Water Resources Research

RESEARCH ARTICLE
10.1029/2025WR040478

A Streamflow Permanence Classification Model for Forested Streams That Explicitly Accounts for Uncertainty and Extrapolation

Jonathan D. Burnett[1], Kristin L. Jaeger[2], Sherri L. Johnson[1], Steven M. Wondzell[1], Jason B. Dunham[3], Matthew I. Barker[2], Emily Heaston[3], Nathan Chelgren[3], Michael G. Wing[4], Brian Staab[5], and Michael E. Brown[6]

[1]Pacific Northwest Research Station, U.S. Forest Service, Corvallis, OR, USA, [2]U.S. Geological Survey, Washington Water Science Center, Tacoma, WA, USA, [3]Forest and Rangeland Ecosystem Science Center, U.S. Geological Survey, Corvallis, OR, USA, [4]College of Forestry, Oregon State University, Corvallis, OR, USA, [5]U.S. Forest Service, Portland, OR, USA, [6]Bureau of Land Management Oregon/Washington, Portland, OR, USA

Abstract Accurate mapping of headwater streams and their flow status has important implications for understanding and managing water resources and land uses. However, accurate information is rare, especially in rugged, forested terrain. We developed a streamflow permanence classification model for forested lands in western Oregon using the latest light detection and ranging-derived hydrography published in the National Hydrography Dataset. Models were trained using 2,518 flow/no flow field observations collected in late summer 2019–2021 across headwaters of 129 sub-watersheds. The final model, the Western Oregon WeT DRy model, used Random Forest and 13 environmental covariates for classifying every 5-m stream sub-reach across 426 sub-watersheds. The most important covariates were annual precipitation and drainage area. Model output included probabilities of late summer surface flow presence and were subsequently categorized into three streamflow permanence classes—Wet, Dry, and Ambiguous. Ambiguous denoted model probabilities and associated prediction intervals that extended over the 50% classification threshold between wet and dry. Model accuracy was 0.83 for sub-watersheds that contained training data and decreased to 0.67 for sub-watersheds that did not have observations of late summer surface flow. The model identified where predictions extrapolated beyond the domain characterized by the training data. The combination of spatially continuous estimates of late summer streamflow status along with uncertainty and extrapolation estimates provide critical information for strategic project planning and designing additional field data collection.

Plain Language Summary Understanding where small streams are located and if they flow year-round or seasonally matters for how these streams and their watersheds are managed. We developed a model trained on simple flow/no flow field observations and 13 variables that describe climate, topography, and land cover conditions to provide predictions of whether a 5-m stream sub-reach is likely to have late summer flow or may go dry. The model is applied to a high resolution, light detection and ranging-derived stream network for 426 sub-watersheds in western Oregon and represents years with average to slightly drier-than-average rainfall. We also evaluated differences in model accuracy between watersheds that did and did not have training data to provide more realistic uncertainty estimates of model predictions. Field observations for these types of models is generally limited both in number of observations and their geographic distribution. Accuracy estimates for watersheds with (83% correct) and without (67% correct) training data can help managers decide if they use model predictions or if they need to collect additional field data. In addition, this model identifies stream locations where the model is extrapolating, which can be used for identifying locations where more data collection is needed for model improvement.

Key Points:

- Empirical model classifies late summer surface water presence in stream sub-reaches as Wet, Dry, or Ambiguous classes in headwater streams
- Model assessment includes separate accuracy estimates for stream locations that are both near and far from training data locations
- Model identifies regions where model predictions are extrapolations beyond conditions captured by training data

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. L. Jaeger,
kjaeger@usgs.gov

Citation:

Burnett, J. D., Jaeger, K. L., Johnson, S. L., Wondzell, S. M., Dunham, J. B., Barker, M. I., et al. (2025). A streamflow permanence classification model for forested streams that explicitly accounts for uncertainty and extrapolation. *Water Resources Research*, *61*, e2025WR040478. https://doi.org/10.1029/2025WR040478

Received 12 MAR 2025
Accepted 4 JUN 2025

Author Contributions:

Conceptualization: Jonathan D. Burnett, Kristin L. Jaeger, Sherri L. Johnson, Steven M. Wondzell, Jason B. Dunham
Data curation: Jonathan D. Burnett, Kristin L. Jaeger, Emily Heaston
Formal analysis: Jonathan D. Burnett
Funding acquisition: Michael E. Brown
Investigation: Jonathan D. Burnett, Kristin L. Jaeger, Matthew I. Barker
Methodology: Jonathan D. Burnett, Kristin L. Jaeger, Nathan Chelgren
Project administration: Sherri L. Johnson
Resources: Sherri L. Johnson, Michael G. Wing
Software: Michael G. Wing

1. Introduction

Headwater streams, the first- and second-order streams that extend to the upper portions of watersheds (Golden et al., 2025), account for most of the channel network by length (Downing et al., 2012), yet accurate mapping of both their location and permanence of streamflow remains a limiting factor (Brinkerhoff, 2024; Messager et al., 2024). The management of headwaters streams, and their adjacent riparian forests, frequently depends on their streamflow permanence classification (Acuña et al., 2017; Boisjolie et al., 2017; Brinkerhoff et al., 2024; Kampf et al., 2021; Messager et al., 2024). A crucial, primary classification is whether the streams have perennial,

Supervision: Sherri L. Johnson
Visualization: Jonathan D. Burnett, Kristin L. Jaeger, Matthew I. Barker
Writing – original draft: Jonathan D. Burnett, Kristin L. Jaeger
Writing – review & editing: Jonathan D. Burnett, Kristin L. Jaeger, Sherri L. Johnson, Steven M. Wondzell, Jason B. Dunham, Matthew I. Barker, Michael G. Wing, Brian Staab, Michael E. Brown

or year-round flow, or are non-perennial, defined as the stream being dry at some point in a typical year (Busch et al., 2020).

The location of streams, hereafter referred to as hydrography, and streamflow permanence is especially difficult to determine in headwater streams in forested regions. Frequently, the complex, rugged terrain can hinder observations, and the channels can lack strong topographic signatures, and can be obscured by vegetation (Benstead & Leigh, 2012; Kim et al., 2023; Metes et al., 2022). Depending on the resolution of the hydrographic data used, the upper extents of perennial headwater streams might not be delineated on stream network maps (Anderson et al., 2024; Messager et al., 2024). Streamflow data and classifications are lacking in part because of the rarity of data collection in the headwaters region of the river network and the bias in data collection in the downstream, perennial regions of the river network (Krabbenhoft et al., 2022; van Meerveld et al., 2020). There remains a persistent need for better mapping and characterization of headwaters and their streamflow permanence (Brinkerhoff, 2024). For the United States, the most widely available data set to determine streamflow permanence, the National Hydrography Dataset, has been shown to have errors for headwater streams by as much as 50% (Fritz et al., 2013) based on data that may be decades out of date (Hafen et al., 2020).

There has been substantial recent momentum in both the development of models and data collection methods to classify and characterize streamflow permanence in headwater streams across a range of time scales and geographical extents. Mapping streamflow permanence status has included approaches that span from physical models of individual catchments at varying temporal resolutions (Hafen et al., 2023; Mahoney et al., 2023; Scheller et al., 2024; Ward et al., 2020); statistical models to identify wetted channel length dynamics (Botter et al., 2021, 2024; Bujak-Ozga et al., 2023; Jensen et al., 2018), or surface flow duration, including changes through time (Peterson et al., 2024; Sauquet et al., 2021). In addition, classification models of streamflow permanence have used empirical approaches for individual catchments (Durighetto et al., 2022; Kaplan et al., 2022; Pate et al., 2020; Whiting & Godsey, 2016), and have been expanded to all streams and rivers at regional (Jaeger et al., 2019; Sando et al., 2022), and global extents (Messager et al., 2021; Sauquet et al., 2021). Hybrid approaches that combine physical and empirical approaches are also evolving (Döll et al., 2024; Mimeau et al., 2024).

Data requirements vary for these different approaches, with intensive data collection efforts generally corresponding to modeling at fine spatial and temporal resolutions. Consequently, there have been advancements in technology to facilitate cost effective data collection opportunities of surface flow presence versus absence, including electrical resistance sensors (ER sensors; Bhamjee et al., 2016; Chapin et al., 2014; Goulsbra et al., 2014), multi-sensor technologies that also leverage the use of field cameras (Assendelft & van Meerveld, 2019), and simple one-time visual observations leveraging applications on mobile devices and crowdsourcing (FLOwPER (Jaeger et al., 2020), Stream Tracker (Kampf, 2018), Crowd Water (Seibert et al., 2019), DRYvER (Truchy et al., 2023)). Simple wet or dry observations during low flow periods are particularly useful given that more extensive data collection can be extremely expensive, yet there remains a desire to provide model estimates over large geographic extents in locations where no hydrologic measurements exist or are rare and observation of streamflow permanence through remote sensing is not a viable option, such as in forested headwater systems (Kim et al., 2023).

In the Pacific Northwest of United States, the U.S. Geological Survey (USGS) developed the $\text{PROSPER}_{\text{PNW}}$ model (PRObability of Streamflow PERmanence; Jaeger et al., 2019), which provides a probability of year-round flow for 30-m-stream reaches. $\text{PROSPER}_{\text{PNW}}$ is a Random Forest (RF) model that used climatic and physiographic covariates trained on visual wet or dry observations. Two main limitations of the existing $\text{PROSPER}_{\text{PNW}}$ model are the relative coarseness of the underlying hydrography on which the streamflow permanence estimates are based and lack of observational data for some portions of the modeling domain, particularly small forested streams in the rugged region of western Oregon. $\text{PROSPER}_{\text{PNW}}$ is based on the National Hydrography Dataset Plus Medium Resolution (NHD MR), which is derived from a combination of USGS 1:100,000 scale maps and 1:62,500-scale maps and has a resolution of 30-m (Johnston et al., 2009). This resolution can underestimate the full extent of the stream network, especially in mountainous terrain, where channel length may be underestimated by as much as 50% (Christensen et al., 2022; Clarke et al., 2008; Colson et al., 2008; Vance-Borland et al., 2009). Conversely, light detection and ranging (LiDAR)-derived hydrography can improve mapping accuracy in headwater streams (Metes et al., 2022; Russell et al., 2015) and there are major efforts across many regions for LiDAR-derived hydrography (Anderson et al., 2024). However, LiDAR-derived hydrography can delineate many

additional headwater channels with unknown flow status. It is challenging for managers to plan management actions that involve these newly mapped channels without some means of classification of which reaches are more or less likely to support perennial flow. Field observations of each reach of each channel is time consuming and often not logistically possible on the timeline of the planning efforts, especially if observations were not collected in advance during the time of minimum annual flow.

The objective of this study was to develop a new streamflow permanence classification model at a sub-regional scale for forested headwaters, using updated, high resolution LiDAR hydrography and new observations of streamflow permanence that span a large geographic area and were collected over three summers. We incorporated additional elements of algorithm comparison, uncertainty analyses, and identification of model extrapolation. We included additional components in the covariate selection process to address land and water management interests in evaluating the importance of drainage area, precipitation, and forest cover on streamflow permanence (Clarke et al., 2008; Kampf et al., 2021; Segura et al., 2020). We evaluated three model algorithms, RF, Logistic Regression (LR), and Extreme Gradient Boosting (XGB) as part of model development. Additionally, we evaluated the accuracy of models when training data were located physically near to the prediction area versus data that is physically further away and thus may not represent local conditions. Finally, the model identifies stream sub-reaches that are outside of the training data set; therefore, streamflow permanence predictions are considered extrapolations, representing first approximations of streamflow permanence.

2. Data and Methods

2.1. Study Site

This study was conducted in forested, mountainous areas of western Oregon (Figure 1). The study area included three ecoregions, the Coast Range, the Cascades, and the Klamath Mountains (Omernik & Griffith, 2014). The lowland Willamette River valley that separates the Coast and Cascade Ranges in the northern half of the study area was excluded from this study due to its high population density, limited forest cover, and human modifications to stream networks. The Coast Range had the lowest average elevations of the three mountain regions with maximum elevations of 1,200 m; maximum elevations were approximately 2,300 m in the Klamath Region and exceeded 3,000 m for the Cascade Range. The climate is maritime temperate with cool, wet winters. More than 2,000 mm average annual precipitation occurs in the Coast Range and higher elevations of the Cascade Range. The Klamath Mountains are drier, receiving only approximately 1,440 mm of annual precipitation (PRISM, 2014). Summers are dry and warm with little to no precipitation in late July and August (Daly et al., 2008; Gaines et al., 2022).

The geology of the study area is complex and we broadly characterize the geology in lithologic provinces following O'Connor et al. (2014); these lithologic provinces generally correspond to the three ecoregion boundaries. The Coast Range and western flank of the Cascade Range are characterized by marine sediments that comprise the Coast Range sedimentary rocks lithologic province interspersed with either more resistant Columbia River basalt and volcanic rocks collectively referred to as Coast Range volcanic rocks and Columbia River basalt lithologic province. Un-weathered basalt and andesite comprise the High Cascades lithologic province and occur along the Cascade Crest. Older aged, deeply dissected volcanic rocks correspond to the Western Cascades lithologic province that flanks the west side of the High Cascades province. The Klamath lithologic province to the south is a complex combination of older Paleozoic and Mesozoic volcanic, sedimentary, and metamorphic rocks with intense deformation history.

The study area extends across 31,443 km^2 and is composed of 871 sub-watersheds at the 12-digit Hydrologic Unit Code (HUC12) scale. The size of individual HUC12s within the study area range from 17 to 194 km^2 with a median of 70 km^2. The modeling domain within the larger study area includes a subset of 426 HUC12s selected based on the following criteria: (a) LiDAR-derived elevation models (DEMs) are available in a publicly accessible repository at the Oregon Department of Geology and Mineral Industries (DOGAMI, 2025) for at least 90% of the area within each HUC12 boundary, and (b) LiDAR-derived hydrography are available as part of the latest updates to the National Hydrography Dataset (NHD; USGS, 2021; DOGAMI, 2025) as of 1 September 2022. Location resolution of the LiDAR ranges between 1 and 3 m, with 20-cm vertical resolution for elevation. The NHD resolution is 1:12,000.

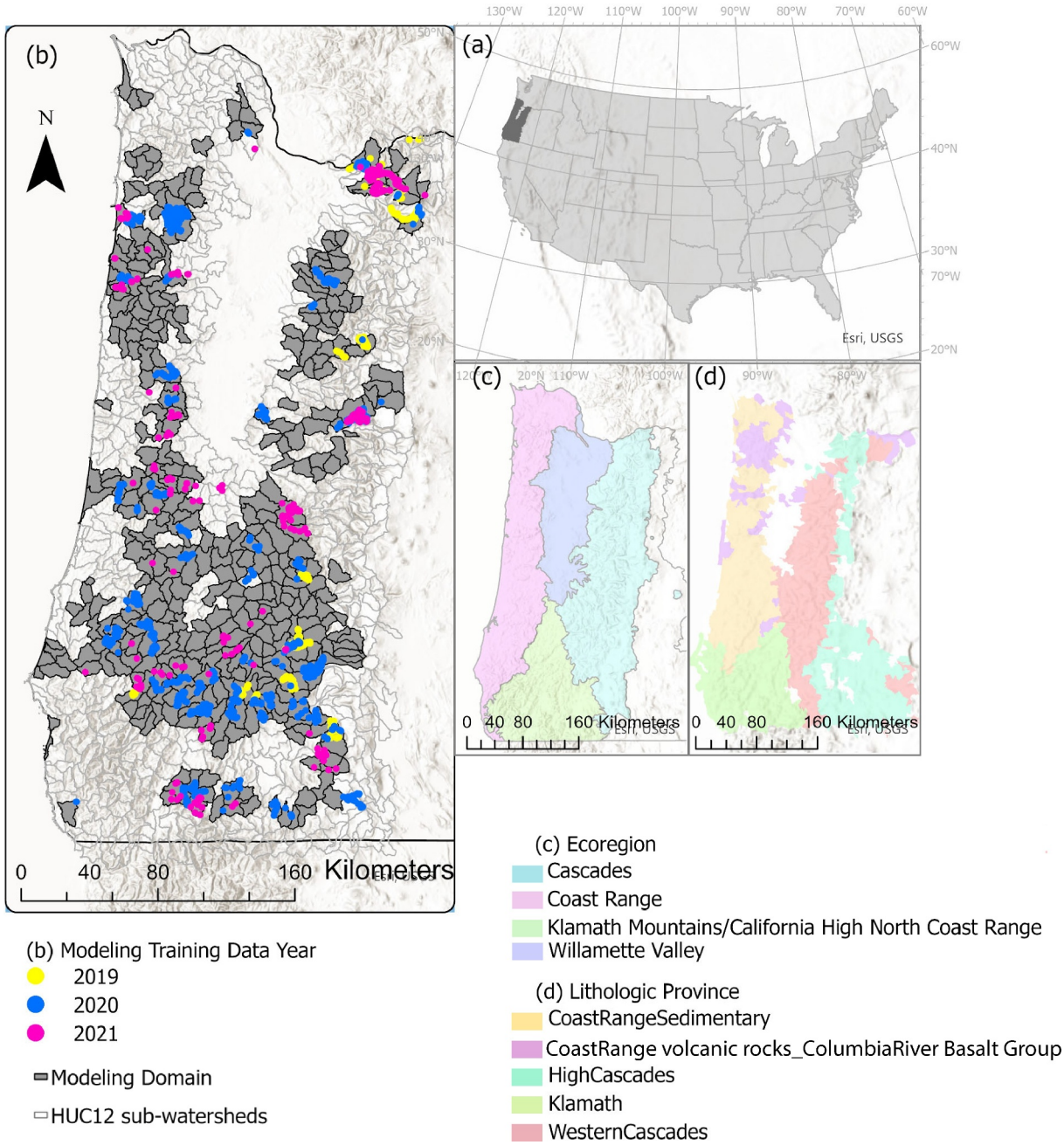

Figure 1. Study area extent in western Oregon (a), study area that includes 871 12-digit Hydrologic Unit Code (HUC12) sub-watersheds and modeling domain of 426 sub-watersheds with 2,518 FLOwPER observations for model training (b), overlay of the Environmental Protection Agency Level Three ecoregions within the study area (Omernik & Griffith, 2014) (c), and overlay of the five major lithologic provinces within the study area (O'Connor et al., 2014) (d). The Willamette Valley ecoregion is excluded from analysis due to our focus on forested sub-watersheds. Base map from Esri and its licensors, copyright 2022.

2.2. Field Observations of Surface Water Presence

We used the FLOwPER feature mapping application (Jaeger et al., 2020) to collect observations of late-summer streamflow status (i.e., wet or dry) in the modeling domain between 15 July and 30 September for the years 2019, 2020, and 2021. Observations represent a single location and did not include repeat observations at the same location within a year. FLOwPER is a survey form hosted within the ArcGIS Survey123 application that establishes a protocol and data entry standard for collecting streamflow permanence observations. Data collection

was conducted by several entities with an emphasis for data collection at road-stream crossings on public lands to ensure rapid data collection across the broadest possible area with minimal access restrictions.

Surface water presence and absence data were collected during the late summer seasons of 2019, 2020, and 2021 across 140 HUC12s in the study area. Annual precipitation for these years was 85% (2019) to 96% (2021) of the 30-year normal (1981–2010 period) for the study area (PRISM, 2014) and late summertime low-flow conditions were apparent by August of each year (National Drought Mitigation Center, 2019). Year of observation was included in the model, with model results characterizing a classification of flow status during low flow periods for each year and reported as a mean across these 3 years. Our study included a mean of 19 observations per HUC12 with a standard deviation of 36. Observations of wet streams occurred across drainage areas ranging from 0.0032 to 52.3 km^2 and dry observations occurred across drainage areas ranging from 0.0008 to 33.8 km^2. A total of 1,412 observations were wet, 36 were discontinuous, and 1,204 observations were dry with 99 repeat locations where observations were taken at the same location in more than one year of the study period. Approximately 19%, 63%, and 18% of the data were collected in 2019, 2020, and 2021, respectively. Based on the NHD LiDAR-derived hydrography, which is equivalent to 1:12,000 (USGS, 2021), approximately 52% of the observations were collected in Strahler first-order streams, 32% were collected in second-order streams, 12% were collected in third order streams, 3% were collected on fourth-order streams, with the remainder being fifth order and larger order streams. Finally, approximately 33% of the observations were collected in the Western Cascades lithologic province followed by 22% of the observations occurring in the High Cascades lithologic province, with the remaining observations collected in the other three lithologic provinces (Figure 1).

2.3. Climatic and Physiographic Covariates

We assembled a spatial data set of 96 hydro-topographic and climatic variables for consideration in the model (Table S1 in Supporting Information S1). Covariates were selected to capture different aspects of environmental and geophysical conditions contributing to surface flow expression (Costigan et al., 2016; Hammond et al., 2021; Shanafield et al., 2021). The primary determinant of covariate inclusion was the availability of spatially continuous coverage across the 426 HUC12s in the modeling domain. Covariates included 5-m resolution hydro-topographic metrics derived from 5-m resolution LiDAR DEMs (DOGAMI, 2025). Climatic covariates included 800-m resolution precipitation and air temperature data for each calendar year extending from 2015 to 2021 and the 30-year climate normal period represented by 1980–2010. Lag periods of 2015–2018 were included to describe antecedent conditions in years prior to the study period.

2.4. Model Workflow

The workflow in Figure 2 shows components of: (a) model preprocessing of the data and developing the training data set (Section 2.4.1), (b) the cross-validation framework (Section 2.4.2), (c) covariate selection (Section 2.4.3), (d) model development (Section 2.4.4), (e) predictions (Section 2.4.5), and (f) subsequent analysis of the final model (Section 2.4.6).

2.4.1. Model Preprocessing

2.4.1.1. Processing of Observation Data

A total of 2570 observation points within 129 sub-watersheds (i.e., HUC12s) were selected for model development from the FLOwPER database by filtering observation data with the following criteria: (a) Points located in natural channels to avoid the potentially confounding influence of artificial diversions and canals on model inference. (b) A streamflow status of either wet or dry; points with discontinuous flow status were removed because the objective was a binomial classification. Furthermore, the low frequency of discontinuous data (<10% overall) would not produce a sufficiently accurate classification model. (c) High geolocation accuracy as defined by having a 10-m or less 95% circular error probable, meaning that estimated position location has a 95% or greater probability of falling within 10-m or smaller radius (Chaitanya et al., 2016). The 10-m distance limit was imposed to be consistent with the 10-m observation distance of the FLOwPER survey and to constrain the influence of measurement error on model calibration. (d) Points within 10 m of the NHD LiDAR-derived hydrography (USGS, 2021).

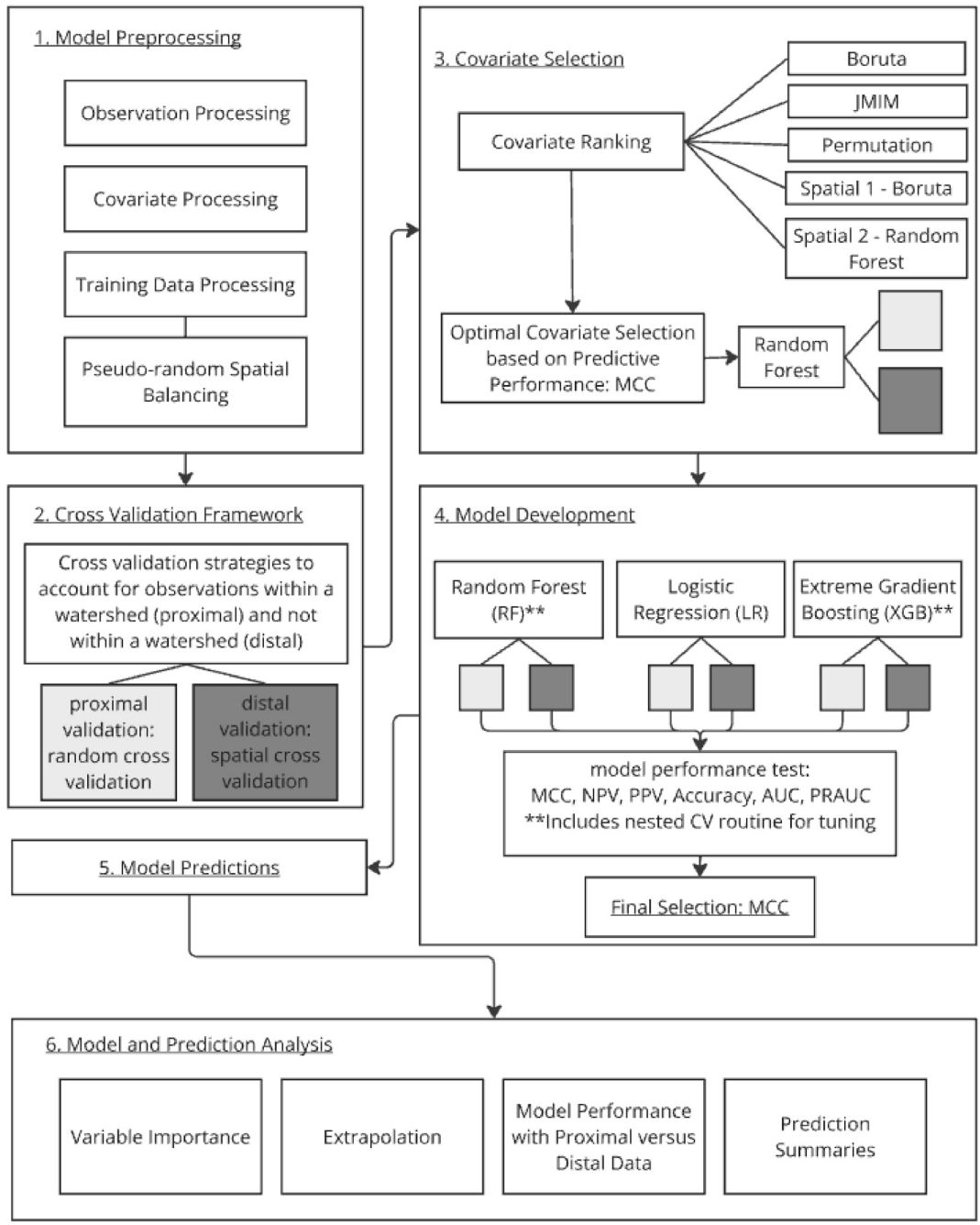

Figure 2. Schematic of workflow to describe steps within (1) model preprocessing, (2) cross validation with proximal versus distal data, (3) covariate selection considering five methods, (4) model development considering three algorithms, (5) model predictions, and (6) model and prediction analysis. Dark and light shading refer to the proximal and distal validation process in Steps 2, 3, and 4.

2.4.1.2. Covariate Processing

Values of the climatic and physiographic covariates considered in the model were extracted from the gridded spatial layers both as local values and non-local values (Table S1.1 in Supporting Information S1). Local covariate values represent conditions at the immediate location of the stream reach (5-m) and correspond to the location of the observation point. FLOwPER observations are taken based on surface flow conditions over a 10-m stream reach. A smaller 5-m stream reach length was chosen to capture the fine scale variability of LiDAR-

derived covariates. Non-local values represent values over varying lengths or drainage areas, e.g., basin average values. Different non-local sizes and types were generated to include down-channel and up-channel conditions over relatively large distances (e.g., 50-m, 500-m, 1,000-m) and upslope conditions (e.g., 100-m, 250-m, 500-m, basin average) to capture dominant influences of topographic conditions at varying scales. The most frequently used non-local size was the basin scale drainage area weighted average of the covariate as estimated with flow conditioned parameter grids (Barnhart et al., 2020), which have been used in streamflow permanence modeling across a range of geographic scales (Jaeger et al., 2019; Jensen et al., 2018; Kaplan et al., 2020; Sando et al., 2022). Basin-scale averages were estimated for covariates where the influence of that covariate on surface flow expression is thought to be due to an accumulation of that condition (e.g., climate, landcover, lithology).

2.4.1.3. Aligning Training Data and Covariates to the Stream Network Modeling Domain

A stream network was generated to serve as the modeling domain on which the model was trained, and predictions were estimated and correspond to the areas represented by 426 HUC12 sub-watersheds (refer to Text S1 in Supporting Information S1 for details). The stream network was generated as stream points to represent 5-m stream sub-reaches within the LiDAR-derived hydrography for the study area.

Model training data were produced by aligning the observation points to the nearest stream point of the stream network in the modeling domain. Additionally, each stream point within the modeling domain outside of the training data locations was also attributed with each of the 96 predictor covariates to allow for prediction at stream points throughout the stream network.

2.4.1.4. Training Data Processing: Filtering and Spatial Balancing

The training data were subsequently filtered to produce a total of 2,518 observations for inclusion in the model. Subsequent filtering included removing observations that had missing values in the covariates, which can occur as a result periodic missing values in underlying source grids of a given covariate. Data were examined for low variation (coefficient of variation < 0.5) with covariates, although none were flagged for removal except for Strahler stream order, where variability was limited by the inherently constrained nature of the data. Training data were then examined for multi-collinearity through pairwise correlation analysis among each of the 96 covariates. Only covariates with extremely high correlation (greater than 0.97) were removed because inclusion of highly correlated variables tends to result in higher predictive performance (Hanberry, 2024). Although parsimony was a consideration, the overall objective of this model was predictive accuracy (Shmueli, 2010). The potential of producing a model containing many features that in turn included potentially redundant information was deemed acceptable if it resulted in maximum predictive accuracy. The filtered data reduced the 2,570 observations to 2,518 observations (1,361 wet observations, 1,157 dry observations; Table S2.1 in Supporting Information S1) from 129 HUC12 sub-watersheds.

The filtered data underwent a first tier of data splitting through pseudo-random spatial balancing to account for different densities of data across the model domain (refer to Text S2 in Supporting Information S1 for details). Random splits were not stratified across years. Spatial balancing used oversampling of both categories at the HUC12 level to ensure equal numbers of wet and dry observation within a spatial group, resulting in a training data set of 3,076 wet/dry (1,538 wet and 1,538 dry) points for model training. More wet points are in the balanced training data set than in the source observation data because in some spatial grouping arrangements dry was the majority class. The training data set were divided into 20 sub-groups, to ensure both a sufficiently large population of groups and sufficient number of observations present in each group subsequent randomization and splitting for training and validation during cross validation procedures (refer to Text S2 in Supporting Information S1 for details).

