	N - -	*Holodiscus discolor*		*Clintonia uniflora*	1+	bedrock	
Cupressus noot.	T - -	*Oplopanax horridus*	2	*Cornus canadensis*	F-	cobble	
Pinus albicaulus	- -	*Phyllodoce empetrif.*		*Erythronium* spp.		pumice	
Tsuga mertensiana	- -	*Rhododendron albi.*		*Lupinus* spp.		soil	
	- -	*Rubus spectabilis*	1	*Lysichiton americanus*		talus / boulder	

Nearby Trace 1- (1-2%) 1 (2-4%) 1+ (4-5%) 2- (5-10%) 2 (10-20%) 2+ (20-25%) 3- (25-35%) 3 (35-70%) 3+ (70-100%) ☐ sheet entered ☐ sheet verified

Nearby opportunistic point name & description:

Figure 18. Completed accuracy assessment data sheet, collected in 2011.

3.3.3. Quality control

Accuracy assessment plot data went through a quality control process similar to that of map training data. During the year following data collection, all plot data were entered into a spreadsheet, and plot locations were verified and occasionally adjusted with reference to NAIP imagery and the field diagram. The remainder of the QC work was delayed some seven years while plot data from OLYM and NOCA—as well as data representing previously unsampled parts of MORA—were collected and analyzed, eventually resulting in an NCCN-wide map classification.

After the map classification became available, we used the species cover data, plot description and diagram, and imagery and map class indicator species (see **Section 2.5.3**) to label each plot with a single best map class call. However, there were two situations in which we allowed some flexibility. Sometimes, we were unable to make a single best call because the floristics field data were perfectly intermediate between two classes. Both calls were treated as legitimate possible answers at the 72 plots where this occurred. In another 21 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 ground truth for a secondary patch.

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-nine 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

In an attempt to alleviate undersampling of some map classes, additional plots used for AA were collected in 2014 and 2019. The 2014 plots were collected under the mapping protocol Q described in Section 2.1.1 and later repurposed for AA. The 2019 plots were collected using a protocol similar to that for the 2011 plots. Both datasets were subjected to quality control as described above. Cumulatively, a total of 734 field plot samples were collected; 705 of them passed the QC process. We determined the reached inference area (RIA) by removing from the AIA any route portions at least 1 km long and lacking sampled plots (refer back to **Figure 17**). The RIA totaled 6,555 hectares, equivalent to 55.7% of the attempted inference area or 6.9% of the full park. **Table 13** 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 two classes, both of which are among the rarest in the park.

Table 13. 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
R72–Colluvial barren	5,452	228	141	61.8	2.6
R71–Alluvial barren	2,409	166	98	58.9	4.1
H58–Bedrock balds & forest openings	542	54	28	52.7	5.2
H62–Alpine sparse herbaceous	1,507	126	80	63.2	5.3
C14–Silver fir & big huckleberry	9,264	983	501	51.0	5.4
C12–Silver fir & Alaska blueberry	7,034	714	388	54.3	5.5
S43–Sitka alder	1,994	191	110	57.6	5.5
H57–Green fescue dry meadow	965	124	54	43.5	5.6
C13–Mtn hemlock & Cascade azalea	9,581	909	545	59.9	5.7
C26–Conifer krummholz & treed cliff	788	67	45	67.3	5.7
S47–Successional huckleberry	1,836	263	119	45.1	6.5
C21–Mtn hemlock & heather	3,207	359	214	59.4	6.7
S49–Alpine heather	1,520	143	111	77.6	7.3
C23–Mt Rainier subalp fir & whitebark pine	1,671	280	129	46.1	7.7
C20–Subalp fir & Sitka valerian	7,296	1,038	570	54.9	7.8
S45–Vine maple	1,118	161	96	59.6	8.6
S48–Subalpine heather	2,040	288	184	64.0	9.0
C05–W hemlock & sword fern	3,098	515	297	57.7	9.6
H51W–Subalpine herbaceous wetland	460	76	45	59.4	9.8
S40W–Low elevation shrub wetland	133	22	13	60.6	9.9
C10–Moist silver fir & foamflower	4,580	849	457	53.8	10.0
H53–Showy sedge & Sitka valerian	1,096	194	114	58.9	10.4
H50W–Lowland marsh & meadow	17	5	2	36.1	10.6
C11–Mesic silver fir & w hemlock	7,790	1,415	864	61.0	11.1
H60W–Black alpine sedge	146	25	16	64.4	11.2
C06–W hemlock & salal	2,396	588	291	49.4	12.1
B31–Broadleaf riparian & swamp forest	799	174	100	57.2	12.5
H63–Alpine buckwheat pumice vegetation	255	59	32	54.1	12.5
B30–Successional gravel bar	705	141	90	64.2	12.8
C04–Moist w hemlock & foamflower	1,721	451	250	55.5	14.5
S41W–Subalpine willow wetland	96	27	21	77.1	21.6
H56–Subalpine summer-dry grass-forb	309	80	69	86.7	22.3

Table 13 (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
H64–Alpine lupine pumice vegetation	146	44	44	100.0	30.0
C03–Sitka spruce & wood-sorrel	140	77	43	55.7	30.5

3.4. Photo-interpretation

An additional 121 plots were randomly generated in five mapped abiotic classes that were not targeted or were inadequately sampled during the field sampling (R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE, R72–COLLUVIAL BARREN, R73–BEDROCK BARREN, W81–FRESH WATER and W82–EXPOSED SNOW AND ICE). These points were photo-interpreted to map class. Although the latter two classes were not included in the AA contingency tables, they 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 826 sample plots passed the QC process or were photo-interpreted (**Figure 19**). The final coordinates for each plot were used to extract the predicted map class label from the final map. **Table 14** 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).

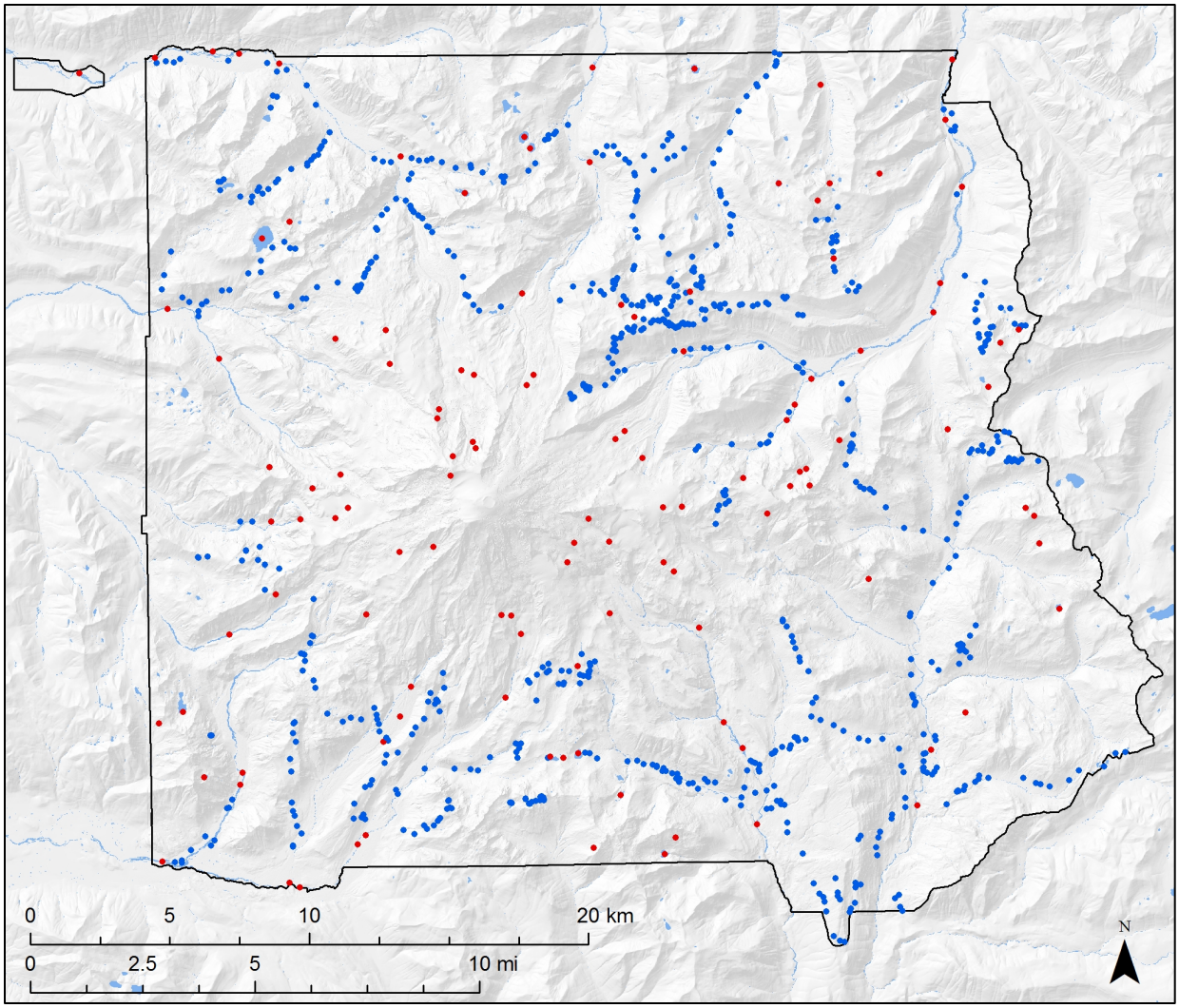


Figure 19. Accuracy assessment (AA) plot locations. 705 quality-controlled field-collected AA plots (blue dots) and 121 photo-interpreted points for unvegetated map classes (red dots) are shown.

