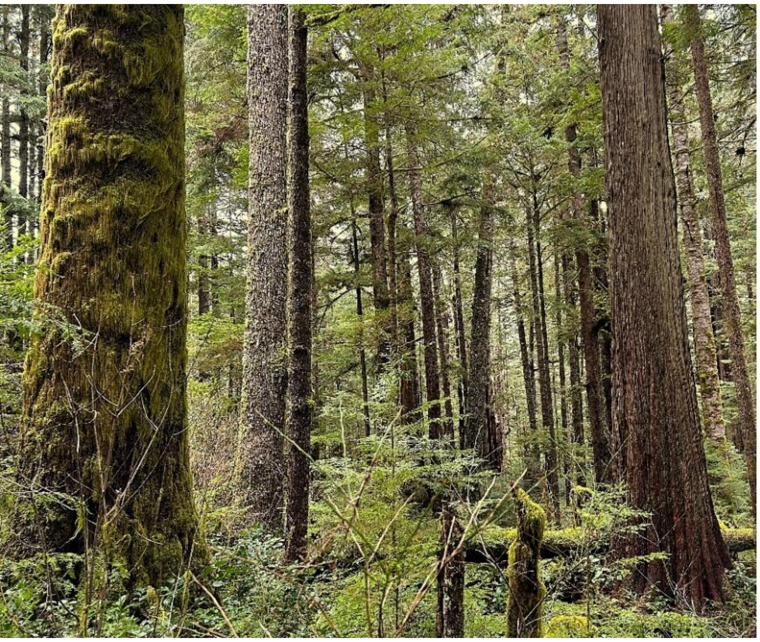
Structurally complex and older forest detectors

Ethan W. Hughes, Allison Bailey, Daniel C. Donato,
Kate McBurney, and Joshua S. Halofsky



Washington State Department
of Natural Resources
July 07, 2025

Cover photos (starting from top): Old cedar/hemlock forest in Coast District of the Olympic Peninsula (Allen Estep/DNR); Late-mature Douglas-fir/hemlock forest in Northwest Region, west Cascades (Daniel Donato/DNR); Old silver fir/hemlock/cedar forest in the Coast District of the Olympic Peninsula (Allen Estep/DNR).

Contents

List of Figures

List of Tables

Acknowledgments
We would like to thank and acknowledge the support of Informatics staff for their time reviewing methodology, code review, and analytical insights to help make this project successful in the short timeframe it was executed. We also thank region staff, especially region foresters and biologists, for their work collecting the ground truth stand development stage data in western WA. We could not have conducted these analyses without their hard work and expertise.

Executive summary

The Washington State Department of Natural Resources (DNR) is tasked with identifying, mapping, and projecting structurally complex forest (SCF) and older forest (OF) conditions across its westside forestlands. OF are a subset of SCF and thus not necessarily exclusive. This report describes the Detectors project, which evaluated whether DNR's remotely-sensed forest inventory system (RS-FRIS)—a height-driven dataset based on aerial photogrammetry—could be used to accurately map and differentiate SCF and OF conditions at the scale of a forest stand, using a logistic regression approach.

The study had two main objectives: (1) assess whether models developed from RS-FRIS could reliably detect SCF and OF across DNR-managed forest lands in western Washington, and (2) compare these RS-FRIS-derived results with other mapping efforts using different methodologies.

Over 2200 field plots were sampled to train and test the SCF and OF models. Results indicated RS-FRIS metrics ascribing canopy layers and a measure of mean tree diameter were most effective in identifying SCF from simpler forest structures, while those same attributes, along with canopy texture, were used to classify OF. The SCF model was found to better predict forest conditions of interest than the OF model. The relatively high accuracy of the SCF model provides DNR with greater confidence in using that dataset moving forward. However, the more equivocal results of the OF model suggest more research and development is necessary prior to using that model in future analyses.

The SCF model compared favorably with pre-existing datasets, performing similarly or better than other available maps. The OF model outperformed one dataset but underperformed compared to another. While future work will explore other methods to classify and improve the SCF and OF models, a more immediate outcome will be to estimate the abundance and location of SCF (i.e., within or outside of existing conservation areas) as part of DNRs Landscape Assessment.

Introduction

Mapping mature and old forests over large landscapes has become a prominent issue in public forest management (Davis et al. 2015, 2022; Gray et al. 2023; Pelz et al. 2023; Woodall et al. 2023). A particular challenge is developing spatial mapping tools that both cover broad areas spanning millions of acres, while also having stand-level accuracy.

To date, available tools have focused primarily on providing estimates of mature and old forest *amounts* over broad spatial domains (Davis et al. 2015, 2022; Pelz et al. 2023; Woodall et al. 2023). These tools, such as the Gradient Nearest Neighbor (GNN, Ohmann and Gregory 2002; Bell et al. 2015) developed by the US Forest Service, are widely used for monitoring broad-scale trends but lack the fine-scale fidelity to evaluate individual stands. In contrast, fine-scale methods, often employing high-resolution LiDAR (Kane et al. 2010a, 2010b), have been used to identify and characterize old-growth forest stands. These fine-scale approaches typically incorporate various combinations of forest height, canopy features such as cover or layering, and canopy ruggedness (e.g., 'rumple factor' or rugosity metrics), to distinguish old from younger forests at the stand scale. High-resolution data can discern fine details of forest canopies, and are well-proven in this application, with the main drawback being such approaches typically only provide a one-time snapshot because LiDAR data is rarely reacquired through time due to cost.

The Washington State Department of Natural Resources (DNR) manages westside forestlands under policies that necessitate the quantification and mapping of certain mature and old forest conditions on the landscape (DNR 1997, 2006; see detailed background below). A key need is the ability to not only map these structure conditions currently, but to also project changes in their abundance and distribution over the next several decades in a manner compatible with existing forest growth and yield models. Ideally, DNR could accomplish this mapping and projecting goal using its currently available, remotely-sensed forest inventory (RS-FRIS; Ricklefs 2025). DNR's inventory data derives wall-to-wall estimates of dozens of forest structure metrics based on structure-from-motion digital aerial photogrammetry (DAP). A key factor is that the structure metrics are generated from statistical regressions mostly driven by aspects of forest height, making the RS-FRIS inventory fundamentally a height-driven dataset.

The primary objective of this project was to evaluate the degree to which a mainly height-driven forest inventory dataset could be used to distinguish certain mature and old forest structural conditions across DNR-managed forestlands in western Washington (Part 1). A secondary objective was to compare any resulting RS-FRIS derived mature and old forest structure maps with other maps that have attempted to capture similar forest conditions using different statistical approaches (Part 2). The project was prioritized and requested to be completed (including data collection, data processing, model development and testing, and reporting) within four months in early 2025.

Part 1: DNR Detectors

1. Background

In western Washington, DNR-managed forestlands fall under a Habitat Conservation Plan (HCP, DNR 1997) and the Policy for Sustainable Forests (DNR 2006), which collectively requires the agency to track 'structurally complex' and 'older' forests (respectively, at least Maturation II and Vertical Diversification structure; Van Pelt 2007; Table 1). These structurally complex forest (SCF) and older forest (OF) thresholds are part of a continuum of forest conditions that change over time as forests grow and develop (Table 1, Figure 1).

Table 1. Van Pelt Stand Development Stages (SDS; Van Pelt 2007).

Van Pelt SDS	Structurally Complex Forest (SCF)	Older Forest (OF)
SI - Stand Initiation	No	No
CC - Canopy Closure	No	No
CE - Competitive Exclusion	No	No
M1 - Maturation I	No	No
M2 - Maturation II	Yes	No
VD - Vertical Diversification	Yes	Yes
HD - Horizontal Diversification	Yes	Yes
PL - Pioneer Cohort Loss	Yes	Yes

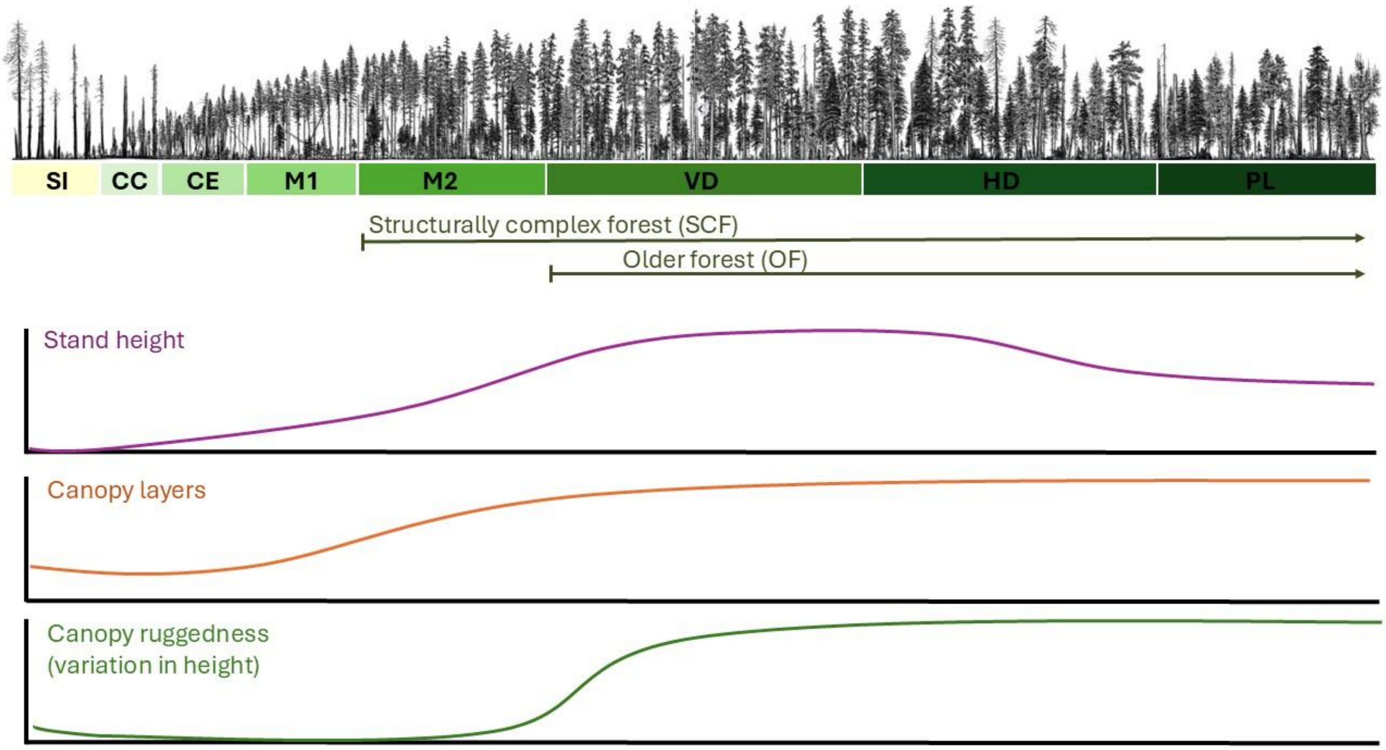


Figure 1. An illustration of forest growth and development beginning with a stand-replacing disturbance (top row; drawing by Robert Van Pelt (in Franklin et al. 2018). Below the illustration are approximations of various stand development stages (SDSs) as defined by Van Pelt (2007, see Table 1 for full SDS names), followed by the SDS thresholds used in this project to delineate structurally complex forests (SCF) and older forests (OF). The bottom three rows represent conceptual curves illustrating how various stand structural elements change with SDS, per published studies (Franklin and Van Pelt 2004; Kane et al. 2010a; Griffey et al. 2021).

Various models of forest succession, or stand development stages (SDSs), have been proposed (e.g., Carey and Curtis 1996; Oliver and Larson 1996; Franklin et al. 2002). DNR contracted with Dr. Robert Van Pelt in 2006, an Affiliate Assistant Professor at the University of Washington, to create the first widely used dichotomous key of SDSs for western Washington, in a field guide he authored called, "Identifying Mature and Old Forests in Western Washington" (Van Pelt, 2007). Dr. Van Pelt is one of the foremost forest ecology experts in the Pacific Northwest, having authored several dozen peer-reviewed scientific papers on old-growth forest structure, and extensively studied old-growth forests across North America, particularly in California and the Pacific Northwest. The field guide and SDS dichotomous key developed by Dr. Van Pelt was the backbone of DNR's data collection for the current effort, known as the Detectors project.

The DNR Detectors project (McBurney et al. 2024) was initiated to address the needs of the Landscape Assessment report, which periodically provides estimates of western Washington DNR-managed forestlands designated as SCF (Maturation II, Vertical Diversification, Horizontal Diversification, and Pioneer Cohort Loss) that can be managed to meet Older Forest (Vertical Diversification, Horizontal Diversification, and Pioneer Cohort Loss) conditions (Table 1, Figure 1). The Detectors project was tasked with developing a new approach to classify DNR-managed forestlands into SCF and/or OF for potential use for the Landscape Assessment, using RS-FRIS data. The Detectors project required that the resulting models be compatible with growth and yield modeling software for projection into future decades. This limited the analysis to variables (e.g., tree heights, diameters, basal area, etc.) that could be modeled by available software packages. Therefore, non-parametric approaches or those using direct canopy measurements (e.g., LiDAR or DAP-derived canopy height models) that could not be projected forward in time were not considered in the Detectors analysis.

Part 1 of this report reviews the methods used for variable selection and model building to estimate SCF and OF. The methods and results of this research are subject to change as other approaches may be explored in future research, and new SDS training plots are collected and incorporated into the models.

2. Methods

2.1 SDS training data

2.1.1 Overview

This section describes the SDS field data collection effort. The training data served as the foundation for establishing SCF and OF Detector model relationships with RS-FRIS metrics.

The field assessments were conducted in over 2200 stands (2.5-acre circles; described in Section 2.1.3) across DNR-managed forestlands in western Washington (Figure 2) by trained and experienced field staff. Stands were field-visited, geolocated, and classified to development stage. The dataset included stands visited specifically for the Detectors project as well as instances where SDS information was gleaned from other work, such as old-growth assessments and pre-timber sale stand evaluations. Once the data were quality-checked for clarity and accuracy, the dataset was routed through subsequent Geographic Information System (GIS) processing and statistical analyses to identify variables and variable thresholds that best distinguish SCF and OF.