2.4.2. Cross Validation That Accounts for Differences Between Spatially Distal and Spatially Proximal Data

A second tier of data splitting through cross validation was imposed in the modeling workflow to evaluate model accuracy for sub-watersheds that either have data or do not have data. Specifically, we evaluate accuracy for (a) sub-watersheds that have training data within the sub-watershed (HUC12) boundary, here referred to as a proximal condition, and (b) sub-watersheds that do not have training data within the sub-watershed boundary, here referred to as a distal condition. The cross validation described in this section was applied in both the

Covariate Selection (Section 2.4.3) and Model Development (Section 2.4.4) components of the workflow (Figure 2). A limitation of conventional cross validation approaches that randomly assign data to training and validation folds without consideration for the potential of spatial autocorrelation is that it can produce optimistic estimations of predictive performance as a result of training and testing with nearby neighbors (Roberts et al., 2017; Tsamardinos et al., 2015). This can result in training subsets that are spatial neighbors to the testing data subset and therefore highly spatially correlated. To mitigate the potential consequence of the conventional cross validation approach, we imposed (a) a completely random cross validation to provide a characterization of predictive performance for training data that are spatially near or proximal to the prediction locations and (b) a spatial cross validation to provide a characterization of predictive performance for training data that are spatially far or distal to the prediction locations (Figure 2) (refer to Text S3 in Supporting Information S1 for details). To evaluate predictive performance for both the proximal and distal strategies, we implemented a 5-fold cross validation approach by taking data from the 20 spatially balanced groups and randomly assigning into a 5-fold (completely random without resampling) where each fold contains approximately the same amount of data (Figure S3.1 in Supporting Information S1). The distal strategy included assigning spatial groupings to folds such that the hold out test data did not have observations from within a sub-watershed (refer to Text S3 in Supporting Information S1 for details).

2.4.3. Covariate Selection

A subset of covariates for fitting streamflow permanence prediction models was pre-selected from the full suite of 96 covariates using covariate selection. Covariate pre-selection prior to model development saves computational time by identifying a subset of covariates that is most important to prediction without requiring all 96 covariates. This would consequently increase model accessibility by reducing model size, complexity, and computation time required for running the final prediction model.

2.4.3.1. Covariate Ranking

Covariate selection was applied to identify and rank the most important variables to aid in computational efficiency (refer to Text S4 in Supporting Information S1 for details). Covariate selection essentially front loads the most important covariates for predicting streamflow permanence and thus allows for identification of parsimonious model without the arduous process of tuning and fitting models to all 96 covariates. A total of five covariate selection methods were implemented (Figure S4.1 in Supporting Information S1). Three conventional methods, Boruta, Joint Mutual Information Maximization (JMIM), and Permutation covariate selection methods, were employed and two additional spatial approaches, Spatial 1–RF and Spatial-2–Boruta, which are modified spatial approaches to mitigate the influence of spatial bias from clustering of observations within sub-watersheds. The five methods generated ranked combinations of a subset of covariates based on their importance for predicting late summer streamflow status.

For each individual covariate selection approach, ranked covariate combinations were generated that included a minimum of four covariates and a maximum of 96 covariates, which were all covariates considered. Combinations were defined by starting with the first four ranked variables of each covariate selection method and increasing to all 96 covariates, with covariates being added sequentially in descending order of importance according to the ranking by each of the five covariate selection methods. To expedite processing time, after 10 covariates, the next three covariates in descending order of importance were added to the covariate combination. A total of 50 covariate permutations for each of the five covariate selection methods were evaluated.

Following the covariate selection, three covariates, drainage area (DA), annual precipitation (P_Annual) and proportion of forest cover, were reordered as the top three important covariates to ensure these three covariates could be evaluated for their importance in the final model. Reordering of the three covariates did not influence their importance in the model in RF and XGB because the model development process uses an internal randomization of covariate splits that occurs during model fitting. Covariate order does not influence LR prediction results due to the commutative property of the model terms. It is possible that relative covariate importance can be influenced by order particularly if multicollinearity exist in the data. However, including these covariates ensured consideration of covariates that described water inputs, overstory vegetation influences, and drainage area. Land and water stewardship agencies within the Pacific Northwest region are interested in understanding the influence of these covariates on streamflow permanence. Drainage area threshold values have

been identified as a regulatory strategy for streamflow permanence (Clarke et al., 2008). Similarly, timber harvest has been shown to potentially have mixed influences over time on low flow conditions in forested streams (Coble et al., 2020; Segura et al., 2020). Therefore, understanding the relative importance of forest cover on streamflow permanence could have land management implications (Kampf et al., 2021). Finally, land and water resource managers were interested in understanding the importance of precipitation given the potential sensitivity of streamflow permanence to changes in precipitation in this region (Ward et al., 2020) and the prospect of future regional changes in precipitation (Luce et al., 2013).

2.4.3.2. Covariate Combination Scoring

RF models were fit to each of the 50 covariate permutations for each of the five covariate selection methods using the Ranger package (Wright & Ziegler, 2017) in R (R Core Team, 2024) with the following hyper-parameterization: (mtry = rounded sqrt (number of covariates), number of random splits = 3, number of trees = 1,000, probability trees = true, classification = true, split rule = extratrees, method = permutation). Hyperparameter tuning was not conducted in an effort to reduce processing time.

Matthews Correlation Coefficient (MCC) (Matthews, 1975; Equation 1) scores from the RF models were compared to determine which covariate combination produced the best scores.

$$\mathrm{MCC} = \frac{\frac{\mathrm{TP}}{N} - S \times P}{\sqrt{\mathrm{PS}(1-S)(1-P)}} \tag{1}$$

where

$$N = \mathrm{TN} + \mathrm{TP} + \mathrm{FN} + \mathrm{FP}$$

$$S = \frac{\mathrm{TP} + \mathrm{FN}}{N}$$

$$P = \frac{\mathrm{TP} + \mathrm{FP}}{N}$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

MCC was chosen to evaluate covariate combinations over simple and balanced accuracy because MCC is more robust to imbalances between wet and dry classification, by accounting for correct classifications of both classes, whereas even balanced accuracy may show inflated performance if predictions are better for a given class (Chicco et al., 2021; Chicco & Jurman, 2020). MCC values range from −1 to 1. MCC of −1 indicates a model that is always wrong; MCC of 0 is correct in half of instances and equivalent to random chance, and MCC of 1 is correct in all instances. Landis and Koch (1977) offer an interpretation of Cohen's Kappa (Cohen, 1960) that can be applied to MCC scores and breaks out the ranges as: <0 “poor,” 0–0.2 “Slight,” 0.21–0.40 “Fair,” 0.41–0.60 “Moderate,” 0.61 to 0.80 “Substantial,” and 0.81 to 1.0 “Almost Perfect.”

MCC scores were estimated for each RF model using spatial cross validations to characterize spatially distal predictive performance and 20 repeats of 5-fold random cross validation to characterize spatially proximal predictive performance. If there was not a spatial aspect to covariate selection, we expected that MCC scores from the distal and proximal evaluations would be similar for each covariate selection method.

2.4.4. Model Development

We developed streamflow permanence prediction models using three different algorithms, LR, RF, and XBG (Figure 2). The three different model development methods were selected because each algorithm has previously been used in hydrological applications, including streamflow permanence (LR: Jensen et al., 2018; Kaplan et al., 2022; RF: Hammond et al., 2021; Messager et al., 2021; Yu et al., 2019; SGB: Sahour et al., 2021; LR, RF, XGB: Papacharalampous et al., 2023; Zounemat-Kermani et al., 2021). Further, each of these methods has different strengths for parameterization and hyper-parameterization complexity, and variance versus bias tradeoff. For example, decision tree models tend to minimize bias at the expense of increased variance whereas

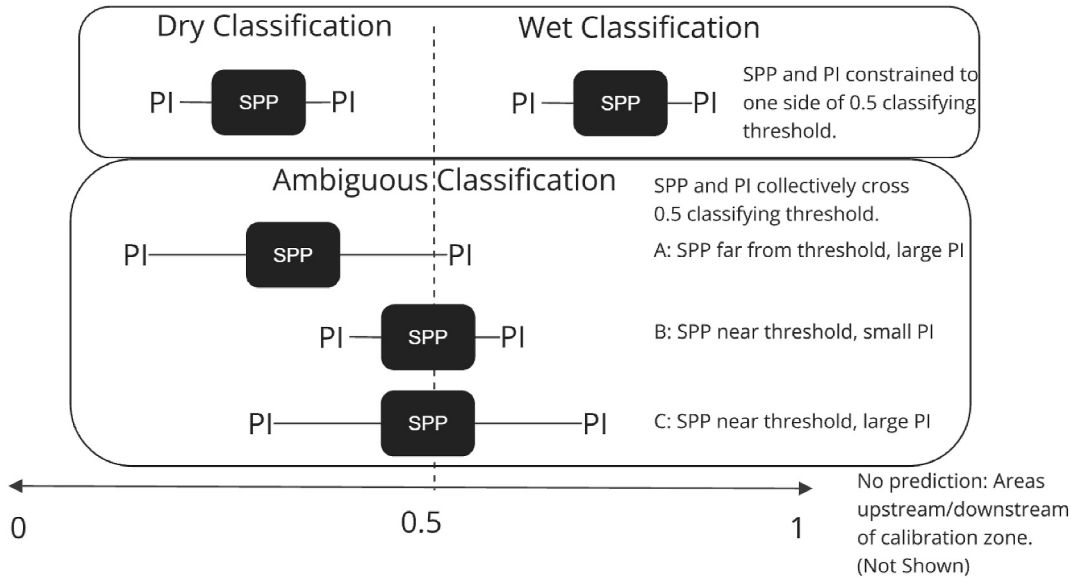

Figure 3. Schematic of streamflow permanence classes in WOWTDR model based on proximity of streamflow permanence probability (SPP) and associated prediction intervals (PI) to 0.5 threshold that determines Wet/Dry class membership and Ambiguous and No prediction classes.

LR models tend to minimize variance at the expense of increased bias (refer to Text S5 in Supporting Information S1 for details). All three algorithms included the random cross validation framework for evaluating predictive performance in areas spatially proximal to the training data and spatial cross validation to evaluate predictive performance in areas spatially distal from the training data (Figure 2). A nested cross validation routine was part of each of the proximal and distal cross validation routines for both RF and XBG to tune selected hyperparameters within these models (Figure S5.1 in Supporting Information S1). Nested cross validation was not necessary in LR because there are no hyperparameters to tune.

For RF, we used a high number of trees (2000), set the resampling to bootstrap, tuned the "mtry" parameter (the number of variables to consider at each decision tree node split) using a nested loop, and used the default hyperparameter of 10 for minimum node size (S 5). For XGB, we tuned nine hyperparameters (Table S5.1 in Supporting Information S1). Hyperparameter tuning methodology was the same for both proximal and distal validation strategies.

Preliminary analyses were conducted to examine interactions, autocorrelation coefficients, and mixed effects for LR compared to RF. Prediction accuracy of LR were low compared to simple RF. In addition, including interactions, autocorrelation coefficients, and mixed effects was computationally expensive, taking days to run and resulting in large covariance matrices. Therefore, these components were not included in the LR algorithm given the tradeoff between model performance and computational cost; final selected covariates were included additively.

Performance tests included MCC, Negative Predictive Value (NPV), Positive Predictive Value (PPV), Accuracy, Receiver Operator Characteristic Area Under the Curve (AUC), and Precision-Recall Area Under the Curve (PRAUC) (Figure 2). The optimal covariate combination for fitting the final model was selected based on the highest median MCC calculated from the 100 scores produced by the repeat cross-validation routine. MCC was chosen because it accounts for classification accuracy and provides a measure of sensitivity and specificity (Chicco & Jurman, 2020; Chicco et al., 2021).

2.4.5. Model Prediction

The final model, which we term the Western Oregon WeT DRy (WOWTDR) model, was then applied to each stream point representing a 5-m sub-reach of stream in 426 HUC12 sub-watersheds. We calculated an estimated probability of late summer surface flow, hereafter termed probability of streamflow permanence (SPP), using

covariates for each year 2019, 2020, and 2021, then averaged SPP across years for a mean SPP. Values of SPP ranged from 0% to 100% with a threshold of 50% determining membership of Dry or Wet classes (Figure 3). Dry classification corresponded to probabilities with their associated 95% prediction interval being less than 50%, which are considered to be sub-reaches that are likely to go dry in late summer. Wet classification corresponds to probabilities and their associated 95% prediction interval being greater than 50%, which are considered to be sub-reaches that are likely to have surface flow during late summer. Upper and lower 95% prediction intervals (PI) were computed by multiplying the standard error (SE) output from the WOWTDR model by 1.96 and adding and subtracting this value from the estimated probability, respectively (Altman & Bland, 2005).

Two additional classifications, "Ambiguous" and "No Prediction," were included as separate classifications in addition to wet and dry classes. An Ambiguous classification corresponded to the cases where the prediction interval around the probability of late summer streamflow crossed the 50% threshold (Figure 3). We interpret these Ambiguous classifications as suggesting that the model did not have enough information to make a definitive class determination between Wet and Dry. "No Prediction" corresponded to locations within the stream network with drainage areas that were substantially smaller or larger than the observational data set, including the extreme upper reaches of the stream network, or downstream portions of the network where rivers are large, which in this environment, in almost all cases, have late summer surface flow. Drainage areas ranged from 0.0008 to 52.3 km^2; however prediction was limited to the range between the 1st and 99th percentile, 0.008 and 7.7 km^2, respectively, to prevent prediction marginal conditions pertaining to the most influential variable in the model. All locations outside of this range were categorized as "No Prediction."

2.4.6. Model and Prediction Analysis

This section details subsequent analyses of the final WOWTDR model including variable importance, WOWTDR accuracies between sub-watersheds with proximal and distal data, identifying where the model is extrapolating, and summarizing predictions across the modeling domain.

2.4.6.1. Variable Importance

Variable Importance scores were estimated using the DALEX package (Biecek, 2018) using a post-hoc permutation approach to evaluating how the MCC of the final model is reduced when each model covariate was removed. Partial Dependence Profiles (PDPs) were also produced using DALEX. PDPs were estimated by evaluating how model probability of Wet changes across the full range of a given covariate, when all other covariates were held constant.

2.4.6.2. Model Accuracy for Sub-Watersheds With and Without Training Data

We evaluated the predictive performance of WOWTDR by comparing classification accuracy for HUC12 sub-watersheds that contain model calibration data (proximal) and those not containing calibration data (distal) (Figure 2).

2.4.6.3. Extrapolation

Model extrapolation is a condition by which a model is used to infer or estimate a condition that is beyond the bounds of the information used to inform the model. This is particularly a concern for models using RF algorithms because these decision tree algorithms do not produce mathematical relations to response data that would facilitate extrapolation beyond the bounds of the input data. However, identifying extrapolated sub-reaches can be useful to understand which locations are well represented or not well represented by the model, providing users an additional level of confidence in model prediction. Also, identification of sub-watershed with extrapolated model predictions can inform future data collection efforts in support of model refinement and improvement. Therefore, to understand where along the stream network WOWTDR was likely to be extrapolating beyond the bounds of the available training data, a single-class support vector machine (SVM) model was fit to training data using the Kernlab package in R (Karatzoglou et al., 2004) (refer to Text S6 in Supporting Information S1 for details). Single-class SVMs work by creating a multi-dimensional convex hull (i.e., envelope) around the data, where the number of dimensions is equal to the number of different covariates in WOWTDR. A given combination of covariate data at a given point along the stream are then fed to the fitted SVM, and the model determines whether the data reside inside or outside the convex hull. Those data determined to be outside of the convex hull are areas

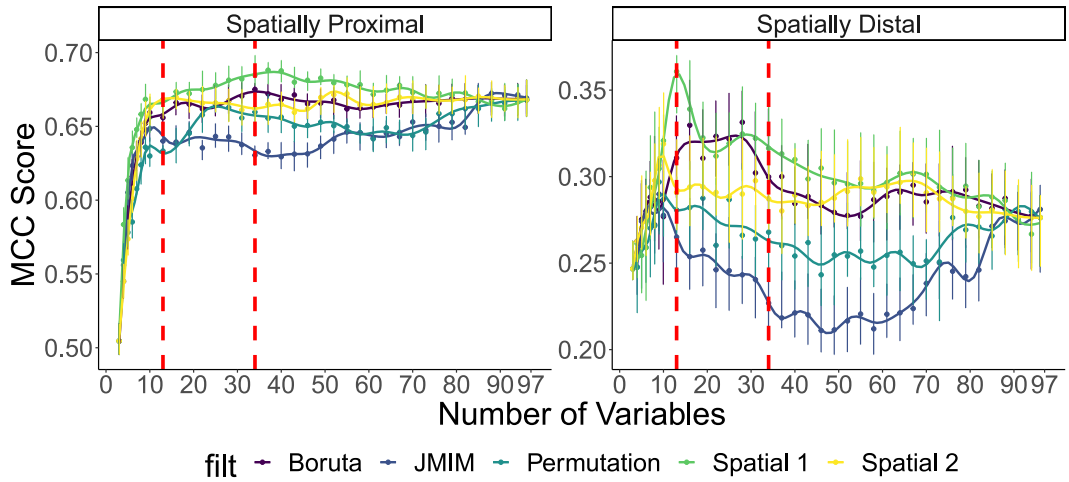

Figure 4. Matthews Correlation Coefficient (MCC) scores with associated 95% confidence interval bars for Random Forest models iteratively fit to increasingly more covariates by covariate selection method where the covariate selection method dictates ranked order of the 96 possible covariates and include manual reordering of DA, P_Annual, and Proportion Canopy. The red dashed lines are annotations of covariate combination resulting in the highest overall MCC across both distal validation strategy and applied to the proximal validation strategy.

where the WOWTDR model is likely to be extrapolating. We report locations where the model is extrapolating for the three streamflow classes: Wet, Dry, Ambiguous, in addition to locations identified as No Prediction. No Prediction locations occur when there are insufficient data to make a prediction, the stream is a fifth Strahler order stream or larger, or the stream is outside of the 1st and 99th percentile drainage area range described previously.

2.4.6.4. Prediction Summaries

WOWTDR prediction results were summarized by ecoregion, lithologic province, and land ownership (State of Oregon, 2016). Land jurisdictions include Bureau of Land Management (BLM), U.S. Forest Service (USFS), additional federal agencies (e.g., National Park Service, U.S. Fish and Wildlife Service), state, private industrial, private non-industrial, and other (e.g., Tribal, municipal, and other lands not captured by jurisdictions).

3. Results

3.1. Optimal Covariate Selection

The five different covariate selection methods showed consistent increases in MCC scores for the proximal validation strategy until the number of covariates used in the model reached 10 (Figure 4). After 10 covariates, Boruta, Spatial −1, and Spatial-2 MCC scores increased more slowly with the inclusion of more covariates. Note that these results incorporate the manual reordering forcing the inclusion of DA, P_Annual, and Proportion Canopy. JMIM and Permutation MCC scores decreased temporarily before a slow increase with Spatial-1 having the highest MCC scores. MCC scores were smaller, and associated 95% confidence intervals were notably larger, for spatially distal covariate selection approaches (Figure 4). Smaller MCC scores in the distal validation strategy were expected given that an absence of observations in the watersheds yielded notably difference orders of importance for the 96 covariates (Table S7.1 in Supporting Information S1), although some variables were consistently identified as important (Table 1). Three covariates (drainage area (DA), total annual precipitation (P_Annual), and proportion canopy cover)) were manually reranked to the top of each covariate selection approach from their original position (Table S7.1 in Supporting Information S1). Eight consistent covariates appeared three, and in one case, four times in the top 13 covariates across the five selection methods. Of these eight consistent covariates, three were representations of precipitation or vapor pressure, three were representations of elevation, and the remaining two were representations of channel length and water storage. Fifteen covariates appeared in only one selection approach. Based on MCC scores, the covariates from Spatial-1 were selected for the final model development. Because the Spatial-1 approach explicitly accounts for spatial dependencies, it is not surprising that this approach performed the best.

Table 1
Covariates Included as the Top 13 for Each of the Five Covariate Selection Methods, Boruta, Permutation, Joint Mutual Information Maximization (JMIM), Spatial-1, and Spatial-2

Covariate	Times of occurrence across five covariate selection approaches	Covariate selection type
DA	5	Spatial-1 (7)
P_Annual	5	Spatial-1 (19)
Proportion Canopy	5	Spatial-1 (88)
WaterStorageCapacity	4	Spatial-1
Elev	3	Spatial-1
Elev_Normalized	3	Spatial-1
Length_Channel	3	Spatial-1
Hydraulic Conductivity	2	Spatial-1
Tmin_Aug	2	Spatial-1
Downstream Channel Slope_50m	1	Spatial-1
Downstream Channel Slope_1km	1	Spatial-1
Length_Hillslope	1	Spatial-1
Tmin_May_previousyear	1	Spatial-1
Elev_Rescaled	3	
P_NormalDifferential	3	
P_annual_previous2years	3	
VaporPressureDeficitAugustMax	3	
P_August	2	
P_August_previousyear	2	
Tmin_May	2	
Curvature_profile	1	
Downstream Channel Slope_100m	1	
Lithologic Province	1	
Local Slope Position relative to Intermediate	1	
TPI_500_m	1	
P_August_normal	1	
P_May	1	
Tmax_August_previousyear	1	
Tmax_May	1	
Tmin_August_previousyear	1	
VaporPressureDeficit_AugustMax_previousyear	1	

Note. Full table ranking for all 96 possible covariates is provided in supplement (Table S7.1 in Supporting Information S1). No fill cells indicate topographic covariates, light gray cells indicate climate covariates, dark gray cells indicate land cover, soil, and storage capacity. The number in parenthesis indicates rank for drainage area (DA), annual precipitation (P_Annual), and Proportion Canopy Cover (Proportion Canopy) prior to reordering for the Spatial-1 selection approach.

3.2. Model Selection

Final model fits for the RF, LR, and XGB algorithms and the 13 covariates identified by the Spatial-1 covariate selection indicated that RF and XGB resulted in the highest median MCC score for both spatially proximal (0.58) and spatially distal (0.26 and 0.32, respectively) evaluations (Table 2). LR resulted in the lowest proximal (0.24) and distal (0.20) median MCC scores, respectively. An evaluation of the LR model's standardized beta coefficients suggest the influence from introducing the three covariates was minor overall given that two of the three covariates, DA and P_Annual, were relatively low for importance (fourth and sixth, respectively). To facilitate easier interpretation and utility for land and water managers, there was a strong preference for a single global model. RF was selected as the final model because although the median MCC scores were the same for RF and

Table 2
MCC Scores for the Spatially Proximal and Spatially Distal Evaluations of the Random Forest (RF), Logistic Regression (LR) and Extreme Gradient Boosting (XGB) Models Fit to the Optimal Covariate Set

Predictive performance evaluation type	Covariates	Model	Median MCC	Lower bound	Upper bound
Spatially Proximal	13	RF	0.58	0.58	0.59
	13	LR	0.24	0.24	0.25
	13	XGB	0.58	0.57	0.59
Spatially Distal	13	RF	0.26	0.24	0.27
	13	LR	0.20	0.17	0.22
	13	XGB	0.32	0.31	0.33

XGB under the proximal validation strategy, Ranger implementation of RF included estimates of standard error of prediction, which provided additional interpretative value to end users. Here, standard error provided by RF was prioritized over performance accuracy in the distal validation strategy, which was higher for XGB.

3.3. WOWTDR Model Performance and Covariate Importance

The final WOWTDR model is a 13-covariate streamflow permanence prediction model that used the RF algorithm fit to FLOwPER observations to train the model (Table 1). WOWTDR model included three precipitation and temperature covariates, nine hydro-topographic covariates that describe the location within the channel network, and a static land cover covariate of percent canopy cover (Table 1). Distribution of values for sub-reaches within the modeling domain is generally well represented by the training data for these 13 covariates although training data has slightly greater representation of higher elevation sites with lower minimum temperature relative to the prediction locations (Figure S7.1 in Supporting Information S1). Total precipitation for the current year of the observation, drainage area, and upstream channel length were the three most important covariates for determining the likelihood of late summer streamflow (Figure 5).

The partial dependence profiles depict generally expected relations between the likelihood of streamflow permanence and individual covariates (Figure 5). As expected, streamflow permanence probability generally increased with increasing total annual precipitation. Probabilities increased sharply at approximately 1,300 mm total annual precipitation and continued to increase until approximately 3,200 mm. Probability values decreased once annual precipitation exceeded 3,200 mm, but this decrease was influenced by a limited number of observations in the training data at high precipitation levels, including some dry stream observations in headwaters that had very small drainage areas and at elevations of less than 1,000 m above sea level (msl, Figure S7.2 in Supporting Information S1). As expected, probabilities of streamflow permanence increased with increases in drainage area and upstream channel length (Figure 5). Also as expected, probabilities of wet sub-reaches decreased with steeper channel slopes, when calculated for short distances of 50 m, but remained level, and varied little with channel slope over longer distances (1 km). We expected streamflow permanence probabilities to increase with increased water storage capacity, and this was the case for water storage capacities between 20 and 35 mm. However, probability of permanence initially decreased with water storage capacity between the values of 0 and 10, and varied little from 10 to 20. Minimum August temperature, hydraulic conductivity, normalized elevation, and minimum May temperature from the previous year all had generally similar magnitudes of covariate importance (Figure 5). The probability of permanence generally decreased when these covariates increased, except for minimum May temperature of the previous year. Probabilities markedly increased as minimum May temperature of the previous year exceeded 5°C. Similar to normalized elevation, probability of permanence decreased with higher elevation, although this relationship was driven by a limited number of observations in the training data for elevations higher than 1,600 m (Figure 5, Figure S7.2 in Supporting Information S1). Probabilities remained relatively level with increases in proportion canopy cover and hillslope length, except for a slight increase in probability of permanence when canopy cover exceeded 0.75 and when hillslope length increased from 0 to 0.25. Hillslope length was calculated as 0.5/drainage density and represented the flow path length prior to reaching the channel (Tucker et al., 2001). We expected that short hillslope lengths would have a rapid response to hydrologic inputs and would substantially contribute to channel

10.1029/2025WR040478

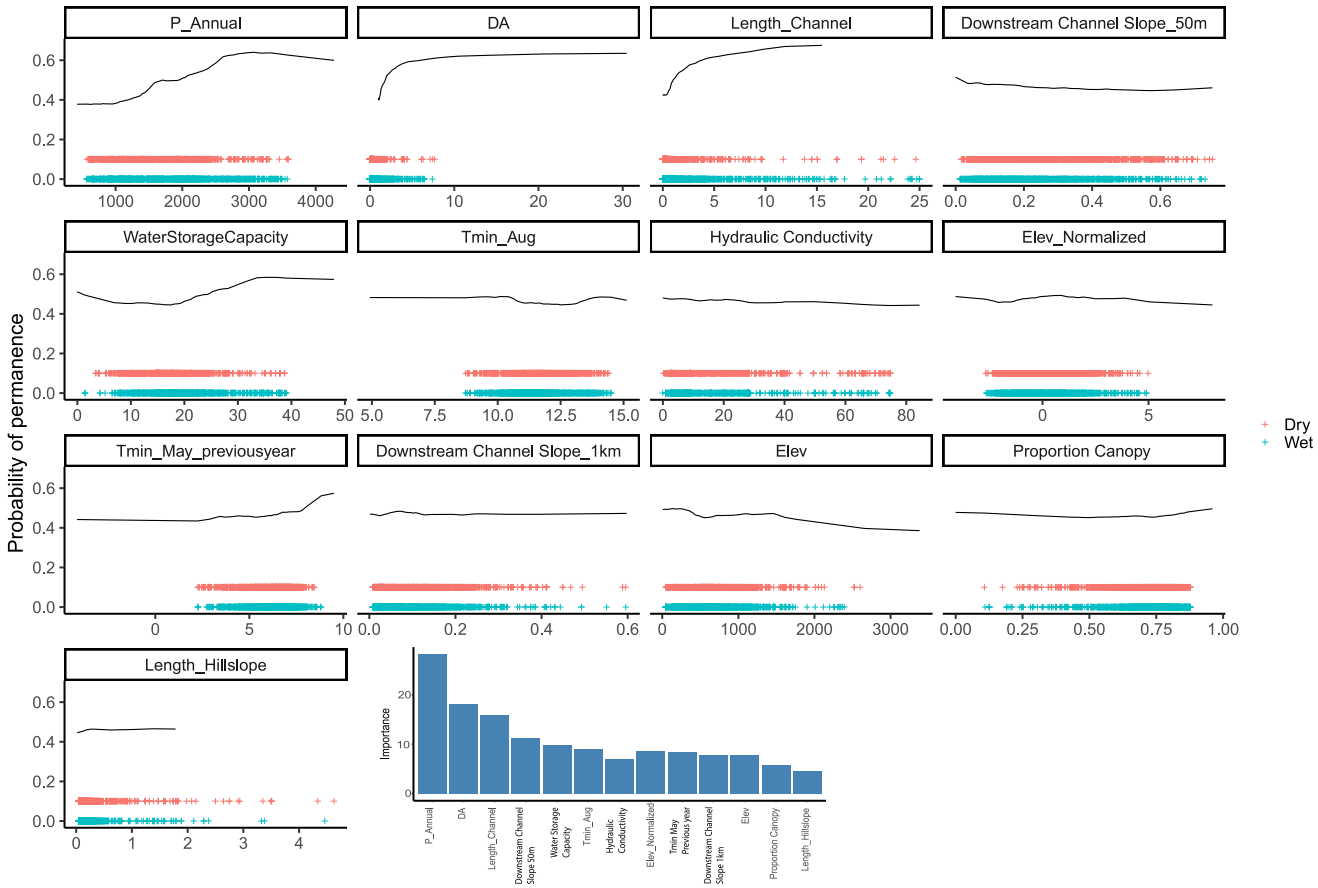

Figure 5. Partial dependence profiles for all covariates in the WOWTDR model are presented in order of descending relative importance. Relative importance score for the given covariate is depicted inside of the associated profile plot. *X* axis values and ranges are dependent on range of the data for the respective covariate. *Y* axis is the likelihood of late-summer streamflow being Wet, with 0.0 being dry and 0.5 being the decision boundary between the Wet and Dry. Relative importance score for the given covariate is displayed in the bar chart.

flow draining small contributing areas (D'Odorico & Rigon, 2003). However, we acknowledge this expectation may vary across the different lithologic provinces (Hale & McDonnell, 2016).

3.4. WOWTDR Proximal Versus Distal Comparison

We evaluated the accuracy of WOWTDR for HUC12s where field observations existed and where they did not exist, termed proximal and distal respectively (Figure 6). In addition, we were also interested in evaluating how data that resulted in Ambiguous classification influenced predictive performance. Therefore, we evaluated performance metrics (a) considering all the data, (b) excluding data that resulted in Ambiguous predictions, and (c) considering only data that resulted in Ambiguous predictions (Figure 6). Evaluation of predictive performance excluding Ambiguous data was conducted to provide the end user greater confidence in Wet and Dry classifications.