Table 14. 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 except for H50W, which required only ten. 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
S40W–Low elevation shrub wetland	133	1	0	0
H60W–Black alpine sedge	146	2	2	7
C03–Sitka spruce & wood-sorrel	140	4	4	13
S41W–Subalpine willow wetland	96	5	7	17
H50W–Lowland marsh & meadow	17	2	2	19
C21–Mtn hemlock & heather	3,207	8	9	27
B30–Successional gravel bar	705	8	9	27
H58–Bedrock balds & forest openings	542	10	9	30
H51W–Subalpine herbaceous wetland	460	11	11	37
H57–Green fescue dry meadow	965	12	14	40
S45–Vine maple	1,118	13	16	43
H62–Alpine sparse herbaceous	1,507	15	14	47
H63–Alpine buckwheat pumice vegetation	255	14	14	47
H56–Subalpine summer-dry grass-forb	309	17	15	50
C23–Mt Rainier subalp fir & whitebark pine	1,671	16	15	50
B31–Broadleaf riparian & swamp forest	799	15	15	50
S48–Subalpine heather	2,040	16	19	53
H64–Alpine lupine pumice vegetation	146	16	16	53
S43–Sitka alder	1,994	22	18	60
C06–W hemlock & salal	2,396	20	23	67
C04–Moist w hemlock & foamflower	1,721	21	24	70
H53–Showy sedge & Sitka valerian	1,096	22	21	70
C26–Conifer krummholz & treed cliff	788	21	21	70
S47–Successional huckleberry	1,836	22	24	73
S49–Alpine heather	1,520	29	23	77
C05–W hemlock & sword fern	3,098	31	29	97
C12–Silver fir & Alaska blueberry	7,034	36	34	113
R71–Alluvial barren	2,409	34	34	113
R73–Bedrock barren	4,442	38	35	117
C14–Silver fir & big huckleberry	9,264	37	45	123
C13–Mtn hemlock & Cascade azalea	9,581	37	39	123
C11–Mesic silver fir & w hemlock	7,790	57	44	147
C20–Subalp fir & Sitka valerian	7,296	46	47	153

Table 14 (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 except for H50W, which required only ten. 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
C10–Moist silver fir & foamflower	4,580	49	52	163
R72–Colluvial barren	5,452	54	57	180

The sampling goal of 30 samples per mapped class, needed for a confident assessment of UA, was achieved for only ten of the 35 classes included in the AA. Four classes (S40W–LOW ELEVATION SHRUB-DOMINATED WETLAND, H60W–BLACK ALPINE SEDGE WETLAND, H50W–LOWLAND MARSH AND MEADOW and C03–SITKA SPRUCE, WESTERN HEMLOCK AND WOOD-SORREL FOREST) had less than five mapped occurrences sampled. The success rate for identified plots, needed for a confident assessment of PA, was similar. Only nine of the classes were identified at least 30 times. The same four classes that were poorly sampled from the UA perspective were also poorly sampled for PA. One class, S40W–LOW ELEVATION SHRUB-DOMINATED WETLAND, was not identified at all.

On the other hand, four classes were significantly oversampled. For R72–COLLUVIAL BARREN, this resulted from adding a PI sample to an already significant field sample. The oversamples of the other classes—C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST, C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST, and C20–SUBALPINE FIR AND SITKA VALERIAN FOREST AND WOODLAND—resulted from fieldwork being redirected to less snowbound areas at montane elevations and in the rain shadow respectively. 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 alliance label attached to each point.

3.6. Analysis

A total of 35 classes were included in the AA contingency tables, including all classes hosting natural vegetation communities except for C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND, which was mapped only in digitized known locations. 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 other two classes for which most AA samples were photo-interpreted[70] were excluded from further analysis. We noted that the 63 samples mapped as these classes indicated a user's accuracy of 100.0%, but felt that because there was little field-sampling and they generally do not host natural vegetation communities, their inclusion would artificially inflate the overall accuracy. The five classes that were

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

mapped via PI or deductive modeling[71] were also excluded from further analysis. Two field-sampled plots fell into one of these classes; both were correctly mapped. A total of 761 plots, 695 of them 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 quality-controlled map class calls. For plots with two best calls, we allowed either as a correct answer; 24 of these 72 plots were counted as correct because the map matched the second of the two best calls. For plots with a secondary patch call within ten meters of the assessed center point, we also allowed that as a correct answer.[72] Of these 21 plots, four 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 86.9%.

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 83.3%. Note that the mapped areas in the table sum to 6,236 hectares rather than the 6,555 hectares contained in the RIA. The seven 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

[71] C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND, M92–BURNED WITH UNCERTAIN VEGETATION, M93–TIMBERLAND WITH UNCERTAIN VEGETATION, M94–DEVELOPMENT and M95–ROADS IN PARK.

[72] 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 15**. The accuracy estimate met the 80% standard for 29 of the 35 map classes. However, for only 14 was the 90% confidence interval entirely above the 80% mark. Three of the six that failed to meet the standard were poorly sampled. Only two classes, C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST and C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST, were conclusively demonstrated to fail to meet the standard via the 90% confidence interval.

Table 15. 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
S40W–Low elevation shrub-dominated wetland	1	0%	NA
C12–Silver fir, hemlock and Alaska blueberry forest	36	61%*	46–76%
C11–Mesic silver fir and western hemlock forest	57	63%*	52–75%
C21–Mountain hemlock, subalpine fir and heather woodland	8	75%	44–100%
H57–Green fescue dry meadow	12	75%	50–100%
S49–Alpine heather shrubland	29	79%	65–93%
S43–Sitka alder shrubland	22	82%	66–98%
H53–Showy sedge and Sitka valerian meadow	22	82%	66–98%
C10–Moist silver fir, western hemlock and foamflower forest	49	84%	74–93%
C06–Western hemlock, Douglas-fir and salal forest	20	85%	69–100%
H58–Bedrock balds and sparsely vegetated forest openings	10	90%	69–100%
S48–Subalpine heather shrubland	16	88%	71–100%
H56–Subalpine summer-dry grass-forb meadow	17	88%	72–100%
C13–Mountain hemlock, silver fir and Cascade azalea forest	37	86%	76–97%
C05–Western hemlock, Douglas-fir and sword fern forest	31	87%	76–99%
H50W–Lowland marsh and meadow	2	100%	75–100%
H60W–Black alpine sedge wetland	2	100%	75–100%
S45–Vine maple shrubland	13	92%	76–100%
C04–Moist western hemlock, Douglas-fir and foamflower forest	21	90%	78–100%
C14–Silver fir, big huckleberry and beargrass forest	37	89%	79–99%
H62–Alpine sparse herbaceous vegetation	15	93%	79–100%
R73–Bedrock barren	38	89%*	80–99%
C23–Mount Rainier subalpine fir and whitebark pine woodland	16	94%*	81–100%
C20–Subalpine fir and Sitka valerian forest and woodland	46	91%*	83–99%
C26–Conifer krummholz and treed cliff	21	95%*	85–100%

Table 15 (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
S47–Successional huckleberry shrubland	22	95%*	86–100%
C03–Sitka spruce, western hemlock and wood-sorrel forest	4	100%*	88–100%
S41W–Subalpine willow wetland	5	100%*	90–100%
B30–Successional gravel bar shrubland	8	100%*	94–100%
R72–Colluvial barren	54	98%*	94–100%
H51W–Subalpine herbaceous wetland	11	100%*	95–100%
H63–Alpine buckwheat pumice vegetation	14	100%*	96–100%
B31–Broadleaf riparian and swamp forest	15	100%*	97–100%
H64–Alpine lupine pumice vegetation	16	100%*	97–100%
R71–Alluvial barren and debris-covered ice	34	100%*	99–100%

The class-specific producer's accuracies, obtained from the PCT, are summarized in **Table 16**. The accuracy estimate met the 80% standard for 27 of the 34 evaluated map classes (PA could not be estimated for S40W–LOW ELEVATION SHRUB-DOMINATED WETLAND, which was not found in the field). However, for only 15 of them was the 90% confidence interval entirely above the 80% mark. Only one class, C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST, was conclusively demonstrated to fail to meet the standard.

Table 16. 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
S40W–Low elevation shrub-dominated wetland	0	NA	NA
C12–Silver fir, hemlock and Alaska blueberry forest	34	62%*	48–75%
H57–Green fescue dry meadow	14	64%	43–85%
S41W–Subalpine willow wetland	7	66%	34–99%
C10–Moist silver fir, western hemlock and foamflower forest	52	71%	61–82%
C06–Western hemlock, Douglas-fir and salal forest	23	76%	61–90%
C14–Silver fir, big huckleberry and beargrass forest	45	76%	66–86%
C04–Moist western hemlock, Douglas-fir and foamflower forest	24	76%	61–92%
H53–Showy sedge and Sitka valerian meadow	21	82%	64–100%
C21–Mountain hemlock, subalpine fir and heather woodland	9	86%	66–100%
S45–Vine maple shrubland	16	83%	69–97%

Table 16 (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
S47–Successional huckleberry shrubland	24	84%	69–100%
C11–Mesic silver fir and western hemlock forest	44	86%	77–94%
C13–Mountain hemlock, silver fir and Cascade azalea forest	39	86%	77–94%
C20–Subalpine fir and Sitka valerian forest and woodland	47	86%	76–96%
H50W–Lowland marsh and meadow	2	100%	75–100%
H60W–Black alpine sedge wetland	2	100%	75–100%
B30–Successional gravel bar shrubland	9	87%*	82–93%
S48–Subalpine heather shrubland	19	89%	79–98%
C05–Western hemlock, Douglas-fir and sword fern forest	29	90%	77–100%
C26–Conifer krummholz and treed cliff	21	92%	77–100%
C03–Sitka spruce, western hemlock and wood-sorrel forest	4	100%*	88–100%
R73–Bedrock barren	35	96%*	89–100%
R72–Colluvial barren	57	95%*	90–99%
H58–Bedrock balds and sparsely vegetated forest openings	9	100%*	94–100%
H51W–Subalpine herbaceous wetland	11	100%*	95–100%
H62–Alpine sparse herbaceous vegetation	14	100%*	96–100%
H63–Alpine buckwheat pumice vegetation	14	100%*	96–100%
S43–Sitka alder shrubland	18	100%*	97–100%
H56–Subalpine summer-dry grass-forb meadow	15	100%*	97–100%
C23–Mount Rainier subalpine fir and whitebark pine woodland	15	100%*	97–100%
B31–Broadleaf riparian and swamp forest	15	100%*	97–100%
H64–Alpine lupine pumice vegetation	16	100%*	97–100%
S49–Alpine heather shrubland	23	100%*	98–100%
R71–Alluvial barren and debris-covered ice	34	100%*	99–100%

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 17** lists these classes as well as the classes with which each is confused.