The objective of the data collection was not to quantify the amounts of SDS conditions on the landscape (that is the objective of the models). Rather, the goal was to provide a robust and diverse set of examples of each stage for the Detector models, to inform how a given stage is typically structured in terms of RS-FRIS metrics. As such, plots needed to represent each *stage* well in terms of their range of characteristics across the landscape, but not necessarily reflect the relative landscape *abundance* of stages (Figure 3).

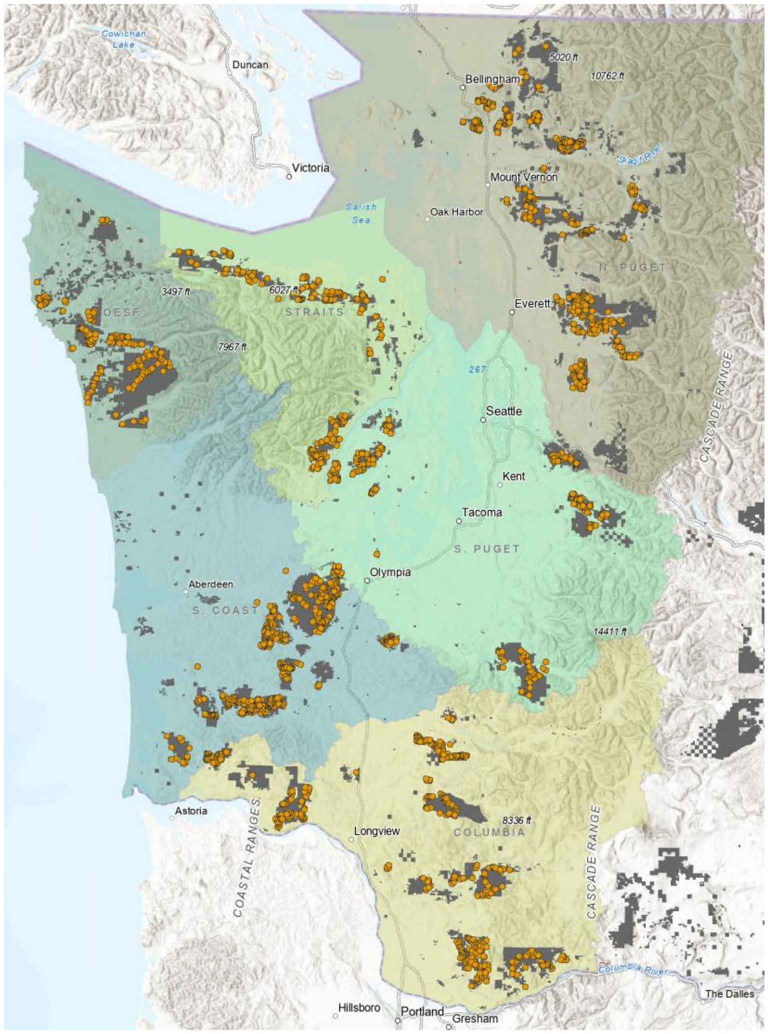


Figure 2. Geographic distribution of the 2217 field plots used to train the Detector models. The six habitat conservation plan (HCP) planning units are shown as colored polygons across western Washington, with DNR-managed forestlands within each HCP unit depicted in dark gray.

Pilot data collection related to this project began in late 2023 through 2024 as part of prior work, but the bulk of data collection occurred during a few weeks in February-March 2025 due to the accelerated timeframe associated with the project. State Uplands personnel from the Forest Resources Division and each westside region participated in field data collection. Because of the array of data sources (further described below), a total of 68 different field observers collected at least one field plot; however, ~ 90% of the sample was accounted for by 20 observers. These personnel were all officially trained on SDS classification of and/or old-growth, instructed by Dr. Robert Van Pelt and/or Dr. Daniel Donato. Dr. Donato, an Affiliate Assistant Professor at the University of Washington, is DNR's designated agency expert on old-growth and forest development. Dr. Donato has collaborated with Dr. Van Pelt on guidance for old-growth assessments, and has taught DNR's old-growth classes for 12 years.

While data collection is largely complete for the Detectors project, the collection of additional SDS training plots will continue through the RS-FRIS inventory program, pre-sale assessments, old-growth assessments, and/or other necessary work to refine and update future Detector models.

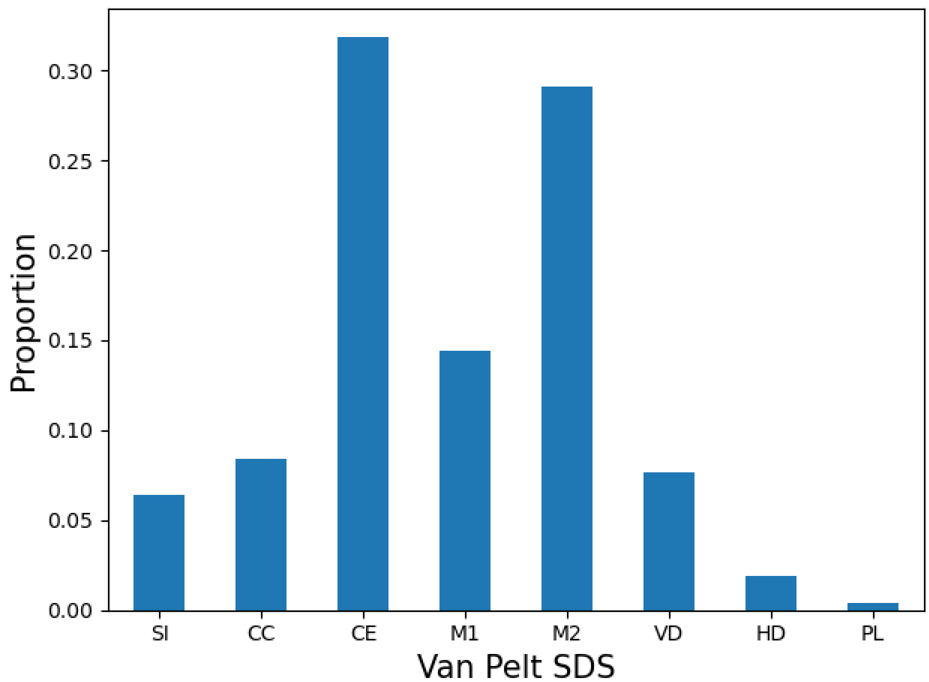


Figure 3. Distribution of the 2217 training plots and corresponding stand development stage (SDS) evaluations. This distribution does not necessarily (nor was it intended to) reflect the actual abundance of various structural stages across the landscape. See Table 1 for full Van Pelt SDS names.

2.1.2 Sampling design and approach

Data collection was stratified to ensure representative distribution across all westside HCP planning units, major land blocks, forest types, riparian and non-riparian areas, and development stages within each of these. Given the compressed timeframe to obtain >2200 plots, a strictly random sample – such as a grid of plots from a random start point – was not practical. Stratification was employed to minimize bias and ensure good coverage of the range of environmental conditions present on westside DNR-managed forestlands. Stratification across all the major gradients was accomplished by using spatial data on land block, riparian areas, and estimated stand height (mean height of the 40 tallest trees from RS-FRIS, applying the *a priori* known generic relationship between development stage and height). Data distributions were checked to compare the characteristics of the training plots to the distribution of major environmental factors present across westside DNR-managed forestland (see below and Appendix A).

2.1.3 Data collection

Project-specific training plots

Most of the data was specifically collected for the purpose of this project. Field personnel collected SDS training plots at pre-mapped points generated in GIS (Figure 2). For efficiency in collecting thousands of points in a short time, plots were mapped primarily in areas accessible by road, but with an aim to be at least 250’ from roads to avoid edge effects. While this constraint meant that 96% of plots were within 860’ of a road, that is also the case for ~85% of westside DNR-managed forestlands, suggesting that data collection near roads (but away from road influence) does not majorly bias the dataset. Training plots were also at least 250’ from resource inventory unit (RIU) boundaries (delineated forested stand polygons maintained by DNR’s Forest Inventory team) to keep plots within a single stand type, and at least 250’ from each other. These minima could not always be met in real-world landscapes (and initial rounds of pilot sampling followed less strict criteria), but subsequent processing steps accounted for these exceptions (see section 2.1.5 for details). Opportunistic training plots in addition to pre-mapped points were allowed, but instructions were to meet the same criteria as listed above. Data collection focused on stands that had not been harvested since the 2021-22 imagery used in RS-FRIS 5.0 (i.e., the training imagery), ensuring that the imagery and RS-FRIS estimates aligned with the conditions of the surveyed stands.

Each field plot was a 2.5-acre circle, as assessment of SDSs should capture stand-scale heterogeneity (Van Pelt 2007; Kane et al. 2010a). The surveyor recorded GPS coordinates at a central point to mark plot center and then evaluated a ~ 60-yard visual radius, classifying the stand with the Key to Stand Development Stages on page 46-47 of Van Pelt (2007). Most plots were keyed with an updated version of the key, currently in revision in collaboration with Dr. Van Pelt (Van Pelt 2025, in prep). The updates do not change the criteria in the key but clarify the originally-intended distinctions among the steps with more quantitative descriptions and thresholds. For example, subjective terms in the old key (e.g., "sparse understory") have been updated to clearer objective terms (e.g., "conifer understory <15% cover"). In addition to making an SDS assessment with the key, surveyors could take notes at each training plot to describe the stand or how their call was made.

Prior old-growth assessments

The entire record of DNR's field old-growth assessments from 2006-2024 was examined to extract SDS information. Not all reports were usable in this respect, as some did not clearly indicate SDS, or the stand was a mix of conditions with spatial locations of each condition not well defined. Where the narrative identified a clear SDS, a single training plot was mapped for the current project near the centroid of the assessed polygon. Old-growth assessments provided 95 of the total training plots.

Prior pre-timber-sale SDS assessments

Since early 2024, DNR has conducted SDS assessments for all proposed westside timber sale units. All pre-sale reports available as of January 2025 were examined for field calls of SDS made by trained staff, and training points placed within stands consistent with narrative pre-sale report descriptions. These pre-sale reports provided 290 of the total training plots.

2.1.4 Representativeness of dataset relative to the landscape

While the objective of all training plots was not to estimate the amount of SDSs on the landscape, it was important to assess whether the set of training plots was well distributed environmentally, rather than covering only a portion of the landscape (e.g., all from one planning unit or forest zone) or missing major portions of key environmental gradients.

The geographic distribution of plots largely corresponded well to the amount of DNR-managed forestland acres among the westside HCP planning units (Table 2). Note that ongoing data collection in the Olympic Experimental State Forest (OESF) will bring the number of plots there closer to that unit's proportion of the westside landscape for future model updates; all added plots will be processed and included in subsequent model updates. Additionally, all eight SDSs were represented in approximately similar proportions in each planning unit (Figure A1). Similarly, plot distribution across riparian areas and environmental gradients (elevation bands, forest types, and site productivity classes) compared well to the distributions of all westside DNR-managed forestlands (Appendix A), indicating the spatial stratification resulted in good representation of conditions on DNR lands, without significant biases.

Table 2. Comparison of training plot distributions among westside DNR-managed forestlands, by HCP planning unit (HCP). The percentages in the second column add up to 99% due to rounding to the nearest percent. Note that ongoing data collection in OESF will bring its plot percentage up for future model updates.

HCP planning unit	% of westside DNR-managed forest lands	% of SDS training plots collected
Columbia	19%	20%
North Puget	28%	26%
Olympic Experimental State Forest (OESF)	17%	7%
South Coast	16%	25%
South Puget	11%	12%
Straits	8%	10%

2.1.5 Field Quality Checking

Each plot was screened for clarity and accuracy before being forwarded for subsequent processing and analysis. Plots were omitted from further analysis if, for example, an apparent GPS error was suspected (e.g., plot not on DNR-managed forestlands), the surveyor did not make a call of a single development stage, or surveyor notes created uncertainty and communication with the surveyor did not clear the uncertainty. After these various quality checks, the final plot count sent to subsequent processing steps was n=2237, which was further reduced to 2217 as described below.

In addition, a key quality assurance component was evaluating SDS training plot consistency among different observers. A subsample of the dataset – 10.1%, or 224 plots – was intentionally visited twice, once for SDS calls originally made by 21 different observers (training points used in the model formulation), and a second time to be reexamined by different field observers, with each observer unaware of the SDS call made by the other. The outcome of this assessment was to get a quantitative sense of observer consistency by DNR-trained staff. If multiple observers frequently assigned different SDSs (or, most importantly for this application, SCF vs not or OF vs not) to the same stand, model training would be compromised. However, if different staff consistently classified a stand to the same development stage (especially SCF vs not and OF vs not), model training would be considered stable and reliable.

Overall mean accuracy of SDS assignment between independent field observers was high (89-96%, Table 3). The distribution of the reexamined training plots closely approximated the distribution of SDSs in the whole sample (Figure 1). Given the inherently variable and fuzzy nature of SDSs in the real world, and the prevalence of truly borderline/transition strands, obtaining the level of consistency found across the 224 plots strengthened confidence of the data collected by DNR-trained staff.

Table 3. Cross-observer consistency in stand development stage (SDS) field classification (i.e., how often two independent observers assigned the same SDS for a given plot) across a random 10.1% subsample of the dataset (n=224 plots).

Same specific SDS call *(among 8 possible stages, see Table 1, Fig 1)*	Structurally complex forest (SCF)	Older forest (OF)
89%	96%	93%

2.2 Training plot polygon delineation in GIS

2.2.1 Overview

This section describes the GIS workflow used to 1) prepare the SDS training plots, and 2) combine the plots with RS-FRIS variables. The outcome of these steps was model-ready data used to develop, train, and test the Detector models.

Briefly, and further described below, SDS training plots were combined with DNR's geospatial layers to create plot polygons, which defined the broader forest stand conditions associated with each SDS training plot. Plot polygons were then used to extract inventory variables from RS-FRIS. Data sources included RS-FRIS (version 5.0) and the LDO (version 20240927). Other GIS data sources used were current as of 2025-03-17. The RS-FRIS data associated with each plot polygon then became the model-ready data for the Detectors analyses.