As expected, performance metrics were higher using proximal validation strategies compared to distal validation strategies (Figure 6). Including all the data in the proximal evaluation, accuracy was 0.83 with a MCC of 0.67 with similar negative and positive predictive values (0.82 vs. 0.84). Negative predictive value indicates the likelihood that a Dry prediction is truly dry and positive predictive value indicates the likelihood that a Wet prediction is truly wet. Accuracy decreased to 0.67 and MCC markedly decreased to 0.34 using the distal validation strategy with corresponding decreases in negative and positive prediction values (0.67, 0.66). Performance metrics increased when data that resulted in Ambiguous classifications were excluded (Figure 6). Accuracy and MCC metrics were 0.90 and 0.80, respectively for the proximal condition, indicating substantial agreement to validation data, and 0.71 and 0.43 for the distal condition, which suggested moderate agreement to validation data. Finally,

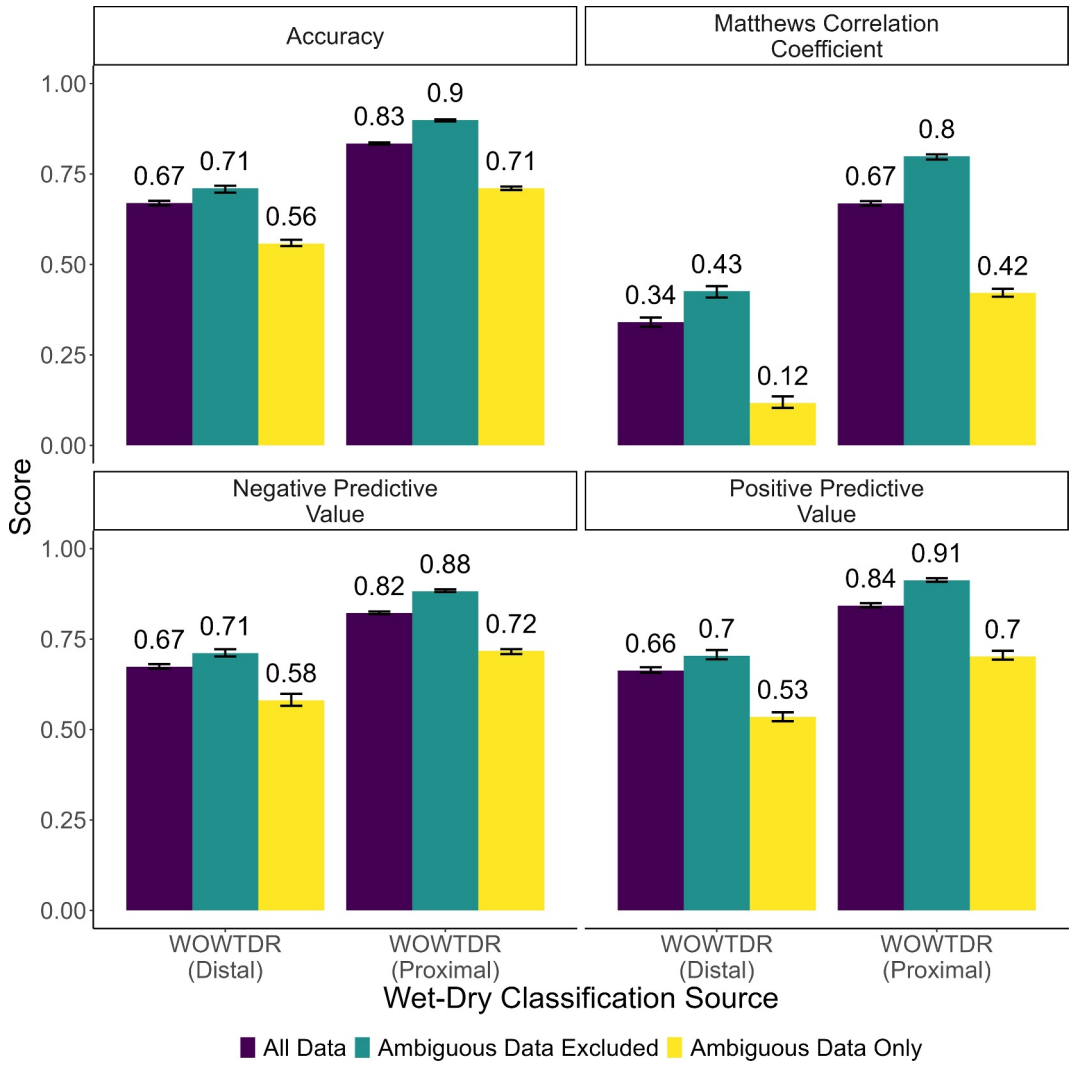

Figure 6. Comparison of four model performance metrics, (a) Accuracy, (b) Matthews Correlation Coefficient, and (c) Negative and (d) Positive Predictive value) for the proximal and distal conditions of WOWTDR predictions while also considering the influence of ambiguous predictions. Negative Predictive Value indicates the likelihood that a Dry prediction is truly dry and Positive Predictive Value indicates the likelihood that a Wet prediction is truly wet. Bars are 95% confidence intervals using the percentile bootstrapping method.

performance metrics were lowest when considering only Ambiguous data, although surprisingly, metrics were higher than 50% for accuracy (0.71) and >0 for MCC (0.42), which indicated that classification accuracy was better than a random coin flip, suggesting moderate prediction strength. Confidence intervals, using the percentile bootstrapping method (refer to Text S3 in Supporting Information S1 for details), were generally small (<0.02) when considering all the data, excluding Ambiguous data, and considering just the Ambiguous data (Figure 6). Confidence intervals were slightly larger for the distal conditions compared to the proximal conditions.

3.5. Extrapolation Detection Model

The SVM model was chosen over the isolation forest (IF) method of detecting extrapolations at any given subreach because it had 99.7% accuracy on hold-out test data and 94.8% on simulated outlier data (Table S6.2 in Supporting Information S1) and achieved better accuracy than IF in 3 of 4 tests and was within 1% of IF on the test 3 (refer to Text S6 in Supporting Information S1 for details).

WOWTDR extrapolated for 23% of sub-reaches in the modeling domain that includes Strahler order one through four (Figure 7). Surprisingly, Ambiguous classifications only accounted for 16% of extrapolated sub-reaches

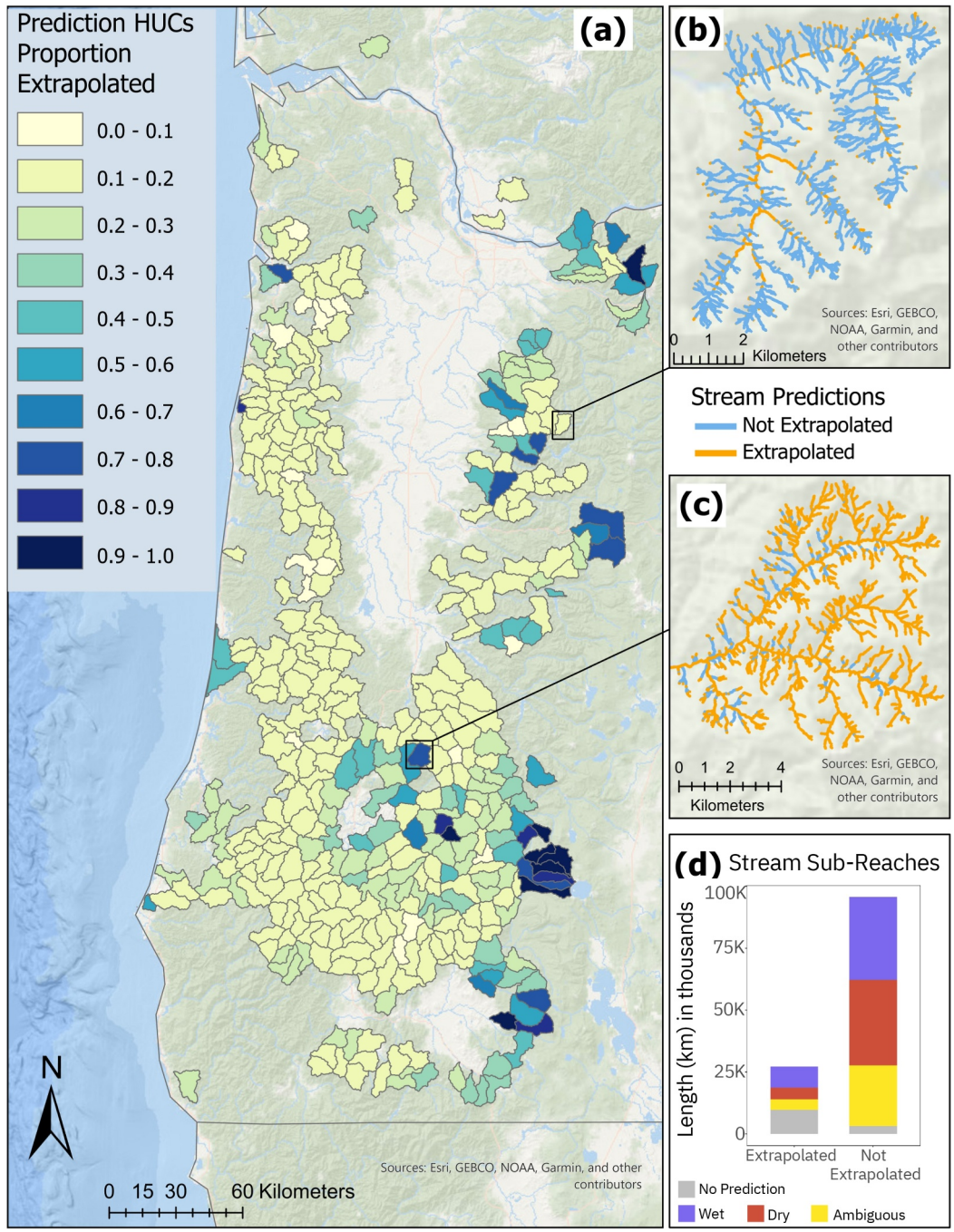

Figure 7. Proportion of stream sub-reaches extrapolated by HUC12 (a), distribution of extrapolated stream sub-reaches in a HUC12 with low proportion of extrapolation (b) and high proportion of extrapolation (c), distribution among the four predictions for the 426 sub-watersheds in the prediction domain (d). Base map from Esri and its licensors, copyright 2022.

indicating that Ambiguous classifications should not necessarily be equated with extrapolation. Consequently, extrapolation detection can be used to identify locations within the model domain that are outside the range of and therefore not represented by the training data. Approximately 31% of extrapolated predictions were classified as Wet and were located in the downstream portions of the network where sub-reaches generally have reliable late summer streamflow but are still upstream of the no prediction zone.

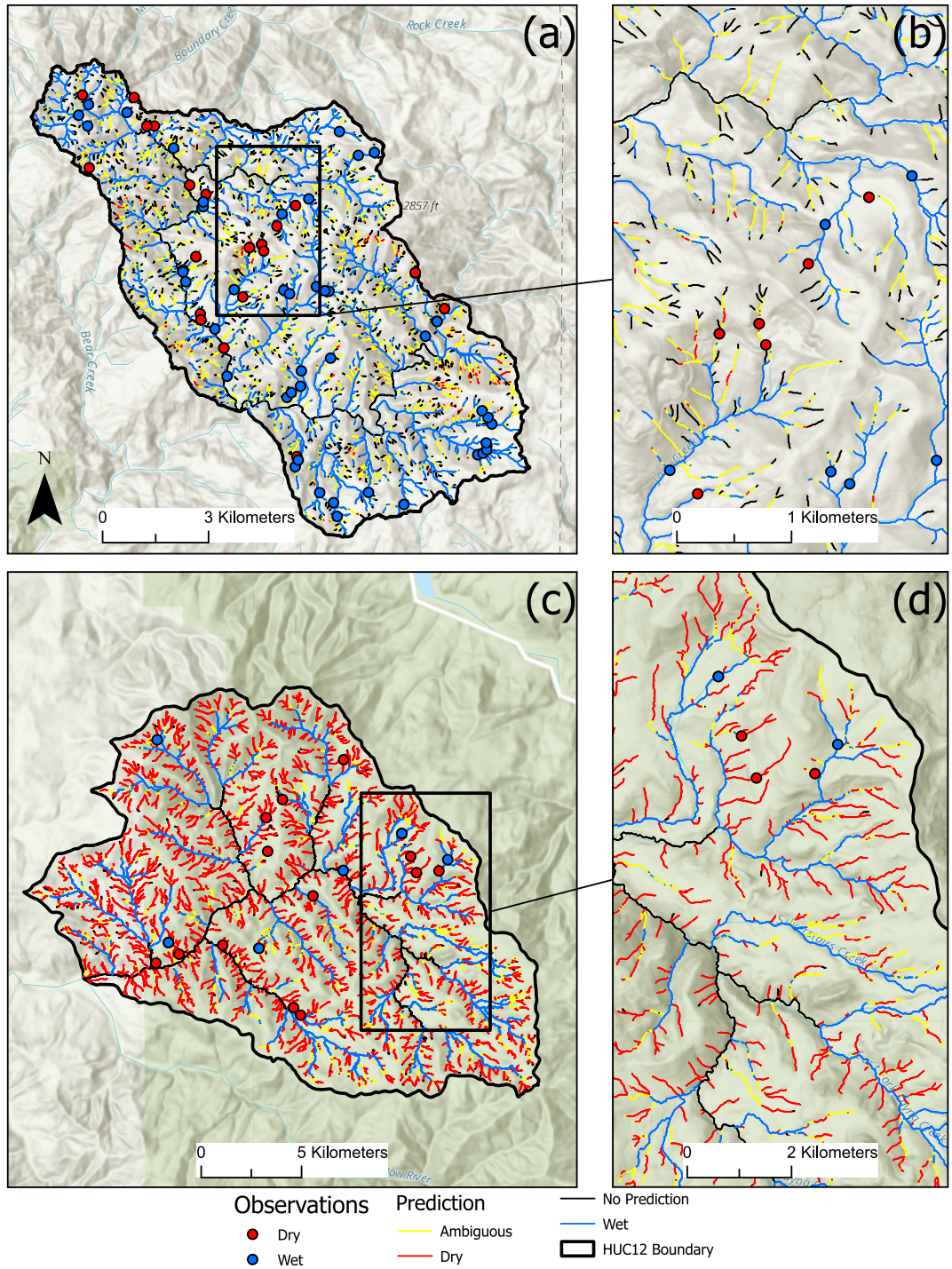

Figure 8. WOWTDR predictions of streamflow permanence classifications of Dry, Wet, Ambiguous, and No Prediction for two representative HUC12 sub-watersheds and observations (a, c). No prediction zones at both upstream and downstream extents of flow network with general downstream sequence of Dry, Ambiguous, and Wet classifications (b, d). Base map from Esri and its licensors, copyright 2022.

3.6. Prediction Summaries

Overall, as expected, classifications transitioned from Dry in the uppermost headwaters to Wet in a downstream direction and were generally consistent across the 3 years (Figure 8, Figure S7.3 in Supporting Information S1).

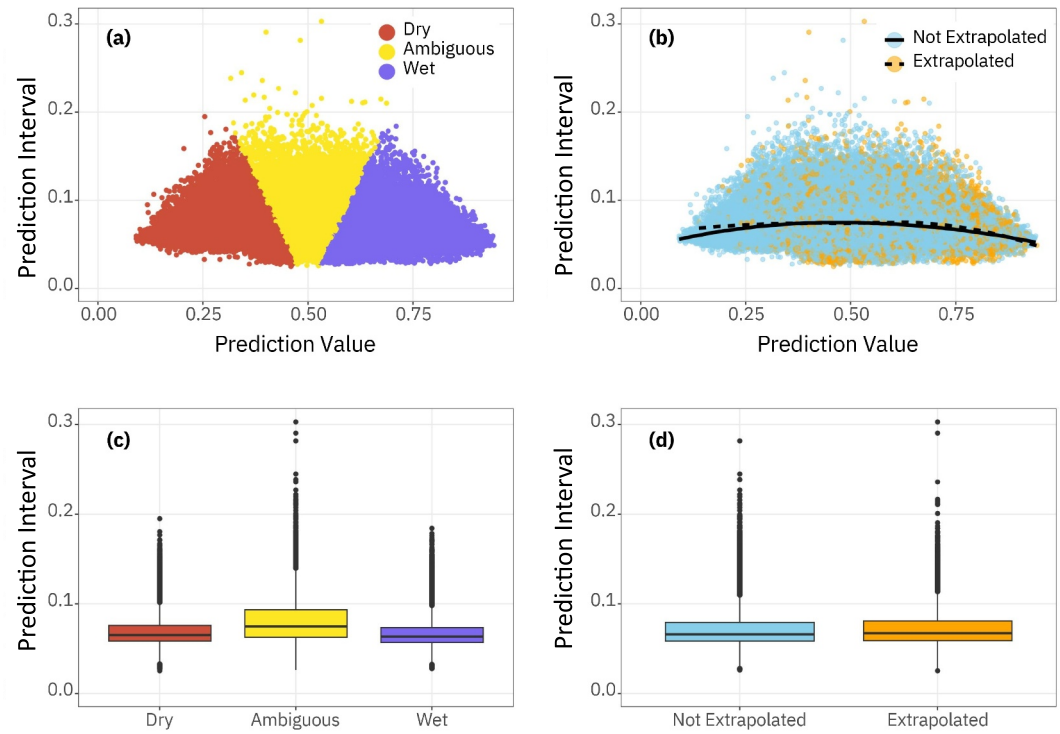

Figure 9. Distribution of probability of permanence and associated 95% Prediction Interval by prediction class of Wet, Dry, and Ambiguous (a) and by extrapolated and not extrapolated sub-reaches(b). Boxplots of distribution of prediction interval by prediction class (c) and extrapolated and not extrapolated sub-reaches (d). Box center lines indicate the median, box edges are the 25th and 75th percentile; lines extend to the 5th and 95th percentiles.

Ambiguous classifications tended to correspond to transition zones between Wet and Dry classifications. There was only slightly higher frequency of both Wet and Dry classifications occurring in 2019 compared to 2020 and 2021 and slightly more Ambiguous classifications in 2020 (Figure S7.3 in Supporting Information S1).

The prediction interval was generally similar for Wet and Dry prediction classes, but was larger and more variable within sub-reaches classified as Ambiguous (Figure 9). Most prediction intervals in sub-reaches with an Ambiguous streamflow permanence class were <0.1; therefore, most of the Ambiguous predictions were a result of model estimates near the 0.5 decision threshold between Wet and Dry as opposed to an estimate far from the threshold but with larger intervals (Figure 3).

The prediction intervals for sub-reaches both within and outside of the extrapolation zones had very similar prediction interval distributions (Figure 9b) such that prediction interval is not a useful identifier of an extrapolated location. Instead, the lack of difference in prediction interval magnitudes between extrapolated and non-extrapolated sub-reaches underscores the need for a separate extrapolation identification model.

Our evaluation of WOWTDR predictions across the 116,494 km of mapped sub-reaches indicated differing levels of Wet, Dry, Ambiguous, and No Prediction classifications by Strahler Order, lithology, ecoregion, and land jurisdictional (Figure 10).

The starkest patterns in streamflow status classification were across Strahler stream orders (Figure 10a) with a progressive shift to a greater proportion of sub-reaches classified as Wet as stream order increased from first- to third-order. Strahler first-order streams are dominated by either Dry (0.5) or Ambiguous (0.28) classes (Figure 10a). Sub-reaches that are classed as No Prediction in first-order streams represent the farthest upstream sub-reaches of the channel network. The proportion of dry sub-reaches decreased for second-order streams to 0.21 as the proportion of both Ambiguous and Wet classes increased (0.32 and 0.46, respectively). Third-order streams were dominated by Wet classifications (0.86) with a small proportion of No Prediction sub-reaches (0.03). The proportion of No Predictions increased to 0.4 for fourth-order streams. All sub-reaches in fifth-order and larger

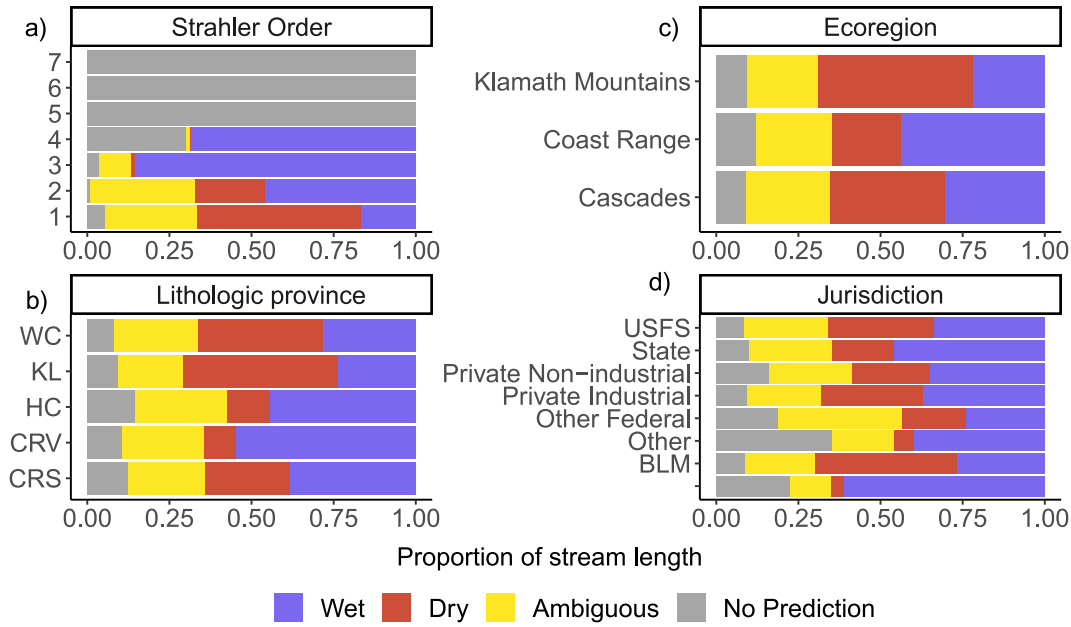

Figure 10. WOWTDR classification prediction across 116,494 km of light detection and ranging stream network smaller than Strahler fifth-order within the modeling domain by (a) Strahler Order, (b) lithologic province, (c) Environmental Protection Agency Level 3 Ecoregion classification, and (d) land jurisdiction. The *y*-axis is the proportion of the total stream length where the relative contributions of the four possible predictions classes are stacked to a total of one for each given category.

were classified as No Prediction. The increase of Ambiguous class and reduction of Dry class in second-order sub-reaches suggests that second- order is where streams in western Oregon transition from Wet to Dry on the LiDAR-derived hydrography used in this study.

Dry classifications were proportionally higher than other classifications for the Coast Range sedimentary rocks (0.4), Klamath (0.53), and Western Cascades (0.45) lithologic provinces (Figure 10b). Conversely, the High Cascades and Coast Range volcanic rocks and Columbia River basalt lithologic provinces had a relatively high proportion of wet classifications (0.32 and 0.30, respectively). Both the High Cascades and Coast Range volcanic rocks and Columbia River basalt lithologic provinces had a relatively small number of observations with which to train the model (213 and 20, respectively) compared to other lithologic provinces.

The distribution of stream status classifications by ecoregion was similar to the distribution by lithologic province (Figure 10c). In general, the distribution of classifications was approximately similar for the Coast Range, Cascades, and Klamath Mountains ecoregions. However, the Coast Range ecoregion had a larger relative proportion of streams classified as Wet and Klamath Mountains ecoregion had a higher relative proportion of streams classified as Dry.

Across our study region, BLM lands contained the largest proportion of Dry sub-reaches (0.53) followed by U.S. Forest Service (0.4) and Private Industrial land (0.4), which was largely composed of private forests (Figure 10d).

4. Discussion

The modeling approach used in this study includes a robust consideration of spatial balancing and several uncertainty components that include (a) an Ambiguous prediction class, (b) model predictive performance in sub-watersheds containing training data and those that do not, and (c) identification of where the model was extrapolating. We suggest that such an approach is critical to helping managers make informed decisions by providing a clear understanding of the level of uncertainty and where the uncertainty occurs in the landscape and within models (Tonkin et al., 2019).

Existing streamflow permanence models for this region include the $\text{PROSPER}_{\text{PNW}}$ model (Jaeger et al., 2019), a regional model based on a relatively coarse hydrography that does not adequately represent small streams, and

individual catchment scale models that are based on higher resolution hydrography and intensive data collection over smaller spatial extents for catchments of generally less than 64 km^2 (Hafen et al., 2023; Ward et al., 2018, 2020). The WOWTDR model bridges this geographic and spatial resolution gap by providing predictions for western Oregon that covered a geographic extent much larger than small individual catchment studies while leveraging high-fidelity LiDAR-derived hydrography and a framework that allows users to infer key influences of processes that drive flow permanence.

4.1. Merits and Tradeoffs of Intensive Model Development

The model development approach used in this study included numerous processing-intensive components that, when considered in aggregate, are novel inclusions for empirical models. Strategies were implemented to solve modeling challenges while achieving both the highest accuracy and highest understanding of uncertainty possible. Aside from the overarching computational challenge of developing a model that provides relatively fine spatial resolution estimates (5-m sub-reaches) on LiDAR-derived hydrography, over a relatively large geographic extent (426 HUC12 sub-watersheds), intensive processing components were incorporated that each come with tradeoffs between computational cost and model accuracy. Components of our empirical modeling approach included (a) additional components of uncertainty, (b) spatial balancing to account for imbalanced data, (c) intensive covariate selection routines to reduce 96 possible covariates, (d) and testing multiple algorithms. The tradeoffs for each of the four components need to be carefully considered in future modeling efforts.

4.1.1. Incorporating Uncertainty in Model Predictions

In the WOWTDR model, Ambiguous classifications are assigned where the model is unable to determine with certainty that late-summer streamflow status is either Wet or Dry. However, the relatively small size of the mean prediction interval associated with WOWTDR model performance estimate (<0.1, Figure 9) suggests that WOWTDR predictions of Ambiguous occur with a high level of precision (Figure 3). The implication of this result is that additional observations in Ambiguous zones may not necessarily improve WOWTDR predictive performance if the Ambiguous classification is the result of the model receiving conflicting response observations over similar ranges of covariate expression. It is likely that conflicting observations are not in error, but instead are a result of a consequence of local-scale conditions not captured by the predictor covariates. A range of local conditions that influence streamflow permanence have been identified in individual catchment studies including covariates that describe geomorphic planform, cross-sectional area, and subsurface characterization (Bush et al., 2023; Jensen et al., 2018; Ward et al., 2018; Warix et al., 2023), yet fine-scale gridded data sets remain coarse and do not capture on-the-ground conditions (Dohman et al., 2021; Hafen et al., 2023; Kaplan et al., 2020; Prancevic & Kirchner, 2019).

Conversely, Ambiguous classifications may indicate real streamflow permanence conditions that are in an intermediate state between wet and dry where, definitive surface flow conditions are not evident, but instead, may include some degree of discontinuous surface water presence or near surface water presence (Gallart et al., 2012). The duration of this intermediate flow data could be extremely short or sustained throughout the late summer streamflow period (Costigan et al., 2017). Although the FLOwPER mobile app includes a streamflow status option of "discontinuous flow" to approximately demarcate the intermediate state of "continuous flow" and "dry," there were very few occurrences of these observations, which were ultimately excluded from the training data set. More observational data of this intermediate flow state is needed to distinguish between Ambiguous classification more definitively in the WOWTDR model that represent this intermediate flow state or represent a limitation in the model to distinguish between Wet and Dry. Ambiguous predictions are less frequent in occurrence than Wet or Dry predictions and can help land managers prioritize field observations to a smaller portion of the stream network.

Similarly, our estimation of areas of extrapolation, where WOWTDR provides predictions beyond the watersheds for which there was training data, is an important component of this work. There is a risk that model predictions that extend to areas beyond training data can result in meaningless predictions if covariates are substantially different than the training data (Meyer & Pebesma, 2021). The distribution of our training data and that of the larger modeling domain indicate that training data generally represented the distribution of covariates in the modeling domain well (Figure S7.1 in Supporting Information S1). Nevertheless, predictions in extrapolation

zones should be considered carefully and identification of these areas can help prioritize additional field observation campaigns for research as well as management planning.

4.1.2. Consideration of Spatial Dependence

Despite thousands of field observations, the spatial distribution of late summer streamflow observations was sparse for the large study area (only 129 out of 426 watersheds had observational data available) with observations often not evenly spatially distributed throughout a given HUC12 watershed. The nature of these data presented two challenges to assessing model performance: (a) the imbalance between wet and dry observations and (b) the spatial autocorrelation inherent to clustered data.

To mitigate these two challenges, we incorporate balancing of wet and dry observations by spatial groups. The even distribution of wet and dry allows the model decision threshold between wet and dry to sit at 0.5 on a scale from 0.0 to 1.0, which allows a straightforward post-hoc classification of Wet, Dry, and Ambiguous. There is a risk of inducing bias by potentially duplicating spurious observation data. However, spatial balancing partially mitigates the risk of biasing the performance estimates by ensuring each spatial group had sufficient observations (at least 73) to accommodate the iterative validation process and evaluating the model performance for proximal and distal to the training data. As a result, data that would be considered spurious with unusually high or low values would have less potential to bias the model. The final number of spatial groups to meet the criteria of sufficient iteration both across and within groups was 20.

Our model performance analysis included a computationally intensive evaluation of predictive performance using a spatial cross-validation strategy, representative of performance distal to training data, and a non-spatial cross-validation strategy, representative of predictive performance proximal to model training data. Roberts et al. (2017) demonstrated that spatial cross-validation tends to produce lower estimates of predictive performance on spatially dependent data when compared to results from a conventional random cross-validation strategy. We acknowledge that the application of spatial cross-validation is not novel to empirical streamflow permanence modeling (refer to Messager et al., 2021; Sando et al., 2022). However, here we explicitly report the results of the distal and the proximal because they exhibit the performance differences between spatial and non-spatial approaches described by Roberts et al. (2017). These differences provide critical insight on the relative performance improvement in sub-watersheds containing training data over those that do not, as well as quantitative information on model generalizability, a feature that was not included in the regional-scale $\text{PROSPER}_{\text{PNW}}$. These insights could be used by land managers who make risk decisions for the purpose of prioritizing resources. For example, overall proximal predictive performance accuracy may be sufficiently high (when Ambiguous is removed) for using WOWTDR predictions for planning purposes. Conversely, for watersheds without training data, median predictive performance that is below a value of interest may result in decision makers determining that field observations are necessary to inform planning. However, our analysis was not able to include estimates of a minimum amount of observation data to achieve a given accuracy. As a different method to identify areas where the model is applicable or not applicable, Meyer and Pebesma (2021) provided a dissimilarity index and a subsequent "area of applicability (AOA)" as a function of covariate importance and distance to training data. They note that the AOA is sensitive to the cross-validation strategy in addition to the sampling design, and more work is needed. Our collective proximal/distal cross-validation strategies and extrapolation detection component is an alternative approach to this concept to inform area of applicability of the WOWTDR model.