Table 17. 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)
S40W–Low elevation shrub wetland	0%	B30–Successional gravel bar (0.21%)
C12–Silver fir & Alaska blueberry	61%	C13–Mtn hemlock & Cascade azalea (1.57%) C11–Mesic silver fir & w hemlock (1.00%) C10–Moist silver fir & foamflower (0.92%) C14–Silver fir & big huckleberry (0.91%) C20–Subalp fir & Sitka valerian (0.37%)
C11–Mesic silver fir & w hemlock	63%	C10–Moist silver fir & foamflower (2.00%) C14–Silver fir & big huckleberry (1.16%) C04–Moist w hemlock & foamflower (1.16%) C12–Silver fir & Alaska blueberry (1.00%) C06–W hemlock & salal (0.97%) C05–W hemlock & sword fern (0.24%)
H57–Green fescue dry meadow	64%	H53–Showy sedge & Sitka valerian (0.17%) S48–Subalpine heather (0.14%) H56–Subalpine summer-dry grass-forb (0.13%) S47–Successional huckleberry (0.07%) S49–Alpine heather (0.06%)
S41W–Subalpine willow wetland	66%	S47–Successional huckleberry (0.09%) H53–Showy sedge & Sitka valerian (0.08%)
C10–Moist silver fir & foamflower	71%	C11–Mesic silver fir & w hemlock (2.00%) C12–Silver fir & Alaska blueberry (0.92%) C06–W hemlock & salal (0.23%) C04–Moist w hemlock & foamflower (0.19%) C05–W hemlock & sword fern (0.15%) C14–Silver fir & big huckleberry (0.15%)
C21–Mtn hemlock & heather	75%	C20–Subalp fir & Sitka valerian (1.25%) C26–Conifer krummholz & treed cliff (0.03%)
C06–W hemlock & salal	76%	C11–Mesic silver fir & w hemlock (0.97%) C05–W hemlock & sword fern (0.54%) C10–Moist silver fir & foamflower (0.23%) C14–Silver fir & big huckleberry (0.23%)
C14–Silver fir & big huckleberry	76%	C11–Mesic silver fir & w hemlock (1.16%) C12–Silver fir & Alaska blueberry (0.91%) C13–Mtn hemlock & Cascade azalea (0.45%) C06–W hemlock & salal (0.23%) C10–Moist silver fir & foamflower (0.15%) S45–Vine maple (0.12%) S43–Sitka alder (0.08%)

Table 17 (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)
C04–Moist w hemlock & foamflower	76%	C11–Mesic silver fir & w hemlock (1.16%) C10–Moist silver fir & foamflower (0.19%) C05–W hemlock & sword fern (0.15%)
S49–Alpine heather	79%	S48–Subalpine heather (0.18%) H53–Showy sedge & Sitka valerian (0.06%) C26–Conifer krummholz & treed cliff (0.06%) H57–Green fescue dry meadow (0.06%)

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 18**. All AA plots that successfully passed through the QC process were used, including those of the seven map classes excluded from the map class level analysis (these fell into the last five categories in the table). The user's and producer's accuracy estimates for all evaluated categories exceeded the 80% standard at 90% confidence.

Table 18. 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	23	100%*	98–100%	24	97%*	90–100%
Conifer	383	100%*	100–100%	386	100%*	99–100%
Tall shrub	35	94%*	86–100%	34	98%*	93–100%
Shrubland	73	93%*	88–99%	73	95%*	92–99%
Herbaceous	121	95%*	91–99%	118	96%*	92–100%
Water	44	100%*	99–100%	44	100%*	99–100%
Rock	126	100%*	100–100%	126	100%*	100–100%
Snow & ice	19	100%*	97–100%	19	100%*	97–100%
Developed	2	100%	75–100%	2	100%	75–100%
Other disturbed	0	NA	NA	0	NA	NA

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 may 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 QC process will have compensated 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, all classes except H50W–LOWLAND MARSH AND MEADOW require 30 samples of mapped occurrences. As documented in **Table 14**, this was only achieved for ten of the 35 classes. Several causes for this failure are described below.

Sixteen of the 25 undersampled classes were rare (totaling 100 or fewer 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. Access difficulties to the north and west sides of the park resulting from the long-lasting snowpack likely contributed to many of these shortfalls. Five other classes (C21–MOUNTAIN HEMLOCK, SUBALPINE FIR AND HEATHER WOODLAND, C23–MOUNT RAINIER SUBALPINE FIR AND WHITEBARK PINE WOODLAND, S48–SUBALPINE HEATHER SHRUBLAND, S49–ALPINE HEATHER SHRUBLAND and H53–SHOWY SEDGE AND SITKA VALERIAN MEADOW) are located at subalpine and alpine elevations. It is likely that the snowpack inhibited the ability of the crews to reach and assess plots of these classes. The remaining four undersampled classes all were sampled at 20 or more mapped occurrences.

It is likely that some of the undersampling problems resulted from changes in the mapped classes made between the time of AA sampling in 2011 and production of the final maps in 2019–20. For example, during the intervening time period, we addressed substantial confusion between H51W–SUBALPINE HERBACEOUS WETLAND and H53–SHOWY SEDGE AND SITKA VALERIAN MEADOW by carving a new map class, S41W–SUBALPINE WILLOW WETLAND, out of plots with substantial *Salix*

commutata presence that had floristics and mapping tendencies intermediate to the two original classes.

Many other changes were made 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 15** and **Table 16**. 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 19**. 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 19. 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	+	76	C11 (77%)	none, no reciprocity
C05–W hemlock & sword fern	+	+	C06 (50%), C11 (22%)	–
C06–W hemlock & salal	+	76	C11 (49%), C05 (27%)	none, no reciprocity
C10–Moist silver fir & foamflower	+	71	C11 (55%), C12 (25%)	consider merge with C11
C11–Mesic silver fir & w hemlock	63*	+	C10 (31%), C14 (18%), C04 (18%)	consider merge with C10
C12–Silver fir & Alaska blueberry	61*	62*	C13 (33%), C11 (21%), C10 (19%), C14 (19%)	consider merge with C13
C13–Mtn hemlock & Cascade azalea	+	+	C12 (64%)	consider merge with C12
C14–Silver fir & big huckleberry	+	76	C11 (37%), C12 (29%)	none, no reciprocity
C20–Subalp fir & Sitka valerian	+	+	C21 (57%), C13 (20%), C12 (17%)	consider merge with C21
C21–Mtn hemlock & heather	75	+	C20 (97%)	consider merge with C20
B30–Successional gravel bar	+	+	S40W (100%)	consider merge with S40W
S40W–Low elevation shrub wetland	0	NA	B30 (100%)	consider merge with B30
S41W–Subalpine willow wetland	+	66	S47 (51%), H53 (49%)	none, no reciprocity

Table 19 (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
S47–Successional huckleberry	+	+	S48 (43%), S41W (20%), H53 (19%), H57 (17%)	–
S48–Subalpine heather	+	+	S49 (26%), S47 (26%), H53 (26%), H57 (21%)	consider merge with S49
S49–Alpine heather	79	+	S48 (50%)	consider merge with S48
H53–Showy sedge & Sitka valerian	+	+	S48 (28%), H57 (25%)	–
H57–Green fescue dry meadow	75	64	H53 (29%), S48 (25%), H56 (23%)	none, no reciprocity

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.***

H57–GREEN FESCUE DRY MEADOW (UA 75%, PA 64%) may be slightly under-mapped in the northeast where it is most abundant and may be somewhat over-mapped in the Paradise vicinity. The class is generally distributed where *Festuca viridula* is abundant, and intergrades substantially with S47–SUCCESSIONAL HUCKLEBERRY SHRUBLAND, H53–SHOWY SEDGE AND SITKA VALERIAN MEADOW and H56–SUBALPINE SUMMER-DRY GRASS-FORB MEADOW where the fescue combines with the dominants of the other classes. Map training data was weak, as reliable identification depends on complete floristic information, and frequent fine-scale patchiness creates additional assessment challenges. Two plots of H57 mapped as H56 along the western portion of the Sunrise road where the two classes intergrade. Two plots of H57 mapped as H53 at the eastern park boundary, near Crystal and Tipsoo Lakes; both had floristics intermediate between the two classes. One plot of obvious S48–SUBALPINE HEATHER SHRUBLAND at Paradise mapped as H57, but this error does not appear to be widespread. Several nearby training plots were labeled as H57, but it is possible that the class is less abundant there than the data imply. Plots collected in 2008 often confused several aster family species, which could have affected the best map class assessment in this area. Though we corrected these errors in QC when we were confident they had occurred, we often had no way of being sure. ***Within the constraints of the floristic and spatial ambiguities detailed above, the class appears to be reasonably well-mapped, although we recommend NPS assess the map in the Paradise area using complete floristics plots and the indicators given in Nielsen et al. (2021c).***

S41W–SUBALPINE WILLOW WETLAND (UA 100%*, PA 66%) may be under-mapped in upland seep settings. One large such occurrence of *Salix commutata* or *S. barclayi* at Paradise was mapped as

S47–SUCCESSIONAL HUCKLEBERRY SHRUBLAND. Although significant S41W is mapped nearby, it appears to all be in lower-lying areas. Merging the two classes is not a good option as that would sacrifice the wetland S41W occurrences—the majority of the map class—which appear to be well-mapped. It is likely this is an uncommon error as no similarly upland training plots were found at MORA, although they do exist at the other NCCN parks. ***We recommend that NPS correct these mapping errors where they are encountered.***

C04–MOIST WESTERN HEMLOCK, DOUGLAS-FIR AND FOAMFLOWER FOREST (UA 90%, PA 76%) is mapped as C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST at four plots across the south side of the park, but none appear to represent significant problems. All plots seem to be in microsites below MMU that are considerably wetter than the surroundings. Furthermore, all have significant presence of *Abies amabilis*, though not enough to establish that C11 (or C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST) would be an appropriate call. Two are adjacent to the Stevens Canyon road on the slopes of Backbone Ridge and may have been impacted by road construction. Furthermore, C04 is mapped within 50 meters of three of the plots. ***Based on the evidence here, we believe this class is well-mapped. Fine-scale intergrading of soil moisture conditions will always result in class ambiguities; 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.***

C06–WESTERN HEMLOCK, DOUGLAS-FIR AND SALAL FOREST (UA 85%, PA 76%) is mapped as C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST at four plots, but none appear to represent significant problems. Three of the sites are along a 2-kilometer stretch of the Stevens Canyon road on the west side of Backbone Ridge; the other is on a steep south-facing slope above the junction of the North and South Mowich Rivers. The Backbone Ridge sites are floristically intermediate—all contain significant amounts of *Abies amabilis* or *A. procera*—and all fell within 20 meters of mapped C06, which is the dominant class mapped on this slope. A GPS malfunction occurred during collection of the Mowich River site; again, C06 is mapped nearby, just 60 horizontal meters up from the nominal plot location on this steep featureless slope. ***Based on the evidence here, we believe this class is well-mapped.***

C14–SILVER FIR, BIG HUCKLEBERRY AND BEARGRASS FOREST (UA 89%, PA 76%) is mapped as C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST at three plots and as C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST at four plots. The errors do not appear to represent significant problems. The confusion with C11 occurred at two plots just 60 meters apart near Fryingpan Creek (one floristically intermediate to C11, both within 40 meters of mapped C14) and at one plot near Pyramid Creek that was floristically clear but located on the boundary of mapped C14). The confusion with C12 occurred at three plots all within 200 meters on the Cowlitz Divide trail (all floristically intermediate to C12, all within 40 meters of mapped C14), and at another intermediate plot on the Moraine Creek trail, also with C14 mapped in the vicinity. ***Although sites floristically intermediate between these classes are not uncommon, there is no evidence here that the classes are poorly mapped.***