2.2.2 Bringing SDS training plots into GIS

Each SDS training plot location included a x, y point location, SDS field assessment classification (Table 1), the date of visit, field observer name, and a unique plot identifier. Training plots and associated attributes were compiled into an ESRI file geodatabase and projected to DNR's standard Washington State Plane South projection (FIPS 4602, NAD 1983 HARN datum).

Additional attributes were assigned to the training plots (n=2237) to aid Detector model development. These variables included the HCP unit name (Table 2), RIU identifier, and occurrence within or outside a riparian area (based on the LDO riparian identifier).

2.2.3 Plot polygons

Two different sized circular plot polygons, 2.5 and 5.0 acres, were created to correspond, respectively, with the SDS field training plot size and a larger stand area to explore the influence of potentially capturing more stand heterogeneity (Van Pelt, personal communication). In many cases, the resulting plot polygons were smaller and irregularly shaped due to the need to constrain the polygon with features that naturally limit the extent of the forested portion of a stand (such as roads and water bodies), as well as limiting the plot to a single RIU polygon to reduce impacts from adjacent stand structures that may be classified differently. These spatial constraints were designed to ensure the 2.5- and 5-acre polygons delineating the training plot represented a single stand (and SDS value) as intended by the field observer.

The first step in generating plot polygons was to buffer the training plot locations by two radii lengths, 186.2 and 263.3 feet, to create two polygon feature classes containing circular plots of approximately 2.5 and 5 acres in area. (Figure 4). Each plot polygon was then clipped to the associated RIU boundary to ensure each 2.5- or 5-acre buffer did not reflect conditions in adjacent RIUs. Finally, non-forested areas (as delineated in the LDO) were erased from the plot polygons (Figure 4).

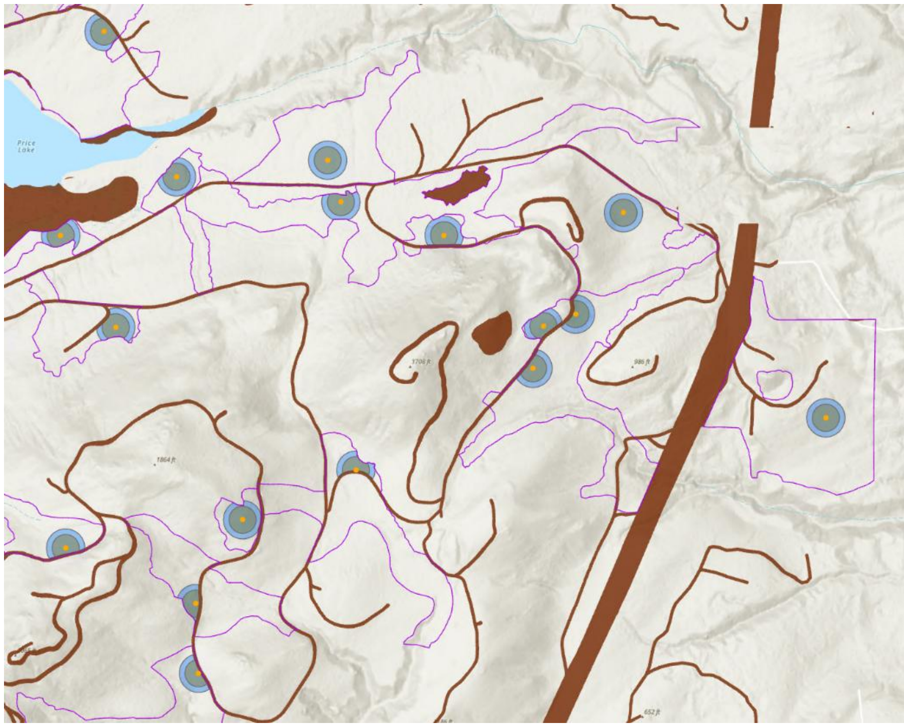


Figure 4. Final 2.5-acre (green) and 5-acre (blue) plot polygons around each SDS field plot (yellow point). Purple outlines are RIU boundaries and brown areas are non- forested areas. Some polygons remained complete circles while other plot polygons were truncated by RIU boundaries or non-forested areas.

2.2.4 Model-ready metrics

For model development, training, and testing, a variety of variables were extracted from RS-FRIS raster (gridded) datasets. The specific variables tested, evaluated, and selected are described in Section 2.3.

For DNR’s Detector models, all relevant variables from RS-FRIS raster pixels (1/10th acre squares) within a plot polygon were extracted to points and then assigned to the plot polygon (Figure 5). Clipped plot polygons with fewer than 12 RS-FRIS pixel points, representing ~half of a 2.5-acre polygon, were excluded from the model training dataset. 12 RS-FRIS pixel points were set as the minimum to ensure the training data had sufficient RS-FRIS coverage to robustly relate stand structure information to the SDS training plot. Fewer than 1% of plot polygons were below the RS-FRIS coverage minimum, resulting in 2217 plot polygons being available for use in training the Detector model. However, for model performance testing against other data sources (see Part 2 below), the full dataset was used that included the smaller polygons (n=2231).

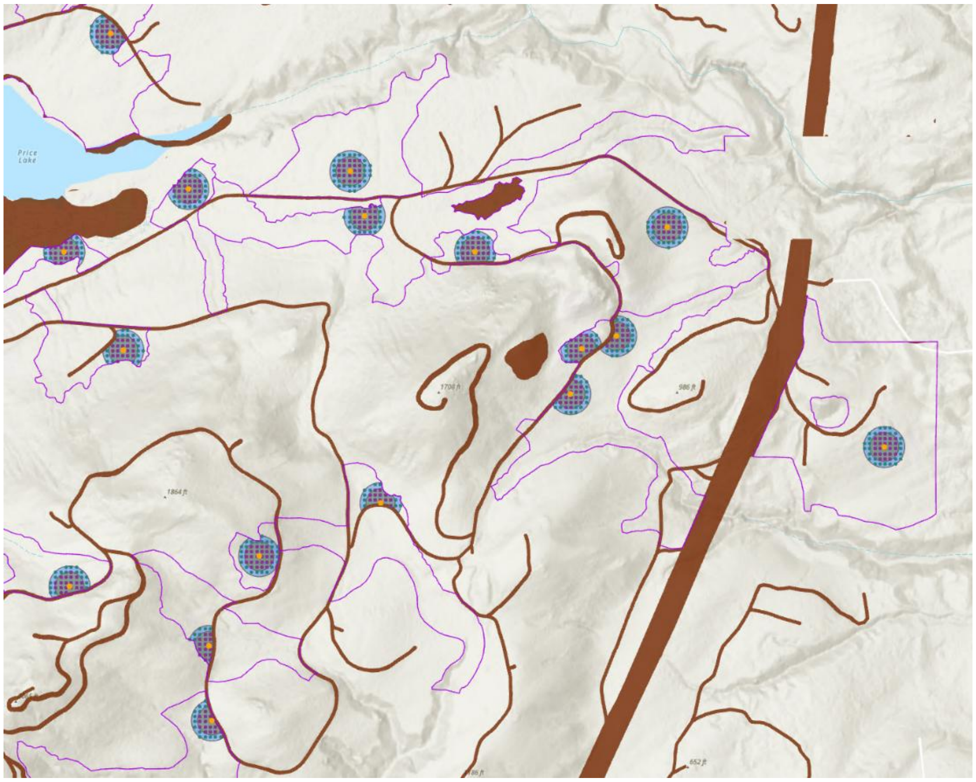


Figure 5. RS-FRIS pixel center locations (purple and green dots) within the two sizes of plot polygons.

2.3 Initial data evaluation for the Detectors analysis

Initial explorations of RS-FRIS metrics and analyses did not discern any difference when using the 2.5- and 5-acre plot polygon scales. The remaining Detector analyses thus used 2.5 acres to ensure consistency with the area assessed and represented by the SDS training plots.

A selection of RS-FRIS attributes, along with a canopy texture metric, were used as variables to investigate class separation in the SCF and OF Detector models (Table 4). Canopy texture (last row of Table 4) was calculated by taking the standard deviation of height of the 40 tallest trees per acre (ht_t40) within each 2.5-acre plot polygon around each SDS training plot. While not as effective as more rigorous canopy texture metrics such as canopy rumple (Kane 2010a), our use of canopy texture was the closest approximation available that would meet our objective of only using readily available RS-FRIS metrics that could also be projected forward in time.

Table 4. RS-FRIS attributes, descriptions, correlation between RS-FRIS and ground inventory plots , and attribute units used in SDS modeling (adapted from Ricklefs (2025)). NA values with no R^2 estimate means there is currently no readily available correlation comparing the remotely-sensed feature to inventory validation plots.

Variable	Description	R^2	Units
ht_t40	Predicted height of the 40 tallest trees per acre	0.94	ft.
ht_lorey	Predicted Lorey height	0.89	ft.
cfvol	Predicted total cubic foot volume per acre	0.88	cu. ft.
ht_max	Predicted maximum tree height	0.87	ft.
baT100	Predicted basal area of the 100 tallest trees per acre	0.86	sq. ft.
ht_t100	Predicted mean height of the 100 tallest trees per acre	0.86	ft.
qmd_t100	Predicted quadratic mean diameter of the 100 tallest trees per acres	0.81	in.
ba	Predicted basal area	0.8	sq. ft.
ba6	Predicted basal area of trees > 6 in. DBH	0.8	sq. ft.
ba4	Predicted basal area of trees > 4 in. DBH	0.79	sq. ft.
tpa30	Predicted number of trees per acre > 30 in. DBH	0.79	trees
tpa31	Predicted number of trees per acre > 31 in. DBH	0.77	trees
qmd6	Predicted quadratic mean diameter for trees > 6 in. DBH	0.7	in.
tpa21	Predicted number of trees per acre > 21 in. DBH	0.64	trees
tpa6	Predicted number of trees per acre > 6 in. DBH	0.62	trees
tpa20	Predicted number of trees per acre > 20 in. DBH	0.61	trees
tpa4	Predicted number of trees per acre > 4 in. DBH	0.59	trees
tpa8	Predicted number of trees per acre > 8 in. DBH	0.56	trees
tpa11	Predicted number of trees per acre > 11 in. DBH	0.53	trees
canopy layers	Predicted count of distinct canopy layers	0.51	layers
qmd	Predicted quadratic mean diameter	0.46	in.
tpa	Predicted number of trees per acre	0.08	trees
cover	Predicted canopy cover (0-100)	NA	%
canopy texture	Standard deviation of ht_t40 within plot polygon	NA	ft.

To estimate relationships among variables, Spearman's correlation coefficient was calculated, which does not require linearity between variables (Dodge 2008; Figure 6). Most of the input variables were highly correlated with height (illustrated by the yellow-white tones in Figure 6). Examining Spearman's correlations was useful because variable inclusion in the SCF and OF Detector models had to meet three criteria: 1) a relatively reduced correlation with other input variables (Table 4, Figure 6), 2) a high correlation with the ground inventory validation plots (Table 4), and 3) the ability to separate SCF and OF classes well. As an example, Figure 6 indicates trees per acre (tpa) is not correlated with other RS-FRIS variables, meeting the first criteria. But Table 4 illustrates tpa is also not as highly correlated with inventory validation plots (R^2=0.08) compared to the other RS-FRIS variables, failing the second criteria. Therefore, tpa is not a strong candidate variable to use in the SCF or OF Detector models.

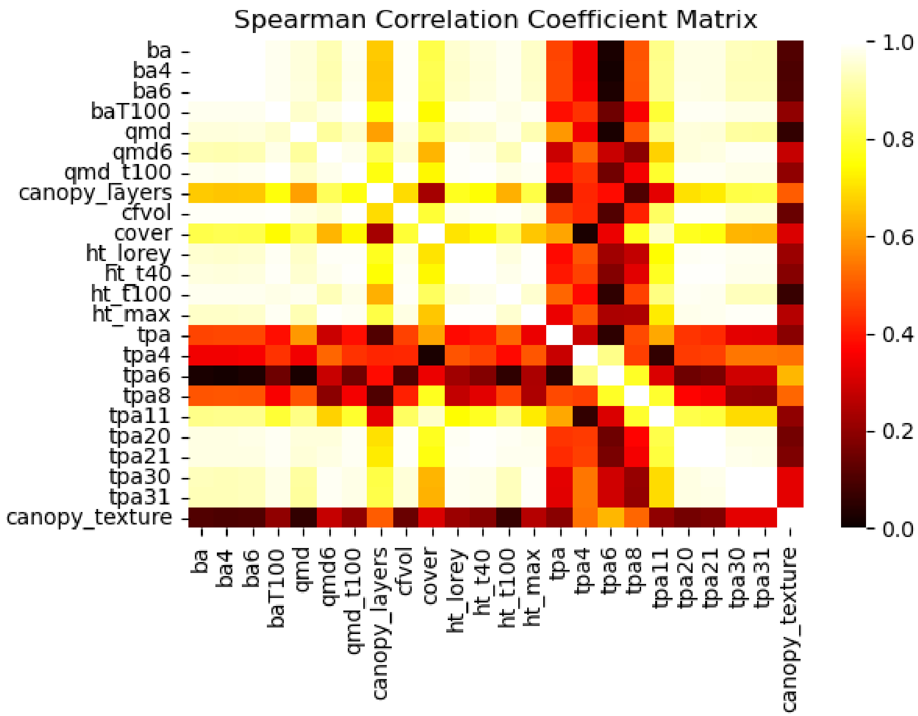


Figure 6. Absolute value of the Spearman correlation coefficients between input variables. Coefficients of 0 and 1 reflect low and high correlations, respectively.

2.4 Developing detector models

2.4.1 Creating training and testing Detector datasets

To develop a Detectors model, two datasets were needed – a dataset to train the model and an independent second dataset to test the accuracy of the model. Each time a model was assessed, multiple training and testing datasets were created through repeated random sampling that required the testing and training datasets to have equivalent number of class samples. It was important that the number of samples in each class be represented equally so the model did not learn to only predict the class with the most data. As an example, the total OF training dataset contained 2217 plot polygons (Not OF = 1998, OF = 219). Each time the training and testing datasets were created, the training dataset contained 218 points (Not OF = 109; OF = 109) and the testing dataset contained 218 points (Not OF = 109; OF = 109). The SCF and OF models were run 500 times, with each model run using a unique random sample of training and testing data. By running the model repeatedly, the variation in model performance was recorded.