4.1.3. Intensive Covariate Selection

Our covariate selection approach required substantial development time and processing, but resulted in a more efficient model than simply using all the available data. Covariate selection showed that including more than 13 covariates did not result in substantial increases in predictive performance metrics for the spatially proximal validation strategy, and for spatially distal performance metrics, there was a marked decline when more than 13 covariates were included (Figure 4). Decreases in performance metrics as more covariates are included past a particular high point is a consequence of sparse training data relative to the number of covariates or dimensions that can be considered in the model. The additional covariates result in overfitting to noisy data and cause reduced performance of average conditions (Theodoridis & Koutroumbas, 2006; Trunk, 1979). However, each of the covariate selection methods resulted in different covariate rankings and predictive performance estimates, which demonstrate there is no one-size-fits-all approach to covariate selection. Given that Spatial-1 and Boruta utilize

variable importance estimates within a repeated resampling routine to choose variables, we expect that these two methods will consistently outperform the other methods considered for predicting categorical wet and dry streamflow classes. However, the Spatial-1 covariate selection approach resulted in the highest performance metrics for both spatially proximal and distal strategies, suggesting that covariate selection approaches that account for spatial dependencies should be considered when selecting variables and evaluating model performance in streamflow permanence classification. Other machine learning modeling have identified the importance of spatial cross-validation to remove spatially correlated covariates that would result in overfitting and limit the model's ability to provide prediction beyond the training locations (Meyer et al., 2019). In our modeling case, Spatial-1 was the preferred covariate selection approach and may be a preferred approach in similar modeling scenarios where there are numerous covariates across a broad range of resolution that must be filtered and where data are sparse, spatially dependent, and highly imbalanced. However, if validating the spatially distal strategy is not a priority, the Boruta covariate selection approach would be acceptable, although the resulting model would not generalize as well to sub-watersheds that do not have training data.

The choice to override the covariate selection and manually re-order drainage area, annual precipitation, and proportion forest canopy cover to ensure their inclusion in the final model was intentional to balance the tradeoff between model performance and the legitimate interests of end users to understand controls on streamflow permanence that can help inform management decisions. Understanding apparent influences of these three covariates can be more easily translated into land management decisions compared to other covariates that may be more difficult to construct management plans around. For example, hillslope length, total wetness index, or vapor pressure deficit may result in more accurate streamflow permanence estimates, but if a goal is to develop management activities related to streamflow permanence, incorporating the influence of these nuanced landscape or climatic characteristics is less straightforward. This tradeoff between taking a purely data driven approach and including these subjective choices for covariate inclusion was a purposeful decision to arrive at a multi-solution objective.

4.1.4. Testing Multiple Model Development Algorithms

We tested multiple model development algorithms to explore best options for the challenging task of delineating wet and dry stream sub-reaches across a region. Many existing flow permanence models have only used one algorithm exclusively without comparison (e.g., RF: Döll et al., 2024; Jaeger et al., 2019; Sando et al., 2022; Yu et al., 2019; LR: Jensen et al., 2018; Kaplan et al., 2022) although Messager et al. (2021) evaluated different RF algorithms. Our results show that RF had a higher mean MCC value than LR, indicating that RF prediction accuracy was higher (Table 2). In contrast, a similar workflow and similar array of covariates in Penaluna et al. (2022) showed that LR more accurately predicted the upper extent of trout in stream networks compared to RF. In that study, LR likely had higher prediction accuracy than RF because the relation between the drainage area covariates and the response variable, presence of fish, was strong and relatively linear. Unsurprisingly, XGB performed better than RF in our distal evaluation, suggesting marginally less overfitting and broader generalizability than RF. This result is generally consistent with the Sahour et al. (2021) finding that XGB more accurately prediction streamflow compared to RF when using correlations to tree ring sizes. The higher prediction accuracy of XGB likely results from its ability to learn from residuals. However, the cost of XGB's increased prediction accuracy is the need to tune multiple hyperparameters in a computationally intensive process. Furthermore, we were not able to locate an XGB implementation that contained native support for estimating standard errors of prediction. Given the improvement in interpretive value provided by inclusion of standard error of prediction estimates and the relative ease of developing a RF model, the results do not provide enough evidence to suggest a deviation from the use of RF for this application to streamflow permanence.

4.2. Streamflow Permanence in Forested Headwaters of Western Oregon

4.2.1. Patterns in Streamflow Permanence

We recognize that stratifying the landscape by Strahler stream order is scale dependent and that the low order headwater regions encompass both perennial and non-perennial sections of streams (Christensen et al., 2022; Golden et al., 2025). Our modeling results corroborate this assessment such that, for this LiDAR-derived hydrography resolution and in the forested headwaters of western Oregon, the transition between the nonperennial and perennial portions of the network frequently occurs in the second-order streams.

4.2.2. Controls on Streamflow Permanence

Surface flow presence demonstrates the balance between the subsurface capacity to transmit water and time-varying delivery of water from upstream (Durighetto & Botter, 2022; Godsey & Kirchner, 2014; Prancevic & Kirchner, 2019; Ward et al., 2018). Topographic variables including drainage area, slope, and curvature, which govern subsurface transmissivity, and discharge, which describes the upstream delivery (Durighetto & Botter, 2022; Godsey & Kirchner, 2014) have been correlated with where surface flow presence expands and contracts across a given network (Prancevic & Kirchner, 2019). In our study, the inclusion of the four covariates, drainage area, channel length, which is highly correlated (Pearson's correlation coefficient = 0.93) with drainage area and may be considered a surrogate, channel slope, and water storage capacity, as important variables following annual precipitation (Figure 5) is consistent with these prior works and the conceptual framework of morphology and lithology controlling how water moves through the landscape and drives streamflow permanence (Durighetto & Botter, 2022; Leach et al., 2024). Stream sub-reaches with larger drainage area, which serves as a proxy variable for discharge, correspond to reliable, persistent flow (Durighetto & Botter, 2022; Godsey & Kirchner, 2014), and sub-reaches with higher channel slope have shorter and/or more rapid subsurface flow paths resulting in more transport in the subsurface and consequently, smaller streamflow permanence probabilities (Prancevic & Kirchner, 2019). These covariates are interpreted as the best available surrogates for key watershed processes of streamflow permanence, but effective characterization of lithologic influences are lacking even though lithologic influence has been identified to be a key driver of streamflow permanence in other studies (Leach et al., 2024; Moidu et al., 2021). Therefore, although the prevalence of variables that describe morphology and lithology in empirical models supports this perceptual idea that morphology and lithology control streamflow permanence, the capacity to identify covariates across a region that accurately influence streamflow permanence at the scale of a 5-m reach remains limited.

Climate may not necessarily control the spatial dynamics of where wet and dry reaches occur in an individual catchment (Durighetto & Botter, 2022), but can influence streamflow volume year-to-year and across broader geographical extents (Botter et al., 2021; Hammond et al., 2021; Messager et al., 2021; Ward et al., 2020). The WOWTDR modeling domain was large enough geographically that precipitation differed substantially among watersheds; therefore, it was not surprising that precipitation was the most important predictor (Figure 5), which is consistent with models at larger regional (Hammond et al., 2021; Jaeger et al., 2019; Sando et al., 2022) and global extents (Botter et al., 2021; Messager et al., 2021). We would expect precipitation to be less important if the modeling domain was smaller and less diverse, consistent with other similarly scoped empirical modeling examples (Jaeger et al., 2023; Kaplan et al., 2022).

4.2.3. Limitations

This model relies on LIDAR DEMs and LiDAR-derived hydrography. While the availability of these types of data worldwide are highly variable by location, in the coming decades, these data are expected to be increasingly available across much of the coterminous United States due to ongoing investment by the USGS 3D elevation program (3DEP) and the 3D hydrography program (3DHP) to produce 1-m resolution LiDAR and complementary hydrography (Anderson et al., 2024).

The goal of proximal and distal cross-validation was to provide the end user with an estimate of model performance in areas lacking observation data. However, the spatial grouping approach can substantially alter estimates. Ecoregion and stream gage locations have been used for spatial groupings for rivers in geographically larger domains (e.g., Messager et al., 2021; Sando et al., 2022). However, these factors were not applicable to this study focusing on headwater stream networks in a geographically smaller modeling domain. Instead, we grouped observations by HUC12s as a means of practical convenience. Whether grouping observations this way is physically meaningful is unknown and would require additional study.

The WOWTDR model provides estimates for conditions that are consistent with approximately normal to below-normal annual precipitation conditions and may not represent wetter than average conditions. Repeated collection of training data at sites across years across a range of hydroclimatic conditions would be needed to provide predictions that represent drier and wetter climatic conditions and the potential influence of year-to-year differences in climate, including changing snowpack on streamflow permanence for a given location.

5. Next Steps

Estimating the extent of perennial flow in headwaters streams will remain an ongoing need for land managers and an ongoing challenge for researchers for numerous reasons. Deriving the extent of perennial streams from LiDAR and remote sensing applications continues to be challenging (Kim et al., 2023), leaving modeling as the best means for understanding flow across large spatial extents. The regional-scale $PROSPER_{PNW}$ model (Jaeger et al., 2019) was an original effort to model streamflow permanence across a large geographic extent using RF. This model provided predictions at relatively fine spatial resolution (30 m) for that point in time. Since then, the $PROSPER_{PNW}$ has been refined (e.g., Sando et al., 2022) and extended (e.g., Messager et al., 2021). Notable distinctions between the sub-regional WOWTDR and the regional-scale $PROSPER_{PNW}$ include WOWTDR providing streamflow classifications at a markedly higher spatial resolution to address needs of end users (Kampf et al., 2021). Additionally, WOWTDR results suggest that intensive spatial balancing, covariate selection, and uncertainty analysis provide worthwhile contributions to streamflow permanence modeling, although more work remains. Sando et al. (2022) showed that driver variables important to streamflow permanence vary across space. Additional study is necessary to understand if this variation across space is highlighting spatial groupings of observation data that are different enough to warrant spatially focused and potentially more accurate prediction models.

Data Availability Statement

Data sets and processing scripts for this research are available in this USFS Data Release: Burnett (2025); these data were collected using funding from the U.S. Government and can be used without additional permissions or fees.

Acknowledgments
The authors would like to acknowledge Bureau of Land Management and U.S. Forest Service for field data collection. Thanks to Amanda Pollock (Oregon State University) for help with figures. This manuscript was improved by helpful comments of peer reviewers. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Funding for this project was provided by the US Department of Interior Bureau of Land Management under agreement with the USFS with additional direct contributions from the US Department of Agriculture Forest Service Pacific Northwest Research Station and the U.S. Geological Survey.

References

Acuña, V., Hunter, M., & Ruhí, A. (2017). Managing temporary streams and rivers as unique rather than second-class ecosystems. *Biological Conservation*, *211*, 12–19. https://doi.org/10.1016/j.biocon.2016.12.025

Altman, D. G., & Bland, J. M. (2005). Standard deviations and standard errors. *BMJ*, *331*(7521), 903. https://doi.org/10.1136/bmj.331.7521.903

Anderson, R., Lukas, V., & Aichele, S. S. (2024). The 3D national topography model call for action—Part 1. The 3D hydrography program. In *Circular (1519)*. U.S. Geological Survey. https://doi.org/10.3133/cir1519

Assendelft, R. S., & van Meerveld, H. J. I. (2019). A low-cost, multi-sensor system to monitor temporary stream dynamics in mountainous headwater catchments. *Sensors*, *19*(21), 21. https://doi.org/10.3390/s19214645

Barnhart, T. B., Schultz, A. R., Siefken, S. A., Thompson, F. E., Welborn, T., Sando, T. R., & McCarthy, P. M. (2020). Flow conditioned parameter grids for mechanistic, statistical, and machine learning hydrologic models: A seamless basin characteristic dataset for the contiguous United States. In *2020, H028-06. AGU fall meeting abstracts*.

Benstead, J. P., & Leigh, D. S. (2012). An expanded role for river networks. *Nature Geoscience*, *5*(10), 678–679. https://doi.org/10.1038/ngeo1593

Bhamjee, R., Lindsay, J. B., & Cockburn, J. (2016). Monitoring ephemeral headwater streams: A paired-sensor approach. *Hydrological Processes*, *30*(6), 888–898. https://doi.org/10.1002/hyp.10677

Biecek, P. (2018). DALEX: Explainers for complex predictive models in R. *Journal of Machine Learning Research*, *19*(84), 1–5.

Boisjolie, B. A., Santelmann, M. V., Flitcroft, R. L., & Duncan, S. L. (2017). Legal ecotones: A comparative analysis of riparian policy protection in the Oregon Coast Range, USA. *Journal of Environmental Management*, *197*, 206–220. https://doi.org/10.1016/j.jenvman.2017.03.075

Botter, G., McNamara, J., & Durighetto, N. (2024). Extending active network length versus catchment discharge relations to temporarily dry outlets. *Water Resources Research*, *60*(1), e2023WR035617. https://doi.org/10.1029/2023WR035617

Botter, G., Vingiani, F., Senatore, A., Jensen, C., Weiler, M., McGuire, K., et al. (2021). Hierarchical climate-driven dynamics of the active channel length in temporary streams. *Scientific Reports*, *11*(1), 21503. https://doi.org/10.1038/s41598-021-00922-2

Brinkerhoff, C. B. (2024). The importance of source data in river network connectivity modeling: A review. *Limnology & Oceanography*, *69*(12), 3033–3060. https://doi.org/10.1002/lno.12706

Brinkerhoff, C. B., Gleason, C. J., Kotchen, M. J., Kysar, D. A., & Raymond, P. A. (2024). Ephemeral stream water contributions to United States drainage networks. *Science*, *384*(6703), 1476–1482. https://doi.org/10.1126/science.adg9430

Bujak-Ozga, I., (Ilja) van Meerveld, H. J., Rinaldo, A., & von Freyberg, J. (2023). Short-term dynamics of drainage density based on a combination of channel flow state surveys and water level measurements. *Hydrological Processes*, *37*(12), e15041. https://doi.org/10.1002/hyp.15041

Burnett, J. D. (2025). *Western Oregon wet dry (WOWTDR) annual predictions of late summer streamflow status for the 2019 through 2021 for western Oregon*. Forest Service Research Data Archive. https://doi.org/10.2737/RDS-2024-0037

Busch, M. H., Costigan, K. H., Fritz, K. M., Datry, T., Krabbenhoft, C. A., Hammond, J. C., et al. (2020). What's in a name? Patterns, trends, and suggestions for defining non-perennial rivers and streams. *Water*, *12*(7), 1980. https://doi.org/10.3390/w12071980

Bush, S. A., Birch, A. L., Warix, S. R., Sullivan, P. L., Gooseff, M. N., McKnight, D. M., & Barnard, H. R. (2023). Dominant source areas shift seasonally from longitudinal to lateral contributions in a montane headwater stream. *Journal of Hydrology*, *617*, 129134. https://doi.org/10.1016/j.jhydrol.2023.129134

Chaitanya, D. E., Ganesh, L., & Rao, G. S. (2016). Performance analysis of hyperbolic multilateration using circular error probability. *Procedia Computer Science*, *85*, 676–682. https://doi.org/10.1016/j.procs.2016.05.253

Chapin, T. P., Todd, A. S., & Zeigler, M. P. (2014). Robust, low-cost data loggers for stream temperature, flow intermittency, and relative conductivity monitoring. *Water Resources Research*, *50*(8), 6542–6548. https://doi.org/10.1002/2013WR015158

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*(1), 13. https://doi.org/10.1186/s13040-021-00244-z

Christensen, J. R., Golden, H. E., Alexander, L. C., Pickard, B. R., Fritz, K. M., Lane, C. R., et al. (2022). Headwater streams and inland wetlands: Status and advancements of geospatial datasets and maps across the United States. *Earth-Science Reviews*, *235*, 104230. https://doi.org/10.1016/j.earscirev.2022.104230

Clarke, S. E., Burnett, K. M., & Miller, D. J. (2008). Modeling streams and hydrogeomorphic attributes in Oregon from digital and field data. *JAWRA Journal of the American Water Resources Association*, *44*(2), 459–477. https://doi.org/10.1111/j.1752-1688.2008.00175.x

Coble, A. A., Barnard, H., Du, E., Johnson, S., Jones, J., Keppeler, E., et al. (2020). Long-term hydrological response to forest harvest during seasonal low flow: Potential implications for current forest practices. *Science of the Total Environment*, *730*, 138926. https://doi.org/10.1016/j.scitotenv.2020.138926

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Colson, T., Gregory, J., Dorney, J., & Russell, P. (2008). Topographic and soil maps do not accurately depict headwater stream networks. *National Wetlands Newsletter*, *30*, 25–28.

Costigan, K. H., Jaeger, K. L., Goss, C. W., Fritz, K. M., & Goebel, P. C. (2016). Understanding controls on flow permanence in intermittent rivers to aid ecological research: Integrating meteorology, geology and land cover. *Ecohydrology*, *9*(7), 1141–1153. https://doi.org/10.1002/eco.1712

Costigan, K. H., Kennard, M. J., Leigh, C., Sauquet, E., Datry, T., & Boulton, A. J. (2017). Chapter 2.2—Flow regimes in intermittent rivers and ephemeral streams. In *Intermittent rivers and ephemeral streams* (pp. 51–78). Academic Press.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

D'Odorico, P., & Rigon, R. (2003). Hillslope and channel contributions to the hydrologic response. *Water Resources Research*, *39*(5). https://doi.org/10.1029/2002WR001708

Dohman, J. M., Godsey, S. E., & Hale, R. L. (2021). Three-dimensional subsurface flow path controls on flow permanence. *Water Resources Research*, *57*(10), e2020WR028270. https://doi.org/10.1029/2020WR028270

Döll, P., Abbasi, M., Messager, M. L., Trautmann, T., Lehner, B., & Lamouroux, N. (2024). Streamflow intermittence in Europe: Estimating high-resolution monthly time series by downscaling of simulated runoff and random forest modeling. *Water Resources Research*, *60*(8), e2023WR036900. https://doi.org/10.1029/2023WR036900

Downing, J. A., Cole, J. J., Duarte, C. M., Middelburg, J. J., Melack, J. M., Prairie, Y. T., et al. (2012). Global abundance and size distribution of streams and rivers. *Inland Waters*, *2*(4), 229–236. https://doi.org/10.5268/IW-2.4.502

Durighetto, N., & Botter, G. (2022). On the relation between active network length and catchment discharge. *Geophysical Research Letters*, *49*(14), e2022GL099500. https://doi.org/10.1029/2022GL099500

Durighetto, N., Mariotto, V., Zanetti, F., McGuire, K. J., Mendicino, G., Senatore, A., & Botter, G. (2022). Probabilistic description of streamflow and active length regimes in rivers. *Water Resources Research*, *58*(4), e2021WR031344. https://doi.org/10.1029/2021WR031344

Fritz, K. M., Hagenbuch, E., D'Amico, E., Reif, M., Wigington P. J., Jr., Leibowitz, S. G., et al. (2013). Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. *JAWRA Journal of the American Water Resources Association*, *49*(4), 867–882. https://doi.org/10.1111/jawr.12040

Gaines, W. L., Hessburg, P. F., Aplet, G. H., Henson, P., Prichard, S. J., Churchill, D. J., et al. (2022). Climate change and forest management on federal lands in the Pacific Northwest, USA: Managing for dynamic landscapes. *Forest Ecology and Management*, *504*, 119794. https://doi.org/10.1016/j.foreco.2021.119794

Gallart, F., Prat, N., García-Roger, E. M., Latron, J., Rieradevall, M., Llorens, P., et al. (2012). A novel approach to analysing the regimes of temporary streams in relation to their controls on the composition and structure of aquatic biota. *Hydrology and Earth System Sciences*, *16*(9), 3165–3182. https://doi.org/10.5194/hess-16-3165-2012

Godsey, S. E., & Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order. *Hydrological Processes*, *28*(23), 5791–5803. https://doi.org/10.1002/hyp.10310

Golden, H. E., Christensen, J. R., McMillan, H. K., Kelleher, C. A., Lane, C. R., Husic, A., et al. (2025). Advancing the science of headwater streamflow for global water protection. *Nature Water*, *3*, 1–11. https://doi.org/10.1038/s44221-024-00351-1

Goulsbra, C., Evans, M., & Lindsay, J. (2014). Temporary streams in a peatland catchment: Pattern, timing, and controls on stream network expansion and contraction. *Earth Surface Processes and Landforms*, *39*(6), 790–803. https://doi.org/10.1002/esp.3533

Hafen, K. C., Blasch, K., Gessler, P. E., Dunham, J., & Brooks, E. (2023). Estimating streamflow permanence with the watershed Erosion Prediction Project Model: Implications for surface water presence modeling and data collection. *Journal of Hydrology*, *622*, 129747. https://doi.org/10.1016/j.jhydrol.2023.129747

Hafen, K. C., Blasch, K. W., Rea, A., Sando, R., & Gessler, P. E. (2020). The influence of climate variability on the accuracy of NHD perennial and nonperennial stream classifications. *JAWRA Journal of the American Water Resources Association*, *56*(5), 903–916. https://doi.org/10.1111/1752-1688.12871

Hale, V. C., & McDonnell, J. J. (2016). Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison. *Water Resources Research*, *52*(2), 1358–1374. https://doi.org/10.1002/2014WR016124

Hammond, J. C., Zimmer, M., Shanafield, M., Kaiser, K., Godsey, S. E., Mims, M. C., et al. (2021). Spatial patterns and drivers of nonperennial flow regimes in the contiguous United States. *Geophysical Research Letters*, *48*(2), e2020GL090794. https://doi.org/10.1029/2020GL090794

Hanberry, B. B. (2024). Practical guide for retaining correlated climate variables and unthinned samples in species distribution modeling, using random forests. *Ecological Informatics*, *79*, 102406. https://doi.org/10.1016/j.ecoinf.2023.102406

Jaeger, K. L., Burnett, J., Heaston, E. D., Wondzell, S. M., Chelgren, N., Dunham, J. B., et al. (2020). FLOwPER user guide—For collection of FLOw PERmanence field observations. In *FLOwPER user guide—For collection of FLOw PERmanence field observations (USGS Numbered Series 2020–1075; Open-File Report, Vols. 2020–1075)*. U.S. Geological Survey. https://doi.org/10.3133/ofr20201075

Jaeger, K. L., Sando, R., Dunn, S. B., & Gendaszek, A. S. (2023). Predicting probabilities of late summer surface flow presence in a glaciated mountainous headwater region. *Hydrological Processes*, *37*(2), e14813. https://doi.org/10.1002/hyp.14813

Jaeger, K. L., Sando, R., McShane, R. R., Dunham, J. B., Hockman-Wert, D. P., Kaiser, K. E., et al. (2019). Probability of streamflow permanence model (PROSPER): A spatially continuous model of annual streamflow permanence throughout the Pacific Northwest. *Journal of Hydrology X*, *2*, 100005. https://doi.org/10.1016/j.hydroa.2018.100005

Jensen, C. K., McGuire, K. J., Shao, Y., & Andrew Dolloff, C. (2018). Modeling wet headwater stream networks across multiple flow conditions in the Appalachian Highlands. *Earth Surface Processes and Landforms*, *43*(13), 2762–2778. https://doi.org/10.1002/esp.4431

Johnston, C. M., Dewald, T. G., Bondelid, T. R., Worstell, B. B., McKay, L. D., Rea, A., et al. (2009). Evaluation of catchment delineation methods for the medium-resolution national hydrography dataset. In *Scientific investigations report (2009–5233)*. U.S. Geological Survey. https://doi.org/10.3133/sir20095233

Kampf, S. K. (2018). Stream Tracker web application. Retrieved from https://www.streamtracker.org/

Kampf, S. K., Dwire, K. A., Fairchild, M. P., Dunham, J., Snyder, C. D., Jaeger, K. L., et al. (2021). Managing nonperennial headwater streams in temperate forests of the United States. *Forest Ecology and Management*, *497*, 119523. https://doi.org/10.1016/j.foreco.2021.119523

Kaplan, N. H., Blume, T., & Weiler, M. (2020). Predicting probabilities of streamflow intermittency across a temperate mesoscale catchment. *Hydrology and Earth System Sciences*, *24*(11), 5453–5472. https://doi.org/10.5194/hess-24-5453-2020

Kaplan, N. H., Blume, T., & Weiler, M. (2022). Event controls on intermittent streamflow in a temperate climate. *Hydrology and Earth System Sciences*, *26*(10), 2671–2696. https://doi.org/10.5194/hess-26-2671-2022

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab—An S4 package for Kernel methods in R. *Journal of Statistical Software*, *11*(9), 1–20. https://doi.org/10.18637/jss.v011.i09

Kim, S., Yoon, S.-K., & Choi, N. (2023). The definition of perennial streams based on a wet channel network extracted from LiDAR data. *Applied Sciences*, *13*(2), 2. https://doi.org/10.3390/app13020704

Krabbenhoft, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., et al. (2022). Assessing placement bias of the global river gauge network. *Nature Sustainability*, *5*(7), 586–592. https://doi.org/10.1038/s41893-022-00873-0

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Leach, J. A., Webster, K. L., Hudson, D. T., Buttle, J., & Nehemy, M. (2024). Zero-flow dynamics for headwater streams in a humid forested landscape. *Hydrological Processes*, *38*(12), e70025. https://doi.org/10.1002/hyp.70025

Luce, C. H., Abatzoglou, J. T., & Holden, Z. A. (2013). The missing mountain water: Slower westerlies decrease orographic enhancement in the Pacific Northwest USA. *Science*, *342*(6164), 1360–1364. https://doi.org/10.1126/science.1242335

Mahoney, D. T., Christensen, J. R., Golden, H. E., Lane, C. R., Evenson, G. R., White, E., et al. (2023). Dynamics of streamflow permanence in a headwater network: Insights from catchment-scale model simulations. *Journal of Hydrology*, *620*, 129422. https://doi.org/10.1016/j.jhydrol.2023.129422

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Messager, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., et al. (2021). Global prevalence of non-perennial rivers and streams. *Nature*, *594*(7863), 391–397. https://doi.org/10.1038/s41586-021-03565-5

Messager, M. L., Pella, H., & Datry, T. (2024). Inconsistent regulatory mapping quietly threatens rivers and streams. *Environmental Science & Technology*, *58*(39), 17201–17214. https://doi.org/10.1021/acs.est.4c01859

Metes, M. J., Jones, D. K., Baker, M. E., Miller, A. J., Hogan, D. M., Loperfido, J. V., & Hopkins, K. G. (2022). Ephemeral stream network extraction from LiDAR-derived elevation and topographic attributes in urban and forested landscapes. *JAWRA Journal of the American Water Resources Association*, *58*(4), 547–565. https://doi.org/10.1111/1752-1688.13012

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*(9), 1620–1633. https://doi.org/10.1111/2041-210X.13650

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Mimeau, L., Künne, A., Branger, F., Kralisch, S., Devers, A., & Vidal, J.-P. (2024). Flow intermittence prediction using a hybrid hydrological modelling approach: Influence of observed intermittence data on the training of a random forest model. *Hydrology and Earth System Sciences*, *28*(4), 851–871. https://doi.org/10.5194/hess-28-851-2024

Moidu, H., Obedzinski, M., Carlson, S. M., & Grantham, T. E. (2021). Spatial patterns and sensitivity of intermittent stream drying to climate variability. *Water Resources Research*, *57*(11), e2021WR030314. https://doi.org/10.1029/2021WR030314

National Drought Mitigation Center; U.S. Department of Agriculture; National Oceanic and Atmospheric Administration. (2019). *United States Drought Monitor*. University of Nebraska-Lincoln. Retrieved from https://droughtmonitor.unl.edu/DmData/GISData.aspx

O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. *Geological Society of America Bulletin*, *126*(3–4), 377–397. https://doi.org/10.1130/B30831.1

Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework. *Environmental Management*, *54*(6), 1249–1266. https://doi.org/10.1007/s00267-014-0364-1

Oregon Department of Geology and Mineral Industries (DOGAMI). (2025). LiDAR digital terrain model mosaic. Retrieved from https://gis.dogami.oregon.gov/maps/lidarviewer/

Papacharalampous, G., Tyralis, H., Doulamis, A., & Doulamis, N. (2023). Comparison of tree-based ensemble algorithms for merging satellite and Earth-observed precipitation data at the daily time scale. *Hydrology*, *10*(2), 2. https://doi.org/10.3390/hydrology10020050

Pate, A. A., Segura, C., & Bladon, K. D. (2020). Streamflow permanence in headwater streams across four geomorphic provinces in Northern California. *Hydrological Processes*, *34*(23), 4487–4504. https://doi.org/10.1002/hyp.13889

Penaluna, B. E., Burnett, J. D., Christiansen, K., Arismendi, I., Johnson, S. L., Griswold, K., et al. (2022). UPRLIMET: UPstream regional LiDAR model for extent of trout in stream networks. *Scientific Reports*, *12*(1), 20266. https://doi.org/10.1038/s41598-022-23754-0

Peterson, D. A., Kampf, S. K., Puntenney-Desmond, K. C., Fairchild, M. P., Zipper, S., Hammond, J. C., et al. (2024). Predicting streamflow duration from crowd-sourced flow observations. *Water Resources Research*, *60*(1), e2023WR035093. https://doi.org/10.1029/2023WR035093

Prancevic, J. P., & Kirchner, J. W. (2019). Topographic controls on the extension and retraction of flowing streams. *Geophysical Research Letters*, *46*(4), 2084–2092. https://doi.org/10.1029/2018GL081799

PRISM, Oregon State University PRISM Climate Group. (2014). *PRISM climate data [Map]*. Oregon State University. Retrieved from https://prism.oregonstate.edu/

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Russell, P. P., Gale, S. M., Muñoz, B., Dorney, J. R., & Rubino, M. J. (2015). A spatially explicit model for mapping headwater streams. *JAWRA Journal of the American Water Resources Association*, *51*(1), 226–239. https://doi.org/10.1111/jawr.12250

Sahour, H., Gholami, V., Torkaman, J., Vazifedan, M., & Saeedi, S. (2021). Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. *Environmental Earth Sciences*, *80*(22), 747. https://doi.org/10.1007/s12665-021-10054-5

Sando, R., Jaeger, K. L., Farmer, W. H., Barnhart, T. B., McShane, R. R., Welborn, T. L., et al. (2022). Predictions and drivers of sub-reach-scale annual streamflow permanence for the upper Missouri River basin: 1989–2018. *Journal of Hydrology X*, *17*, 100138. https://doi.org/10.1016/j.hydroa.2022.100138

Sauquet, E., Shanafield, M., Hammond, J. C., Sefton, C., Leigh, C., & Datry, T. (2021). Classification and trends in intermittent river flow regimes in Australia, northwestern Europe and USA: A global perspective. *Journal of Hydrology*, *597*, 126170. https://doi.org/10.1016/j.jhydrol.2021.126170

Scheller, M., van Meerveld, I., & Seibert, J. (2024). How well can people observe the flow state of temporary streams? *Frontiers in Environmental Science*, *12*. https://doi.org/10.3389/fenvs.2024.1352697

Segura, C., Bladon, K. D., Hatten, J. A., Jones, J. A., Hale, V. C., & Ice, G. G. (2020). Long-term effects of forest harvesting on summer low flow deficits in the Coast Range of Oregon. *Journal of Hydrology*, *585*, 124749. https://doi.org/10.1016/j.jhydrol.2020.124749

Seibert, J., Strobl, B., Etter, S., Hummer, P., & (Ilja) van Meerveld, H. J. (2019). Virtual staff gauges for crowd-based stream level observations. *Frontiers in Earth Science*, *7*. https://doi.org/10.3389/feart.2019.00070

Shanafield, M., Bourke, S. A., Zimmer, M. A., & Costigan, K. H. (2021). An overview of the hydrology of non-perennial rivers and streams. *WIREs Water*, *8*(2), e1504. https://doi.org/10.1002/wat2.1504

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

State of Oregon. (2016). Ownership Land Management: Oregon Department of Forestry web map data set. Retrieved from https://oregon-department-of-forestry-geo.hub.arcgis.com/datasets/geo::ownership-land-management/about

Theodoridis, S., & Koutroumbas, K. (2006). *Pattern recognition*. Elsevier.