Map classes for which merging may be a viable option

There are five pairs 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.***

S40W–LOW ELEVATION SHRUB-DOMINATED WETLAND (UA 0%, PA NA) and B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND (UA 100%*, PA 87%*) could be merged; the combined class B30+S40W would have UA 100%* and PA 100%*. The only mapped area of S40W that was visited during AA was found to be B30 in an overflow channel of the Ohanapecosh River, although other nearby areas appear in NAIP imagery to be correctly mapped as S40W. Examination of the map does give the impression that the UA of S40W may be quite low (e.g., it appears to *map over*[73] small openings in riparian forests that are probably best considered forest) but that most of the excess is mapped near major rivers and could be easily corrected, perhaps after some field inspection. It also appears evident that merging S40W with B30 would accomplish nothing positive and that the mapped distribution of S40W can provide some useful information about locations of possibly unknown shrub-dominated wetlands that might be useful. ***Given the significant ecological distinction between these classes and the relative ease by which excess mapped S40W might be hand-corrected, we recommend keeping the classes as they are.***

C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST (UA 61%*, PA 62%*) and C13–MOUNTAIN HEMLOCK, SILVER FIR AND CASCADE AZALEA FOREST (UA 86%, PA 86%) could be merged; the combined class C12+C13 would have UA 85% and PA 78%. The AA found *two-way confusion*[74] between these classes, although C12 has confusion with several other forested classes as well. Five plots where C13 was found in the field were mapped as C12; all were in the elevational transition zone between the classes, with C13 was mapped nearby. Four of the five were in cold-air drainage basins. The three plots mapped as C13 where C12 was found were generally higher on slopes. Two had intermediate floristics, and all were within a short distance of mapped C12 just above the elevation-based transition. There is some evidence here that microclimatic conditions may cause occasional mapping errors, and certainly that floristic ambiguity can be an issue, but the errors do not appear to affect areas distant from the transition zone. C12 is one of the most poorly mapped classes in each of the NCCN maps, occupying a variably moist niche between the lower and upper montane zones where it may be confused with several other classes. However, the pattern of confusion varies from park to park. At NOCA, C12 and C13 are somewhat confused with each other, but each is primarily confused with other classes. At OLYM, they were not confused with each other at all. ***In the interest of maintaining a consistent NCCN classification, we recommend keeping the classes separate at MORA. Because unmappable micro-climatic conditions may often determine***

[73] 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.

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

the fine-scale distribution of these classes, 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.

C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST (UA 63%*, PA 86%) and C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST (UA 84%, PA 71%) could be merged; the combined class C10+C11 would have UA 81% and PA 76%. The most significant issue here is that seven plots identified as C10 in the field mapped as C11. Three plots occurred in similar settings lower east-facing slopes above the Ohanapecosh River or its extension upstream as Chinook Creek. Floristics were intermediate on one plot, and all three were just a short distance—between eight and 30 meters—upslope from mapped C10 closer to the channel. At three other plots where this error occurred (near Martha Falls in Stevens Canyon, above the Paradise River near its junction with the Nisqually, and above Cataract Creek), microhabitats related to upland seeps seem likely to be responsible for the mapping mistakes. For all but one of the seven plots, C10 was mapped within 30 meters, and at four of them it was mapped within ten meters. In all, the AA plots provide some evidence that C10 may extend slightly higher above valley bottoms on some aspects than it is mapped, but in most cases the errors again seem to reflect microhabitat issues and in any case they do not appear to affect areas distant from the transition zone. C11 is also a class that has mapping issues at the other NCCN parks, but again it was confused with different classes at each park. ***Again, in the interest of maintaining a consistent NCCN classification, we recommend keeping the classes separate at MORA. Because unmappable microhabitats may often determine the fine-scale distribution of these classes, 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.***

C21–MOUNTAIN HEMLOCK, SUBALPINE FIR AND HEATHER WOODLAND (UA 75%, PA 86%) and C20–SUBALPINE FIR AND SITKA VALERIAN FOREST AND WOODLAND (UA 91%*, PA 86%) could be merged; the combined class C20+C21 would have UA 96%* and PA 94%*. Two-way confusion affected these classes; two plots were confused in each direction. This represents a more significant issue for the mapping of C21 as only eight mapped plots of that class were visited. The C21 plots that mapped as C20 both had intermediate floristics; one, by Hidden Lake in the Palisades, was adjacent to a training plot that had been called to C20 because of a variety of understory indicators that were found there and may have been missed in the AA plot. The other plot may represent a more significant error, as it was above Three Lakes on the eastern park boundary where little C21 is mapped. High elevations east of highway 123 in the southeastern quadrant of the park were very poorly sampled during training collection, and floristics appear to be ambiguous in this region, as mountain-heather communities often seem to include *Rhododendron albiflorum* and *Vaccinium membranaceum* as well. It is possible that C21 is substantially under-mapped here, but one ambiguous AA plot does not provide enough evidence. ***We recommend the classes be kept separate, but that NPS investigate the mapping of subalpine forests in this part of the park more thoroughly; a sampling trip along the Pacific Crest Trail between Chinook Pass and Three Lakes would be very useful.***

S49–ALPINE HEATHER SHRUBLAND (UA 79%, PA 100%*) and S48–SUBALPINE HEATHER SHRUBLAND (UA 88%, PA 89%) could be merged; the combined class S48+S49 would have UA 89%* and PA 95%*. Three plots identified as S48 in the field were mapped as S49. One plot was in Spray Park, 15 meters from mapped S48. This plot may have been mislocated as it appears quite sparse in NAIP imagery but had abundant *Lupinus latifolius* documented in the field. The other two plots were at Paradise. One had intermediate and atypical floristics and was located on the boundary of mapped S48. The other was a good bit higher, above the elevation at which the map shows significant amounts of S48, and perhaps taking advantage of a topographically sheltered location. ***S48 and S49 were kept separate at the other NCCN parks and there does not appear to be convincing evidence to merge them at MORA. It is possible that small amounts of S48 occur in sheltered microclimates at higher elevations than the map shows.***

3.7.3. Other known mapping issues

As mentioned above, C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST experiences significant confusion with several classes. In addition to the upper montane classes discussed above (C13–MOUNTAIN HEMLOCK, SILVER FIR AND CASCADE AZALEA FOREST and C14–SILVER FIR, BIG HUCKLEBERRY AND BEARGRASS FOREST), it may also be confused with the lower montane classes C10–MOIST SILVER FIR, WESTERN HEMLOCK AND FOAMFLOWER FOREST and C11–MESIC SILVER FIR AND WESTERN HEMLOCK FOREST. ***Complete floristics and careful diagnosis of indicator species is essential for correct identification of this class.***

4. Vegetation of Mount Rainier National Park

4.1. Vegetation map

The vegetation map (**Figure 20**, and a higher-resolution version at Nielsen et al. 2021d) illustrates the distribution of the 42 map classes across the park and surrounding buffer. The map contains over 210 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 20**. The map classes vary widely in abundance, with most of them limited in extent. The most common six classes—five of which are montane and subalpine conifer types—collectively occupy half of the park, while 18 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.

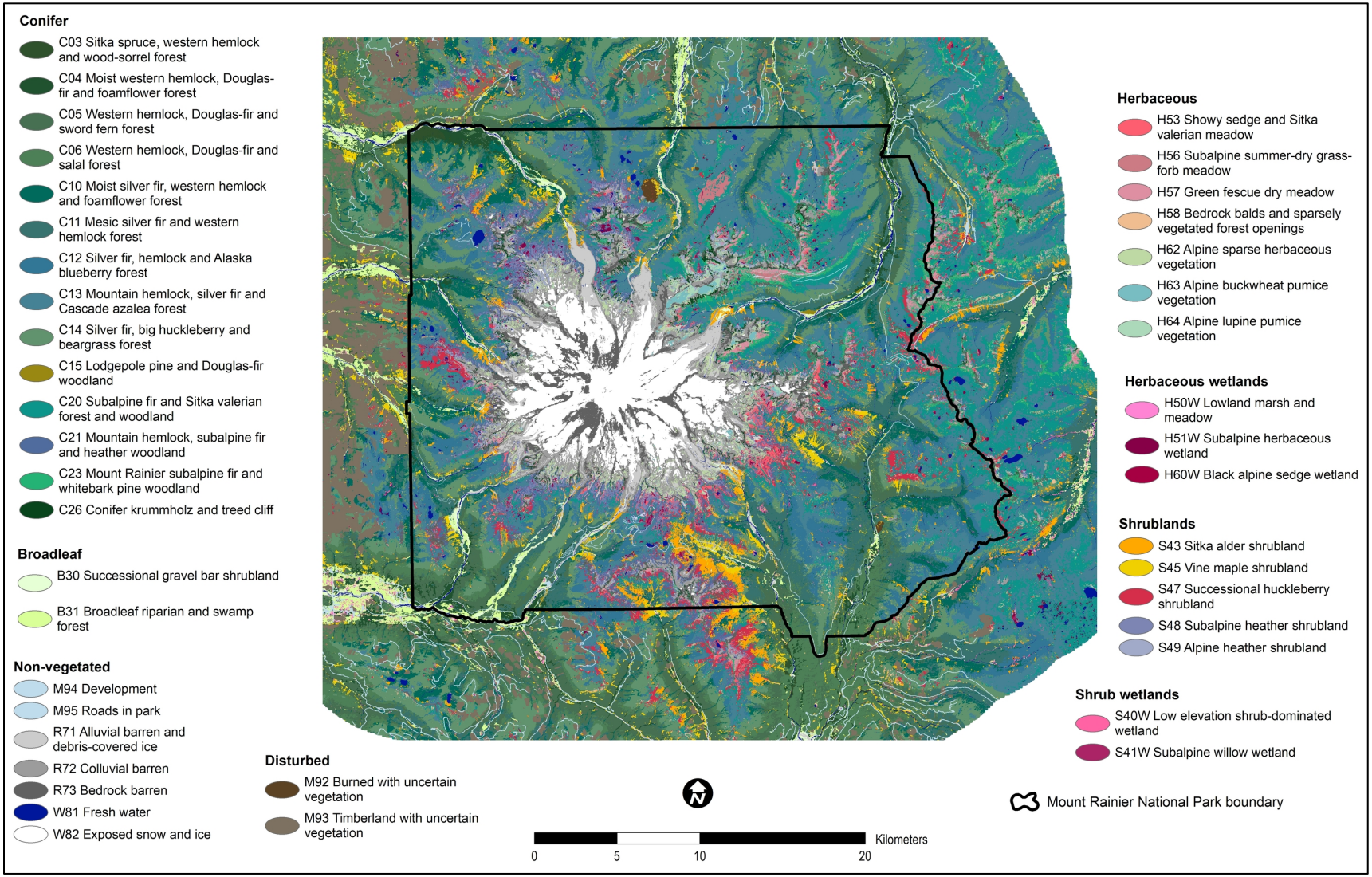


Figure 20. Vegetation map of Mount Rainier National Park.

Table 20. Map class estimated area and proportion of park, listed by area.