2.4.2 Variable selection

Because many RS-FRIS metrics (i.e., variables) are highly correlated with height (Figure 6), not all RS-FRIS metrics were included in the Detector models. Instead, an analysis was done to investigate how well each individual input variable separated SCF and OF classes to identify appropriate variable(s) to include in the final Detector models. To estimate how well each variable separated SCF and OF classes, a single-variable logistic regression was fit for each variable and then assessed against various model performance accuracy metrics (see Section 2.4.3). Each variable was fit 500 times with the randomly sampled testing and training method described above to help assess how well it separated SCF and OF classes and how much mean accuracy varied. Akaike Information Criteria (AICc, Burnham and Anderson 2011) was also calculated on each single-variable logistic regression model as another variable selection line of evidence. AICc values help compare models, identifying which candidate model(s) best fit the data while avoiding model overfitting.

Models with more than one variable were also developed and tested using the most important variables from the single-variable logistic regression, selecting additional variables with relatively moderate to lower correlations with height ($R < 0.85$) that were also consistent with other attributes found to be useful in distinguishing more structurally complex SDS conditions in the peer-reviewed literature (Kane 2010a; Figure 1).

2.4.3 Metrics to evaluate model performance and accuracy

Mean accuracy, balanced accuracy, true positive rate, true negative rate, f-score, and j-statistic were computed to evaluate model performance. These metrics can be calculated using the true positive (tp), true negative (tn), false positive (fp), and false negative (fn) values from a confusion matrix, a table that compares model predictions with actual data (see Figure 7 for example and definition of tp, tn, fp, and fn).

		Predicted	
		p	**n**
Actual	**p’**	True positive	False negative
	n’	False positive	True negative

Figure 7. Confusion matrix example. True positive is a correct identification of the positive condition (is SCF/OF, predicts SCF/OF); false negative predicts the negative condition, but is actually the positive condition (is SCF/OF, predicts Not SCF/OF); false positive predicts the positive condition, but is actually the negative condition (is Not SCF/OF , predicts SCF/OF); and true negative is the correct identification of the negative condition (is Not SCF/OF, predicts Not SCF/OF).

All measures were selected to examine different aspects of model accuracy. Balanced accuracy is a good metric to use for datasets where the class samples are unbalanced. For example, if class 1 had 90 samples and class 2 had 10, then a model could predict all the samples were in class 1 and obtain a mean accuracy of 90% while the balanced accuracy would be 50%. True positive and negative rates respectively evaluate how well a model detects actual positive (e.g., SCF/OF) and negative (e.g. Not SCF/Not OF) values. F-scores evaluate model accuracy while accounting for false positive and negative values. The f-score can be useful when the positive condition is rare. J-statistics evaluate model accuracy using the true positive and negative rates and is useful when these rates are of equal importance. See equations 1-6 below for the various model accuracy calculations.

$$\textit{mean accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \tag{1}$$

$$\textit{balanced accuracy} = \frac{tpr + tnr}{2} \tag{2}$$

$$\textit{true positive rate (tpr)} = \frac{tp}{tp + fn} \tag{3}$$

$$\textit{true negative rate (tnr)} = \frac{tn}{tn + fp} \tag{4}$$

$$f\text{-}score = \frac{2tp}{2tp + fp + fn} \quad (5)$$

$$j\text{-}statistic = tpr + tnr - 1 \quad (6)$$

2.5 Calculating class probabilities from model parameters and identifying probability thresholds

The coefficients and intercepts (model parameters) fit from the logistic regression were used to estimate the probability that a plot was in one of two classes. Model parameters estimated by logistic regression take the form

$$y = \beta_0 x_0 + \beta_1 x_1 + \cdots + b \quad (7)$$

where β_i are the coefficients and x_i are the variables, and *b* is the intercept. Eq. 7 was then used in a sigmoid function (Eq. 8) to calculate class probabilities. An example for estimating the probability using a single variable is provided below.

$$prob(class) = \frac{e^{b + \beta_0 * x_0}}{1 + e^{b + \beta_0 * x_0}} \quad (8)$$

Because the resulting models assigned a probability that a stand was SCF/OF ranging from zero to one, a method was needed to identify the appropriate probability threshold to estimate the amount of SCF/OF classified acres on DNR-managed forestlands. While such a threshold could have been selected subjectively, which could be perceived as DNR biasing the amount of area identified as SCF/OF, probability thresholds were selected by optimizing for the 10 largest j-statistics (equation 6) and then minimizing false positives (as directed by internal project leadership).

3. Results

Two investigations were conducted using the methods described above. The investigations aimed to estimate the variables that best identified SCF and OF, and how well those variables performed defining each class.

Details of the modeling results are described below in Sections 3.1-3.3., but model accuracy (in many ways the most relevant metric) is summarized here. Overall SCF and OF accuracy were respectively 81 ± 1% and 72 ± 3%, with overlap between variables selected for the SCF and OF models (Table 5).

Table 5. Selected SCF and OF variables in the 2.5-acre plot polygon models. R^2 is the correlation between the remotely-sensed variable and inventory validation plot data (see Table 4). R^2 with NA means there is currently no correlation comparing the remotely-sensed variable to inventory validation plots. To help illustrate the relationships amongst variables, the Spearman correlation coefficients listed in the table are in relation to qmd6. The SCF and OF class separation columns represent the mean accuracy in distinguishing between SCF/OF using single variable logistic regressions. Canopy layers and qmd6 variables were selected in both SCF and OF Detector models.

Variables	R^2	Spearman Correlation to qmd6	Class Separation SCF	Class Separation OF
canopy layers	0.51	0.83	80 ± 1%	74 ± 3%
qmd6	0.70	1.00	81 ± 1%	66 ± 3%
canopy texture	NA	0.28	NA	62 ± 3%

3.1 Structurally complex forest (SCF)

Results from single variable logistic regression models indicated qmd6 and canopy layers had the highest mean accuracies (>0.8, Figure 8), suggesting these two attributes might be the strongest single predictors of SCF, even with a moderate correlation between the two variables (Figure 6, Table 5) relative to the suite of RS-FRIS metrics examined (Table 4).

Despite its lower support in a single-variable AIC_c model selection, canopy layers (Table 6) was also included in the final SCF model due to its ecological relevance and moderate correlation with qmd6 compared to other RS-FRIS variables. Canopy layering is a key characteristic of mature forest conditions in western Washington (see Figure 1 illustration), and when combined with qmd6, may help identify additional SCF stands beyond those captured by tree diameter alone. Further, previous work (Franklin et al. 2002, Franklin and Van Pelt 2004, Van Pelt 2007, Kane et al. 2010a) have highlighted its prominence as an important feature of many developed forests in western Washington and Oregon.

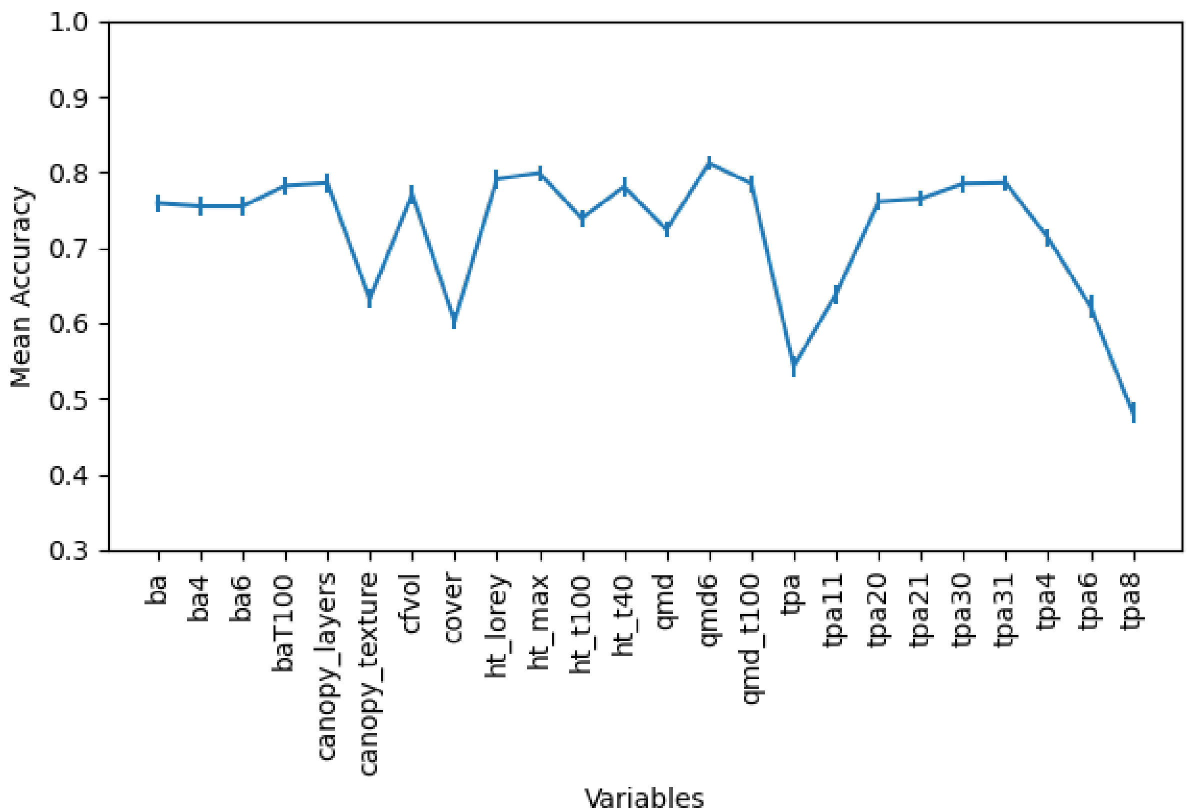


Figure 8. Mean accuracy of single variable logistic regression of SCF. Variables are shown on the x-axis and mean accuracy on the y-axis. Some high performing variables such as canopy layers and quadratic mean diameter of trees 6" DBH and greater (qmd6) have mean accuracies around 0.80. The small standard deviation, shown as error bars, illustrates minimal variation in model performance over the 500 iterations. See Table 4 for variable definitions and units. Note mean and balanced accuracy are similar to each other because models were trained on balanced samples.

Table 6. AIC_c results ranking all single variable SCF logistic regression models from strongest to weakest. Delta AIC_c values describe the relative strength of each model. AIC_c values > 10 suggest low support of single variable models relative to the top model (qmd6). Because most variables were derivations of height (Figure 6), correlations with height were a relative categorical assignment, where Spearman's correlation (R) with absolute values ≥ 0.85 was considered strong correlation, values between 0.7 and 0.84 moderate correlation, and values <0.7 low correlation.

Variable	AIC_c	Delta AIC_c	Correlation with height
qmd6	733.7	0	Strong
ht_max	749.14	15.44	Strong
ht_lorey	779.7	45.99	Strong
canopy layers	783.02	49.31	Moderate
qmd_t100	791.75	58.05	Strong
baT100	799.22	65.52	Strong
ht_t40	804.83	71.13	Strong
tpa31	815.26	81.56	Strong
tpa30	819.84	86.13	Strong

cfvol	832.58	98.88	Strong
ba	863.96	130.26	Strong
tpa21	868.23	134.53	Strong
ba6	871.64	137.93	Strong
tpa20	871.68	137.98	Strong
ba4	871.83	138.13	Strong
ht_t100	887.57	153.87	Strong
qmd	906.6	172.9	Strong
tpa11	1076.97	343.26	Low
cover	1085.95	352.25	Low
tpa4	1122.8	389.1	Low
canopy texture	1156.11	422.41	Low
tpa8	1192.26	458.56	Low
tpa	1193.62	459.91	Low
tpa6	1196.43	462.73	Low

Using both qmd6 and canopy layers in a two-variable model resulted in a model with a mean accuracy of 81% (Figure 9).

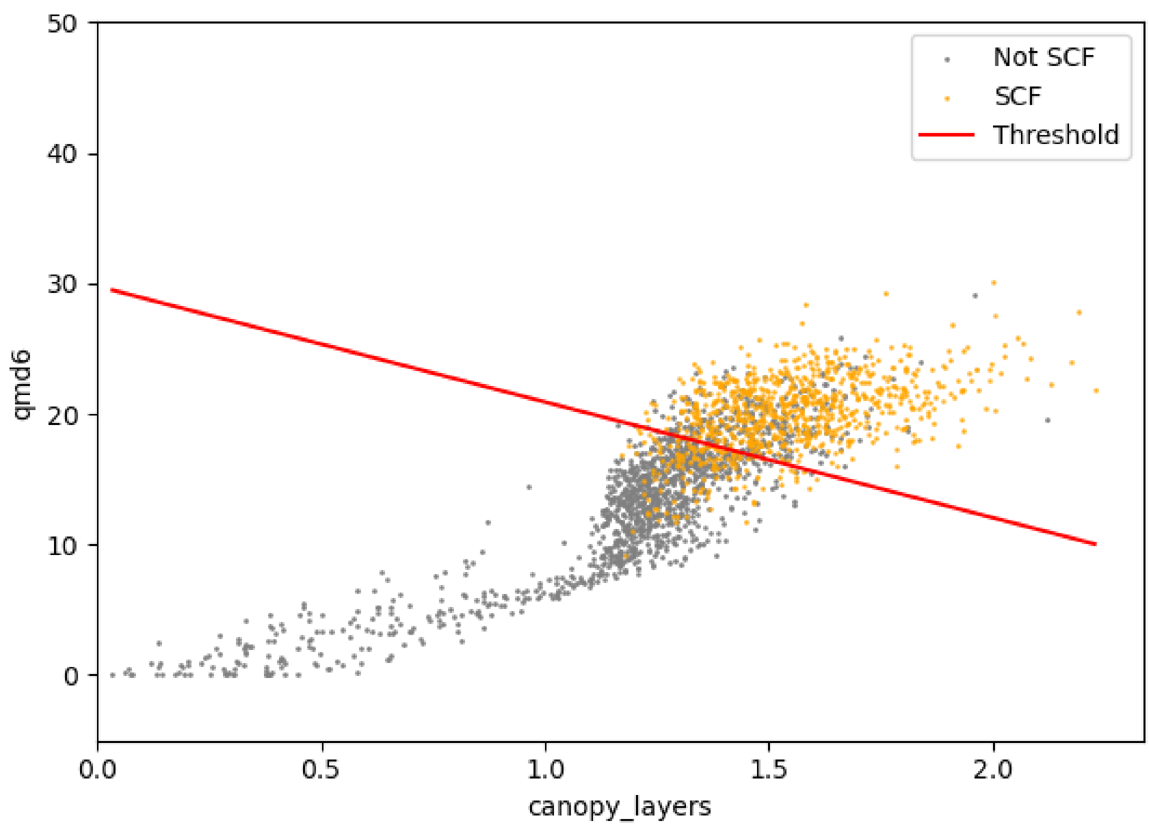


Figure 9. Two-variable logistic regression using canopy layers (x-axis) and qmd6 (y-axis), both of which had the highest mean accuracies when examined individually (Figure 8). The threshold (red line) separating SCF (orange) and Not SCF (grey) had a mean accuracy of 81 ± 1%.