Tonkin, J. D., Poff, N. L., Bond, N. R., Horne, A., Merritt, D. M., Reynolds, L. V., et al. (2019). Prepare river ecosystems for an uncertain future. *Nature*, *570*(7761), 301–303. https://doi.org/10.1038/d41586-019-01877-1

Truchy, A., Csabai, Z., Mimeau, L., Künne, A., Pernecker, B., Bertin, W., et al. (2023). Citizen scientists can help advance the science and management of intermittent rivers and ephemeral streams. *BioScience*, *73*(7), 513–521. https://doi.org/10.1093/biosci/biad045

Trunk, G. V. (1979). A problem of dimensionality: A simple example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(3), 306–307. https://doi.org/10.1109/TPAMI.1979.4766926

Tsamardinos, I., Rakhshani, A., & Lagani, V. (2015). Performance-estimation properties of cross-validation-based protocols with simultaneous hyper-parameter optimization. *The International Journal on Artificial Intelligence Tools*, *24*(5), 1540023. https://doi.org/10.1142/S0218213015400230

Tucker, G. E., Catani, F., Rinaldo, A., & Bras, R. L. (2001). Statistical analysis of drainage density from digital terrain data. *Geomorphology*, *36*(3–4), 187–202. https://doi.org/10.1016/s0169-555x(00)00056-8

U.S. Geological Survey (USGS). (2021). National hydrography dataset. Retrieved from https://prd-tnm.s3.amazonaws.com/index.html?prefix=StagedProducts/Hydrography/NHD/HU4/GPKG/

Vance-Borland, K., Burnett, K., & Clarke, S. (2009). Influence of mapping resolution on assessments of stream and streamside conditions: Lessons from coastal Oregon, USA. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *19*(3), 252–263. https://doi.org/10.1002/aqc.967

van Meerveld, H. J. I., Sauquet, E., Gallart, F., Sefton, C., Seibert, J., & Bishop, K. (2020). Aqua temporaria incognita. *Hydrological Processes*, *34*(26), 5704–5711. https://doi.org/10.1002/hyp.13979

Ward, A. S., Wondzell, S. M., Schmadel, N. M., & Herzog, S. P. (2020). Climate change causes river network contraction and disconnection in the H.J. Andrews experimental forest, Oregon, USA. *Frontiers in Water*, *2*. https://doi.org/10.3389/frwa.2020.00007

Ward, A. S., Schmadel, N. M., & Wondzell, S. M. (2018). Simulation of dynamic expansion, contraction, and connectivity in a mountain stream network. *Advances in Water Resources*, *114*, 64–82. https://doi.org/10.1016/j.advwatres.2018.01.018

Warix, S. R., Navarre-Sitchler, A., Manning, A. H., & Singha, K. (2023). Local topography and streambed hydraulic conductivity influence riparian groundwater age and groundwater-surface water connection. *Water Resources Research*, *59*(9), e2023WR035044. https://doi.org/10.1029/2023WR035044

Whiting, J. A., & Godsey, S. E. (2016). Discontinuous headwater stream networks with stable flowheads, Salmon River basin, Idaho. *Hydrological Processes*, *30*(13), 2305–2316. https://doi.org/10.1002/hyp.10790

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*, 1–17. https://doi.org/10.18637/jss.v077.i01

Yu, S., Bond, N. R., Bunn, S. E., & Kennard, M. J. (2019). Development and application of predictive models of surface water extent to identify aquatic refuges in eastern Australian temporary stream networks. *Water Resources Research*, *55*(11), 9639–9655. https://doi.org/10.1029/2019WR025216

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, *598*, 126266. https://doi.org/10.1016/j.jhydrol.2021.126266

References From the Supporting Information

Abedi, R., Costache, R., Shafizadeh-Moghadam, H., & Pham, Q. B. (2022). Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*, *37*(19), 5479–5496. https://doi.org/10.1080/10106049.2021.1920636

Barnhart, T. B., Sando, R. R., Siefken, S. A., McCarthy, P. M., & Rea, A. H. (2020). Flow-conditioned parameter grid tools: U.S. Geological survey software release (version 1.0) [Computer software]. *U.S. Geological Survey*. https://doi.org/10.5066/P9W8UZ47

Bell, D. M., Gregory, M. J., Palmer, M., & Davis, R. (2023). *Guidance for forest management and landscape ecology applications of recent gradient nearest neighbor imputation maps in California, Oregon, and Washington*. Gen. Tech. Rep. PNW-GTR-1018 (p. 41). U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. (Online Only). https://doi.org/10.2737/PNW-GTR-1018

Gendaszek, A. S., Dunham, J. B., Torgersen, C. E., Hockman-Wert, D. P., Heck, M. P., Thorson, J., et al. (2020). Land-cover and climatic controls on water temperature, flow permanence, and fragmentation of Great Basin Stream Networks. *Water*, *12*(7), 7. https://doi.org/10.3390/w12071962

Harar, P., Elbrächter, D., Dörfler, M., & Johnson, K. D. (2022). Redistributor: Transforming empirical data distributions (arXiv:2210.14219). *arXiv*. https://doi.org/10.48550/arXiv.2210.14219

Henderson, S. G., & Nelson, B. L. (2006). *Handbooks in operations research and management science: Simulation*. Elsevier.

Hosmer, D. W., & Lemeshow, S. (1992). Confidence interval estimation of interaction. *Epidemiology*, *3*(5), 452–456. https://doi.org/10.1097/00001648-199209000-00012

Signorell, A. (2023). RobScale function—RDocumentation (version 0.99) [Computer software]. https://www.rdocumentation.org/packages/DescTools/versions/0.99.54/topics/RobScale

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., et al. (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. https://doi.org/10.1016/j.isprsjprs.2018.09.006

Supplemental Information for

A streamflow permanence classification model for western Oregon that explicitly accounts for uncertainty and extrapolation

Jonathan D. Burnett[1], Kristin L. Jaeger[2], Sherri Johnson[1], Steven M. Wondzell[1], Jason B. Dunham[3], Matthew I. Barker[2], Emily Heaston[3], Nate Chelgren[3], Michael G. Wing[4], Brian Staab[5], Michael E. Brown[6],

Affiliations

[1] U.S. Forest Service, Pacific Northwest Research Station, Corvallis Oregon, 97331, USA

[2]U.S. Geological Survey, Washington Water Science Center, Tacoma, WA 98402, USA

[3]U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center, Corvallis, Oregon, 97331, USA

[4]College of Forestry, Oregon State University, Corvallis, Oregon, USA.

[5]U.S. Forest Service, Pacific Northwest Region, Portland, OR, USA

[6]Bureau of Land Management Oregon/Washington, Portland OR, USA

Corresponding author: Kristin Jaeger (kjaeger@usgs.gov)

Table of Contents

S1 Preparation of LiDAR DEM and stream network modeling domain

This study required processing of light detection and ranging (LiDAR) elevation data (DOGAMI, [2025]) that not only comprise the topographic covariates, but is the basis of the flow direction layer required for processing all covariates as flow conditioned values. Additionally, a stream network was generated to serve as the modeling domain on which the model was trained, and predictions were estimated. All processing for this step occurred in Python ver. 3.7 (Python Software Foundation, 2018) and utilized multiple ArcGIS Pro ver. 2.8 (Esri, 2021) functions within the Python environment.

Covariates flagged as "patch scale" in Table S1 were processed as drainage-area weighted values to produce Flow Conditioned Parameter Grids (FCPGs). FCPGs account for change in average basin value with downstream accumulation (Jaeger et al., 2019; Barnhart et al., 2020; Sando et al., 2022). FCPG processing requires an eight direction (*i.e.,* D8) flow direction grid that is derived from the LiDAR-derived digital elevation models (DEMs) (DOGAMI, [2025]) to accumulate values in the downstream direction for each gridded covariate dataset. When gridded covariate datasets are categorical the result is a drainage-area weighted proportion of presence and when the datasets are categorical the gridded values are continuous values representing drainage-area weighted means of the gridded values over the accumulation area. Therefore, this modeling approach requires that the underlying LiDAR-derived elevation data are consistent throughout the modeling domain, which contributes to limitations of applying this approach to the contiguous western Oregon study area. The modeling domains of previous streamflow permanence models have included contiguous areas that comprise larger watersheds (Jaeger et al., 2019, 2023; Sando et al., 2022)

LiDAR-derived DEM data (DOGAMI, [2025]) were downloaded, pit-filled, smoothed, and reprojected to Universal Transverse Mercator (UTM) Zone 10N (North American Datum, 1983). Vertical units representing orthometric height above mean sea level are in units of meters. DEMs were resampled to a 5-m spatial resolution. The steps ensure the elevation and spatial characterization of hydro-topographic conditions represented by DEMs were consistent across the study domain that included many 12-digit Hydrologic Unit Code (HUC12) sub-watersheds that are not geographically adjacent.

The associated flowline network was processed into stream points that each represent a 5- to 7-m sub-reach. These stream points served as the stream network on which the model was developed. National Hydrography Dataset (NHD) flowlines (U.S. Geological Survey, 2021) representing the stream network were aligned to the DEM using the Arc Hydro Tools Create Drainage Line structures tool as described in the Arc Hydro reference manual for use where flowlines may not have been derived from the available source DEM (Esri, 2019). Aligned flowlines were split into 5- to 7-m sub-reaches to ensure spatial agreement with the predictor covariates produced from the DEM. Subreach length varied depending on the direction a given flowline crossed the corresponding DEM grid cells (7 m along diagonals). Sub-reaches were then simplified to points (hereafter stream points to disambiguate with other points) representing the center of the sub-

reach. Stream points and sub-reaches are synonymous, indicating that modeling occurs on points that represent the sub-reach.

Table S1.1. Covariate table that describes the 96 covariates considered for the creation of the WOWTDR model. DataShortname is the name of the covariate as it appears in the scripts and within the data tables of training and prediction data. ManuscriptShortname is the covariate name as it appears in the text. Descriptive name provides a more comprehensive description of the covariate. Type indicates whether the covariate describes landcover, physiography, jurisdiction, or climatic conditions. Scale indicates whether the covariate value is 'local' and drawn from the location of the given sub-reach or if it is at the patch scale that captures the average of condition over a larger area upstream of the sub-reach. Scaling Values to Subreach describes how covariate data were scaled to the sub-reach. Data Description and Derivation Method describes what covariate values represent and when applicable, how they were derived from other data sources. Map Source indicates the raw data source. Source Resolution describes the native resolution of the raw source data in terms of 2-dimensional spatial resolution. When described as "mapped" this indicates a data source drawn from a categorical map of polygons.

Data Shortname	Manuscript Shortname	Descriptive Name	Rank	Scale	Source Resolution (m)	Parameter-ization	Potential Driver	Derivation Method	Data Source	Citation
aspect_cos	N/A	COS Aspect (Northness)	23	Non-local	1	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	a = aspect, northness = cos(a)	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Stage, A. R. (1976). An Expression for the Effect of Aspect, Slope, and Habitat Type on Tree Growth. Forest Science, 22(4), 457â€“460. https://doi.org/10.1093/forestscience/22.4.457
aspect_sin	aspect_sin	SIN Aspect (Westness)	39	Non-local	2	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	a = aspect, westness = sin(a)	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Stage, A. R. (1976). An Expression for the Effect of Aspect, Slope, and Habitat Type on Tree Growth. Forest Science, 22(4), 457â€“460. https://doi.org/10.1093/forestscience/22.4.457

aspect_trasp	aspect_trasp	TRASP Aspect	34	Non-local	3	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	a = aspect in degrees, TRASP aspect = (1 - cos((pi/180) (a-30)) / 2	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Roberts, D. W., & Cooper, S. V. (1989). Concepts and techniques of vegetation mapping. General Technical Report INT - U.S. Department of Agriculture, Forest Service, Intermountain Research Station (USA). https://scholar.google.com/scholar_lookup?title=Concepts+and+techniques+of+vegetation+mapping&author=Roberts%2C+D.W.+%28Utah+State+University%2C+Logan%2C+UT%29&publication_year=1989
Bedrock_Depth_fix	Bedrock_Depth_fix	Depth to Bedrock (m)	47	Non-local	4	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	None	https://www.arcgis.com/home/item.html?id=cdc49bd63ea54dd2977f3f2853e07fff	Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey Metadata. Available online at SSURGO-Metadata-Tables-and-Columns-Report.pdf. Last accessed 23 Feb 2025.

cancov_2017	Proportion Canopy	Proportion Canopy Cover	88	Non-local	5	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	None	https://lemma.forestry.oregonstate.edu/data	Bell, David M., Matthew J. Gregory, Marin Palmer, and Raymond Davis. "Guidance for forest management and landscape ecology applications of recent gradient nearest neighbor imputation maps in California, Oregon, and Washington." *Gen. Tech. Rep. PNW-GTR-1018. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 41 p.(Online only).* 1018 (2023).
cancov_con_2017	cancov_con_2017	% Coniferous Canopy Cover	46	Non-local	6	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	None	https://lemma.forestry.oregonstate.edu/data	Bell, David M., Matthew J. Gregory, Marin Palmer, and Raymond Davis. "Guidance for forest management and landscape ecology applications of recent gradient nearest neighbor imputation maps in California, Oregon, and Washington." *Gen. Tech. Rep. PNW-GTR-1018. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 41 p.(Online only).* 1018 (2023).

cancov_hdw_2017	cancov_hdw_2017	% Hardwood Canopy Cover	69	Non-local	7	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	None	https://lemma.forestry.oregonstate.edu/data	Bell, David M., Matthew J. Gregory, Marin Palmer, and Raymond Davis. "Guidance for forest management and landscape ecology applications of recent gradient nearest neighbor imputation maps in California, Oregon, and Washington." *Gen. Tech. Rep. PNW-GTR-1018. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 41 p.(Online only).* 1018 (2023).
curve	curve	Combined Profile and Planimetric Curvature	38	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	profile curvature + planimetric curvature	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Alkhasawneh, M. Sh., Ngah, U. K., Tay, L. T., Mat Isa, N. A., & Al-batah, M. S. (2013). Determination of Important Topographic Factors for Landslide Mapping Analysis Using MLP Network. The Scientific World Journal, 2013, 415023. https://doi.org/10.1155/2013/415023

d_slp100_m	d_slp100_m	100 mDownstream Channel Slope	14	Non-local	5	Average over distance	Topographic control on hydrologic condition over very large extent;	e0 = elevation at location, e100 = elevation at location 100 m downstream abs(e0 - e100) / 100,	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None
d_slp1000_m	G_{1km}	1000m Downstream Channel Slope	5	Non-local	5	Average over distance	Topographic control on hydrologic condition over small extent;	e0 = elevation at location, e1000 = elevation at location 1000 m downstream abs(e0 - e1000) / 1000,	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None

d_slp20_m	d_slp20_m	20m Downstream Channel Slope	29	Non-local	5	Average over distance	Topographic control on hydrologic condition over small-moderate extent;	e0 = elevation at location, e20 = elevation at location 20 m downstream abs(e0 - e20) / 20,	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None
d_slp30_m	d_slp30_m	30m Downstream Channel Slope	44	Non-local	5	Average over distance	Topographic control on hydrologic condition over moderate;	e0 = elevation at location, e30 = elevation at location 30 m downstream abs(e0 - e30) / 30,	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None

d_slp50_m	G_{50m}	50m Downstream Channel Slope	4	Non-local	5	Average over distance	Topographic control on hydrologic condition over moderate;	e0 = elevation at location, e50 = elevation at location 50 m downstream abs(e0 - e50) / 50,	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None
elev	E	Elevation (m)	1	Local	5	Value at sub-reach	Topographic control on hydrologic condition and biological suitability	None	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None

elev_norm	$E_{normalized}$	Median-normalized Elevation	2	Local	5	Value at sub-reach	Topographic control on hydrologic condition and biological suitability	Median-normalized Elevation calculated by taking elevation value at given cell and dividing by median elevation of the HUC12 sub-watershed the point of interest is contained within.	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None

elev_rescale	elev_rescale	Rescaled Elevation	33	Local	5	Value at sub-reach	Topographic control on hydrologic condition and biological suitability	Rescaled Elevation calculated by taking elevation value at given cell and dividing by median elevation of the HUC12 sub-watershed the point of interest is contained within.	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None

FAC_sqkm	DA	Drainage Area (sqkm)	7	Local	5	Value at sub-reach	Topographic control on hydrologic condition	Upstream accumulated drainage area in units of square kilometers using Esri D8 flow path routing and subsequent Esri D8 flow accumulation tools in ArcGIS Pro	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None
HAND_1	HAND_1	HAND (Basin Weighted Average)	25	Non-local	5	Drainage Area Average (FCPG)	Indicator of vertical distance to water which relates to surface flow expression	Flow Distance (Spatial Analyst)â€" ArcGIS Pro | Documentation	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Nobre, A.D., Cuartas, L.A., Hodnett, M., Rennó, C.D., Rodrigues, G., Silveira, A. and Saleska, S., 2011. Height Above the Nearest Drainage–a hydrologically relevant new terrain model. Journal of Hydrology, 404(1-2), pp.13-29.

hload	hload	Heatload Index	56	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	theta = aspect in degrees, HLI = 1-cos(theta-45)/2	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	McCune, B., & Keon, D. (2002). Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science, 13(4), 603â€“606. https://doi.org/10.1111/j.1654-1103.2002.tb02087.x
HslopeLen	$L_{Hillslope}$	Hillslope Length	6	Non-local	5	Value at sub-reach	Geomorphic indicator of hydrologic condition	0.5* 1 / DrainDens	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Tucker, Gregory E., Filippo Catani, Andrea Rinaldo, and Rafael L. Bras. "Statistical analysis of drainage density from digital terrain data." *Geomorphology* 36, no. 3-4 (2001): 187-202.
Hydro_Conductivity	H	Saturated Hydraulic Conductivity (Ksat) in µm/s	11	Non-local	30	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	None	https://www.arcgis.com/home/item.html?id=cdc49bd63ea54dd2977f3f2853e07fff	Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey Metadata. Available online at SSURGO-Metadata-Tables-and-Columns-Report.pdf. Last accessed 23 Feb 2025.

L3region	L3region	EPA Level 3 Ecoregion	90	Non-local	Mapped	Multi-level Factor or Categorical	Spatial grouping of the combination of vegetative and geologic conditions that influence hydrologic condition	None	https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states	Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. Environmental Management, 54(6), 1249–1266. https://doi.org/10.1007/s00267-014-0364-1
landstewardd	landstewardd	Landownership (Private, Private Industrial, Federal, State, Tribal)	75	Non-local	Mapped	Category	Spatial grouping of political boundaries	none	https://gis.blm.gov/orarcgis/rest/services/Land_Status/BLM_OR_Ownership/MapServer	None

Lith_Prov	Lith_Prov	Lithologic Province (as Fixed Effect Factor)	74	Non-local	Mapped	Category	Geologic control on hydrologic condition	The dominant lithologic province within the area of the associated HUC12 subwatershed	https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/126/3-4/377/126005/Geologic-and-physiographic-controls-on-bed?redirectedFrom=fulltext&casa_token=qvXUDswMxe8AAAAA:XpVvjweiMaGvZAzyqjliydT1mDIM4mLNz7EeZ-fHXUpdg-Ax2AL2JCgMbM9odD5P7W7TlA	O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. Geological Society of America Bulletin, 126(3–4), 377–397. https://doi.org/10.1130/B30831.1

Lith_Prov1	Lith_Prov1	Lithologic Province 1 - Coast Range Sedimentary (Proportion of Basin)	94	Non-local	Mapped	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/126/3-4/377/126005/Geologic-and-physiographic-controls-on-bed?redirectedFrom=fulltext&casa_token=qvXUDswMxe8AAAAA:XpVvjweiMaGvZAzyqjliydT1mDIM4mLNz7EeZ-fHXUpdg-Ax2AL2JCgMbM9odD5P7W7TlA	O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. Geological Society of America Bulletin, 126(3–4), 377–397. https://doi.org/10.1130/B30831.1

Lith_Prov2	Lith_Prov2	Lithologic Province 2-Coast Range volcanic rocks/ Columbia River Basalt Group (Proportion of Basin)	95	Non-local	Mapped	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/126/3-4/377/126005/Geologic-and-physiographic-controls-on-bed?redirectedFrom=fulltext&casa_token=qvXUDswMxe8AAAAA:XpVvjweiMaGvZAzyqjliydT1mDIM4mLNz7EeZ-fHXUpdg-Ax2AL2JCgMbM9odD5P7W7TlA	O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. Geological Society of America Bulletin, 126(3–4), 377–397. https://doi.org/10.1130/B30831.1

Lith_Prov3	Lith_Prov3	Lithologic Province 3 - High Cascades (Proportion of Basin)	72	Non-local	Mapped	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/126/3-4/377/126005/Geologic-and-physiographic-controls-on-bed?redirectedFrom=fulltext&casa_token=qvXUDswMxe8AAAAA:XpVvjweiMaGvZAzyqjliydT1mDIM4mLNz7EeZ-fHXUpdg-Ax2AL2JCgMbM9odD5P7W7TlA	O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. Geological Society of America Bulletin, 126(3–4), 377–397. https://doi.org/10.1130/B30831.1

Lith_Prov4	Lith_Prov4	Lithologic Province 4 - Klamath (Proportion of Basin)	96	Non-local	Mapped	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/126/3-4/377/126005/Geologic-and-physiographic-controls-on-bed?redirectedFrom=fulltext&casa_token=qvXUDswMxe8AAAAA:XpVvjweiMaGvZAzyqjliydT1mDIM4mLNz7EeZ-fHXUpdg-Ax2AL2JCgMbM9odD5P7W7TlA	O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. Geological Society of America Bulletin, 126(3–4), 377–397. https://doi.org/10.1130/B30831.1

Lith_Prov5	Lith_Prov5	Lithologic Province 5 - Quaternary Sediment (Proportion of Basin) - Note that QS was removed from training data and predictions due to low sample size but is retained in here because it was part of the initial processing sequence	37	Non-local	Mapped	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/126/3-4/377/126005/Geologic-and-physiographic-controls-on-bed?redirectedFrom=fulltext&casa_token=qvXUDswMxe8AAAAA:XpVvjweiMaGvZAzyqjliydT1mDIM4mLNz7EeZ-fHXUpdg-Ax2AL2JCgMbM9odD5P7W7TlA	O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. Geological Society of America Bulletin, 126(3–4), 377–397. https://doi.org/10.1130/B30831.1

Lith_Prov7	Lith_Prov7	Lithologic Province 7 - Western Cascades (Proportion of Basin)	91	Non-local	Mapped	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://pubs.geoscienceworld.org/gsa/gsabulletin/article-abstract/126/3-4/377/126005/Geologic-and-physiographic-controls-on-bed?redirectedFrom=fulltext&casa_token=qvXUDswMxe8AAAAA:XpVvjweiMaGvZAzyqjliydT1mDIM4mLNz7EeZ-fHXUpdg-Ax2AL2JCgMbM9odD5P7W7TlA	O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K. (2014). Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. Geological Society of America Bulletin, 126(3–4), 377–397. https://doi.org/10.1130/B30831.1

NLCD10	NLCD10	NLCD Class 10 - Wetland (Proportion of Basin)	42	Non-local	30	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://data.usgs.gov/datacatalog/data/USGS:60cb3da7d34e86b938a30cb9#:~:text=The%20NLCD%202019%20design%20aims,2%E2%80%933%2Dyear%20intervals.	Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., and others, 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, v. 146, p. 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006.

NLCD2	NLCD2	NLCD Class 2 - Barren (Proportion of Basin)	89	Non-local	30	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://data.usgs.gov/datacatalog/data/USGS:60cb3da7d34e86b938a30cb9#:~:text=The%20NLCD%202019%20design%20aims,2%E2%80%933%2Dyear%20intervals.	Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., and others, 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, v. 146, p. 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006.

NLCD4	NLCD4	NLCD Class 4 - Developed (Proportion of Basin)	83	Non-local	30	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://data.usgs.gov/datacatalog/data/USGS:60cb3da7d34e86b938a30cb9#:~:text=The%20NLCD%202019%20design%20aims,2%E2%80%933%2Dyear%20intervals.	Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., and others, 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, v. 146, p. 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006.

NLCD5	NLCD5	NLCD Class 5 - Disturbed/burned (Proportion of Basin)	31	Non-local	30	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://data.usgs.gov/datacatalog/data/USGS:60cb3da7d34e86b938a30cb9#:~:text=The%20NLCD%202019%20design%20aims,2%E2%80%933%2Dyear%20intervals.	Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., and others, 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, v. 146, p. 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006.

NLCD6	NLCD6	NLCD Class 6 - Forest (Proportion of Basin)	48	Non-local	30	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://data.usgs.gov/datacatalog/data/USGS:60cb3da7d34e86b938a30cb9#:~:text=The%20NLCD%202019%20design%20aims,2%E2%80%933%2Dyear%20intervals.	Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., and others, 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, v. 146, p. 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006.

NLCD7	NLCD7	NLCD Class 7 - Grass (Proportion of Basin)	53	Non-local	30	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	FCPGtool Categorical to Binary Processing (cat2bin) for one-hot encoding	https://data.usgs.gov/datacatalog/data/USGS:60cb3da7d34e86b938a30cb9#:~:text=The%20NLCD%202019%20design%20aims,2%E2%80%933%2Dyear%20intervals.	Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., and others, 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, v. 146, p. 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006.

plan_curve	plan_curve	Planimetric Curvature	32	Non-local	5	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	w = window size, e0 = elevation at a given cell, n = {neighboring points along a line perpindicular to the direction of steepest slope}, prof_curve = (mean(n) - e0)/ 2w	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Alkhasawneh, M. Sh., Ngah, U. K., Tay, L. T., Mat Isa, N. A., & Al-batah, M. S. (2013). Determination of Important Topographic Factors for Landslide Mapping Analysis Using MLP Network. The Scientific World Journal, 2013, 415023. https://doi.org/10.1155/2013/415023
ppt_nrml_sum	ppt_nrml_sum	30 Year Normal Total Annual Precipitation (mm)	84	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the 30 year normal (1981 - 2011)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

ppt_ratio_aug_yr	ppt_ratio_aug_yr	Ratio of August Precipitation (mm) Reference Calendar Year to Annual Precipitation for the Reference Calendar Year	35	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	ppt8_yr_m_0 / ppt_sum_yr_m_0	https://prism.oregonstate.edu/	None
ppt_sum_yr	P_{Year}	Total Precipitation (mm) Reference Calendar Year	15	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the given calendar year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

ppt_sum_yr_previousyear	ppt_sum_yr_previous year	Total Precipitation (mm) One Calendar Year Prior	73	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the calendar year (CY) occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
ppt_sum_yr_previous2years	ppt_sum_yr_previous 2years	Total Precipitation (mm) Two Calendar Years Prior	68	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the calendar year (CY) occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

ppt5_yr_m_0	ppt5_yr_m_0	Total May Precipitation (mm) for Specified Calendar Year	58	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the given calendar year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
ppt5_yr_previousyear	ppt5_yr_previousyear	Total May Precipitation (mm) One Calendar Year Prior	43	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the calendar year (CY) occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

ppt5_yr_previous2years	ppt5_yr_previous2years	Total May Precipitation(mm) Two Calendar Years Prior	63	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the calendar year (CY) occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
ppt8_nrml	ppt8_nrml	30 Year Normal August Precipitation (mm)	54	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	30-year normal total precipitation for August (mm)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

ppt8_yr_m_0	ppt8_yr_m_0	Total August Precipitation (mm) for Specified Calendar Year	40	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the given calendar year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
ppt8_yr_previousyear	ppt8_yr_previousyear	Total August Precipitation (mm) One Calendar Year Prior	17	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the calendar year (CY) occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

ppt8_yr_previous2years	ppt8_yr_previous2years	Total August Precipitation(mm) Two Calendar Years Prior	86	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	total precipitation for the calendar year (CY) occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
pptdiffnrml	pptdiffnrml	Difference between August Precipitation (mm) for Reference Calendar Year and the 30-year Normal	16	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	ppt8_yr_m_0 - ppt8_nrml	https://www.arcgis.com/home/item.html?id=cdc49bd63ea54dd2977f3f2853e07fff	None
pptpropnrml	pptpropnrml	Ratio of Total Annual Precipitation (mm) for the Reference Calendar Year to the	64	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological	ppt_yr_m_0/pptpropnrml	https://www.arcgis.com/home/item.html?id=cdc49bd63ea54dd2977f3f2853e07fff	None

		30-year Normal					suitability			
prof_curve	prof_curve	Profile Curvature of Topography	50	Local	5	Value at sub-reach	Topographic control on hydrologic condition	w = window size, e0 = elevation at a given cell, n = {neighboring points along the direction of steepest slope}, prof_curve = (mean(n) - e0)/ 2w	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Alkhasawneh, M. Sh., Ngah, U. K., Tay, L. T., Mat Isa, N. A., & Al-batah, M. S. (2013). Determination of Important Topographic Factors for Landslide Mapping Analysis Using MLP Network. The Scientific World Journal, 2013, 415023. https://doi.org/10.1155/2013/415023

ProfCurve	ProfCurve	Profile Curvature of Topography (Basin Weighted Average)	81	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	Curvature functionâ€" ArcGIS Pro | Documentation	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Alkhasawneh, M. Sh., Ngah, U. K., Tay, L. T., Mat Isa, N. A., & Al-batah, M. S. (2013). Determination of Important Topographic Factors for Landslide Mapping Analysis Using MLP Network. The Scientific World Journal, 2013, 415023. https://doi.org/10.1155/2013/415023
slope_pct_taudem	slope_pct_taudem	Slope (%) (FCPG)	30	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	slope = rise / run between given cells	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Tarboton, D.G., 2005. Terrain analysis using digital elevation models (TauDEM). Utah State University, Logan, 3012, p.2018.
SlpDrainDens	SlpDrainDens	Drainage Density upstream of a given sub-reach (corrected for slope lengths)	12	Non-local	5	Value at sub-reach	Geomorphic indicator of hydrologic condition	sumupslpln_km / DA	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Carlston, Charles William. *Drainage density and streamflow*. US Government Printing Office, 1963.