Class code and full name	Est area in park (ha)	Est area in park (ac)	Proportion of park (%)
C14–Silver fir, big huckleberry and beargrass forest	10,103	24,965	10.57
C13–Mountain hemlock, silver fir and Cascade azalea forest	9,672	23,901	10.12
C20–Subalpine fir and Sitka valerian forest and woodland	8,022	19,824	8.39
W82–Exposed snow and ice	7,368	18,206	7.71
C11–Mesic silver fir and western hemlock forest	6,276	15,508	6.57
C12–Silver fir, hemlock and Alaska blueberry forest	6,225	15,383	6.51
R72–Colluvial barren	5,819	14,378	6.09
C10–Moist silver fir, western hemlock and foamflower forest	5,286	13,062	5.53
R73–Bedrock barren	4,076	10,071	4.26
C05–Western hemlock, Douglas-fir and sword fern forest	2,955	7,302	3.09
C06–Western hemlock, Douglas-fir and salal forest	2,783	6,877	2.91
C21–Mountain hemlock, subalpine fir and heather woodland	2,760	6,820	2.89
R71–Alluvial barren and debris-covered ice	2,409	5,953	2.52
C04–Moist western hemlock, Douglas-fir and foamflower forest	2,204	5,446	2.31
S48–Subalpine heather shrubland	2,103	5,198	2.20
S47–Successional huckleberry shrubland	2,052	5,070	2.15
S43–Sitka alder shrubland	1,631	4,031	1.71
C23–Mount Rainier subalpine fir and whitebark pine woodland	1,567	3,871	1.64
H62–Alpine sparse herbaceous vegetation	1,407	3,476	1.47
S45–Vine maple shrubland	1,359	3,357	1.42
S49–Alpine heather shrubland	1,206	2,979	1.26
H53–Showy sedge and Sitka valerian meadow	1,178	2,910	1.23
W81–Fresh water	960	2,373	1.00
H57–Green fescue dry meadow	912	2,254	0.95
C26–Conifer krummholz and treed cliff	803	1,985	0.84
B31–Broadleaf riparian and swamp forest	799	1,975	0.84
B30–Successional gravel bar shrubland	771	1,906	0.81
M95–Roads in park	532	1,315	0.56
H58–Bedrock balds and sparsely vegetated forest openings	488	1,206	0.51
H51W–Subalpine herbaceous wetland	460	1,137	0.48
H56–Subalpine summer-dry grass-forb meadow	273	674	0.29
H63–Alpine buckwheat pumice vegetation	255	630	0.27
S41W–Subalpine willow wetland	188	464	0.20
H64–Alpine lupine pumice vegetation	146	361	0.15
H60W–Black alpine sedge wetland	146	360	0.15
C03–Sitka spruce, western hemlock and wood-sorrel forest	140	346	0.15

Table 20 (continued). Map class estimated area and proportion of park, listed by area.

Class code and full name	Est area in park (ha)	Est area in park (ac)	Proportion of park (%)
M92–Burned with uncertain vegetation	127	314	0.13
S40W–Low elevation shrub-dominated wetland	67	165	0.07
C15–Lodgepole pine and Douglas-fir woodland	37	91	0.04
H50W–Lowland marsh and meadow	17	43	0.02
M94–Development	12	29	0.01

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**.

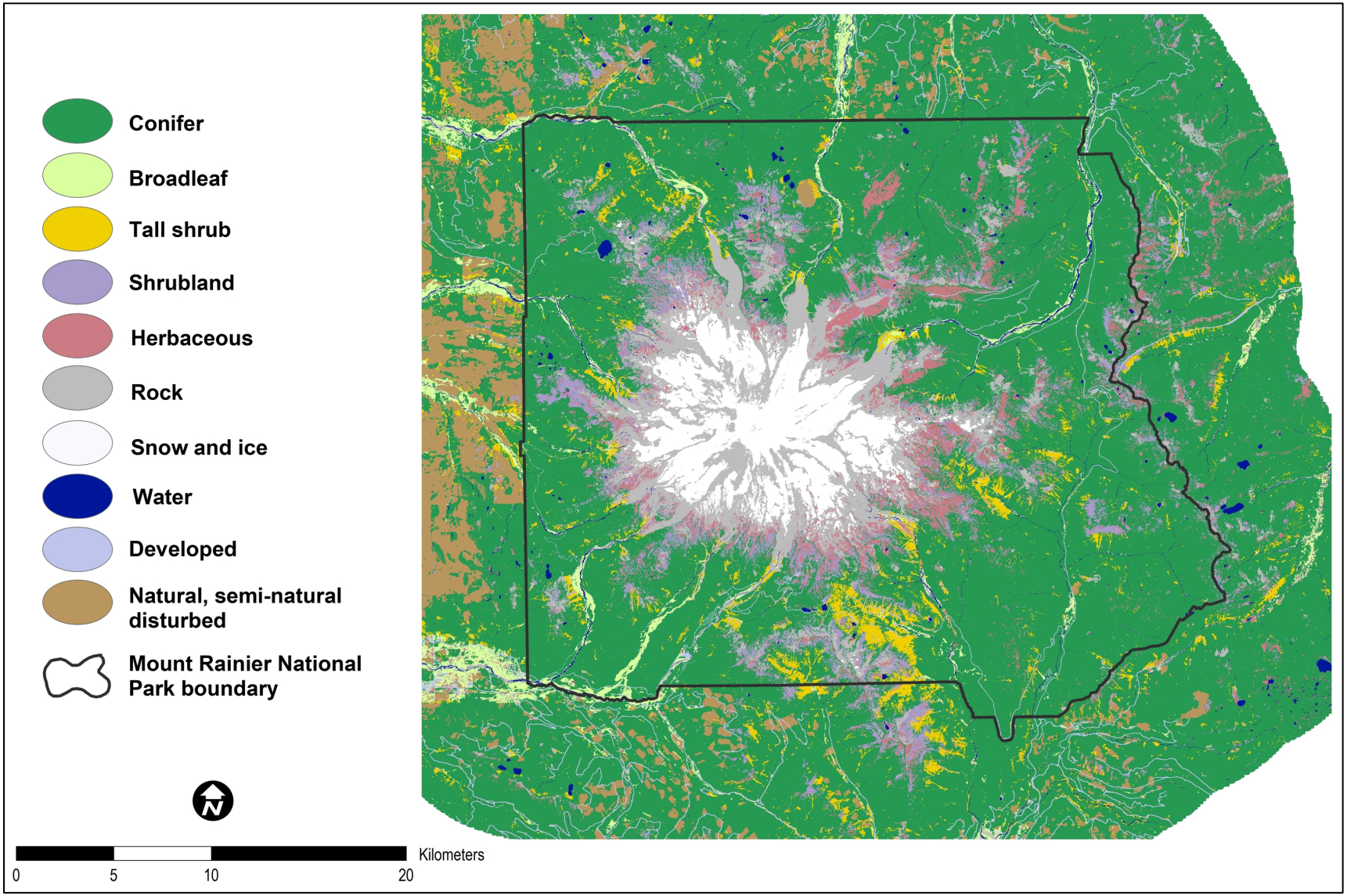


Figure 21. Lifeform map of Mount Rainier National Park.

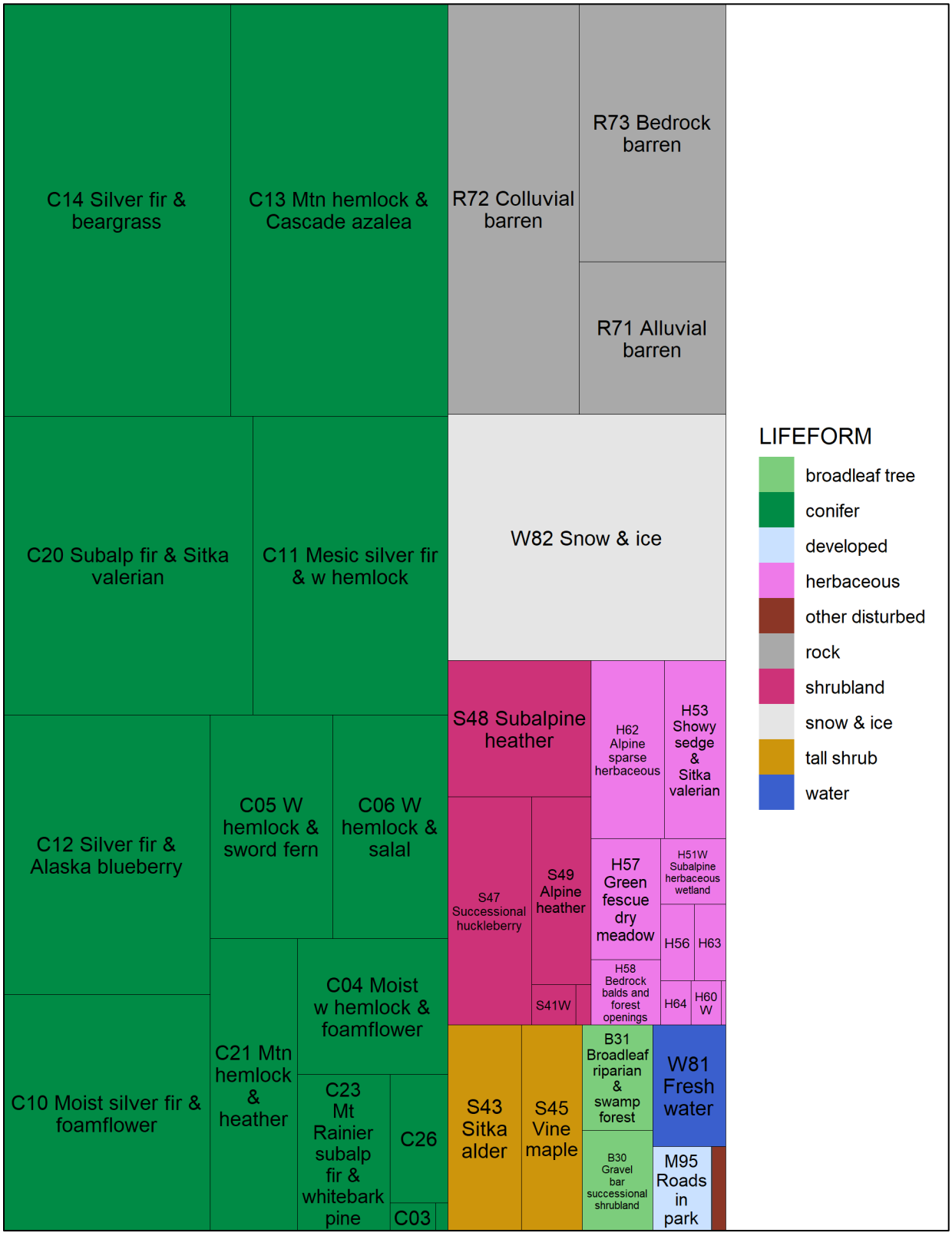


Figure 22. Relative abundance of map classes, grouped by lifeform/land-use category.