3.2 Older forest (OF)

Single variable OF logistic regression models suggested canopy layers best classified OF stands (Figure 10), although with a lower mean accuracy (~73%) than top performing SCF variables (Figure 8). AIC_c results also indicated canopy layers was the single best OF variable (Table 7). The next suite of variables (qmd6, ht_max, qmd_t100, and baT100) were all similarly ranked and strongly correlated with height. The first variable with a low height correlation was canopy texture (Table 7), whose AIC_c value suggested low support although its mean accuracy was moderate (Figure 10).

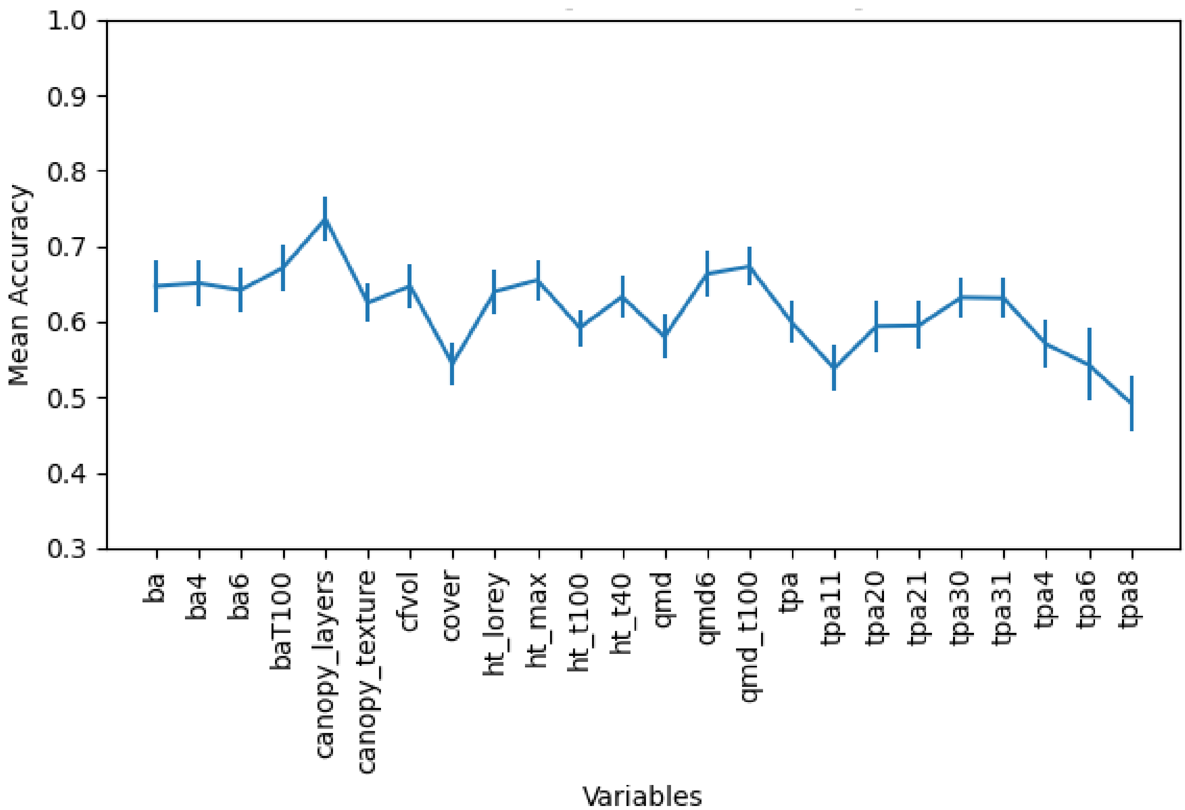


Figure 10. Single variable logistic regression for older forest (OF) classification. Variables are shown on the x-axis and mean accuracy on the y-axis. Canopy layers had the highest mean accuracy around 73% while other relatively high performing variables had mean accuracies around 65%. The standard deviations, shown as error bars, for each variable were larger than in Figure 8, indicating larger variation in model performance over the 500 iterations. Note mean and balanced accuracy are similar to each other because models were trained on balanced samples.

Table 7. AIC_c results ranking all single variable OF logistic regression models from strongest to weakest. Delta AIC_c values describe the relative strength of each model. AIC_c values > 10 suggest low support of single variable models relative to the top model (canopy layers). Because all variables were derived from height, correlations with height were a relative categorical assignment where Spearman's correlation (Table) of $R^2 \geq 0.85$ was considered strong correlation, values between 0.7 and 0.84 moderate correlation, and values <0.7 low correlation.

Variable	AIC_c	Delta AIC_c	Correlation with height
canopy_layers	246.33	0	Moderate
qmd6	268.15	21.81	Strong
ht_max	269.28	22.95	Strong
qmd_t100	269.36	23.03	Strong
baT100	269.89	23.56	Strong
ba4	276.76	30.42	Strong
ht_lorey	277.72	31.38	Strong
ba	277.85	31.52	Strong
ba6	278.91	32.57	Strong
cfvol	279.88	33.54	Strong
ht_t40	280.07	33.73	Strong
tpa31	287.73	41.39	Strong
tpa30	288.09	41.75	Strong
ht_t100	289.79	43.46	Strong
canopy_texture	290.36	44.02	Low
qmd	292.59	46.26	Strong
tpa21	294.58	48.24	Strong
cover	294.89	48.56	Low
tpa20	295.28	48.95	Strong
tpa	300.57	54.23	Low
tpa11	302.1	55.77	Low
tpa4	303.61	57.27	Low
tpa6	307.28	60.95	Low
tpa8	307.31	60.98	Low

Combining the top single variable (canopy layers) with the remaining top variables with varying correlations with height (qmd6 and canopy texture) resulted in an OF model with an overall accuracy of ~72%, which was lower than SCF model. Ecologically, the inclusion of a height-related variable with canopy layers was important to help distinguish later from earlier SDSs. Although canopy texture ranked low from an AICc perspective (Table 7), it was ultimately included in the model because rumple (i.e., canopy complexity) is a key characteristic of older forests (Van Pelt and Franklin 2004, Van Pelt 2007, Kane et al. 2010a, Griffey et al. 2021); initial observations also suggested it may capture some stands with a high standard deviation of height despite being shorter in stature that might otherwise be excluded from OF consideration by the model. The tilted threshold plane in Figure 11 also indicated both qmd6 and canopy texture helped capture some aspects of OF conditions missed by canopy layers alone.

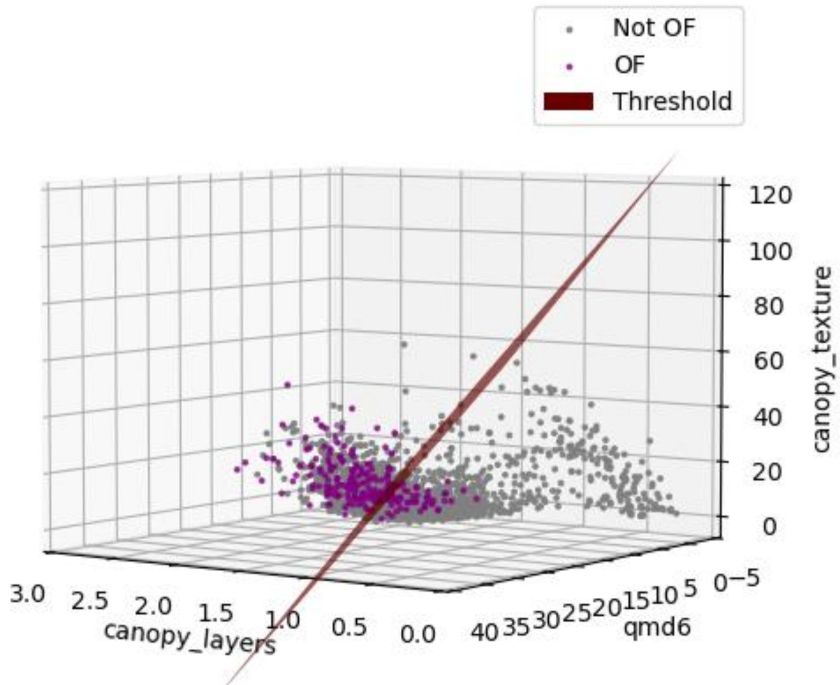


Figure 11. Three variable logistic regression fit using variables canopy layers (x-axis), canopy texture (y-axis), and qmd6 (z-axis). The threshold (brown plane) separating OF (purple) and Not OF (grey) had a mean accuracy of 72 ± 3%.

3.3 Applying the models to western WA

Both the two and three variable SCF and OF Detector models were applied to RS-FRIS rasters (version 5.2) across all ~1.5 million acres of DNR-managed forestlands in western WA. This technique implemented a sliding window across the landscape that took the average or standard deviation of all the forested data within a 2.5-acre circle around the center pixel. The outcome was a score for the center pixel based on the characteristics of that pixel and its surrounding stand neighborhood (within the same RIU, excluding roads or other non-forested areas). The model coefficients (Table 8) were then applied to the results of the sliding window. Once the model coefficients were applied using Eq. 8, the outputs were continuous probability maps indicating the probability a pixel was likely SCF or OF (see Figure 12).

Table 8. Model coefficients applied to RS-FRIS layers to develop SCF and OF probability maps.

	canopy layers (β_0)	**qmd6 (β_1)**	**canopy texture (β_2)**	**intercept (b)**
SCF	3.217 ± 0.340	0.362 ± 0.027	NA	-10.790 ± 0.512
OF	2.255 ± 0.379	0.098 ± 0.036	0.039 ± 0.020	-5.500 ± 0.610

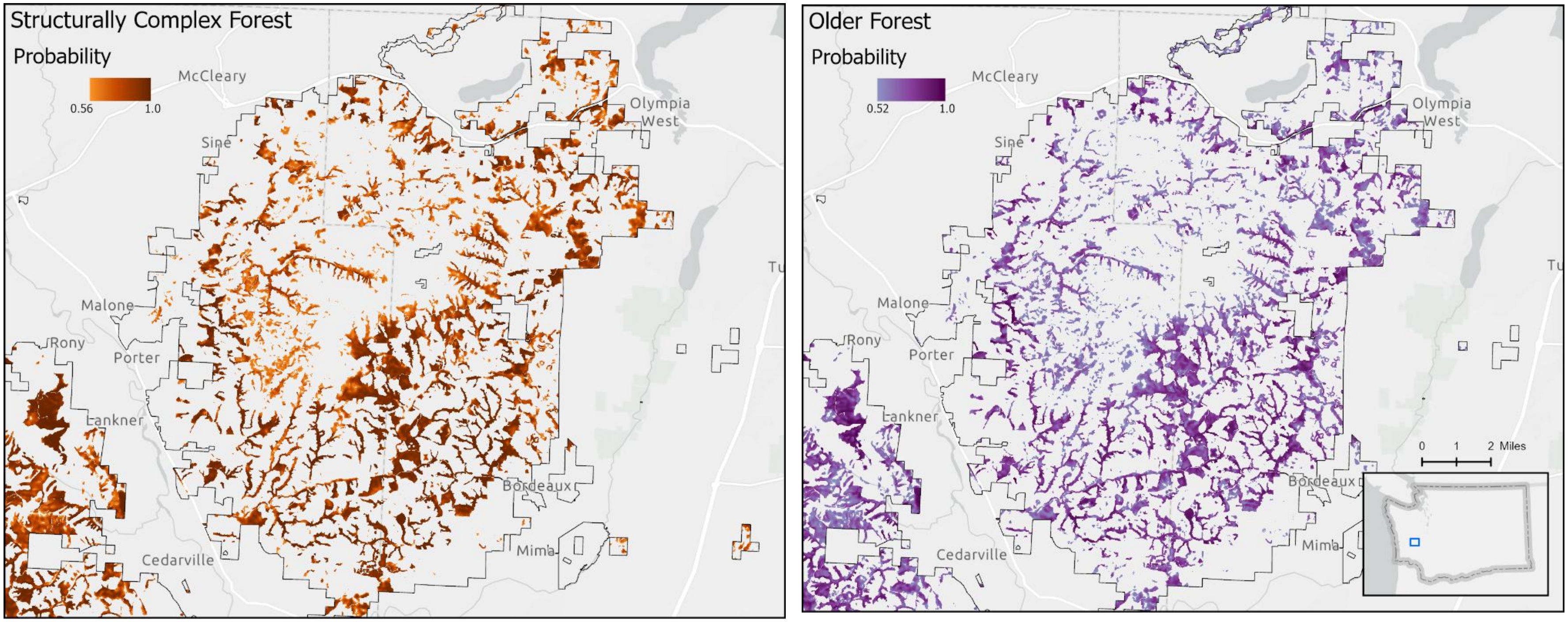


Figure 12. An example of SCF (left) and OF (right) probabilities for Capitol State Forest. Only probabilities above the optimized probability threshold of 0.56 (for SCF) and 0.52 (for OF) are displayed (see Table 9). Darker colors reflect a higher probability of being SCF or OF.

Based on optimizing the j-statistic and minimizing false positives, a probability threshold of 0.56 and 0.52 was used to delineate SCF and OF, respectively (Figure 12). Use of 0.56 and 0.52 thresholds resulted in minimal discrimination between the SCF and OF models (Table 9), resulting in high correlations between the two maps (Figure 13). A high overlap in variables with near-identical thresholds (Table 9) and high map correlations resulted in similar SCF and OF estimates (Table 10).