SPI	SPI	Stream Power Index	67	Local	5	Value at sub-reach	Topographic control on hydrologic condition	CA = catchment area, G = slope gradient (degrees), SPI = ln(CA*tan(G))	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrological Processes, 5(1), 3â€“30. https://doi.org/10.1002/hyp.3360050103
SumUpSlpLen_km	$L_{Channel}$	Total Upstream Channel Length (km)	8	Non-local	5	Upstream Condition	Geomorphic indicator of hydrologic condition	Sum of upstream segment lengths in the NHD flowline hydrography; includes correction for topographic slope along which the stream flows	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None

tmax5_yr_m_0	tmax5_yr_m_0	May Tmax Reference Year	78	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily maximum temperature for August of a given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmax5_yr_previousyear	tmax5_yr_previousyear	May Tmax August of Year Previous to Reference	55	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily maximum temperature for August occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmax5_yr_previous2years	tmax5_yr_previous2years	May Tmax August Two Years Previous to Reference Year	79	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily maximum temperature for August occuring two years prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmax8_yr_m_0	tmax8_yr_m_0	August Tmax Reference Year	92	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily maximum temperature for August of a given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmax8_yr_previousyear	tmax8_yr_previousyear	August Tmax August of Year Previous to Reference	80	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily maximum temperature for August occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmax8_yr_previous2years	tmax8_yr_previous2years	August Tmax August Two Years Previous to Reference Year	70	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily maximum temperature for August occuring two years prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmin5_yr_m_0	tmin5_yr_m_0	May Tmin Reference Year	19	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily minimum temperature for August of a given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmin5_yr_previousyear	T_{Min_May-1}	May Tmin of Year Previous to Reference	9	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily minimum temperature for August occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmin5_yr_previous2years	tmin5_yr_previous2years	May Tmin Two Years Previous to Reference Year	28	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily minimum temperature for August occuring two years prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmin8_yr_m_0	T_{Min_Aug}	August Tmin Reference Year	3	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily minimum temperature for August of a given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmin8_yr_previousyear	tmin8_yr_previousyear	August Tmin of Year Previous to Reference	13	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily minimum temperature for August occuring one year prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

tmin8_yr_previous2years	tmin8_yr_previous2years	August Tmin Two Years Previous to Reference Year	22	Non-local	800	Drainage Area Average (FCPG)	Climatic control on hydrologic condition and biological suitability	mean daily minimum temperature for August occuring two years prior to the given year	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
tp_diss_ratio	tp_diss_ratio	Local Dissection relative to Intermediate	65	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	tp_dissection/tip_dissection 2	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None

tp_dissection	tp_dissection	Topographic Dissection - Local Scale (5 m)	57	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	e_max = maximum elevation in a given area, e_min = minimum elevation in a given area, e_rel = e_max - e_min, tp_dissection = e_rel/e_max	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Nir, D. (1957). The Ratio of Relative and Absolute Altitudes of Mt. Carmel: A Contribution to the Problem of Relief Analysis and Relief Classification. Geographical Review, 47(4), 564â€“569. https://doi.org/10.2307/211866
tp_dissection2	tp_dissection2	Topographic Dissection - Intermediate Scale (50 m)	61	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	e_max = maximum elevation in a given area, e_min = minimum elevation in a given area, e_rel = e_max - e_min, tp_dissection = e_rel/e_max	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Nir, D. (1957). The Ratio of Relative and Absolute Altitudes of Mt. Carmel: A Contribution to the Problem of Relief Analysis and Relief Classification. Geographical Review, 47(4), 564â€“569. https://doi.org/10.2307/211866

tp_rough_ratio	tp_rough_ratio	Local Roughness relative to Intermediate	51	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	tp_roughness/tp_roughness2	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None
tp_roughness	tp_roughness	Average Topographic Roughness - Local Scale (5 m)	52	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	e0 = elevation of a given cell, e_neighbor = elevation of a given neighboring cell, topographic rougness = square root(mean(e0 - e_neighbor of 8 neighbors)^ 2)	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Riley, Shawn J., Stephen D. DeGloria, and Robert Elliot. "Index that quantifies topographic heterogeneity." *intermountain Journal of sciences* 5, no. 1-4 (1999): 23-27.

tp_roughness2	tp_roughness2	Average Topographic Roughness - Intermediate Scale (50 m)	45	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	e0 = elevation of a given cell, e_neighbor = elevation of a given neighboring cell, topographic rougness = square root(mean(e0 - e_neighbor of 8 neighbors)^ 2)	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Riley, Shawn J., Stephen D. DeGloria, and Robert Elliot. "Index that quantifies topographic heterogeneity." *intermountain Journal of sciences* 5, no. 1-4 (1999): 23-27.
tp_sarelratio	tp_sarelratio	Surface Relief Ratio	85	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	surface relief = elevation of stream's source - elevation at stream mouth/stream length	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Evans JS, Oakleaf J, Cushman SA (2014) An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling, version 2.0-0. URL: https://github.com/jeffreyevans/GradientMetrics Accessed: 23 Feb 2025

tp_sarelratio2	tp_sarelratio2	Surface Relief Ratio - Intermediate Scale (50 m)	21	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	surface relief = elevation of stream's source - elevation at stream mouth/stream length	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Evans JS, Oakleaf J, Cushman SA (2014) An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling, version 2.0-0. URL: https://github.com/jeffreyevans/GradientMetrics Accessed: 23 Feb 2025
tp_slppos_ratio	tp_slppos_ratio	Topographic Slope Position Ratio	93	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	tp_slpposidx/tp_slpposidx2	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	None
tp_slpposidx	tp_slpposidx	Slope (%) Position Index - Local Scale (5 m)	71	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	elevation / focal mean of elevation	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Evans et al. 2014; ArcGIS Gradient Metrics Toolbox | evansspatial (evansmurphy.wixsite.com);
tp_slpposidx2	tp_slpposidx2	Slope (%) Position Index - Intermediate Scale 50 m)	77	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic	elevation / focal mean of elevation	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MO	Evans et al. 2014; ArcGIS Gradient Metrics Toolbox | evansspatial (evansmurphy.wixsite.com);

							condition		SAIC/ImageServer	

tp_vrm	tp_vrm	Vector Ruggedness Measurement	24	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	a= cell aspect, b = cell slope, orthogonal unit vector (v_hat) for each cell is broken into x, y, z components, xy=1*sin(a), x = xy*sin(b), y=xy*cos(b), z=1*cos(a), |r|= magnitude of neighborhood of vectors, |r|=square root of ((sum of x)^2 +(sum of y)^2+(sum of z)^2), VRM= 1-|r|/number of cells in neighborhood	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Hobson, R.D., 2019. Surface roughness in topography: quantitative approach. In *Spatial analysis in geomorphology* (pp. 221-246). Routledge.

tph_ge_3_2017	tph_ge_3_2017	Trees Per Hectare	41	Non-local	5	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	None	https://lemma.forestry.oregonstate.edu/data	Bell, David M., Matthew J. Gregory, Marin Palmer, and Raymond Davis. "Guidance for forest management and landscape ecology applications of recent gradient nearest neighbor imputation maps in California, Oregon, and Washington." *Gen. Tech. Rep. PNW-GTR-1018. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 41 p.(Online only).* 1018 (2023).
tphc_ge_3_2017	tphc_ge_3_2017	Trees Per Hectare Coniferous	66	Non-local	5	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	None	https://lemma.forestry.oregonstate.edu/data	Bell, David M., Matthew J. Gregory, Marin Palmer, and Raymond Davis. "Guidance for forest management and landscape ecology applications of recent gradient nearest neighbor imputation maps in California, Oregon, and Washington." *Gen. Tech. Rep. PNW-GTR-1018. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 41 p.(Online only).* 1018 (2023).

tphh_ge_3_2017	tphh_ge_3_2017	Trees per Hectare Hardwood	20	Non-local	5	Drainage Area Average (FCPG)	Landcover control on hydrologic condition	None	https://lemma.forestry.oregonstate.edu/data	Bell, David M., Matthew J. Gregory, Marin Palmer, and Raymond Davis. "Guidance for forest management and landscape ecology applications of recent gradient nearest neighbor imputation maps in California, Oregon, and Washington." *Gen. Tech. Rep. PNW-GTR-1018. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 41 p.(Online only).* 1018 (2023).
TPI_50_m	TPI_50_m	Topographic Position Index 50 m	26	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition - average small extent	TPI = elevation at given cell - mean elevation of neighborhood cells (50m neighborhood radius)	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Weiss, A.D., 2001. Topographic position and landforms analysis. Poster Presentation,Esri Users Conference, San Diego, CA.

TPI_500_m	TPI_500_m	Topographic Position Index 500 m	82	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition - average moderate extent	TPI = elevation at given cell - mean elevation of neighborhood cells (500m neighborhood radius)	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Weiss, A.D., 2001. Topographic position and landforms analysis. Poster Presentation,Esri Users Conference, San Diego, CA.
TPI_5000_m	TPI_5000_m	Topographic Position Index 5000 m	18	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition - average large extent	TPI = elevation at given cell - mean elevation of neighborhood cells (5000m neighborhood radius)	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	Weiss, A.D., 2001. Topographic position and landforms analysis. Poster Presentation,Esri Users Conference, San Diego, CA.

TWI	TWI	Topographic Wetness Index	36	Non-local	5	Drainage Area Average (FCPG)	Topographic control on hydrologic condition	a=flow accumulation area at a given point , tan(b)= local slope (radians), TWI= ln(a/tan(b))	https://gis.dogami.oregon.gov/arcgis/rest/services/LiDAR/DIGITAL_TERRAIN_MODEL_MOSAIC/ImageServer	BEVEN, K. J., & KIRKBY, M. J. (1979). A physically based, variable contributing area model of basin hydrology / Un modÃ¨le Ã base physique de zone dâ€™appel variable de lâ€™hydrologie du bassin versant. Hydrological Sciences Bulletin, 24(1), 43â€“69. https://doi.org/10.1080/02626667909491834
vpdmax5_yr_m_0	vpdmax5_yr_m_0	Average Maximum Vapor Pressure Deficit May of Reference Year	87	Non-local	5	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	mean of daily maximum vapor pressure deficit for the Current Calendar Year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

vpdmax5_yr_previousyear	vpdmax5_yr_previous year	Average Maximum Vapor Pressure Deficit May of Year Previous to Reference	59	Non-local	5	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	mean of daily maximum vapor pressure deficit for the Current Calendar Year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
vpdmax5_yr_previous2years	vpdmax5_yr_previous 2years	Average Maximum Vapor Pressure Deficit May Two Years Previous to Reference Year	62	Non-local	5	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	mean of daily maximum vapor pressure deficit for the year occuring prior to the Current Calendar Year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

vpdmax8_yr_m_0	vpdmax8_yr_m_0	Average Maximum Vapor Pressure Deficit August of Reference Year	60	Non-local	5	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	mean of daily maximum vapor pressure deficit for the Current Calendar Year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
vpdmax8_yr_previousyear	vpdmax8_yr_previous year	Average Maximum Vapor Pressure Deficit August of Year Previous to Reference	49	Non-local	5	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	mean of daily maximum vapor pressure deficit for the year occuring prior to the Current Calendar Year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

vpdmax8_yr_previous2years	vpdmax8_yr_previous2years	Average Maximum Vapor Pressure Deficit August Two Years Previous to Reference Year	76	Non-local	5	Drainage Area Average (FCPG)	Climatic control on hydrologic condition	mean of daily maximum vapor pressure deficit for the Current Calendar Year (CY)	https://prism.oregonstate.edu/	Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688
Water_Depth_fix	Water_Depth_fix	Depth to Water (m)	27	Non-local	30	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	None	https://www.arcgis.com/home/item.html?id=cdc49bd63ea54dd2977f3f2853e07fff	Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey Metadata. Available online at SSURGO-Metadata-Tables-and-Columns-Report.pdf. Last accessed 23 Feb 2025.
Water_Storage_fix	$W_{Storage}$	Water Storage (SSURGO) of the upper 150 cm of soil (mm)	10	Non-local	30	Drainage Area Average (FCPG)	Geologic control on hydrologic condition	None	https://www.arcgis.com/home/item.html?id=cdc49bd63ea54dd2977f3f2853e07fff	Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey Metadata. Available online at SSURGO-Metadata-Tables-and-Columns-Report.pdf. Last accessed 23 Feb 2025.

S2 Pseudo random spatial balancing

HUC12s were initially weighted based on the number of observations within a given HUC12. Spatial groups that include several HUC12s were then formed by randomly combining observations within HUC12s, accounting for their weighting, in order to prevent a spatial group of many HUC12s that have a large number of observations (Figure S2.1). Processing for this step and all subsequent steps outlined in Figure 2 occurred within the R ver. 4.3 (R Core Team, 2024). HUC12s boundaries were not used as the sole basis for spatial grouping because the software only allows 80 groups and there are 129 HUC12s. Furthermore, because observations per HUC12 vary from 1 to 376, there was a risk that using HUCs as individual groups could present a situation where models are evaluated against a very small test set, which could produce a very optimistic or very pessimistic estimation of predictive performance.

Pseudo-randomly grouping observations into 20 sub-groups of multiple HUC12s allowed for a sufficiently large population of groups for the purpose of randomly reshuffling the groups of training data during the spatial cross validation routine described later, while ensuring that minimum data per group were sufficiently large for producing representative test folds. Spatial balancing ensured there was a minimum of 76 observations within a group (maximum observations for a group was 624) which we assumed to be a sufficiently large dataset for the purposes of producing cross-validation performance metrics. The high variability of observations within groups was due to the non-uniform opportunistic sampling of observations within HUC12s that prohibited uniform group sizes while also having at least 20 spatial groups.

Spatial balancing was conducted with the oversampling function in the mlr3 package in R (Lang et al., 2019) by iterating over training data in each of the spatial groups. We termed this process spatial balancing because oversampling was applied on a group basis to achieve balance between numbers of wet and dry training data at the relatively small spatial scales of the groups in comparison to the overall size of the study area. Oversampling randomly duplicates minority class data to ensure the resulting model will produce a decision boundary between the two classes at 50% everywhere across space. Oversampling was chosen over the undersampling alternative (i.e., removal of majority class data) to ensure the fitted model captures as much information about the diverse landscape of the modeling domain as possible. We considered employing the Synthetic Minority Oversampling TEchnique (SMOTE), a method that synthesizes minority data resulting in outcomes that are less biased than random oversampling (Chawla et al. 2002); however, this approach was not integrated into the mlr3 package we employed. Adaptive Synthetic Sampling (ADASYN) is similar to SMOTE and generates synthetic minority class data with small errors to induce variation (He et al., 2008). Future iterations of this research will likely consider employment of ADASYN because it was recently integrated into the mlr3 package.

A potential consequence of oversampling the minority class is the possibility of inducing bias into the model by over representing a condition that is less prevalent on the landscape than the oversampled data suggests. Oversampling at the HUC12 level was considered but deemed problematic because some HUC12s contained fewer than 10 total observations, which may have

overbiased the model when replicating such a small dataset. Additionally, some HUC12s only contained observations from a single late-summer streamflow status (i.e., wet, or dry but not both).

Table S2.1. Summary of FLOwPER observations across years, Strahler Order, Ecoregion (Omernick and Griffith, 2014), and lithologic province (O'Connor et al., 2014).

	Wet	Dry	% Wet	% Dry	Total	% of Total
Year						
2019	328	206	61.4%	38.6%	534	21.2%
2020	906	833	52.1%	47.9%	1739	69.1%
2021	127	118	51.8%	48.2%	245	9.7%
	1361	1157	54.1%	45.9%	2518	
Strahler Order						
1	529	766	40.8%	59.2%	1295	51.4%
2	511	307	62.5%	37.5%	818	32.5%
3	249	71	77.8%	22.2%	320	12.7%
4	72	13	84.7%	15.3%	85	3.4%
	1361	1157	54.1%	45.9%	2518	
L3 Ecoregion						
Cascades	757	576	56.8%	43.2%	1333	52.9%
Klamath Mountains	204	300	40.5%	59.5%	504	20.0%
Coast Range	400	281	58.7%	41.3%	681	27.0%
	1361	1157	54.1%	45.9%	2518	
Lithologic Province						
High Cascades	359	194	64.9%	35.1%	553	21.8%
Klamath	182	261	41.1%	58.9%	443	17.5%
Western Cascades	409	410	49.9%	50.1%	819	32.3%
Coast Range sedimentary rocks	262	236	52.6%	47.4%	498	19.6%
Coast Range volcanic rocks and Columbia River basalt	149	56	72.7%	27.3%	205	8.1%
	1361	1157	56.2%	43.8%	2518	

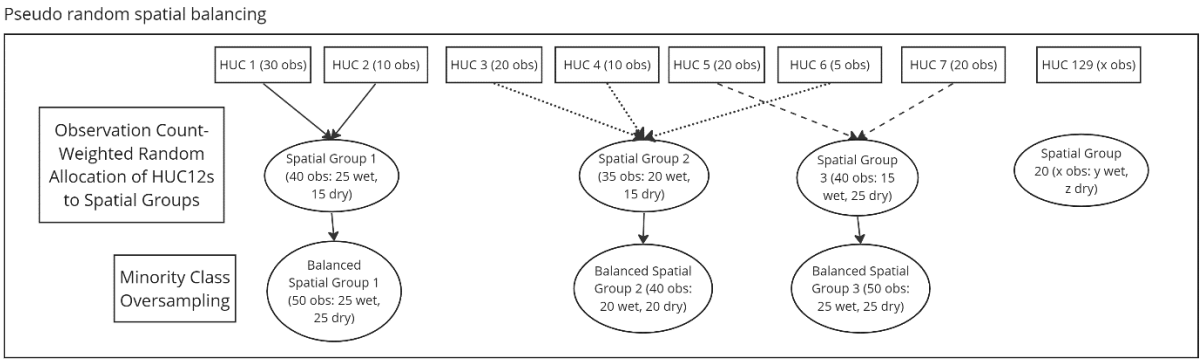

Figure S2.1. Schematic of pseudo random spatial balancing of HUC12s as part of training data processing. Allocation of HUC12s to spatial groups are approximately random but account for differences in data densities among different HUC12s to allow for relatively even number of observations within a spatial group.

S3 Cross validation methods that accounts for differences between spatially distal and spatially proximal data

There are numerous approaches to evaluating model performance, with k-fold cross validation and data splitting being among the most frequently used (Kuhn et al., 2013). We chose cross-validation for predictive performance evaluation over more conventional data splitting (e.g., 20% and 80% validation and testing split, respectively) because cross validation has been shown to be as effective as data splitting (Kuhn et al., 2013). Additionally, the repeated resampling of the data during cross validation prevents a single spurious grouping of spatial groups from overbiasing the result while also ensuring the full spatial variability of the data are used to evaluate model performance. The consequence of spurious selection that is possible with simple data splitting is the possibility of overly pessimistic or optimistic estimation of predictive performance that does not represent the average performance across the model domain (Roberts et al., 2017). One criticism of cross validation approaches is that they tend to overestimate variation resulting in large confidence intervals around a measure of central tendency; however, repeated cross-validation combined with bootstrapping of performance metrics have been shown to effectively mitigate this issue (Kuhn et al., 2013). The next section details the cross-validation routine that includes bootstrapping, accounting for proximal and distal validation strategies.

Cross validation for sub-watersheds with proximal data

We implemented a 5-fold cross validation approach to evaluating model predictive performance by ignoring spatial grouping assignment and randomly assigning training data into a 5-fold (completely random without resampling), ignoring spatial grouping, where each fold contains approximately the same amount of data (Figure S3.1). A model is fitted to 4 folds then tested on the holdout fold. This is repeated until all folds have been held out as an independent test set. In both the covariate selection procedure and the model development, independent tests were evaluated using the Matthews Correlation Coefficient (MCC). In this case, a test on the holdout

fold for 5 folds results in 5 MCC estimates. This process is repeated 30-more times with the random assignment inducing a stochasticity to the resulting MCC estimates. The result is 150 predictive performance estimates from which a median MCC and corresponding 95% confidence interval was assessed with 10,000 bootstrap resamples (with replacement). A criticism of this approach is that it is likely to be optimistic when data are spatially correlated (Tsamardinos et al., 2015; Roberts et al., 2017). This is because of the random assignment of data to folds that allows proximal data (i.e., neighbors) to be used to train and then test the model. However, the conventional cross validation approach may provide interesting insights from the standpoint of predictive performance because users of the resulting model will want to understand how the model is doing in areas that are proximal to the model training data. Because the cross-validation routine is free to choose train and test data the model, we interpret the results as quantifying predictive performance proximal to the model training points. We use sub-watershed (HUC12) as a proxy for proximal data meaning predictions are considered proximal when they occur within a sub-watershed containing data that constrained the model. This assumes conditions within a watershed are more similar than conditions between watersheds.

Cross validation for sub-watersheds with distal data

To provide insight on model predictive performance outside of sub-watersheds containing data, i.e., predictive performance in spatially distal settings, we implemented a spatial cross validation approach using grouping. Spatial groupings of data are held out for testing during the cross validation, and as such, predictive performance on those holdouts is representative of performance across components of space for which the training data did not inform the model. Spatial cross validation was implemented exactly like the conventional cross validation of the proximal data with the exception being that data were randomly assigned to each of the 5-folds based on their spatial grouping described in SI 2 such that data within an individual balanced spatial group was assigned to only one fold compared to the potential for data from a spatial group being assigned to more than one fold as is the case with the proximal data cross evaluation (Figure S3.1). The same routine as the proximal data of model fitting using one holdout fold including repeats and bootstrapping methods to achieve a median and 95% confidence test evaluation was applied. This resulted in the hold out test data not having observations from within a sub-watershed.

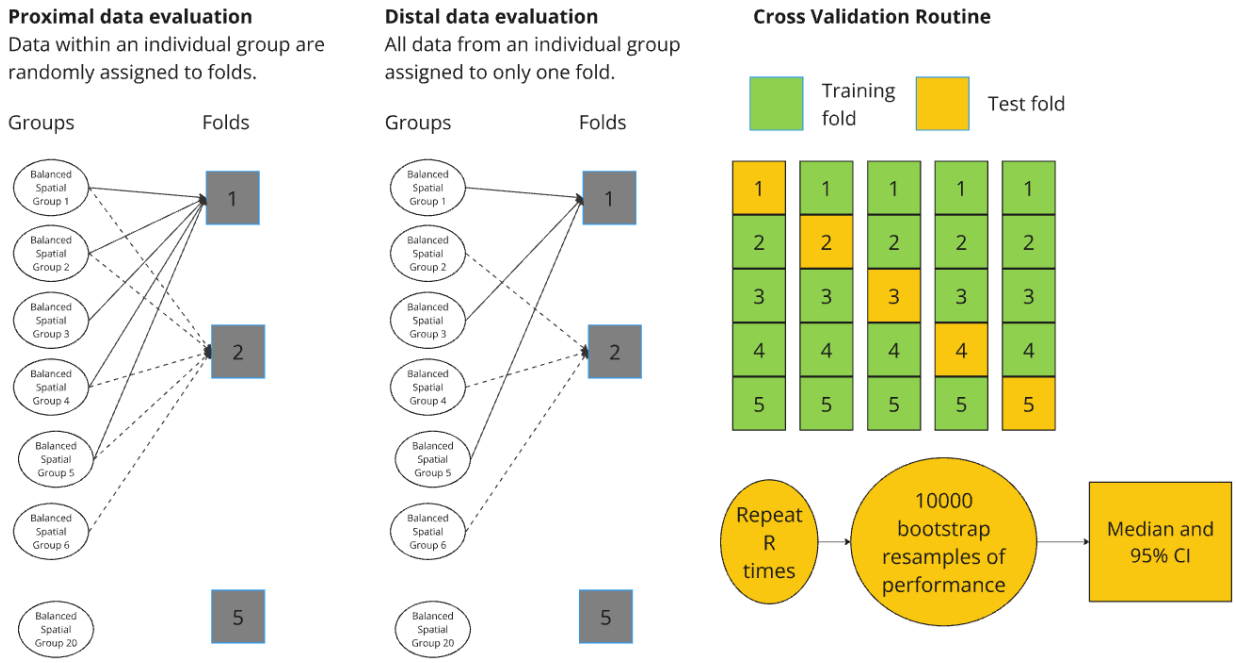

Figure S3.1. Assignment of spatially balanced groups into folds to conduct a distal validation strategy that mitigates the influence of spatial dependence on estimates of prediction accuracy by ensuring model fit data are from different groups than the model test data.

S4 Covariate Selection Methods

Covariate selection is a method for identifying and ranking covariates in order of importance to reduce model dimensionality (Bommert et al., 2020). Covariate selection provides selection of a subset of the most important covariates prior to the more computationally intensive model fitting and evaluation phase. Covariate selection approaches vary depending on the objective of the filter and can result in vastly different rankings of covariates. In our application, covariate selection establishes an order of covariate importance that ideally front loads the most important covariates for predicting late summer streamflow status and thus allows for identification of parsimonious model without the arduous process of tuning and fitting models to all variables.

Five covariate selection methods were implemented and were ranked using Random Forest and MCC (Figures S4.1, S4.2). Three conventional methods were employed: Boruta, Joint Mutual Information Maximization (JMIM), and Permutation covariate selection (Kursa and Rudnicki, 2010; Speiser et al., 2019; Bommert et al., 2020). We implemented two additional ranking approaches that are modified spatial approaches to mitigate the influence of spatial bias from clustered observations within sub-watersheds. For example, HUC12 170900040401 has over 300 observations. A first approach, Spatial-1- Random Forest, more explicitly considers spatial autocorrelation (Pihur et al., 2009). The second approach also accounts for spatial autocorrelation but uses the Boruta algorithm (Pihur et al., 2009).

For each of the two additional spatial covariate selection approaches, 25 resamples were drawn without replacement by random selection of 80% of the 20 spatial groups of training data so that each resample evaluates slightly different combinations of spatial groupings of data. Only 25 resamples were chosen to keep processing time on the rank aggregation stage to a reasonable level because this stage utilizes Monte Carlo simulation. The Spatial-1- Random Forest method generates a Random Forest model for each resample. The variable importance rankings are then extracted from the 25 models and aggregated ranking (Pihur et al., 2009) is performed to establish a final ranking of all covariates. In Spatial-2 – Boruta, the Boruta algorithm (Kursa and Rudnicki, 2010) is run on each of the 25 resamples to produce 25 rankings of covariates. Final rankings of the 96 covariates are derived from the 25 Boruta rankings by using rank aggregation (Pihur et al., 2009). Boruta utilizes bootstrap resampling and Random Forest to conduct an evaluation of variable importance, however, the internals of Boruta do not account for spatial grouping, so substantial differences between Spatial-1 and Spatial-2 are expected. Twenty-five rankings of the 96 variables are then evaluated with the aggregated ranking function 'RankAggreg' ver. 0.6 (Pihur et al., 2009) in R over 100 Monte Carlo simulations to produce a ranking of the 96 features.

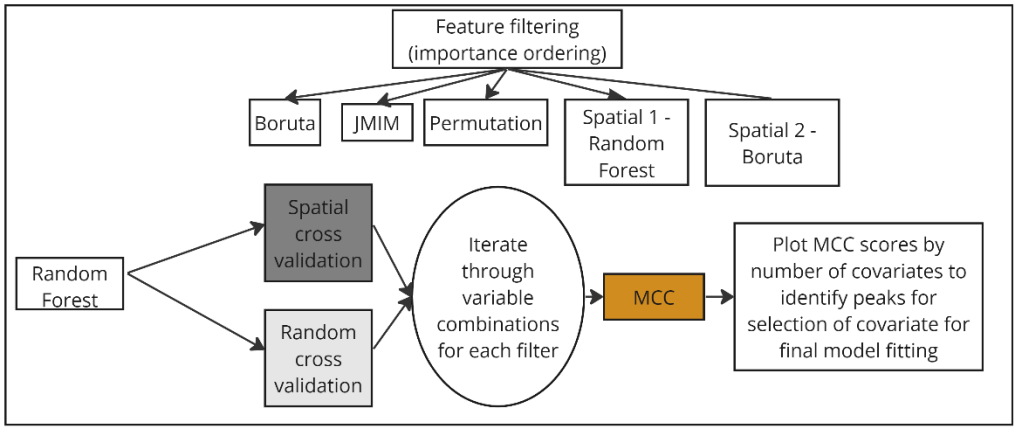

Figure S4.1. Workflow for covariate selection.