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 21** 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 21. 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	47.8%	1	•	X	X	•	–
Abies grandis	grand fir	2.1%	149	X	–	–	–	–
Abies lasiocarpa	subalpine fir	31.9%	4	–	•	X	X	X
Abies procera	noble fir	7.4%	68	–	•	X	–	–
Acer circinatum	vine maple	17.5%	23	X	•	–	–	–
Achlys triphylla	vanillaleaf	24.0%	10	X	X	•	–	–
Aconogonon davisiae	Davis' knotweed	2.9%	122	–	–	–	–	X
Alnus viridis	Sitka alder	5.8%	80	•	X	X	–	–
Antennaria lanata	woolly pussytoes	8.4%	61	–	–	–	X	X
Arnica latifolia	broad-leaved arnica	10.6%	49	–	–	X	X	–
Athyrium filix-femina	lady fern	15.0%	32	X	X	–	–	–
Berberis nervosa	dwarf Oregon-grape	17.7%	22	X	•	–	–	–
Bistorta bistortoides	American bistort	13.9%	36	–	–	–	X	X
Callitropsis nootkatensis	Alaska-cedar	23.1%	12	–	•	X	X	•
Carex kelloggii	lakeshore sedge	1.1%	207	–	–	–	X	–
Carex nigricans	black alpine sedge	3.0%	120	–	–	–	•	X
Carex spectabilis	showy sedge	13.1%	41	–	–	–	X	X
Cassiope mertensiana	white mountain-heather	9.1%	56	–	–	–	•	X
Castilleja parviflora	mountain Indian paintbrush	8.8%	58	–	–	–	X	X
Chimaphila menziesii	little pipsissewa	14.2%	35	X	X	•	–	–
Chimaphila umbellata	pipsissewa	13.3%	40	X	X	X	–	–
Clintonia uniflora	queen's cup	19.2%	20	•	X	X	–	–
Cornus unalaschkensis	bunchberry	15.1%	30	•	X	–	–	–
Erigeron glacialis	wandering daisy	8.9%	57	–	–	–	X	•
Eriogonum pyrolifolium	dirty socks	1.6%	171	–	–	–	•	X
Erythronium montanum	avalanche lily	12.1%	45	–	–	X	X	–
Eucephalus ledophyllus	Cascade aster	9.7%	51	–	–	•	X	–
Festuca viridula	green fescue	13.8%	37	–	–	•	X	–

Table 21 (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
Galium triflorum	fragrant bedstraw	7.4%	69	X	X	•	–	–
Gaultheria shallon	salal	10.2%	50	X	•	–	–	–
Goodyera oblongifolia	rattlesnake-plantain	19.8%	19	X	X	X	–	–
Gymnocarpium dryopteris	western oak fern	13.0%	42	X	X	•	–	–
Hieracium albiflorum	white hawkweed	9.2%	54	X	•	X	X	–
Hylocomium splendens	splendid feather moss	8.4%	62	X	X	–	–	–
Juncus parryi	Parry's rush	3.3%	114	–	–	–	X	X
Ligusticum grayi	Gray's lovage	12.7%	43	–	–	–	X	–
Linnaea borealis	twinflower	21.3%	17	X	X	•	–	–
Luetkea pectinata	partridgefoot	11.7%	47	–	–	•	X	X
Lupinus latifolius	subalpine lupine	31.7%	5	–	–	•	X	•
Lupinus lepidus	prairie lupine	3.4%	112	–	–	–	–	X
Luzula hitchcockii	smooth woodrush	9.7%	52	–	–	•	X	•
Maianthemum stellatum	star-flowered Solomon's plume	9.3%	53	X	X	•	–	–
Micranthes tolmiei	Tolmie's saxifrage	1.6%	172	–	–	–	–	X
Neottia cordata	heartleaf twayblade	8.3%	64	X	X	X	–	–
Oplopanax horridus	devil's club	14.8%	33	X	X	–	–	–
Oreostemma alpigenum	alpine aster	7.0%	73	–	–	–	X	X
Orthilia secunda	one-sided wintergreen	13.4%	39	•	X	X	–	–
Paxistima myrsinites	Oregon-box	8.4%	63	X	X	X	X	•
Pectiantia breweri	Brewer's miterwort	6.7%	77	–	X	X	X	–
Pedicularis contorta	coiled-beak lousewort	5.7%	82	–	–	–	X	X
Phlox diffusa	spreading phlox	8.5%	59	–	–	•	X	X
Phyllodoce empetriformis	pink mountain-heather	21.8%	16	–	–	•	X	X
Picea sitchensis	Sitka spruce	0.9%	228	X	–	–	–	–
Pinus contorta	lodgepole pine	1.1%	213	•	X	•	–	–
Pinus monticola	western white pine	3.7%	110	X	X	X	–	–
Polystichum munitum	sword fern	17.5%	24	X	•	–	–	–
Potentilla flabellifolia	fan-leaf cinquefoil	14.6%	34	–	–	–	X	•
Pseudotsuga menziesii	Douglas-fir	32.5%	3	X	X	•	–	–
Rhododendron albiflorum	Cascade azalea	9.2%	55	–	–	X	•	–

Table 21 (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
Rhododendron menziesii	false azalea	15.1%	31	•	X	X	–	–
Rubus lasiococcus	dwarf bramble	26.5%	8	–	•	X	•	–
Rubus pedatus	trailing raspberry	16.8%	26	–	X	X	–	–
Rubus spectabilis	salmonberry	15.6%	28	X	X	–	–	–
Salix barclayi	Barclay's willow	1.2%	202	–	–	–	X	•
Sambucus racemosa	red elderberry	8.3%	65	X	X	–	–	–
Sorbus sitchensis	Sitka mountain-ash	15.6%	29	–	–	X	X	–
Streptopus lanceolatus	rose twisted-stalk	11.4%	48	•	X	X	–	–
Struthiopteris spicant	deer fern	12.7%	44	X	X	•	–	–
Taxus brevifolia	Pacific yew	7.5%	67	X	X	–	–	–
Thuja plicata	western redcedar	22.1%	15	X	X	–	–	–
Tiarella trifoliata	foamflower	22.7%	13	X	X	X	–	–
Trillium ovatum	Pacific trillium	16.0%	27	X	X	–	–	–
Tsuga heterophylla	western hemlock	43.2%	2	X	X	•	–	–
Tsuga mertensiana	mountain hemlock	20.2%	18	–	–	X	X	•
Vaccinium deliciosum	Cascade blueberry	22.6%	14	–	–	•	X	X
Vaccinium membranaceum	big huckleberry	30.1%	6	•	•	X	X	–
Vaccinium ovalifolium	Alaska blueberry	24.2%	9	•	X	X	–	–
Vaccinium parvifolium	red huckleberry	27.7%	7	X	X	–	–	–
Vaccinium scoparium	grouse whortleberry	8.3%	66	–	–	X	X	–
Valeriana sitchensis	Sitka valerian	23.3%	11	–	–	X	X	–
Veratrum viride	corn lily	17.3%	25	–	•	X	X	–
Viola glabella	pioneer violet	12.1%	46	X	X	•	X	–
Viola orbiculata	roundleaf violet	13.7%	38	•	•	X	X	–
Viola sempervirens	evergreen violet	8.5%	60	X	•	•	–	–
Xerophyllum tenax	beargrass	18.5%	21	–	–	X	•	–

The distribution patterns of the most important species and map classes within each of the lifeform/land-use categories are discussed below.[75] For purposes of discussion, we treat natural

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

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 park, covering three-fifths of 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. Coniferous vegetation is concentrated in the montane zone: lowland forests occupy 9% of the park, lower montane forests occupy 12%, upper montane 27%, subalpine forests and woodlands 13%, and krummholz less than one percent. The seventeen species of conifers[76] are supplanted by other vegetation only where disturbance, snowpack, saturated soils or lack of soil development prevent their establishment and persistence.

Lowland forests

The pioneer species Douglas-fir (*Pseudotsuga menziesii*) and the shade-tolerant western hemlock (*Tsuga heterophylla*) dominate lowland conifer forests. Douglas-fir, the third most encountered plant in the park, is present throughout the lowland and montane zones—often as ancient emergent trees scattered over a younger canopy—and is missing from only the wettest sites. Western hemlock, the second most common plant, is the major successional tree in 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 and C06–WESTERN HEMLOCK, DOUGLAS-FIR AND SALAL FOREST. In moister settings, western redcedar is often codominant with western hemlock in impressive old-growth stands of C04–MOIST WESTERN HEMLOCK, DOUGLAS-FIR AND FOAMFLOWER FOREST, such as at the Grove of the Patriarchs near the Ohanapecosh River. The lower Carbon River valley hosts an unusual inland population of Sitka spruce (*Picea sitchensis*). Though similar to C04, the forests in the range of the spruce are treated here as C03–SITKA SPRUCE, WESTERN HEMLOCK AND WOOD-SORREL FOREST. Other lowland conifer forest associates, none of which are normally more than prominent, include the small understory tree Pacific yew (*Taxus brevifolia*), grand fir (*Abies grandis*) and western white pine (*Pinus monticola*). Small amounts of silver fir (*Abies amabilis*) are often present in the understory.

Red huckleberry (*Vaccinium parvifolium*) and dwarf Oregon-grape (*Berberis nervosa*) are prominent in the understory of most lowland conifer stands except in the moistest settings. Dense successional C05 forests often have very sparse understories, sometimes limited to scattered dwarf Oregon-grape, rattlesnake plantain (*Goodyera oblongifolia*) and little pipsissewa (*Chimaphila menziesii*). C06

[76] There are plausible reports of subalpine larch (*Larix lyallii*) growing with Sitka alder on morainal deposits at 4800’ elevation just below the Emmons Glacier. This would represent the 18th conifer in the park, more than are documented at either OLYM or NOCA.

forests always contain salal (*Gaultheria shallon*) and usually twinflower (*Linnaea borealis*), while C04 forests have a distinctive lush understory of foamflower (*Tiarella trifoliata*), sword fern (*Polystichum munitum*), devil's club (*Oplopanax horridus*), vanillaleaf (*Achlys triphylla*) and western oak fern (*Gymnocarpium dryopteris*). In addition to Sitka spruce, C03 is notable for another coastal affiliate, deer fern (*Struthiopteris spicant*).

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 devil's club, foamflower, trailing raspberry (*Rubus pedatus*) and queen's cup (*Clintonia uniflora*). Mesic midslope stands may have well-developed understories with twinflower, Alaska blueberry and red huckleberry, but may consist of dense younger stands with sparse understories of pipsissewa (*Chimaphila umbellata*) and one-sided wintergreen (*Orthilia secunda*). Another lower montane forest type, very infrequent in the park, is C15–LODGEPOLE PINE AND DOUGLAS-FIR WOODLAND. It is found only on lahar deposits near Longmire and on glacial outwash adjacent to the White River.