Table 9. Variable threshold used to delineate SCF and OF using the optimized probability threshold of 0.56 for SCF and 0.52 for OF. NA indicates a variable was not used in the SCF model. In both models, near-identical qmd6 and canopy layer thresholds were found by the j-statistic.

	qmd6	canopy layers	canopy texture
SCF	17.4	1.47	NA
OF	17.2	1.47	15

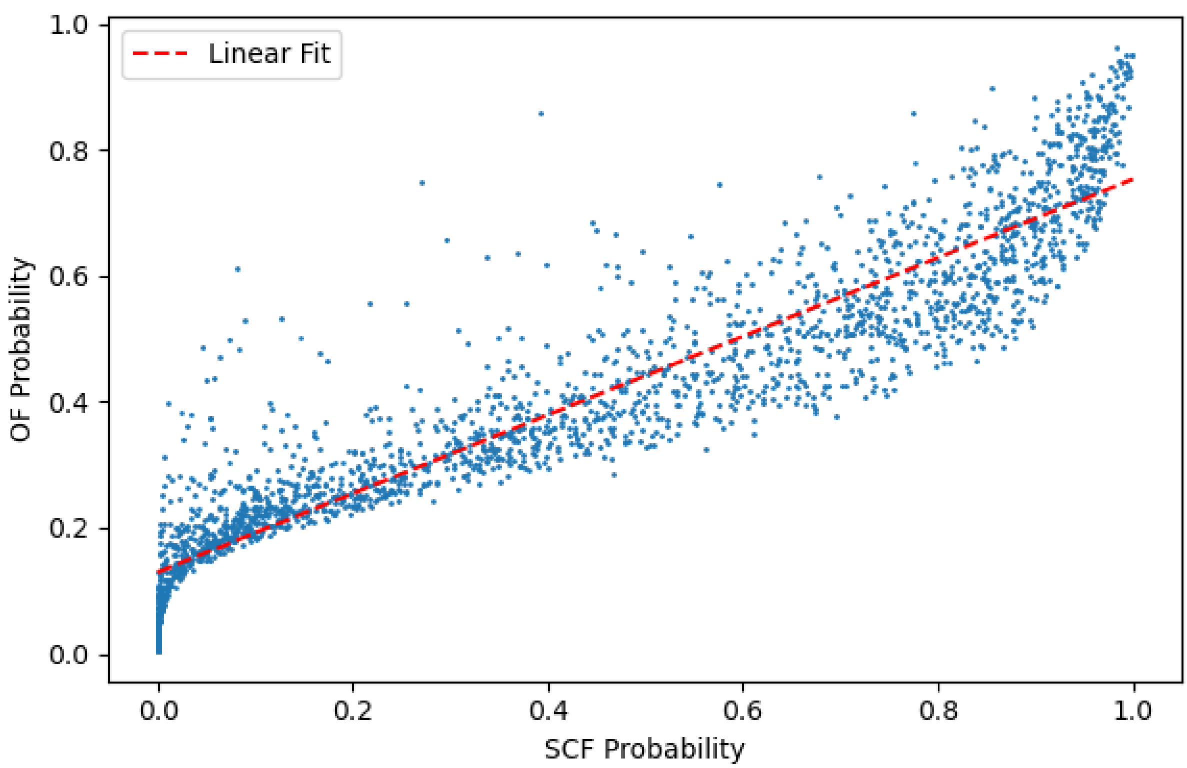


Figure 13. Correlation between SCF and OF maps based on 2237 training plots.

Table 10. Estimates of total SCF and OF as area and percentage of DNR-managed forestlands in western WA. Probabilities greater than or equal to 0.56 and 0.52 were classified as SCF and OF, respectively, in the current analysis, based on a quantitative analysis of the j-statistic. Estimated SCF and OF area would have differed (with varying levels of confidence) if other probabilities had been subjectively selected. The acres and percentages calculated in this table are draft and subject to change. For the most up-to-date data, consult the future Landscape Assessment Analysis.

Model	Threshold	Acres	%
SCF	0.56	345,710*	23
OF	0.52	349,014*	23

*While there is high spatial overlap and convergence between SCF and OF, not all OF acres are also SCF acres, and vice versa. Therefore, none of the values in this table can be used to derive Maturation 2 acres as a specific stand-alone class.

4. Discussion

The logistic regression approach chosen for data exploration in the Detectors project revealed both a promising method for identifying certain forest conditions and an inherent limitation. Detector results indicated SCF could be identified with reasonable accuracy (81 ± 1%). However, for the OF model, the 1) lower accuracy (72 ± 3%), 2) the similarity of variables and thresholds to SCF (Tables 4 and 9), 3) that the vast majority of SCF acres were also classified as OF at the optimized thresholds (Table 10), and 4) the high correlation with SCF (R^2 = 0.87; Figure 13) collectively reduced confidence in the current OF methodology and resulting map.

Additionally, initial examination, field checking of map outputs, and local knowledge of the land base suggests the SCF map validates reasonably well for stands that are at least Maturation II, whereas the OF model is likely overpredicting the older forest condition. Related, Vertical Diversification, Horizonal Diversification, and Pioneer Cohort Loss SDSs (Table 1; Figure 1) compose ~10% of all the SDS plots, despite efforts to target these stages to increase its representation in training the models (field staff reported difficulty in locating much more of those conditions), suggesting OF conditions are less prevalent than predicted by the OF model.

Fundamentally, these models based on RS-FRIS data are height-driven, and as such, appeared to work well in detecting late mature stage stands (i.e. Maturation II) and/or older forests that are tall (refer to Figure 1). Inspection of map outputs suggested the models and maps could miss SCF and OF that were not tall (higher false negative rates), such as those on low-productivity sites and/or have experienced moderate canopy disturbance (e.g., wind damage, breakage). This limitation applied to both models; however, the SCF model evidently had better fidelity in dividing Maturation I from Maturation II at the optimized threshold (Figures 9 and 11), while also being able to capture tall OF at higher probability levels, similar to what the OF model could do.

The SCF model likely performed better because the transition between Maturation II (the first stage of SCF) and earlier stages follows a gradient of stand height growth (Figure 1), allowing a height-driven model to more easily differentiate stands on either side of the divide. In contrast, the separation between Vertical Diversification (the first stage of OF) and earlier stages is less distinct in terms of height, as stands have often already reached their maximum by that point (Figure 1), making it more challenging for a height-driven model to pinpoint the division.

Another possible reason for lower OF accuracy was the canopy texture metric used. Canopy texture is a known critical feature to distinguish old forests from young (Kane et al. 2010a). Standard deviation of height was the best proxy we had for this approach using the RS-FRIS data; however it may not have sufficient resolution for the task of distinguishing older forest canopy structure. The variable was kept in the OF model because it was desired that the model have the ability to identify shorter OF that may not meet the height thresholds (see above), but it may not be identifying the condition consistently enough to capture all or most such stands.

The Detectors modeling effort was a success in collecting a large amount of training data representative of the land base in a short timeframe (Table 2, Figures A1-A3). Indeed, acquiring nearly 2,400 2.5-acre training plots is rare in most ecological studies. And yet the area represented by the training plots is less than 0.5% of DNR-managed forestlands in western Washington. Further, only ~10% of the ~2200 training plots were categorized as OF, in part because surveyors were unable to locate substantially more OF stands to sample, despite seeking as many as possible. With the approach used in the Detectors project, it is possible more plots could yield a different outcome, an assessment that will be possible to investigate in future model iterations as training plots continue to be collected in the years to come.

While more training plots will be of use, the Detectors project provides opportunities for future model exploration with the current dataset. For example, logistic regressions could be rerun with the inclusion of other environmental variables such site index or potential vegetation types (PVT) (Figure A3), and variable interactions. Past silvicultural treatments (e.g., thinnings) that accelerate stand development could also be factored in. Other statistical approaches that can be projected forward, such as boosted regression trees, could be an alternative modeling approach to better handle data nonlinearities than logistic regression. Alternatively, if model requirements are relaxed such that the best current map is developed even if it cannot be projected in a growth and yield model, other non-parametric statistical methods could be explored. For example, removing the projection requirement could result in the inclusion of other variables, distributions, and characteristic curves derived directly from the canopy height model (CHM), an interim RS-FRIS product. It's possible that examining the range or distribution of RS-FRIS values within a plot polygon, rather than the mean RS-FRIS value as done in the Detectors project, could result in different models for both SCF and OF. It was also observed that old forests in the OESF and Straits HCP units tended to be shorter than those on other DNR-managed forestlands, suggesting region-specific models could improve overall model accuracy with a sufficient sample size and the inclusion of other variables. In short, there are many opportunities to potentially improve the Detectors model given sufficient interest and time.

Part 2: Model Comparisons

5. Background

Earlier efforts by other groups have also assessed mature and older forests conditions in western Washington. However, these earlier efforts had different objectives than the Detectors project. Resilient Forestry (RF) was contracted by DNR in 2023 to develop models to separate SDSs into 5 and 8 classes. This work involved reviewing literature, finding and creating variables that separated SDS, delivering a report (RF 2023), and R scripts that processed, fit, and applied models. RF's models used RS-FRIS variables, as well as variables extracted from the CHM. Although the CHM offered more detailed canopy texture information than was used in constructing the Detector models, it could not be incorporated into the Detectors' work because it violated the project constraint for inclusion of only metrics that could be projected forward in time. Another key difference from the current investigation is that, at the time of the RF contract, model development was limited to pre-existing field-verified old-growth polygons—also utilized in the Detectors project—and Weighted Old Growth Habitat Index (WOGHI) scores derived from DNR's earlier plot-based inventory, which the Detectors project excluded. As a result, the available data for RF provided only a partial view of the range of SDSs present on DNR-managed forestlands. Lastly, because the RF contract required the creation of 5 and 8 classes, their model was never tested against the Detectors binary SCF and OF classifications. Nonetheless, RF's work served as proof-of-concept and a first step in using remotely-sensed data to detect forest structure variability on DNR-managed forestlands. The work from DNR's contract with RF inspired DNR to begin developing the Detector models.

In contrast to the RF work, gradient nearest neighbor (GNN) maps display imputed estimates of forest structure and composition derived from the relationship between forest inventory plots, topo-climatic variables, and Landsat satellite pixel reflectance values (Ohmann and Gregory 2002). Developed ~two decades ago by the Landscape Ecology Modeling, Mapping, and Analysis Team (LEMMA 2025), GNN has a long and rich history of use in the Pacific Northwest. GNN has been extensively used to examine forested landscapes (e.g., Ohmann et al. 2007, 2012), and is the central dataset used to specifically examine forest structure trends across the Northwest Forest Plan area (Davis et al. 2015, 2022). While GNN maps are generated at 30-m cell resolution, the developers of GNN recommend the data be summarized at broader scales, such as watersheds, to reduce model inaccuracy. Thus, GNN data quantifies the *amount* of a forest condition, but does not have the spatial resolution to map forest structure at the stand scale, unlike the Detectors project and RF contract.

Given the various models and maps of forest structure, the Part 2 objective of this report was to compare available models that could be reclassified to represent SCF and OF conditions. Specifically, DNR's Detector models were compared to models created by RF, and to GNN data.

6. Methods

6.1 RF input data

SDS training data collected by DNR staff (detailed in Section 2.1) was then processed in GIS (detailed in Section 2.2) and resulted in a plot polygon around each SDS training point within which RS- FRIS pixel variables were summarized (Table 11). To prepare data for the RF model, the mean for each RS-FRIS variable was summarized by RIU polygon (the original scale of the RF model) and by 2.5-acre and 5-acre plot polygon. RF's model also used variables extracted via the CHM (clipped at 2.5-acre, 5-acre, and RIU scale) and processed using code created and delivered by RF.

Table 11. RS-FRIS and Resilient Forest (RF) variables used in stand development stage (SDS) modeling. Variable origin indicates if the variable was derived by RS-FRIS or RF. R^2 is the correlation between RS-FRIS variables and DNR inventory plots. NA values indicate variables developed specifically for the RF model that cannot be correlated with DNR's inventory plots. See RF (2023) for RF variable origin descriptions.

Variable name	Variable origin	R^2	Units
Mean height of the 40 largest trees per acre	RS-FRIS	0.94	ft.
Gross board foot volume	RS-FRIS	0.91	bf.
Lorey's height	RS-FRIS	0.89	ft.
Aboveground biomass	RS-FRIS	0.89	metric tons/acre
Number of trees > 6 in. DBH	RS-FRIS	0.62	trees/acre
Number of trees > 20 in. DBH	RS-FRIS	0.61	trees/acre
Number of conifer trees per acre > 4 in. DBH	RS-FRIS	0.59	trees/acre
Count of distinct canopy layers	RS-FRIS	0.51	layers
Maximum gap patch area	RF	NA	
Mean gap patch area	RF	NA	
Percent of area in gap patches	RF	NA	
Maximum tall tree patch area	RF	NA	
Mean tall tree patch area	RF	NA	
Percent of area in tall tree patches	RF	NA	
Vector ruggedness measure	RF	NA	
Percent of area in rugged patches	RF	NA	
Mean VRM x CHM patch area	RF	NA	
Shannon's diversity index of VRM x CHM patches	RF	NA	

Relative mutual information of VRM x CHM patches	RF	NA

Because the RF models were originally fit with only pre-existing field-verified old–growth polygons and Weighted Old Growth Habitat Index (WOGHI) scores from DNR's previous plot-based inventory, RF recognized the training data did not represent the full spectrum of SDSs present on DNR-managed forestlands. The report by RF recommended validating model performance with field verified SDS assessments. That data now being available, for this work, RF's logistic regression models were refit using the new SDS dataset collected by DNR. Results of the original fits created by RF can be found in the Appendix B.

6.2 GNN input data

One attribute included in the GNN dataset is a synthetic metric called the Old-Growth Structural Index (OGSI). OGSI assigns a score to a pixel based on a combination of up to four of the following variables: 1) the density of large live trees, 2) the diversity of live tree diameters, 3) large snag density, and 4) the percent of ground cover of coarse woody debris (Davis et al. 2015, 2022). Additional work by the same research group has applied thresholds to the OGSI data to approximate different stages of forest development (Davis et al. 2015, 2022), which have been used to periodically assess forest trends across the Northwest Forest Plan (Davis 2022).