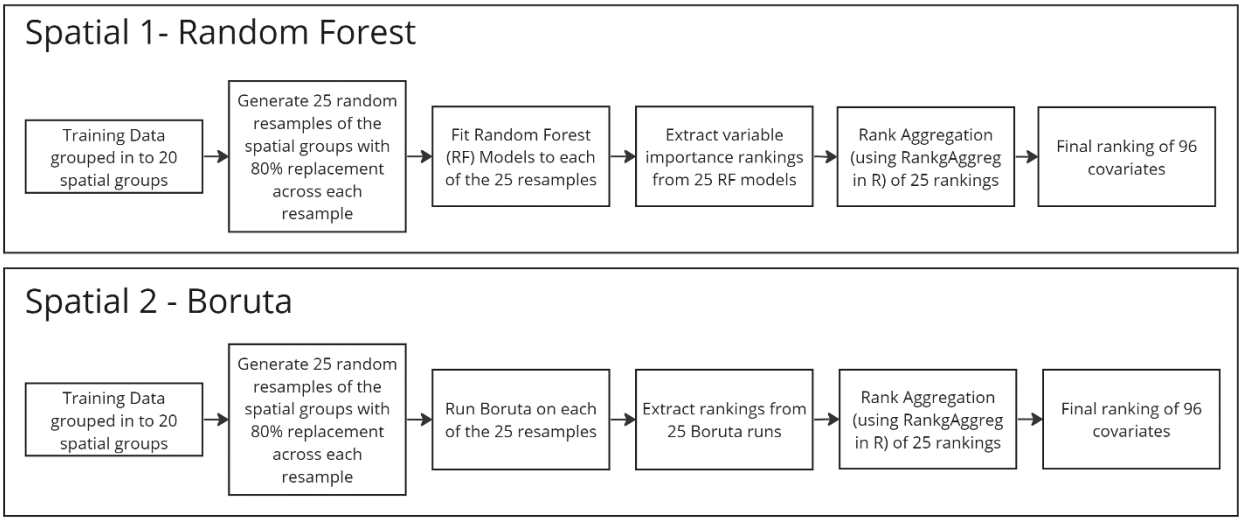

Figure S4.2. Process for Spatial-1 – Random Forest (top) and Spatial-2 – Boruta (bottom) covariate selection methods.

S5 Model Development

Model Algorithms

Logistic Regression (LR) was implemented with the 'glmnet' package (Friedman et al., 2010) and is a parametric statistical model used for binomial classification problems. The approach uses a logit transformation to cast a categorical response as continuous 0 to 1 as a function of log of the odds (Hosmer and Lemeshow, 1992). It has been used for modeling streamflow permanence in previous work (Jensen et al., 2018; Gendaszek et al., 2020).

Random Forest (RF) was incorporated because it has seen increased application in streamflow permanence modeling across a range of scales (Moidu et al., 2021; Penaluna et al., 2022; Sando et al., 2022). Random forest is a machine learning model algorithm that harnesses the power of randomization and large amounts of data to implement a technique called bagging that uses bootstrapped resampling of the available training data to generate an ensemble of weakly correlated decision trees that estimate the response variable (Breiman, 2001). Each decision tree contains different combinations of covariates and different decision splits for each covariate. The use of randomization and resampling minimizes variance and overfitting. Furthermore, RF trees account for interactions among covariates and due to the heavy use of randomization and bagging, are robust to multi-collinearity.

The Ranger implementation of random forest was used here because it uses parallelization to reduce processing times (Wright and Ziegler, 2017) and includes a jackknife procedure to estimate standard error of prediction, from which prediction interval can be computed (Wager et al., 2014). To induce additional randomization and further reduce the potential for overfitting, we parameterized the algorithm to build extremely random trees that choose random decision tree splits rather than the optimal split method in the conventional RF algorithm (Geurts et al., 2006). Furthermore, we also parameterized Ranger to use probability machines that use conditional probability functions to produce outcomes in terms of probability of class membership ranging from 0.0 to 1.0 instead of the more conventional binary outcome (Malley et al., 2012). To minimize the potential for overfitting, we used a high number of trees (2000), set the resampling to bootstrap, tuned the 'mtry' parameter (the number of variables to consider at each decision tree node split) using a nested loop, and established a minimum node size of ten, which is the default hyperparameter value for probability RF. Tuned 'mtry' values were 3 and 12 for the distal and proximal approaches, respectively.

The broader class of algorithms under the gradient boosted machines umbrella have seen use in ecological applications (Abedi et al., 2022; Naghibi et al., 2020; Sahin, 2020; Yu et al., 2021), but as of this writing, there is no evidence of use in streamflow permanence modeling. XGB was implemented with the 'xgboost' package (Chen et al., 2022) and is a formulation of the concept of regularized gradient decent boosting which is an ensemble of decision trees. Rather than using bagging to produce decision trees in parallel as in Random Forest, XGB uses gradient boosting to produce a series of shallow decision trees where each iteration uses the residual errors from

the previous decision tree to fit the next (e.g. gradient descent). The objective of XGB is to minimize bias and model underfitting (Friedman, 2001, 2002; Natekin and Knoll, 2013). XGB differs from conventional gradient boosted machines (Friedman, 2001) by implementing a regularization method to mitigate overfitting (Vinayak and Gilad-Bachrach, 2015). Our nested cross validation routine tuned nine hyperparameters that evaluated all possible combinations of values between the lower threshold and upper threshold, increasing at fixed step sizes (Table S5.1). Our parameterization further minimized the overfitting potential by incorporating an early stopping routine within the Nested Spatial Cross-validation routine. The early stopping routine stops searching for optimal combinations of hyperparameters for a given tree when accuracy of the training subset converges with accuracy on an independently withheld test set. Compared to Random Forest, XGB tends to be computationally faster, more robust to imbalances in the training response data, and due to the use of gradient descent, has the potential to achieve higher prediction accuracy if the dataset is sufficiently large.

Table S5.1 – XGB Hyperparameter Search Table

Hyperparameter	**Lower Threshold**	**Upper Threshold**	**Step Size**
nrounds	700	5000	100
max_depth	1	20	1
colsample_bytree	0.2	1	0.1
colsample_bylevel	0.2	1	0.1
gamma	0	1	0.1
min_child_weight	1	10	1
eta	0.01	0.1	0.01
lambda	0.01	1000	0.01
subsample	0.01	1	0.01

Model Performance Tests

Model performance was evaluated under both proximal and distal validation that each included a nested routine within the cross-validation process (Figure S5.1). Nesting allows for tuning hyperparameters as part of model optimization but is distinct from evaluation of the model through performance tests. Nesting substantially increases processing time but is necessary to preserve the separation between the cross-validation testing and the model evaluation (Bischl et al., 2012). Performance tests included MCC, Negative Predictive Value (NPV), Positive Predictive Value (PPV), Accuracy (ACC), Receiver Operator Characteristic Area Under the Curve (AUC), and Precision-Recall Area Under the Curve (PRAUC) (Figure 2).

Negative Predictive Value is defined as.

$$NPV = \frac{TN}{TN+FN} \qquad \text{Eq. S1}$$

Positive predictive value is precision and defined as

$$PPV = \frac{TP}{TP+FP} \qquad \text{Eq. S2}$$

Where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

Accuracy is defined as

$$ACC = \frac{TP+TN}{P+N} \qquad \text{Eq. S3}$$

Where P is the number of real positive cases in the training data, and N is the number of real negative cases in the training data.

Receiver Operator Characteristic Area Under the Curve (AUC) is a plot of the True Positive Rate (TPR; sensitivity) and the False Positive Rate (FPR; 1- specificity) and a measure of discriminative ability at all thresholds. An AUC of 1 indicates no overlap between the TPR and FPR. An AUC of 0.5 indicates complete overlap of TPR and FPR and that the model has no ability to discern between wet and dry classes. An AUC of 0 indicates that the model predicts wet for all dry classes and dry for all wet classes.

$$TPR = \frac{TP}{TP+FN} \qquad \text{Eq. S4}$$

$$FPR = \frac{FP}{TN+FP} \qquad \text{Eq. S5}$$

Finally, Precision-Recall Area Under the Curve is a plot of the PPV (Precision, Eq. S2) and Recall is TPR (Sensitivity, Eq. S4). Although PRAUC scalar values are on the 0-1 scale as the AUC, PRAUC is better suited to imbalanced data where false positives are more significant than false negatives because precision is the numerator. Since the positive value here is “wet” PRAUC is the proportion of Wet, precision is the proportion of “wet” predictions that were truly “wet and recall is the proportion of “wet” predictions out of all positive instances in the data. A value of 1 suggests that the model has perfect precision and recall at all thresholds

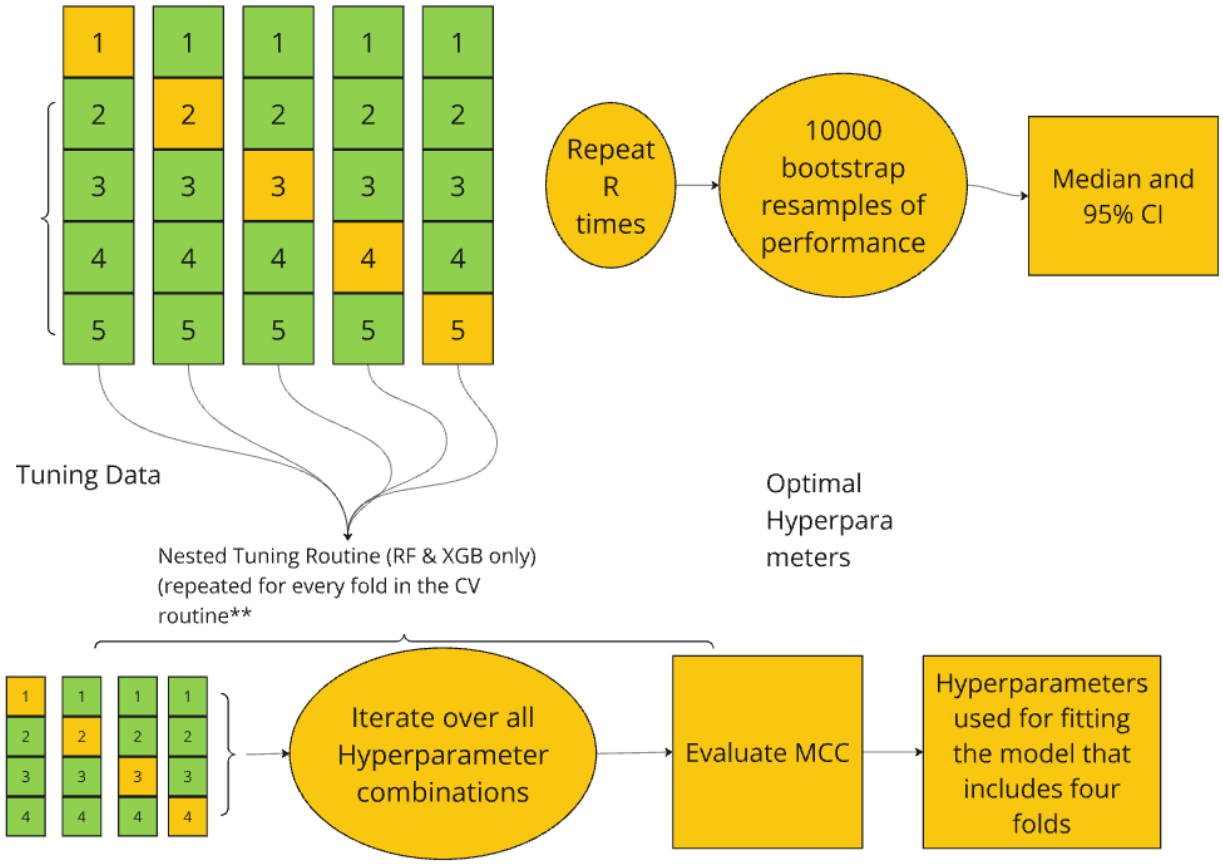

Figure S5.1. Nested routine within the random (proximal data) and spatial (distal data) cross validation.

S6 Extrapolation Detection Model

Extrapolation of the WOWTDR model is a potential cause for concern. The underlying Random Forest decision trees that comprise WOWTDR do not extrapolate because decision trees are not functions and thus are not able to describe how probability of a response condition might change as a function of the increase or decrease of an associated covariate condition. As such, interpretations of WOWTDR model predictions in extrapolation zones need to be considered more carefully. We tested the use of both Isolation Forests (IFs) and Support Vector Machines (SVMs) (Al Farizi et al., 2021) to determine where on the landscape WOWTDR predictions are occurring outside of the model training data. While these models are designed to identify data outliers, in our use case, we consider outlier detection with extrapolation detection because it indicates the level of similarity between the simultaneous expression of covariates at a given point and the overall training dataset. Hereafter, use of the word outlier is considered synonymous with extrapolation.

IFs split variables at random locations across the range of randomly selected covariates, isolating the point. The number of splits it takes to isolate a point from the rest of the data is an indication of 'outness' and therefore considered an outlier, such that a small number of splits represents an outlier; many splits represent a smaller likelihood of being an outlier. SVMs work by creating a

‘hull’ in n-dimensional space to encapsulate the covariate expression ranges described by the training data. The SVM algorithm then calculates eigenvalues and eigenvectors as the distance between a given location characterized by a combination of covariate information and the edge of the hull. Larger positive distances indicate more ‘inness’ inside the hull and thus are not outliers and increasingly negative distances exhibit more ‘outness’, indicating data are outside of the range captured by the training data and thus represent locations where the WOWTDR model is extrapolating.

To evaluate how often and where extrapolation occurs in the WOWTDR model, we developed an extrapolation detection model (Figure S6.1). The flow permanence model training data were split into a 90%/10% training and testing dataset. Training data were rescaled to reduce extrapolation detection bias on any particular variable by centering data on the median and scaling by median absolute deviation.

SVM and IF extrapolation detection models were fit to the training dataset for the purpose of extrapolation identification. SVM and IFs were tested against different variations of the available data designed to evaluate the effectiveness of each model for extrapolation identification under different scenarios as follows, additional details are provided in the Methods section below:

> (1) 10,000 random resamples with replacement of the training data to examine accuracy when no new information is considered, (2) Data from 10,000 resamples with replacement of the hold out test data to evaluate accuracy on data from the same population the training data were drawn from but excluding the training data, (3) 10,000 resamples with replacement of the original training data where covariate data associated with data points are randomly reshuffled amongst the data to examine accuracy when new combinations of covariates are presented without changing numerical values, (4) 10,000 simulated data points generated using the inverse transformation sampling method on the empirical cumulative distribution function of the training data to examine accuracy with a new dataset simulated on covariate composition and distribution from the original training data, and (5) 10,000 simulated data points that contain a 50% split of non-outliers and outliers where outliers were generated by increasing values by 0.25 median absolute deviations.

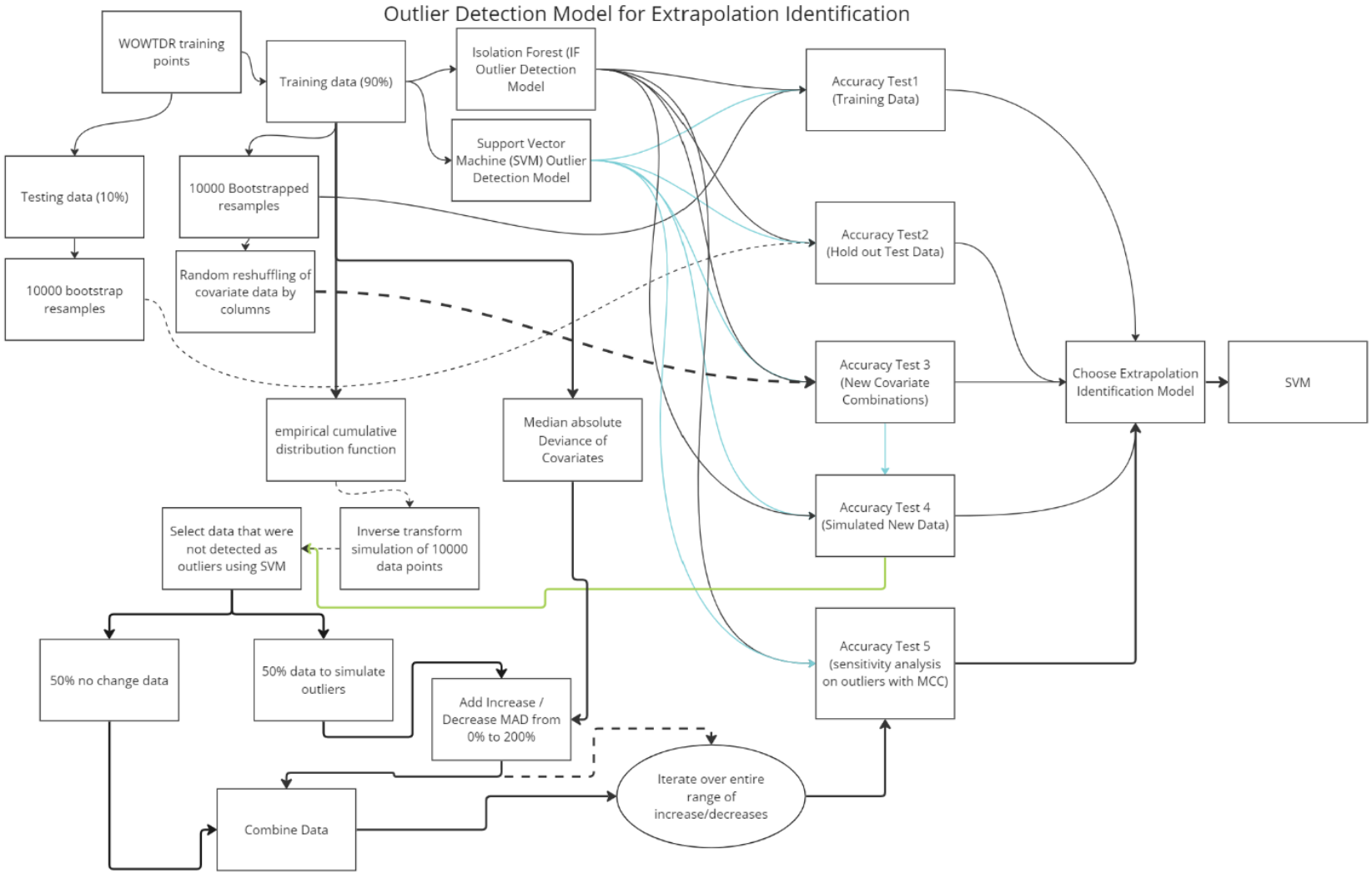

Figure S6.1. Workflow of extrapolation model.

Methods

Fitting the models

Unlike isolation forests (IFs), SVMs operate on distributional assumptions of the underlying data. As a result, the training data for IFs and SVMs were transformed by centering to median of the respective covariate and scaling to the median absolute deviation (MAD) of the respective covariate (Table S6.1). The result is transformed data where values are distance from median measured in units of median absolute deviation using the RobScale function in r (Signorell, 2023). This is a more robust approach than mean centering and standard deviation scaling, particularly when data are non-normally distributed as with the data in this study. Log transformation of specific variables exhibiting significant skew (e.g., flow accumulation) was considered but determined to be unnecessary because outlier detection performance during preliminary testing produced similar results with and without transformation. Covariate-specific centers and scaling coefficients produced during robust scaling were retained for the purpose of rescaling data used to generate predictions.

Table S6.1. Robust Scaling Values centered to median and scaled to median absolute deviation (MAD).

Covariate	**Median**	**MAD**
Proportion Canopy	0.74	0.09
P_Annual	1636.09	727.57
DA	0.15	0.17
Elev	625.23	350.03
Elev_Normalized	-0.06	1.6
Tmin_Aug	11.52	0.92
Downstream Channel Slope_50m	0.21	0.15
Downstream Channel Slope_1km	0.09	0.08
Length_Hillslope	0.13	0.07
Length_Channel	0.54	0.64
Tmin_May_previousyear	6.33	1.08
WaterStorageCapacity	17.57	4.69
Hydraulic Conductivity	9	10.1

SVM parameters:
Nu = 0.002
Kernel = rbfdot
Tol = 0.00001

Support Vector Machine (SVM) and Isolation Forest (IF) extrapolation detection models were fit to the training dataset for the purpose of extrapolation identification. The IF tree is too complex to provide here, but the SVM parameters are provided in the table above.

Both the SVM and the IF model were evaluated for extrapolation detection suitability using five different tests. The first four tests use different data drawn from the original observation data using different methods to determine if the models were generalized enough to adequately discriminate outliers in new data. In test 1 through 4, no outlier data are presented so errors are representative of the false negative rate (how frequently data that should be 'in' are determined by the model to be 'out'). All test outcomes are evaluated with simple accuracy to simplify interpretation on the scale of 0.0 to 1.0 with scores of 0.5 representing truly random performance, and scores below 0.5 representing model outcomes that are more consistently inverted from truth. Simple accuracy is appropriate here because for the first four tests, all data should be "in" and in the case of test 5, the data are balanced.

Test 1 and Test 2 examine how frequently the model misclassifies 'in' data as 'out'. Because none of the data are outliers, a low accuracy on either would indicate a poorly parameterized model that detects outliers within the 'in' distribution and a lower accuracy Test 2 than Test 1 would suggest that the models are highly constrained to the training data and likely not suitable for evaluating whether new data represent extrapolations beyond the original training domain.

Test 3 uses the training data to show the influence, if any, of randomly reshuffling covariate expression combinations. All data were labeled 'in' for this. Low accuracy on Test 3 suggests a model that is calibrated to the unique multi-variate combinations in the training data and not be generalizable to new information. On the other hand, a high accuracy score indicates the extrapolation detection model is largely insensitive to unique multi-variate combination and instead captures a more generalized boundary around the data.

Test 4 uses simulated data drawn from the training set distribution to demonstrate outlier detection performance on new data drawn from the same distribution. All data were labeled 'in' because they were ultimately derived from the training data. Data not labeled 'in' A low accuracy on this test would be indicative of an outlier detection model that is over-fit to the data because it does not accept data from a similar distribution. The simulation methodology to generate data for Test 4 is described in more detail below.

Test 5 uses simulated data drawn from the training dataset distribution where half of the simulated data are artificially altered by incrementally increasing and incrementally decreasing median absolute deviations from the median. A relatively high split of 50% was chosen to ensure the outlier detection was being rigorously challenged and ensure a balanced accuracy assessment for the purpose of discerning a given model's ability to discriminate between outliers and non-outliers. High accuracy in Test 5 (arbitrarily established at 85%) would indicate an outlier model with good discriminatory power. Data from test 5 were incorporated into a sensitivity analysis for the purpose of understanding how sensitive each outlier detection model is to varying magnitudes and directionality of outliers, i.e., extreme conditions outside the upper and lower bounds of the original data distribution. Test 5 includes 50% outlier data at increasing positive

and negative magnitudes to characterize accuracy when new data are presented with half of the data known to be positive ('in') and half the data are known to be negative ('out').

Simulating data

Simulated data for tests 4 and 5 were generated using the inverse transformation sampling method (Henderson and Nelson, 2006; Harar et al., 2022) on the empirical cumulative distribution function (ECDF) of the training data. The ECDF was fitted using the 'ecdf' function in R 'Stats' package. Because the ECDF is non-continuous and a series of steps, the inverse transformation sampling method required the use of interpolation to generate continuous values along the ECDF (Harar et al., 2022). Inversion occurs first by fitting a linear interpolation model using the ECDF as the X values and the original data to the Y and taking 100000 evaluations of the interpolation model at 100000 randomly selected locations along the ECDF. Note that simulating data from a multivariate normal distribution was evaluated, but data were determined to not conform to multivariate normal distribution with sufficient agreement to produce meaningful simulations. Use of data simulated from a Generative Adversarial Network was tested, but is not shown here because resulting data were less congruent to the original data than using the inversion method described above. There was a high level of agreement between distribution of the simulated data for 13 predictor covariates (colored density histogram) compared to the original data (density plot indicated with black line). Figure S6.2 depicts the high level of agreement between distribution of the simulated data for four of the 13 covariates and is representative of the close agreement for the other nine covariates.

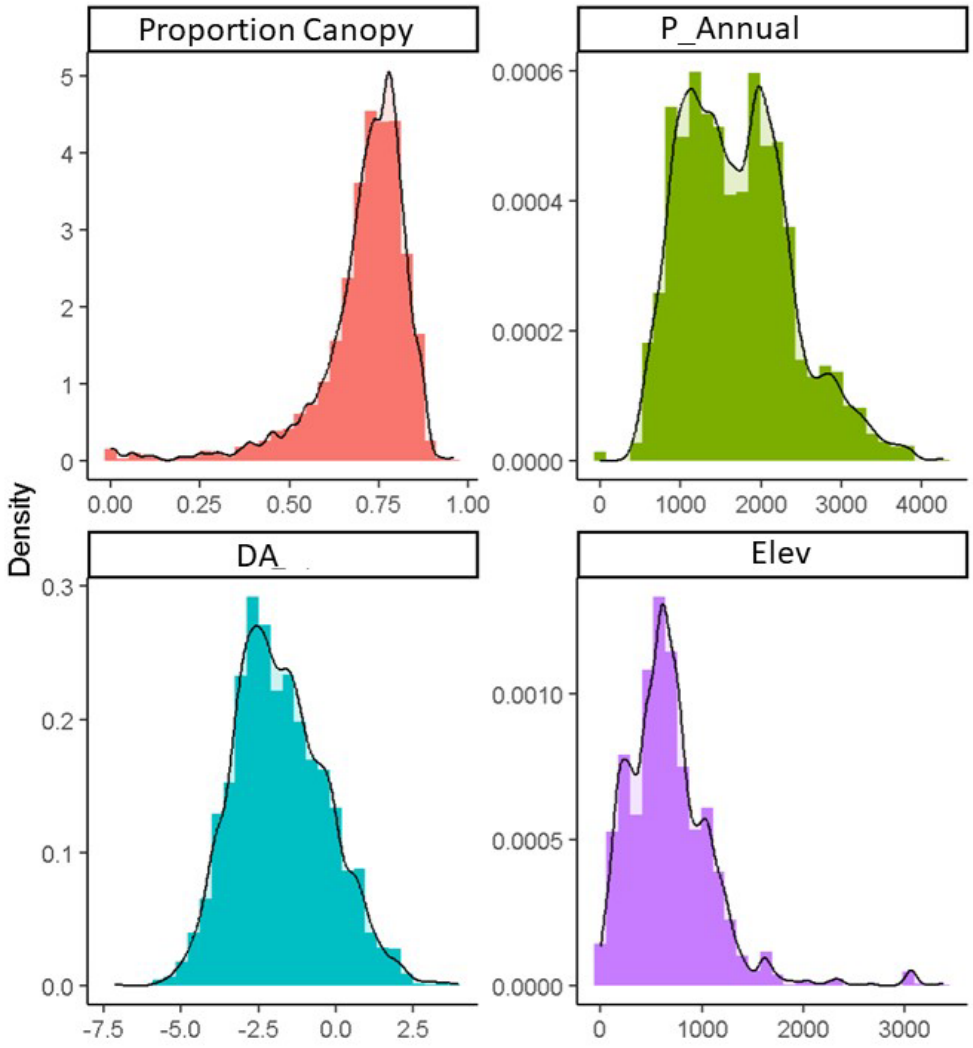

Figure S6.2. Density histograms of simulated data are overlaid with density plots of the original data for four of the most important covariates to demonstrate that simulated data are drawn from the same distribution.

Simulating outliers

Outlier simulation data for Test 5 were generated by identifying and isolating the simulated datapoints used in test 4 that were not erroneously identified as outliers, then splitting the available data in half. Erroneous outlier conditions in test 4 were removed for this evaluation, reducing the 10000 candidate datapoints to 9443 to prevent those outcomes from confounding the analysis. Half of the data points was unchanged and thus not outliers. The other half of the data points was multiplied by a specified percentage of each covariates respective MAD to produce artificially extreme data then added to the original data values to examine sensitivity under increasing magnitudes and subtracted from the original values to examine sensitivity to decreasing value magnitudes. To examine sensitivity of both the SVM and IF outlier detection models to outliers, increases to the process was repeated from 0 to 200% in increments of 1%. The result was then plotted to characterize accuracy as a function of a multiplier of MAD for

confirmation that simulated distributions are moving to the right without substantially changing the shape of the distribution. Figure S6.3 shows the simulated distribution for four of the predictor covariates when values are increased by 50% of the respective MAD. Simulations depict the expected shift to the right along the x-axis caused by the value increasing, with some deviation from the original shape due to covariates having discrete ranges due to naturally limiting upper ranges (e.g., 100% forest canopy cover) and in the case of drainage area, underlying transformations of the data.

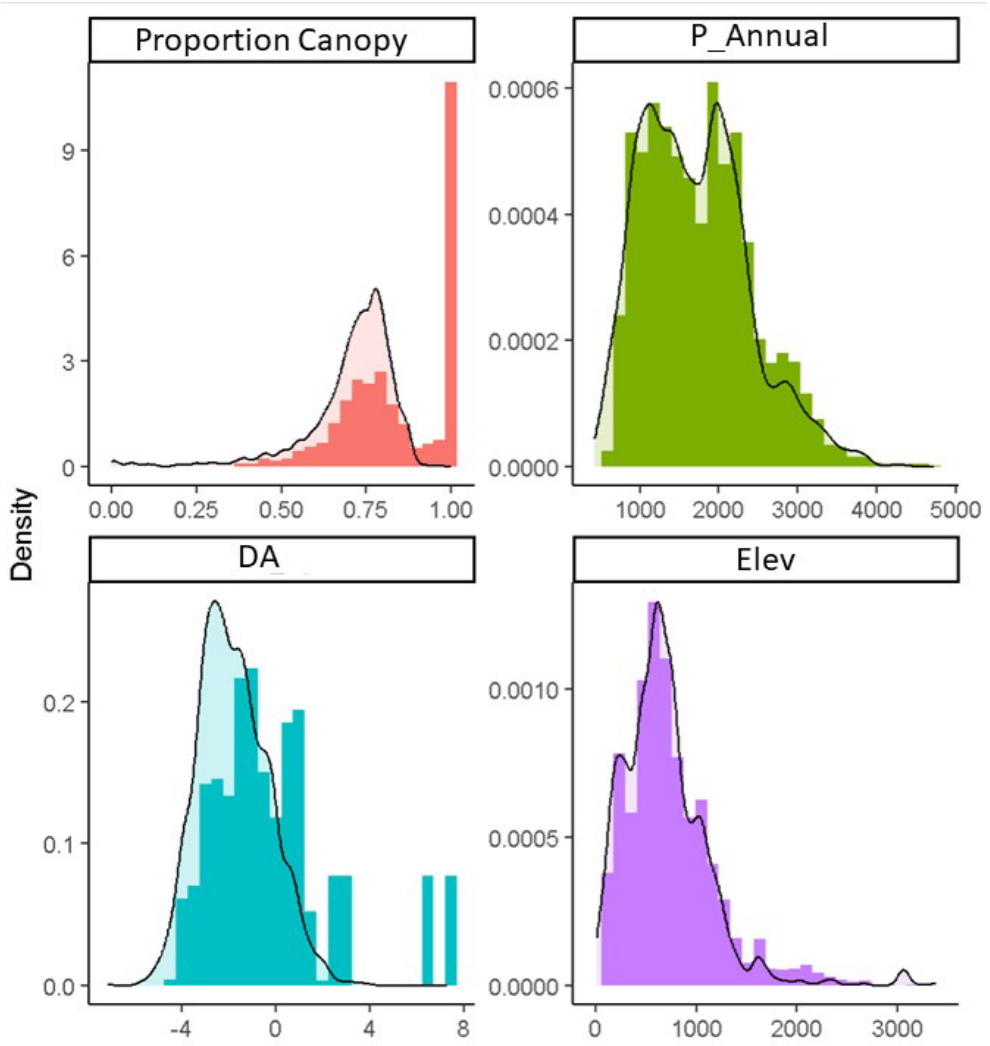

Figure S6.3. Density histograms of simulated outlier data are overlaid with density plots of the original data for four of the most important covariates to demonstrate that simulated data are drawn from the same distribution. For this figure the simulation where 50% of data include the addition of two times the median absolute deviation of the respective covariate appears. The most extreme of the outlier simulations is depicted to better visualize the shifting of the distribution to be outside of the original data distribution. Drainage area is log transformed here for visualization purposes only.