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 usually dominated by silver fir. C12–SILVER FIR, HEMLOCK AND ALASKA BLUEBERRY FOREST and C14–SILVER FIR, BIG HUCKLEBERRY AND BEARGRASS FOREST are transitional from the lower montane, with high elevation tree species usually subordinate to silver fir, western hemlock and/or Douglas-fir. C12 is a moist forest with Alaska-cedar and both hemlock species in the overstory and a lush understory characterized by Alaska blueberry and trailing raspberry. C14 is drier and often associated with past fire. It is the most abundant in the park, occupying a mid-montane elevation band especially on slopes with warmer aspects. Here, noble fir (*Abies procera*) and Alaska-cedar often join the canopy with silver fir, western hemlock and Douglas-fir. The understory is dominated by big huckleberry (*Vaccinium membranaceum*) and beargrass (*Xerophyllum tenax*) is usually present. At higher elevations both C12 and C14 give way to the second most abundant class, C13–MOUNTAIN HEMLOCK, SILVER FIR AND CASCADE AZALEA FOREST. This class is characterized by the near total replacement of western hemlock by mountain hemlock, a shrub-dominated understory consisting primarily of big huckleberry and Cascade azalea (*Rhododendron albiflorum*), and a ground layer usually dominated by dwarf bramble (*Rubus lasiococcus*).

Subalpine forests and woodlands

The subalpine zone is characterized by the increasing presence of large canopy gaps in and between treed areas. The shade-intolerant subalpine fir becomes much more prominent throughout the park in this zone. It is the undisputed dominant subalpine tree species in the northeast, where C20–

SUBALPINE FIR AND SITKA VALERIAN FOREST AND WOODLAND is found on middle to upper slopes at all but the most sheltered north-facing sites. Silver fir is usually present in these sites, at least as regeneration, and the understory is characterized by a lush herbaceous layer typically including dwarf bramble, Sitka valerian (*Valeriana sitchensis*), subalpine lupine (*Lupinus latifolius*) and smooth woodrush (*Luzula hitchcockii*). Drier woodlands above that—especially on steep south- and west-facing slopes—are C23–MOUNT RAINIER SUBALPINE FIR AND WHITEBARK PINE WOODLAND, with whitebark pine (*Pinus albicaulis*) usually present and a dry meadow-like understory of subalpine lupine, green fescue (*Festuca viridula*), Cascade aster (*Eucephalus ledophyllus*) and spreading phlox (*Phlox diffusa*). Engelmann spruce (*Picea engelmannii*) is a frequent but rarely dominant associate found especially in the east-side upper montane and subalpine zones. Outside the northeast, C21–MOUNTAIN HEMLOCK, SUBALPINE FIR AND HEATHER WOODLAND often signals the entry to the subalpine zone. These woodlands, transitional to mountain-heather shrublands above, are characterized by mountain hemlock and subalpine fir with openings occupied by pink mountain-heather (*Phyllodoce empetriformis*) and Cascade blueberry (*Vaccinium deliciosum*). C20 remains common in places where mountain-heathers are less prominent throughout the park's subalpine zone. 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.

4.2.2. Broadleaf trees

Communities dominated by deciduous broadleaf trees occupy two percent of the park. They are primarily a feature of 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*). Without repeated disturbance, these successional shrublands mature into B31–BROADLEAF RIPARIAN AND SWAMP FOREST. The B31 canopy is composed of red alder, usually with codominant western hemlock and western redcedar. Salmonberry (*Rubus spectabilis*), lady fern (*Athyrium filix-femina*) and fragrant bedstraw (*Galium triflorum*) are common and abundant in the understory of these forests. Upland deciduous or mixed forests are rare at MORA and are included with C05–WESTERN HEMLOCK, DOUGLAS-FIR AND SWORD FERN FOREST or mapped as the tall shrubland class S45–VINE MAPLE SHRUBLAND. Bigleaf maple (*Acer macrophyllum*) is nearly always the main broadleaf component at these sites, which usually have developed following landslides.

4.2.3. Upland tall shrubs

Tall shrublands in uplands cover three percent of the park, occupying avalanche tracks, montane talus slopes and toe-slope debris aprons. 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, moist talus and toe slopes in the montane zone. Although the species is not extremely common in the dataset, it is generally dominant where it occurs, and S43–SITKA ALDER SHRUBLAND is fairly abundant in the park. Salmonberry, lady fern, Brewer's miterwort (*Pectiantia breweri*) and red elderberry (*Sambucus racemosa*) are common associates here. Drier talus and toe slopes at somewhat lower elevations host S45–VINE MAPLE SHRUBLAND, in which vine maple (*Acer circinatum*) is often joined by young

bigleaf maple, a variety of smaller shrubs and bedstraw (*Galium* spp.). Vine maple is also very common in lowland coniferous forest understories.

4.2.4. Upland shrublands

Shorter shrublands in uplands cover six percent of the park, occupying disturbed areas at montane elevations and exposed areas with poor soil development at higher elevations. In the upper montane and subalpine zones, shrublands recovering from fire or otherwise undergoing succession toward conifers are often dominated by a mix of big huckleberry, Cascade blueberry, Sitka mountain-ash (*Sorbus sitchensis*) and Oregon-box (*Paxistima myrsinites*), often with prominent beargrass. These successional shrublands, mapped as S47–SUCCESSIONAL HUCKLEBERRY SHRUBLAND, are often found between forests and subalpine meadows. The iconic subalpine mountain-heather shrublands, S48–SUBALPINE HEATHER SHRUBLAND, are dominated by pink mountain-heather, white mountain-heather (*Cassiope mertensiana*) and Cascade blueberry, usually with prominent subalpine lupine and other forbs. They are most extensive clockwise from the south side through the northwest of the park, and are found just above C21–MOUNTAIN HEMLOCK, SUBALPINE FIR AND HEATHER WOODLAND. 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. It is associated with north-facing slopes and most extensive on the north side of the park.

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 six percent of the park's area. Herbaceous communities often transition across short distances, responding to finer-scale changes in topography, substrate and soil moisture than adjacent forests.

<u>Lowland, montane and subalpine elevations</u>

The lowest elevation herbaceous communities in the park occur in forest openings on bedrock-limited soils and are mapped as H58–BEDROCK BALDS AND SPARSELY VEGETATED FOREST OPENINGS. The most accessible congregation of bedrock balds is near the Muddy Fork Cowlitz in and below Box Canyon, but they are found in a variety of settings in the park. Many sparsely vegetated shrub-dominated areas on colluvial and glacial deposits below the closed forest line are also mapped as H58; there may be little that unifies these sites apart from their disturbed and early successional character. Lush meadows of subalpine lupine, showy sedge (*Carex spectabilis*), Sitka valerian (*Valeriana sitchensis*) and a variety of other herbaceous species are mapped as H53–SHOWY SEDGE & SITKA VALERIAN MEADOW. They are usually not far above the continuous forest line—in and around C20–SUBALPINE FIR AND SITKA VALERIAN FOREST AND WOODLAND—and are most abundant in the parklands within a few kilometers of Indian Bar, and nearly absent in the northwest.

Two subalpine meadow types are associated with the driest part of the park on Mount Rainier's northeast side. H57–GREEN FESCUE DRY MEADOW occurs in large expanses on steep south-facing slopes and is particularly abundant near Sunrise. These dry meadows are characterized by green fescue (*Festuca viridula*), spreading phlox and Cascade aster, and are found throughout the park but are much more limited in extent outside the northeast, where they are closely linked to C23–MOUNT RAINIER SUBALPINE FIR AND WHITEBARK PINE WOODLAND. H56–SUBALPINE SUMMER-DRY GRASS-FORB MEADOW is much more geographically constrained, being limited to within a few kilometers of Sunrise. It covers most of Grand Park and other smaller flats and gentle slopes nearby, and is particularly characterized by woolly pussytoes (*Antennaria lanata*), fan-leaf cinquefoil (*Potentilla flabellifolia*) and alpine aster (*Oreostemma alpigenum*). Its habitat is marked by significant winter snow accumulation and is moist in the spring, but becomes quite dry by midsummer.

Alpine elevations

Variations in soil moisture, substrate and length of time since colonization contribute to the wide range of herbaceous alpine communities at Mount Rainier. We have subsumed a good bit of this variety into a widespread map class, H62–ALPINE SPARSE HERBACEOUS VEGETATION, which is characterized by partridgefoot (*Luetkea pectinata*), Parry's rush (*Juncus parryi*) and exclusively alpine plants such as Piper's woodrush (*Luzula piperi*) and Tolmie's saxifrage (*Micranthes tolmiei*). This is the most abundant herbaceous map class in the park, found throughout the alpine zone on rocky upper slopes where snow lingers late into the growing season. Two alpine classes are strongly associated with pumice substrates, mostly in the northeast and especially on and around Burroughs Mountain. H64–ALPINE LUPINE PUMICE VEGETATION is found on west-facing alpine slopes and has reasonably high vegetative cover, typically including prairie lupine (*Lupinus lepidus*), spreading phlox and coiled-beak lousewort (*Pedicularis contorta*). It has strong similarities to H61–SPREADING PHLOX AND COMMON JUNIPER DRY MEADOW at OLYM, and the two classes should perhaps be combined. The more desolate H63–ALPINE BUCKWHEAT PUMICE VEGETATION is associated with colder and snowier east-facing slopes. The most commonly found plants there are dirty socks (*Eriogonum pyrolifolium*) and Davis' knotweed (*Aconogonon davisiae*).

4.2.6. Wetlands

Less than one percent of the park is mapped as one of the five wetland map classes. Low-gradient areas conducive to wetland formation are mostly limited in the Cascades to lowland valleys and high-elevation headwaters basins. MORA has a higher base elevation than the other NCCN parks and the opportunities for extensive lowland wetlands are correspondingly limited; consequently, most of its wetlands are in the subalpine zone. There are riparian communities and occasional midslope seeps in the intermediate montane zones, but few large wetlands.

Lowlands

Lowland herbaceous wetlands, mapped as H50W–LOWLAND MARSH AND MEADOW, are usually dominated by water sedge (*Carex aquatilis*) and other graminoids and usually are found along pond and lake margins. The very limited such wetlands in the park are on the south side, mostly near the Nisqually and Ohanapecosh Rivers. S40W–LOW ELEVATION SHRUB SWAMP wetlands, usually dominated by Sitka willow (*Salix sitchensis*) and rose spirea (*Spiraea douglasii*), may ring

herbaceous vegetation adjacent to a pond or occur on wet toe slopes or areas where drainage is impeded. They appear to be more common in the park but are still quite limited in extent.

Subalpine and alpine

Herbaceous subalpine wetlands, mapped as H51W–SUBALPINE HERBACEOUS WETLAND, may be composed of a mix of fan-leaf cinquefoil, black alpine sedge (*Carex nigricans*), lakeshore sedge (*Carex kelloggii*) and a variety of other forbs and sedges. These wetlands are found in high-elevation basins throughout the park. Occasionally dense patches of Barclay's willow (*Salix barclayi*) may dominate these environments; these are mapped as S41W–SUBALPINE WILLOW WETLAND. The two types often form mosaics, with 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. H60W–BLACK ALPINE SEDGE WETLAND is rarely connected to the water table; patches are often inclusions within subalpine or alpine heather shrublands. Rock barrens are usually nearby.