Earlier iterations of the dataset included two OGSI thresholds: OGSI80 and OGSI200. OGSI80 tends to correspond with the Van Pelt key for Maturation I (Van Pelt 2007) while OGSI200 is more akin to the Vertical Diversification stage, recognizing such structural conditions may be present outside of the OGSI200 designation. Newer iterations of the GNN data, such as the dataset used in the present analysis, included two additional thresholds to represent conditions older than Maturation I and generally younger than Vertical Diversification: OGSI120 and OGSI160 (i.e., the beginning stages of SCF).

6.3. Methods for model comparison

6.3.1 The RF model

Like the Detectors project, RF models were fit using all available SDS training plots (n = 2231) at the 2.5- and 5-acre scales. For the RF model developed at the RIU scale, only RIUs with a single SDS plot were used in the training and testing of the model. Also like the Detectors project, model predictions were compared with the SDS field training data - a confusion matrix was used to calculate the true negatives (tn), true positives (tp), false negatives (fn), and false positives. From here the true positive rate (tpr), true negative (tnr), f-score, mean accuracy (accuracy), and balanced accuracy were calculated (see derivation details in Section 2.4.3 in Part 1).

6.3.2 The GNN model

Because GNN data does not have the same spatial specificity as the Detector models, total GNN area that was SCF (OGSI120 + OGSI160 + OGSI200) and OF (OGSI200) were compared against Detector project estimates of those same conditions.

7. Results

7.1 Detectors and RF comparison

From a balanced accuracy perspective, the SCF Detector model (81 ± 1%) performed similarly to all equivalent RF models (82%; Table 12). Using the same accuracy metric, the Detector OF model (71% balanced accuracy) outperformed all similar RF models (50-64% balanced accuracy), regardless of RF scale (2.5 acres, 5 acres, RIU). Broadly, RF OF models predicted more of the negative condition (Not OF), which led to a higher mean accuracy, but a lower balanced accuracy (Table 12). As an example, RF 5 Stage RIU Scale OF model (last row of Table 12) did not predict any positive conditions (OF). This led to a high mean accuracy (90%), but a balanced accuracy of 50%. If the model was applied to the landscape, it would underpredict OF area.

Table 12. DNR and RF SCF and OF model performance. Balanced accuracy provides one of the more useful summaries of model performance. Abbreviations: tn = true negative; fp = false positive; fn = false negative; tp = true positive; tpr = true positive rate; tnr= true negative rate; accuracy = mean accuracy (further described in section 2.4.3). Total number of RF training plots at 2.5- and 5-acre scales: 2,231 (SCF = 869; Not SCF = 1,362; OF = 220; Not OF = 2,011). Total number of RF training plots at RIU scale (corresponding to total number of RIUs): 1,452 (SCF = 520; Not SCF = 932; OF = 143; Not OF = 1,309).

Model	tn	fp	fn	tp	tpr	tnr	f-score	accuracy	balanced accuracy
DNR SCF	1083	279	147	722	0.83	0.8	0.77	0.81	0.81
DNR OF	1367	644	57	163	0.74	0.68	0.32	0.69	0.71
RF 8 Stage 5-acre Scale SCF	1136	226	166	703	0.81	0.83	0.78	0.82	0.82
RF 8 Stage 5-acre Scale OF	1986	25	154	66	0.3	0.99	0.42	0.92	0.64
RF 8 Stage 2.5-acre Scale SCF	1136	226	166	703	0.81	0.83	0.78	0.82	0.82
RF 8 Stage 2.5-acre Scale OF	1986	25	154	66	0.3	0.99	0.42	0.92	0.64
RF 5 Stage 5-acre Scale OF	2010	1	216	4	0.02	1	0.04	0.9	0.51
RF 5 Stage 2.5-acre Scale OF	2010	1	216	4	0.02	1	0.04	0.9	0.51
RF 8 Stage RIU Scale SCF	753	179	92	428	0.82	0.81	0.76	0.81	0.82
RF 8 Stage RIU Scale OF	1307	2	120	23	0.16	1	0.27	0.92	0.58
RF 5 Stage RIU Scale OF	1309	0	143	0	0	1	0	0.9	0.5

7.2 Detectors and GNN comparison

Across all DNR-managed forestlands, GNN data estimated ~424,000 acres, or 28%, of the land base was SCF, with roughly half of that area classified as OF (Table 14). Detector estimates suggested a roughly similar amount of the land base (345,710 acres) was SCF. However, nearly all detector SCF was also classified as OF (Table 10, Table 14).

Table 13. Comparison of area classified as SCF and OF in the Detector models and GNN dataset. Consistent with Table 1 and Figure 1, SCF acreage includes all SDSs from Maturation II (M2) through Pioneer Cohort Loss (PL). OF acreage includes all SDSs from Vertical Diversification (VD) through PL. While there is high spatial overlap and convergence between SCF and OF, not all OF acres are also SCF acres, and vice versa. Therefore, none of the Detector values in this table can be used to derive Maturation 2 acres as a specific stand-alone class.

	SCF	OF
Detectors*	345,710 (23%)	349,014 (23%)
GNN	423,830 (28%)	214,697 (14%)

*While there is high spatial overlap and convergence between SCF and OF, not all OF acres are also SCF acres, and vice versa. Therefore, none of the Detector values in this table can be used to derive Maturation 2 acres as a specific stand-alone class.

8. Discussion

Recognizing a different question at a different spatial scale (RIU) was posed when developing the RF models, model comparisons suggest the simpler Detector models performed equally well, if not better. The DNR SCF and OF models balanced the true positive and negative rates (see section 2.4.3 for definitions), an intentional design of the Detector models. A positive outcome of the modeling approach for the DNR Detectors were near-equal mean and balanced accuracy values. Comparable balanced accuracies, along with the inability to project the RF methods within a growth and yield model, support the use of Detector models at this stage of model development. This assertion is made recognizing the RF statistical approach, and the thresholds identified by RF, may have differed if RF had similar objectives, and the far larger training plot sample size available, to the Detectors project.

From an area estimate perspective, both Detector and GNN classifications identified similar abundances of SCF on DNR-managed forestlands. Where the two approaches diverged was in their OF estimation. GNN identified an amount of OF equivalent to half of SCF area, whereas the Detectors classified an amount of OF almost equivalent to SCF. The similarity in SCF estimates from two independent methods strengthens confidence in the Detector model for that condition, especially as it provides a finer spatial scale of classification than GNN. However, the discrepancy in OF area continues to indicate low confidence in the current Detector OF model. It is uncertain if the GNN estimate of OF was also higher than what the land base supports. However, it is much more consistent with local knowledge of DNR-managed forestlands, suggesting an alternative modeling approach or a different complement of attributes that cannot be projected forward in time may result in a better OF map.

Conclusion

The Detectors project provided an initial classification and map of structurally complex forest and older forest probabilities. This effort showed that applying logistic regressions to a primarily height-driven inventory dataset, at 1/10th acre resolution, can yield a promising approach to mapping structurally complex forest (late mature through later stages), while older forest (old-growth structure only) proved difficult with the data and approach employed. The models developed, particularly the SCF Detector, compared well with other old-forest modeling/mapping products.

Like most scientific endeavors, the Detectors project showed promise and suggested ways such modeling might be improved. Further development proposals are anticipated, and retraining the current Detector models is anticipated with new versions of the remotely-sensed data and further training data collection (e.g., OESF planning unit). The current results support the use of the SCF model and map in the forthcoming Landscape Assessment, while caution should be used in any applications of the OF model and map documented in this report.

References

Bell, D.M., Gregory, M.J. and Ohmann, J.L., 2015. Imputed forest structure uncertainty varies across elevational and longitudinal gradients in the western Cascade Mountains, Oregon, USA. Forest Ecology and Management, 358, pp.154-164.

Burnham, K.P., Anderson, D.R. and Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral ecology and sociobiology, 65: 23-35.

Carey, A.B. and Curtis, R.O., 1996. Conservation of biodiversity: A useful paradigm for forest ecosystem management. Wildlife Society Bulletin. 24 (4): 610-620.

Davis, R.J., Ohmann, J.L., Kennedy, R.E., Cohen, W.B., Gregory, M.J., Yang, Z., Roberts, H.M., Gray, A.N. and Spies, T.A., 2015. Northwest Forest Plan–the first 20 years (1994-2013): status and trends of late-successional and old-growth forests. Gen. Tech. Rep. PNW-GTR-911. Portland, OR: US Department of Agriculture, Forest Ser- vice, Pacific Northwest Research Station. 82 p.

Davis, R.J., Bell, D.M., Gregory, M.J., Yang, Z., Gray, A.N., Healey, S.P., and Stratton, A.E., 2022. Northwest Forest Plan—the first 25 years (1994–2018): status and trends of late-successional and old-growth forests. Gen. Tech. Rep. PNW-GTR-1004. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 112 p.

[DNR] Department of Natural Resources, 1997. Final habitat conservation plan. Washington State Department of Natural Resources, Olympia, WA. 455 p.

[DNR] Washington State Department of Natural Resources, 2006. Policy for sustainable forests. Washington State Department of Natural Resources, Olympia, WA. 72 p.

Dodge, Y., 2008. The concise encyclopedia of statistics. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-32833-1 379

Franklin, J.F., Spies, T.A., Van Pelt, R., Carey, A.B., Thornburgh, D.A., Berg, D.R., Lindenmayer, D.B., Harmon, M.E., Keeton, W.S., Shaw, D.C. and Bible, K., 2002. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forests as an example. Forest ecology and management, 155(1-3), pp.399-423.

Franklin, J.F. and Van Pelt, R., 2004. Spatial aspects of structural complexity. Journal of Forestry 102(3):22-28.

Franklin, J.F., Johnson, K.N. and Johnson, D.L., 2018. Ecological forest management. Waveland Press.

Gray, A.N., Pelz, K., Hayward, G.D., Schuler, T., Salverson, W., Palmer, M., Schumacher, C. and Woodall, C.W., 2023. Perspectives: The wicked problem of defining and inventorying mature and old-growth forests. Forest Ecology and Management, 546, p.121350.

Griffey, V., Kellog, B., Haugo, R., Kane, V., 2021. Ownership patterns drive multi-scale forest structure patterns across a forested region in southern coastal Oregon, USA. Forests 12:47.

Halofsky, J.E., Creutzburg, M.K. and Hemstrom, M.A., 2014. Integrating social, economic, and ecological values across large landscapes. Gen. Tech. Rep. PNW-GTR-896. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 216 p.

Halofsky, J.S. and Donato, D.C., 2024. Assessing the Efficacy of a Large-Scale Habitat Conservation Plan in Western Washington, USA. Northwest Science, 97(4), pp.236-248.

Kane, V.R., Bakker, J.D., McGaughey, R.J., Lutz, J.A., Gersonde, R.F. and Franklin, J.F., 2010a. Examining conifer canopy structural complexity across forest ages and elevations with LiDAR data. Canadian Journal of Forest Research, 40(4), pp.774-787.

Kane, V.R., McGaughey, R.J., Bakker, J.D., Gersonde, R.F., Lutz, J.A. and Franklin, J.F., 2010b. Comparisons between field-and LiDAR-based measures of stand structural complexity. Canadian Journal of Forest Research, 40(4), pp.761-773

Landscape Ecology Modeling, Mapping, and Analysis (LEMMA) Team. 2025. Gradient Nearest Neighbor (GNN) raster dataset. Modeled forest vegetation data using direct gradient analysis and nearest neighbor imputation. Retrieved from https://lemma.forestry.oregonstate.edu/data.

McBurney, K., Estep, A, and C. Crump, 2024. Stand Development Stages Model Project Initiation. Washington Department of Natural Resources, Olympia, WA. Appendix 2.

Ohmann, J.L. and Gregory, M.J., 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. Canadian Journal of Forest Research, 32(4), pp.725-741.

Ohmann, J.L., Gregory, M.J. and Spies, T.A., 2007. Influence of environment, disturbance, and ownership on forest vegetation of coastal Oregon. Ecological Applications, 17(1), pp.18-33.

Ohmann, J.L., Gregory, M.J., Roberts, H.M., Cohen, W.B., Kennedy, R.E. and Yang, Z., 2012. Mapping change of older forest with nearest-neighbor imputation and Landsat time-series. Forest Ecology and Management, 272, pp.13-25.

Oliver, C.D. and Larson, B.C., 1996. Forest Stand Dynamics: Updated Edition. John Wiley and Sons, New York.

Pelz, K.A., Hayward, G., Gray, A.N., Berryman, E.M., Woodall, C.W., Nathanson, A. and Morgan, N.A., 2023. Quantifying old-growth forest of United States Forest Service public lands. Forest Ecology and Management, 549, p.121437.

[RF] Resilient Forestry, 2023. Identifying stand development stages in western Washington Review, synthesis, and methodology. Seattle, WA.

Ricklefs, J., 2025. Remotely Sensed Forest Resource Inventory System (RS-FRIS), Washington Department of Natural Resources, Olympia, WA.

Van Pelt, R., 2007. Identifying mature and old forests in western Washington. Washington State Department of Natural Resources.

Van Pelt, R., 2025 (in preparation). Identifying mature and old forests in western Washington. 2nd edition. Washington State Department of Natural Resources.

Appendices

Appendix A – Additional assessments of sampling bias

As noted elsewhere in the report (see section 2.1.4), the geographic distribution of plots generally aligned with the amount of DNR-managed forestland area across the westside HCP planning units (Figure A1).

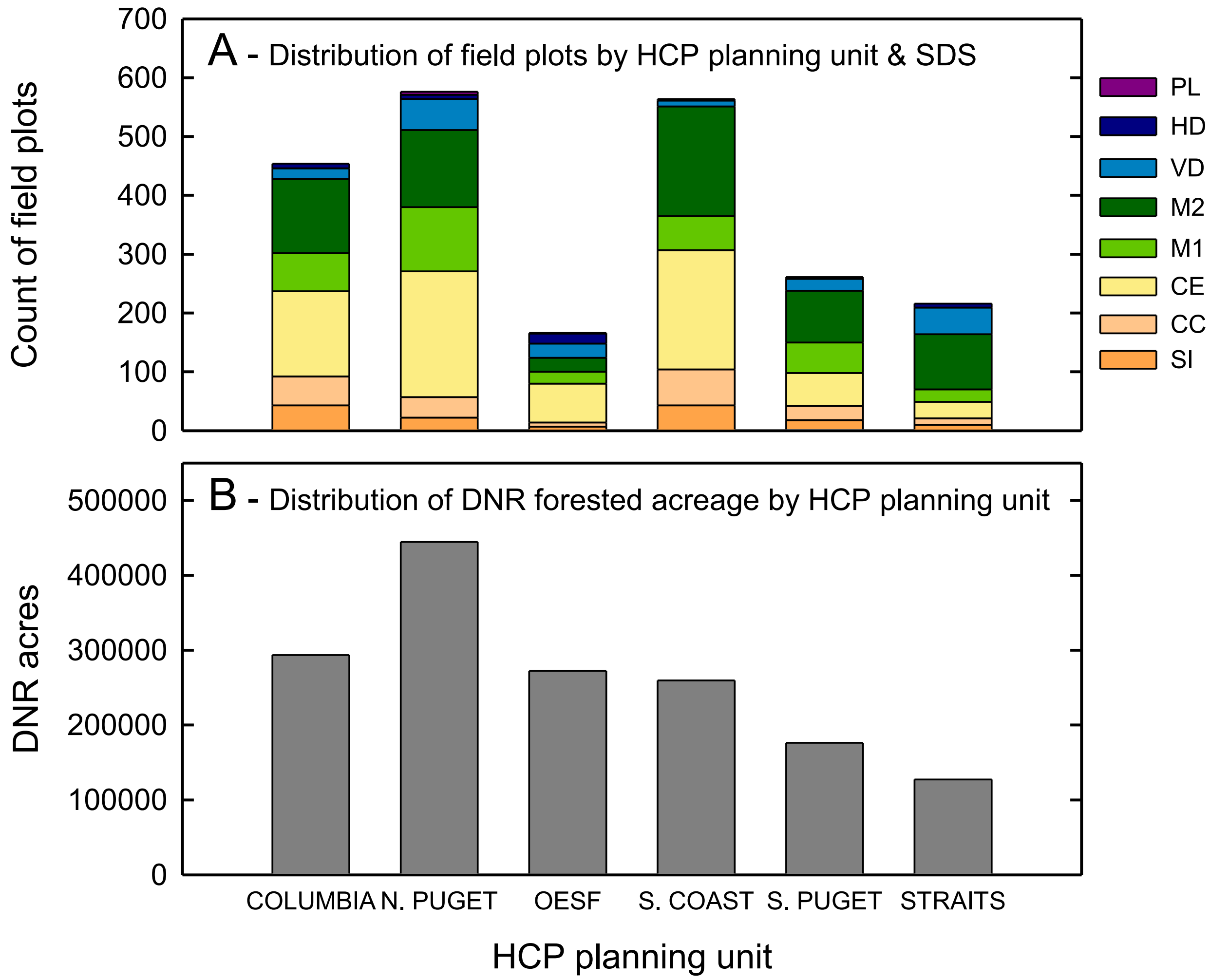


Figure A1. Distribution of (A) training plots and (B) DNR-managed forestlands by westside habitat conservation plan (HCP) planning unit. Panel A parses each planning unit's plots into stand development stage (SDS; see Table 1 for full names of each stage). The plot distribution largely corresponded well to the distribution of acreage among the planning units (See Table 2). All SDSs were also represented in approximately similar proportions in each planning unit. Note that ongoing data collection will eventually increase the number of OESF training plots to ~350 for subsequent model updates. See appendix A for additional assessments examining the distribution of plots across riparian areas and other environmental gradients.

Since riparian management zones compose approximately 1/3 of westside DNR-managed forestlands, an aim was to obtain a similar proportion in the training plot dataset. The DNR's Large Data Overlay (LDO, version 20240927) riparian area was used as a stratification tool. Currently the training plot dataset is composed of 35% riparian stands, compared to 31% of westside DNR-managed forestlands (Figure A2). Riparian areas were more represented in mid-to late development stages since they generally are not subject to variable retention harvest and are becoming older than much of the landscape (Halofsky and Donato 2024).

The distributions of training plots across major environmental gradients compared well to the distributions of all westside DNR-managed forestlands (Figure A2). Plots were distributed proportionally across elevation bands, forest types (potential vegetation types, PVT), and site productivity classes (Figure A3), suggesting no major biases in relation to these key factors. Not surprisingly, given the winter timeframe for the bulk of data collection (snow affecting access), the highest-elevation (which often tend to be low-productivity sites) were less represented in the dataset. Although these are a minor fraction of the land base, ongoing and future field effort will obtain additional training plots in these less represented locales.

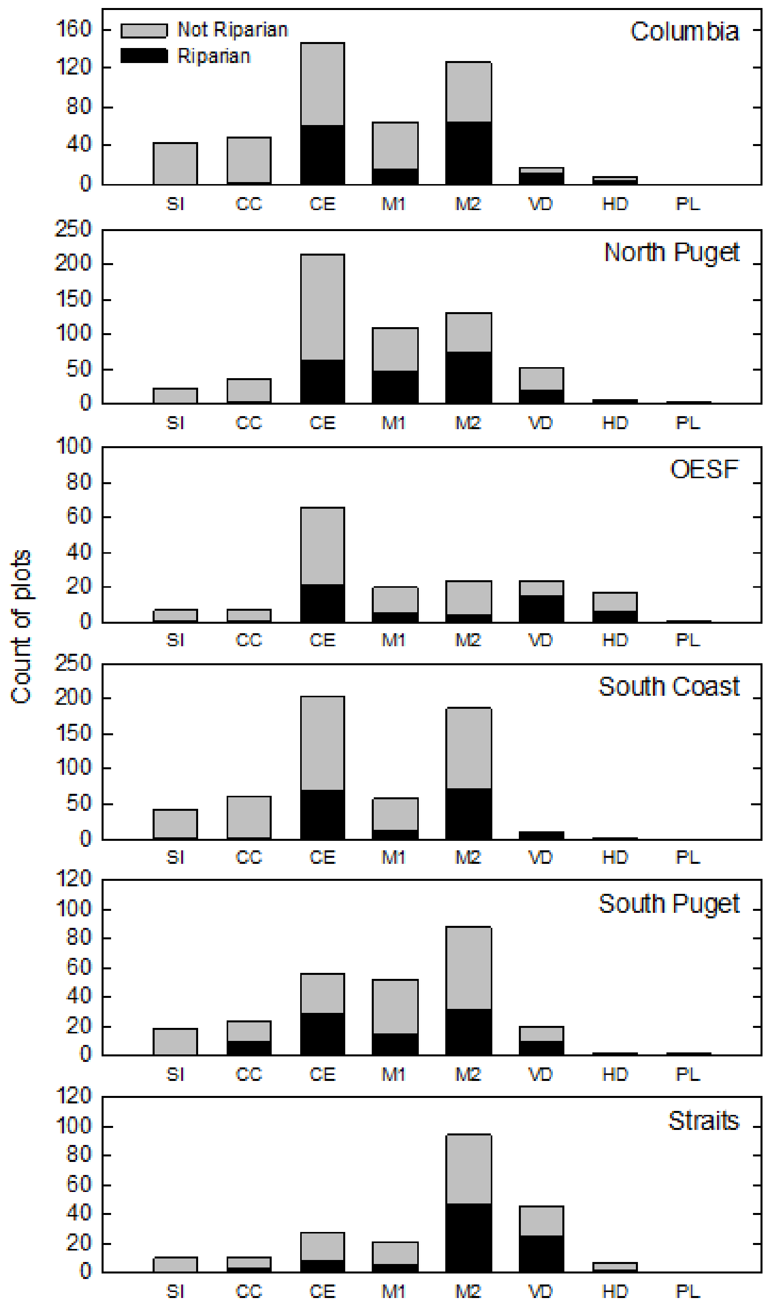


Figure A2. Distribution of training plots by riparian/not-riparian and stand development stage (SDS) in each westside habitat conservation plan (HCP) unit. The proportion of sampled riparian plots (35%) approximated the proportion mapped for all westside DNR-managed forestlands (31%). Riparian stands composed each SDS in generally similar proportions, with riparian areas more represented in mid- to late development stages since they are not subject to variable retention harvest and are becoming older than much of the landscape (Halofsky and Donato 2024).

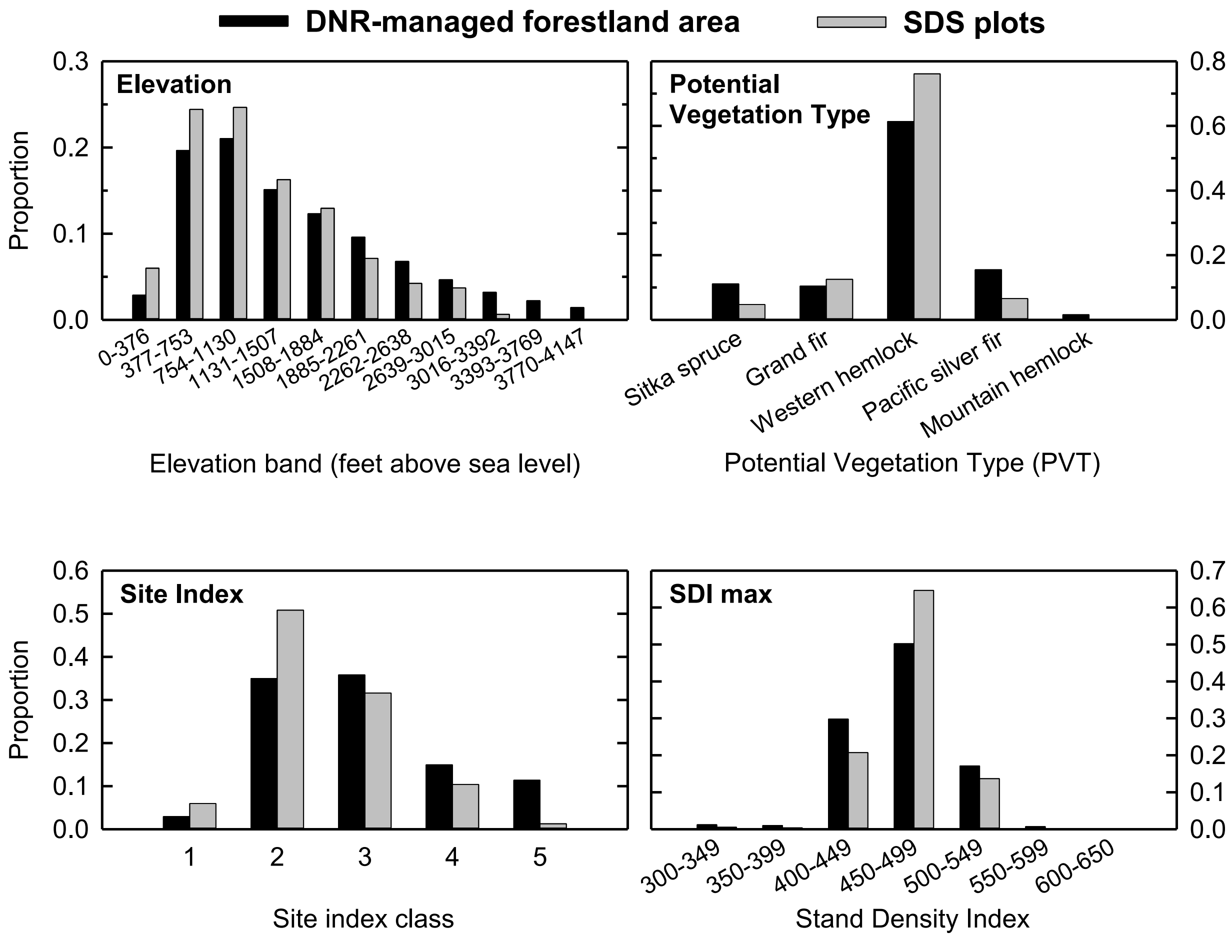


Figure A3. Distributions of elevation, potential vegetation type (PVT), site index class, and stand density index max (SDI max 10) among the training plots compared to all westside DNR-managed forestlands. The distributions were overall similar, suggesting no major bias in relation to these key gradients. Elevations were extracted from a digital elevation model. PVT is from the Integrated Landscape and Assessment Project (Halofsky et al. 2014). Site index class and SDI max 10 are for westside Douglas-fir.

Appendix B – RF SCF and OF performance without refitting

Table B1. Balanced accuracy provides the best summary of model performance. Abbreviations: tn = true negative; fp = false positive; fn = false negative; tp = true positive; tpr = true positive rate; tnr= true negative rate; accuracy = mean accuracy. * Total number of plots = 2,231 (SCF = 869; Not SCF = 1,362; OF = 220; Not OF = 2,011) and at the RIU scale the total number of RIUs = 1,452 (SCF = 520; Not SCF = 932; OF = 143; Not OF = 1,309).

Model	tn	fp	fn	tp	tpr	tnr	f-score	accuracy	balanced accuracy
RF 8 Stage 5-acre Scale SCF	808	554	54	815	0.94	0.59	0.73	0.73	0.77
RF 8 Stage 5-acre Scale OF	1638	373	148	72	0.33	0.81	0.22	0.77	0.57
RF 8 Stage 2.5-acre Scale SCF	808	554	54	815	0.94	0.59	0.73	0.73	0.77
RF 8 Stage 2.5-acre Scale OF	1638	373	148	72	0.33	0.81	0.22	0.77	0.57
RF 5 Stage 5-acre Scale OF	1643	368	162	58	0.26	0.82	0.18	0.76	0.54
RF 5 Stage 2.5-acre Scale OF	1621	390	158	62	0.28	0.81	0.18	0.75	0.54
RF 8 Stage RIU Scale SCF	629	303	41	479	0.92	0.67	0.74	0.76	0.8
RF 8 Stage RIU Scale OF	1005	304	69	74	0.52	0.77	0.28	0.74	0.64
RF 5 Stage RIU Scale OF	1182	127	120	23	0.16	0.9	0.16	0.83	0.53