4.2.7. Natural abiotic areas

Unvegetated natural areas are extensive in the park, cumulatively occupying nearly a quarter of the landscape. W82–EXPOSED SNOW AND ICE is the most abundant class, covering eight percent of the park. Ice is also present beneath the surface at a significant fraction of the three percent of the park mapped as R71–ALLUVIAL BARREN AND DEBRIS-COVERED ICE. R72–COLLUVIAL BARREN and R73–BEDROCK BARREN occupy six and four percent of the park, respectively. Although most of these barrens may entirely lack vascular plants, many include sparse vegetation; unfortunately, they were not well-sampled at MORA so cannot be described floristically. W81–FRESH WATER in the form of rivers and lakes covers one percent of the park. Mowich is by far the largest of the lakes, all of which are natural.

4.2.8. Natural and semi-natural disturbed landscapes

Only one-tenth of one percent of the park 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.

4.2.9. Development

Roads and developed areas cumulatively occupy 0.6% of the park. Unsurprisingly, this is substantially more than in OLYM or NOCA.

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 park. 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.

The most common landscape change agent in the park is periodic severe flooding resulting from storms and glacial outbursts (Antonova et al. 2014). 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.

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.

Glaciers in the park have receded significantly, mostly over the Holocene, but also more recently (Beason 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 unlike any map classes, but are mapped as their best match: for alpine moraines, this may be S49–ALPINE HEATHER SHRUBLAND or H62–ALPINE SPARSE HERBACEOUS VEGETATION; on terminal moraines in valleys it is often B30–SUCCESSIONAL GRAVEL BAR SHRUBLAND or S43–SITKA ALDER SHRUBLAND.

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.

Although fire has left its legacy in the map (see Hemstrom 1979), only two significant fires have occurred in the park in the last 35 years. The Redstone fire (115 hectares, west of the West Fork White River) and the Panther fire (21 hectares, just east of Highway 123 in the Ohanapecosh Valley) both burned in 2003. Areas not resembling fire-compatible vegetation classes within those burns

were mapped instead 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.

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 Longmire as well as various farmed, residential, or industrial lands outside the park. 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 16, 2016. 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 20**.

Literature Cited

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

Antonova, N., C. Copass, and S. Clary. 2014. Landsat-based monitoring of landscape dynamics in Mount Rainier National Park: 1985–2009. Natural Resource Data Series NPS/NCCN/NRDS—2013/637. 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.

Beason, S. R. 2017. Change in glacial extent at Mount Rainier National Park from 1896 to 2015. Natural Resource Report NPS/MORA/NRR—2017/1472. National Park Service, Fort Collins, Colorado. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2242142.

Biek, D. 2000. Flora of Mount Rainier National Park. Oregon State University Press. Corvallis, Oregon.

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

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.

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.

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. Mount Rainier National Park map class key. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

Buckingham, N. M., E. G. Schreiner, T. N. Kaye, J. E. Burger, and E. L. Tisch. 1995. Flora of the Olympic peninsula. Northwest Interpretive Association and Washington Native Plant Society, Seattle, Washington.

Burke Herbarium. 2020. Burke herbarium image collection. Available at: https://biology.burke.washington.edu/herbarium/imagecollection.php.

Burtchard, G. C. 2007. Holocene subsistence and settlement patterns: Mount Rainier and the montane Pacific Northwest. Archaeology in Washington 13:3-44. Available at: https://www.nps.gov/mora/learn/historyculture/upload/AIW-Burtchard2007_acess.pdf.

Burtchard, G. C., J. Y. Cheung, and R. McIntyre. 2017. Mining Glacier Basin: History of the Glacier Basin Mining District, Mount Rainier National Park. Available at: https://www.nps.gov/mora/learn/historyculture/upload/2017-mining-glacier-basin-burtchard_web.pdf.

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.

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.

Crandell, D. R. 1969. The geologic story of Mount Rainier. Geological Survey Bulletin 1292. U.S. Department of the Interior, Washington, DC. Available at: https://pubs.er.usgs.gov/publication/b1292.

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.

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).

Dryden, C. 1968. Dryden's history of Washington. Binfords and Mort, Portland, Oregon.

Edwards, O. M. 1980. The alpine vegetation of Mount Rainier National Park: structure, development and constraints. University of Washington Press, Seattle, Washington.

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).

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, G. W. Douglas and C. Wiberg. 1971. Invasion of subalpine meadows by trees in the Cascade Range, Washington and Oregon. Arctic and Alpine Research 3:215–224.

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.

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.

Hamann, M. J. 1972. Vegetation of alpine and subalpine meadows of Mount Rainier National Park, Washington. Thesis, Botany. Washington State University, Pullman, Washington.

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.

Hemstrom, M. A. 1979. A recent disturbance history of forest ecosystems at Mount Rainier National Park. Thesis, Plant Ecology. Oregon State University, Corvallis, Oregon. Available at: https://ir.library.oregonstate.edu/downloads/5x21tj29s.

Hemstrom, M. A., and J. F. Franklin. 1982. Fire and other disturbances of the forests in Mount Rainier National Park. Quaternary Research 18:32–51.

Henderson, J. A. 1974. Composition, distribution and succession of subalpine meadows in Mount Rainier National Park. Thesis, Plant Ecology. Oregon State University, Corvallis, Oregon. Available at: https://ir.library.oregonstate.edu/downloads/n870zw403.

Hijmans, R. J. 2018. R package 'raster': geographic data analysis and modeling, version 2.8-4. Available at: https://cran.r-project.org/package=raster.

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.

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. Mount Rainier National Park accuracy assessment contingency tables. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

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.

Isaaks, E. H., and R. M. Srivastava. 1988. Spatial continuity measures for probabilistic and deterministic geostatistics. Mathematical Geology 20:313–341.

Jones, G. N. 1938. The flowering plants and ferns of Mount Rainier. University of Washington, Seattle, Washington.

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.

Kroiss, S. J., and J. Hille Ris Lambers. 2015. Recruitment limitation of long-lived conifers: implications for climate change responses. Ecology 96:1286–1297.

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.

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).

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.

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

Menenti, M., and J. C. Ritchie. 1994. Estimation of effective aerodynamic roughness of Walnut Gulch watershed with laser altimeter measurements. Water Resources Research 30:1329–1337.

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

Moore, P. L., L. I. Nelson, and T. M. D. Groth. 2019. Debris properties and mass-balance impacts on adjacent debris-covered glaciers, Mount Rainier, USA. Arctic, Antarctic, and Alpine Research 51:70–83. Available at: https://lib.dr.iastate.edu/nrem_pubs/335.

Mullineaux, D. R. 1974. Pumice and other pyroclastic deposits in Mount Rainier National Park, Washington. Geological Survey Bulletin 1326. U. S. Department of the Interior, Washington, DC. Available at: https://pubs.er.usgs.gov/publication/b1326.

National Park Service (NPS). 2012. Inventory and monitoring at Mount Rainier National Park. Available at: https://www.nps.gov/im/nccn/mora.htm.

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

National Park Service (NPS). 2020a. Annual visitation: Mount Rainier National Park. Available at: https://www.nps.gov/mora/learn/management/annual-visitation.htm.

National Park Service (NPS). 2020b. Mount Rainier National Park: Nature notes author index. Available at: https://www.nps.gov/parkhistory/online_books/mora/notes/nn-auth.htm#b.

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. 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., C. Copass, R. L. Brunner, and L. K. Wise. 2021b. 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. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2286418.

Nielsen, E. M., R. L. Brunner, C. Copass, and L. K. Wise. 2021c. Mount Rainier National Park map class descriptions. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

Nielsen, E. M., C. Copass, R. L. Brunner and K. Braun. 2021d. Mount Rainier National Park vegetation map. National Park Service, Port Angeles, Washington. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2285187.

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.

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

Parker, G. G., and M. E. Russ. 2004. The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry. Forest Ecology and Management 189:307–315.

Pike, R. J., and S. E. Wilson. 1971. Elevation-relief ratio, hypsometric integral and geomorphic area-altitude analysis. Geological Society of America Bulletin 82:1079–1084.

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.

PRISM Climate Group. 2019. PRISM climate data. Available at: https://prism.oregonstate.edu/.

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/.

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.

Riedel, J., and M. A. Larrabee. 2011. Mount Rainier National Park glacier mass balance monitoring annual report, water year 2009. Natural Resource Technical Report NPS/NCCN/NRTR—2011/484. National Park Service, Fort Collins, Colorado.

Rochefort, R. M. 2010. Vascular plant inventory of Mount Rainier National Park. Natural Resource Technical Report NPS/NCCN/NRTR—2010/347. National Park Service, Fort Collins, Colorado. Available at: https://irma.nps.gov/DataStore/Reference/Profile/2124857.

Rochefort, R. M., and D. L. Peterson. 1996. Temporal and spatial distribution of trees in subalpine meadows of Mount Rainier National Park, Washington, U.S.A. Arctic and Alpine Research 28:52–59. Available at: https://s3.wp.wsu.edu/uploads/sites/902/2010/04/2010summer_article-rochefort-peterson1996.pdf.

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.

Sethi, M. L., E. J. Theobald, I. Breckheimer, and J. Hille Ris Lambers. 2020. Early snowmelt and warmer, drier summers shrink post-flowering transition times in subalpine wildflowers. Ecology 101:1–11. Available at: https://doi.org/10.1002/ecy.3171.

Sigafoos, R. S., and E. L. Hendricks. 1961. Botanical evidence of the modern history of Nisqually Glacier, Washington. Geological Survey Professional Paper 387-A. U. S. Department of the Interior, Washington, DC. Available at: https://pubs.er.usgs.gov/publication/pp387A.

Stueve, K. M., D. L. Cerney, R. M. Rochefort, and L. L. Kurth. 2009. Post-fire tree establishment patterns at the alpine treeline ecotone: Mount Rainier National Park, Washington, USA. Journal of Vegetation Science 20:107–120.

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.

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.

United States Geological Survey (USGS). 2014. Holocene or post-glacial eruptions of Mount Rainier. Available at: https://volcanoes.usgs.gov/volcanoes/mount_rainier/mount_rainier_geo_hist_76.html.

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.

United States Geological Survey (USGS). 2020. Mount Rainier. Available at: https://www.usgs.gov/volcanoes/mount-rainier.

United States National Vegetation Classification (USNVC). 2019. United States National Vegetation Classification Database, V2.03. Federal Geographic Data Committee, Vegetation Subcommittee, Washington, DC. Available at: http://usnvc.org.

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.

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.

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.

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.

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 105/175300, May 2021

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