Results

Testing results show that both the SVM and IF outlier detection models accurately identified non-outlier data with greater than 90% accuracy with one exception (Table S6.2). Test 3 is the only test where SVM receives a lower score than IF, and the score is only 1% lower, indicating that SVM is slightly less capable at discerning outliers. However, the relatively high accuracy indicates that the multivariate combinations are accounted for in both models, with the SVM having detected a clearer boundary and thus being more sensitive to reshufflings of the covariates relative to the other covariates. In contrast, IF less accurately identified outliers during Test 4. Test 5 is the only test where known outliers were introduced to the detection models; SVM accurately identified 100% of outliers, and IF accurately identified 90% of outliers (Table S6). The high accuracy (> 85%) of both models for discriminating 'in' data suggests the simulated data is capturing the covariate domains expressed in the training data, and thus both models are appropriate for evaluating the extrapolation models. The lower outcomes of IF in Tests 4 and 5 are likely a consequence of the underlying decision trees in the IF being more constrained to the training data than the SVM. A comparison of the confusion matrices from Test 5 (Table S6.4; Table S6.5) shows that more IF prediction errors were false positives than false negatives, suggesting the IF model is more likely to erroneously identify a condition as an extrapolation than it is to erroneously identify a condition as not an extrapolation. The SVM false positive and false negative rates were much lower and more similar to each other. With SVM outcomes, trends are similar between the additive and subtractive cases. With IF outcomes, accuracy is much lower when reducing covariate values, likely because the IF is more narrowly constrained to specific combinations of covariate conditions observed in the training data. The 100% accuracy of SVM in Test 5 is consistent with the sensitivity analysis results (Figure S9) that depict SVM consistently achieving 100% extrapolation detection accuracy once covariate values are uniformly increased by 6.5% and is an artifact of removing data that appeared to be confusing for the SVM from Test 5 to prevent these datapoints from confounding the outcomes.

Table S6.2. Outlier detection accuracies of the Isolation Forest (IF) and Support Vector Machine (SVM) models for the four testing conditions using different variations on WOWTDR model training data as specified in the 'Data' column. For Test 5 results are shown for the 25% increase in Median Absolute Deviation.

Test	Model	Accuracy	True Outliers	Data
1	IF	0.95	0	Train
1	SVM	0.99	0	Train
2	IF	0.97	0	Test
2	SVM	0.99	0	Test
3	IF	0.96	0	Reshuffle
3	SVM	0.95	0	Reshuffle

4	IF	0.86	0	Simulated
4	SVM	0.95	0	Simulated
5	IF	0.90	50%	Simulated
5	SVM	1.00	50%	Simulated

Table S6.3. Isolation Forest Confusion Matrix output for Test 5 where 50% of the data are known outliers that have had values increased by 25% of the Mean Absolute Deviance from the Median for each covariate.

	Reference	
Prediction	FALSE	TRUE
FALSE	4322	532
TRUE	402	4191

Table S6.4. SVM Confusion Matrix output for Test 5 where 50% of the data are known outliers that have had values increased by 25% of the Mean Absolute Deviance from the Median for each covariate.

	Reference	
Prediction	FALSE	TRUE
FALSE	4724	0
TRUE	0	4723

Sensitivity analysis

The outlier detection capability of the SVM exceeds the capability of IF (Figure S6.4), which is likely because the IF needs additional tuning to improve performance. IF was not evaluated for sensitivity due to the acceptable level of outlier detection performance from the SVM. The SVM plateaus in performance when reaching a marked inflection in performance at an approximately 0.2 median absolute deviation increase then plateaus at an accuracy of 0.97 when 0.4 median absolute deviations have been added to the data. Trends are similar in the case of subtraction but with a more defined decline of accuracy when decreasing by less than 0.25 median absolute deviations. IF reaches a major inflection at about 0.45 median absolute deviations but continues to increase in accuracy to 2 median absolute deviations. This trend does not mirror in the case of subtraction. The highest accuracy IF achieves in the subtraction case is approximately 0.83 in the extreme case of values decreasing by 2 median absolute deviations, and likely because of IFs more sensitive nature evidenced in tests 1-5. Accuracy would likely continue to improve at larger magnitudes of change, but we only examined increases and decreases up to two MADs to keep the processing time reasonable while characterizing extrapolation detection across a broad range of conditions.

The testing methods presented here were intended to provide an indication of outlier detection performance across a broad range of potential covariate expressions. Test 2 suggests reasonably

high detection accuracy on "in" data, and Test 4 suggests very good detection of outliers when all covariate expressions are higher or lower than what is normal for the training data. However, the high accuracy score of Test 3 suggests that novel covariate expression combinations that are within the normal range of the training data may not be accurately characterized as 'out'. This concern is somewhat alleviated by the relatively low change in covariate values (+/- 0.25 MAD) necessary to be 'out' (Fig S6.5), which suggests this approach offers more fidelity on extrapolation detection over simply establishing an envelope from the minimum and maximum values of each covariate.

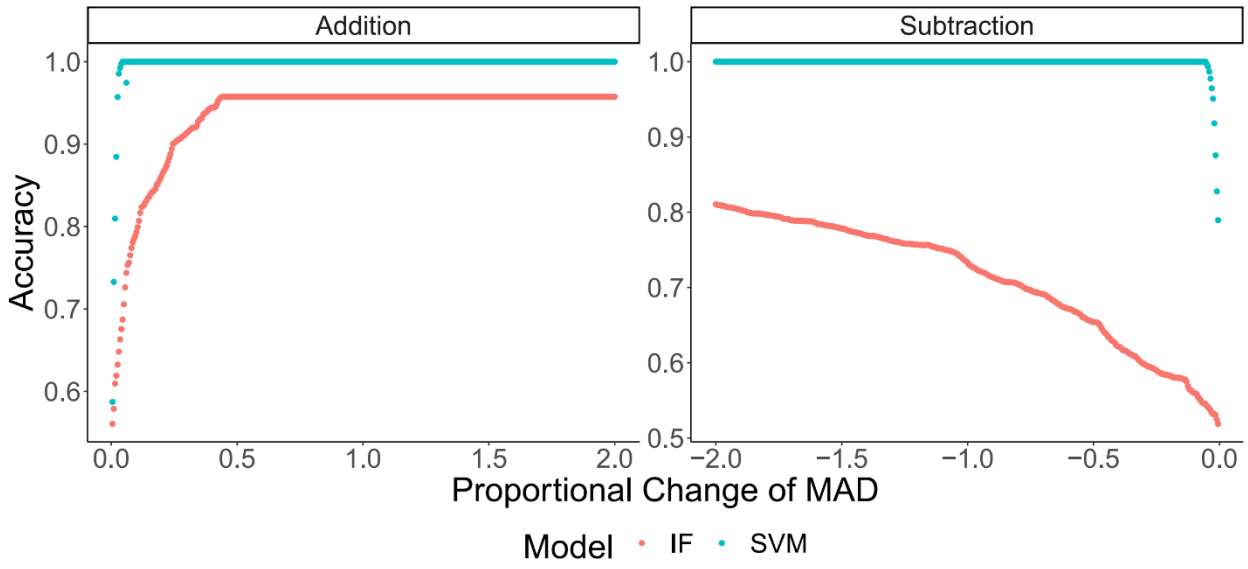

Figure S6.4. Sensitivity analysis of the two outlier detection models from the addition and subtraction of multiples of median absolute deviations (MAD) ranging from 0.005 to 2 (i.e. 0.5% to 200%).

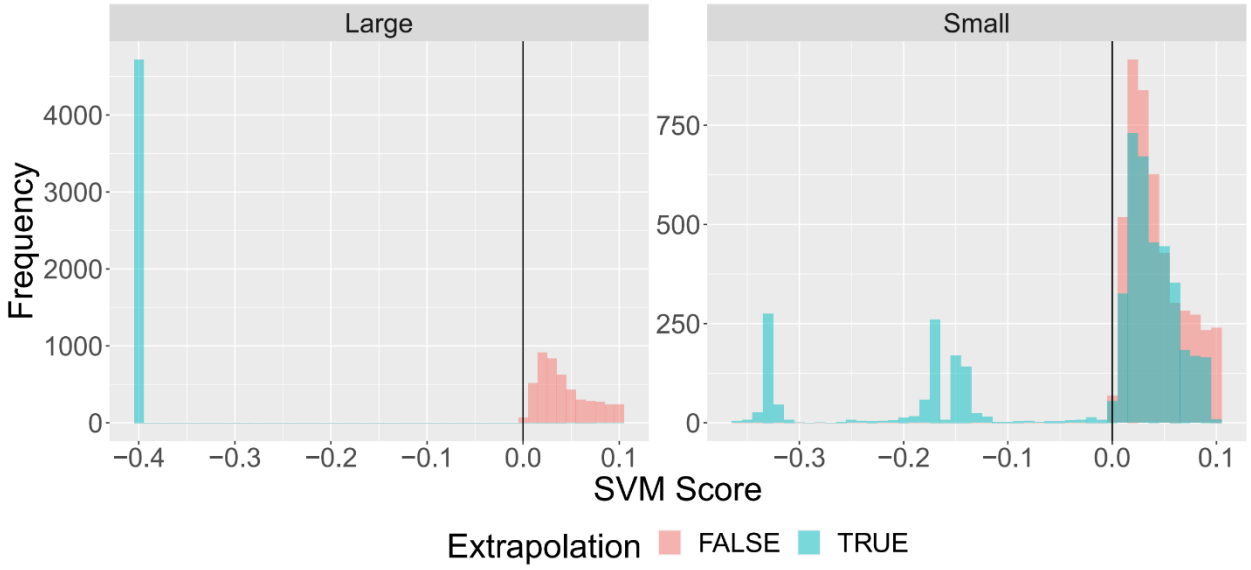

Figure S6.5. Histograms of the simulated data containing 50% outliers where outliers are simulated from original data by uniformly increasing covariate values by 0.25 Median Absolute Deviations (MADs) (left panel) and 0.005 Median Absolute Deviations (right panel). X axis is SVM estimated similarity score to original training data, where values < 0 are outliers. Histogram is colored blue for data points that are truly outliers and red where data points are not outliers. Note the overlap of blue and red on the left panel indicating that the outlier detection model incorrectly identified some outliers as being 'in' whereas the right panel contains no overlap, suggesting that the SVM outlier detection model can correctly identify all data as outliers when it is different by 0.25 or more MADs.

Extrapolation Identification in WOWTDR

The SVM model became the basis for extrapolation detection for WOWTDR because we determined that SVM was the best outlier detection model for our data. Extrapolation detection was conducted on every sub-reach during this process by rescaling the covariate data with the robust scaling method described previously to prepare it for introduction into the SVM using the SVM to estimate extrapolation by classifying the resulting score with the following logic: ISEXTRAPOLATION = TRUE when scores are < 0 and FALSE when scores are ≥ 0.

S7 WOWTDR Model Results

Table S7.1. Covariate selection results for all 96 covariates.
Column headings indicate the spatial covariate selection method, rank is the relative numerical ordering of covariates sequenced for the given covariate selection method. Cell values are covariate ManuscriptShortname from Table S1.

Rank	Boruta	Permutation	JMIM	Spatial 1	Spatial 2
1	DA	Tmin_Aug	ppt_sum_yr_m_2	E	Lchannel
2	Lchannel	TPI_500_m	elev_rescale	Enormalized	DA
3	elev_rescale	prof_curve	vpdmax8_yr_m_0	Tmin_Aug	Enormalized
4	E	vpdmax8_yr_m_0	ppt5_yr_m_0	G50m	ppt_sum_yr_m_2
5	ENormalized	ppt8_yr_m_0	pptdiffnrml	G1km	E
6	tmin5_yr_m_0	tmin5_yr_m_0	ppt8_yr_m_1	Lhillslope	elev_rescale
7	pptdiffnrml	tmin8_yr_m_1	ppt8_nrml	DA	ppt8_yr_m_0
8	PYear	tmax5_yr_m_0	Wstorage	Lchannel	d_slp100_m
9	ppt_sum_yr_m_2	Wstorage	PYear	TMin_May-1	pptdiffnrml
10	ppt8_yr_m_1	tp_slppos_ratio	vpdmax8_yr_m_1	Wstorage	PYear
11	H	ppt8_nrml	Lith_Prov	H	tmax8_yr_m_1
12	Wstorage	tp_dissection	Tmin_Aug	SlpDrainDens	vpdmax8_yr_m_0
13	TMin_May-1	H	ppt8_yr_m_0	tmin8_yr_m_1	tmin5_yr_m_0
14	Lhillslope	curve	ppt_nrml_sum	d_slp100_m	vpdmax8_yr_m_2
15	G50m	tph_ge_3_2017	E	PYear	ppt5_yr_m_1
16	SlpDrainDens	tphc_ge_3_2017	ppt5_yr_m_2	pptdiffnrml	tmax5_yr_m_2
17	vpdmax8_yr_m_0	elev_rescale	vpdmax5_yr_m_2	ppt8_yr_m_1	vpdmax5_yr_m_2
18	tmin5_yr_m_2	ppt_ratio_aug_yr	ppt_sum_yr_m_1	TPI_5000_m	ppt5_yr_m_0
19	vpdmax5_yr_m_2	hload	ppt8_yr_m_2	tmin5_yr_m_0	G1km
20	ppt5_yr_m_1	tmin8_yr_m_2	tmax8_yr_m_1	tphh_ge_3_2017	vpdmax5_yr_m_0
21	Tmin_Aug	d_slp100_m	tphc_ge_3_2017	tp_sarelratio2	tmin5_yr_m_2
22	ppt5_yr_m_2	TPI_50_m	tmax5_yr_m_0	tmin8_yr_m_2	vpdmax5_yr_m_1
23	G1km	vpdmax5_yr_m_1	aspect_trasp	aspect_cos	LHillslope
24	ppt_ratio_aug_yr	tp_roughness2	ppt_ratio_aug_yr	tp_vrm	tmax5_yr_m_1
25	pptpropnrml	NLCD4	Enormalized	HAND_1	G50m
26	vpdmax8_yr_m_1	ppt5_yr_m_1	slope_pct_taudem	TPI_50_m	ppt5_yr_m_2
27	d_slp100_m	SlpDrainDens	pptpropnrml	Water_Depth_fix	ppt8_nrml
28	cancov_hdw_2017	vpdmax5_yr_m_0	hload	tmin5_yr_m_2	ppt8_yr_m_1
29	ppt_nrml_sum	tmin5_yr_m_2	tmin5_yr_m_0	d_slp20_m	ppt8_yr_m_2
30	ppt_sum_yr_m_1	cancov_con_2017	vpdmax5_yr_m_1	slope_pct_taudem	vpdmax8_yr_m_1
31	d_slp30_m	aspect_cos	tmax8_yr_m_0	NLCD5	ppt_ratio_aug_yr
32	vpdmax8_yr_m_2	tp_dissection2	vpdmax8_yr_m_2	plan_curve	tphh_ge_3_2017
33	ppt8_nrml	Lith_Prov	tmax8_yr_m_2	elev_rescale	ppt_sum_yr_m_1
34	tmax8_yr_m_1	vpdmax8_yr_m_1	vpdmax5_yr_m_0	aspect_trasp	aspect_trasp
35	ppt5_yr_m_0	ppt5_yr_m_0	Lith_Prov3	ppt_ratio_aug_yr	d_slp20_m
36	d_slp20_m	SPI	H	TWI	Tmin_Aug

37	tmax8_yr_m_2	C%	Lith_Prov2	Lith_Prov5	tp_slpposidx
38	tmin8_yr_m_1	landsteward	Lith_Prov5	curve	ppt_nrml_sum
39	tmax5_yr_m_0	vpdmax8_yr_m_2	tp_vrm	aspect_sin	d_slp30_m
40	tphc_ge_3_2017	d_slp20_m	aspect_cos	ppt8_yr_m_0	tmin8_yr_m_2
41	tmax5_yr_m_2	DA	tmax5_yr_m_2	tph_ge_3_2017	tmax5_yr_m_0
42	tp_rough_ratio	tp_roughness	cancov_hdw_2017	NLCD10	H
43	vpdmax5_yr_m_0	Bedrock_Depth_fix	ppt5_yr_m_1	ppt5_yr_m_1	hload
44	tmax5_yr_m_1	TMin_May-1	aspect_sin	d_slp30_m	pptpropnrml
45	ppt8_yr_m_0	tp_rough_ratio	cancov_con_2017	tp_roughness2	tph_ge_3_2017
46	tmin8_yr_m_2	plan_curve	Bedrock_Depth_fix	cancov_con_2017	cancov_hdw_2017
47	tmax8_yr_m_0	tp_slpposidx2	tmax5_yr_m_1	Bedrock_Depth_fix	tp_sarelratio
48	tphh_ge_3_2017	ppt5_yr_m_2	Lith_Prov7	NLCD6	tp_dissection
49	vpdmax5_yr_m_1	vpdmax5_yr_m_2	tmin5_yr_m_2	vpdmax8_yr_m_1	tphc_ge_3_2017
50	tp_vrm	NLCD10	SlpDrainDens	prof_curve	tp_rough_ratio
51	cancov_con_2017	Water_Depth_fix	d_slp30_m	tp_rough_ratio	tp_slppos_ratio
52	tp_sarelratio2	ppt8_yr_m_1	TPI_50_m	tp_roughness	TPI_500_m
53	aspect_trasp	pptpropnrml	tph_ge_3_2017	NLCD7	tp_dissection2
54	hload	tphh_ge_3_2017	tmin8_yr_m_2	ppt8_nrml	plan_curve
55	tp_roughness	NLCD2	C%	tmax5_yr_m_1	tp_diss_ratio
56	tph_ge_3_2017	Lith_Prov7	SPI	hload	NLCD6
57	TPI_5000_m	NLCD7	Water_Depth_fix	tp_dissection	tmax8_yr_m_2
58	tp_sarelratio	cancov_hdw_2017	TMin_May-1	ppt5_yr_m_0	slope_pct_taudem
59	slope_pct_taudem	d_slp30_m	tp_sarelratio2	vpdmax5_yr_m_1	TWI
60	ppt8_yr_m_2	NLCD5	tmin8_yr_m_1	vpdmax8_yr_m_0	Wstorage
61	Bedrock_Depth_fix	TPI_5000_m	plan_curve	tp_dissection2	tp_slpposidx2
62	HAND_1	tmax5_yr_m_1	NLCD6	vpdmax5_yr_m_2	HAND_1
63	tp_dissection	tmax5_yr_m_2	tphh_ge_3_2017	ppt5_yr_m_2	Bedrock_Depth_fix
64	tp_dissection2	Lith_Prov2	tp_dissection	pptpropnrml	ProfCurve
65	aspect_cos	Lith_Prov1	d_slp20_m	tp_diss_ratio	tp_sarelratio2
66	aspect_sin	G50m	TPI_5000_m	tphc_ge_3_2017	tmin5_yr_m_1
67	TPI_50_m	Lith_Prov4	TWI	SPI	Lith_Prov5
68	tp_diss_ratio	tmax8_yr_m_2	tp_rough_ratio	ppt_sum_yr_m_2	tp_vrm
69	tp_roughness2	NLCD6	d_slp100_m	cancov_hdw_2017	TPI_50_m
70	tp_slppos_ratio	Lith_Prov3	Lith_Prov4	tmax8_yr_m_2	tmax8_yr_m_0
71	tp_slpposidx2	tp_sarelratio	G50m	tp_slpposidx	C%
72	tp_slpposidx	ppt_sum_yr_m_1	tp_dissection2	Lith_Prov3	SlpDrainDens
73	plan_curve	tp_vrm	landsteward	ppt_sum_yr_m_1	aspect_sin
74	TPI_500_m	aspect_trasp	prof_curve	Lith_Prov	tmin8_yr_m_1
75	SPI	LChannel	L3region	landsteward	TPI_5000_m
76	curve	PYear	TPI_500_m	vpdmax8_yr_m_2	cancov_con_2017
77	ProfCurve	ppt8_yr_m_2	tp_sarelratio	tp_slpposidx2	Lith_Prov3
78	NLCD6	tmax8_yr_m_0	HAND_1	tmax5_yr_m_0	Water_Depth_fix
79	C%	slope_pct_taudem	Lith_Prov1	tmax5_yr_m_2	aspect_cos

80	TWI	tmax8_yr_m_1	DA	tmax8_yr_m_1	tp_roughness
81	Lith_Prov2	L3region	tp_roughness2	ProfCurve	Lith_Prov2
82	prof_curve	tp_sarelratio2	curve	TPI_500_m	Lith_Prov7
83	Water_Depth_fix	G1km	ProfCurve	NLCD4	prof_curve
84	Lith_Prov4	ProfCurve	tp_slpposidx2	ppt_nrml_sum	tp_roughness2
85	Lith_Prov7	Lith_Prov5	Lchannel	tp_sarelratio	NLCD5
86	NLCD5	HAND_1	G1km	ppt8_yr_m_2	SPI
87	Lith_Prov3	E	tp_slppos_ratio	vpdmax5_yr_m_0	Lith_Prov4
88	Lith_Prov5	ppt_sum_yr_m_2	tp_roughness	C%	NLCD10
89	Lith_Prov1	TWI	tp_diss_ratio	NLCD2	landsteward
90	Lith_Prov	ppt_nrml_sum	NLCD10	L3region	Lith_Prov
91	NLCD10	tp_slpposidx	Lhillslope	Lith_Prov7	curve
92	NLCD7	pptdiffnrml	tp_slpposidx	tmax8_yr_m_0	NLCD7
93	NLCD4	aspect_sin	NLCD2	tp_slppos_ratio	L3region
94	landsteward	Enormalized	NLCD4	Lith_Prov1	Lith_Prov1
95	L3region	Lhillslope	NLCD5	Lith_Prov2	NLCD4
96	NLCD2	tp_diss_ratio	NLCD7	Lith_Prov4	NLCD2

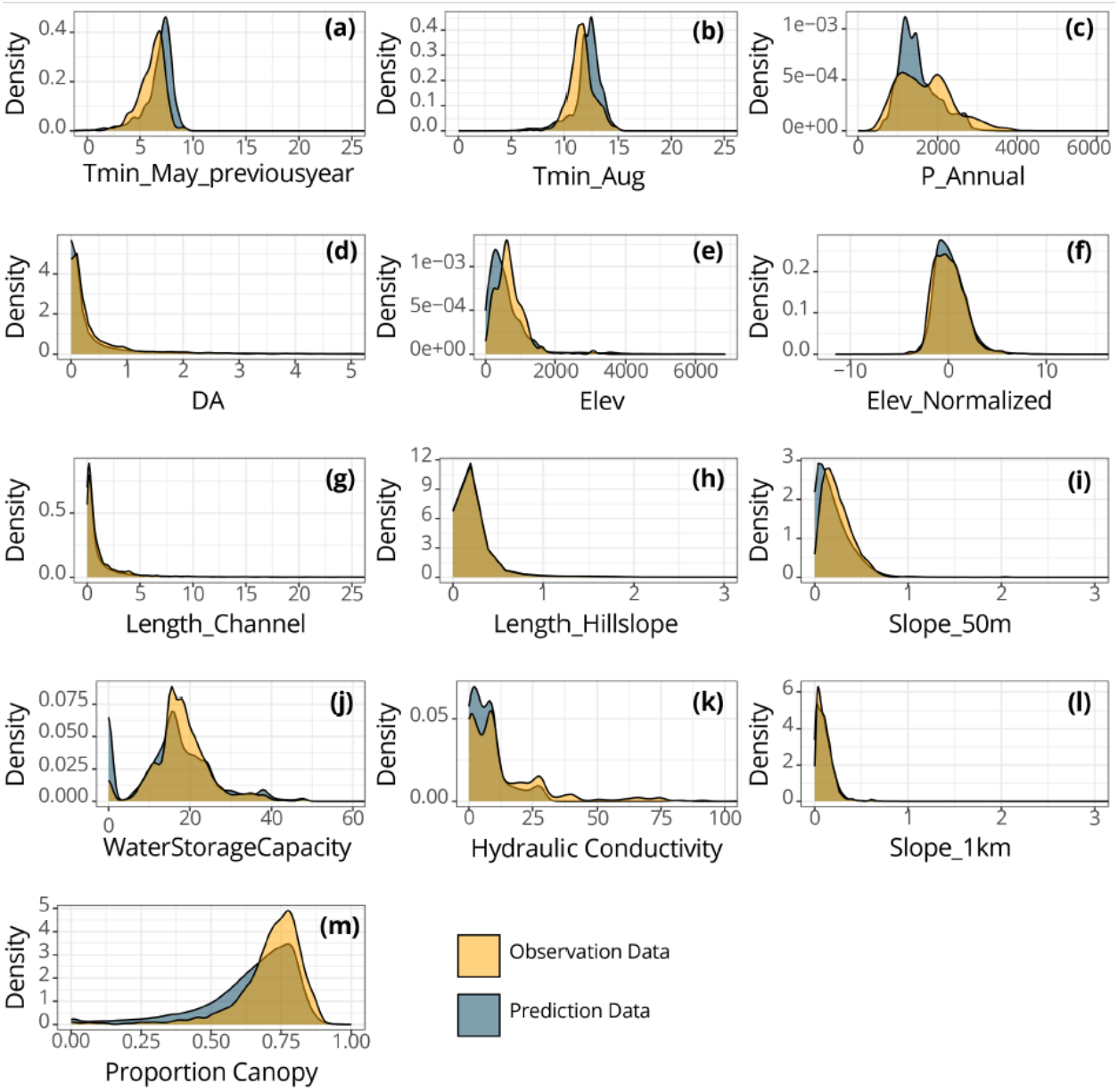

Figure S7.1. Distribution of values of 13 covariates included in the final model for training data and a subset of predictions sub-reaches in the model domain that includes the maximum and minimum values in the sub-reaches.

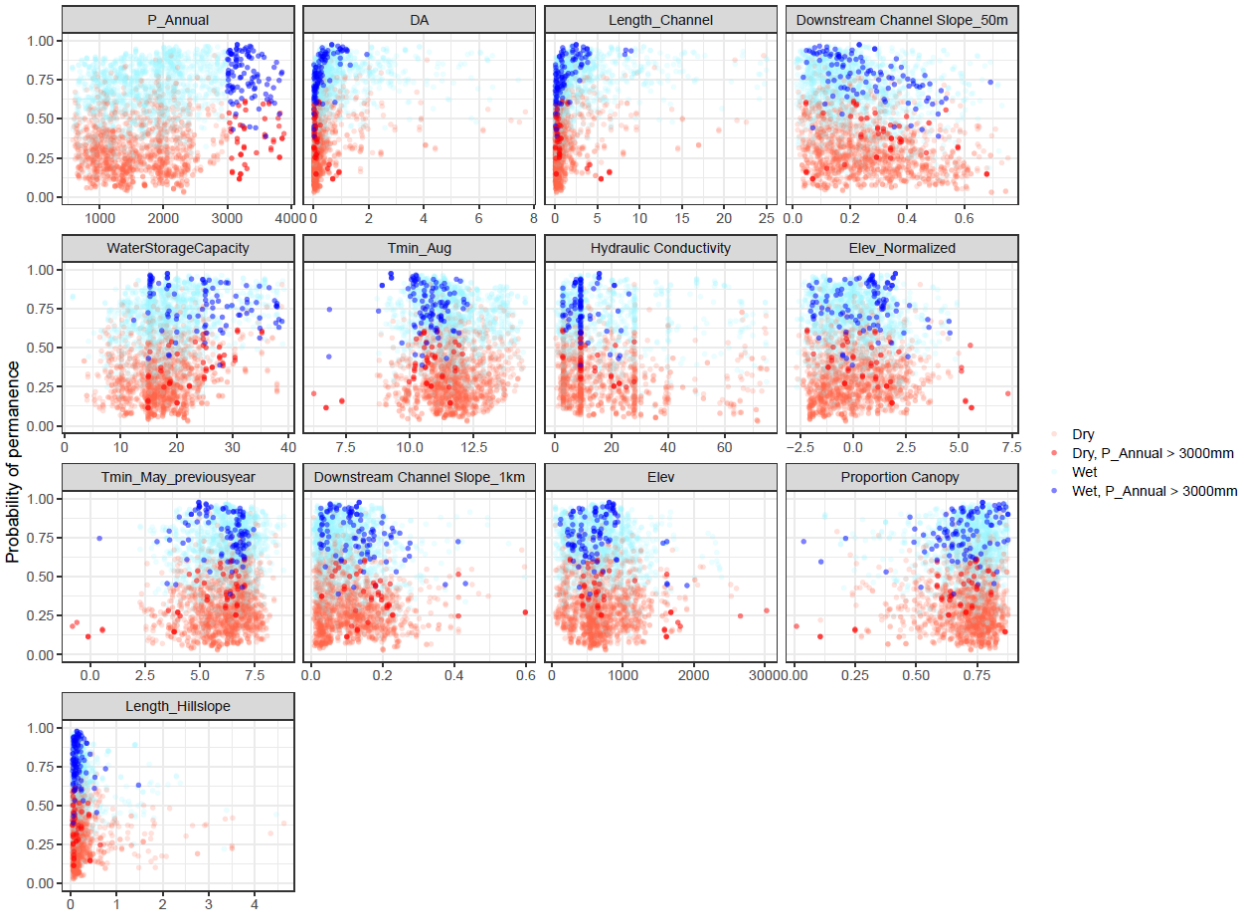

Figure S7.2. Out-Of-Bag streamflow permanence probabilities for training data that include Wet and Dry observations and a subset of Wet and Dry observations where annual precipitation (P_Annual) is more than 3,000mm.

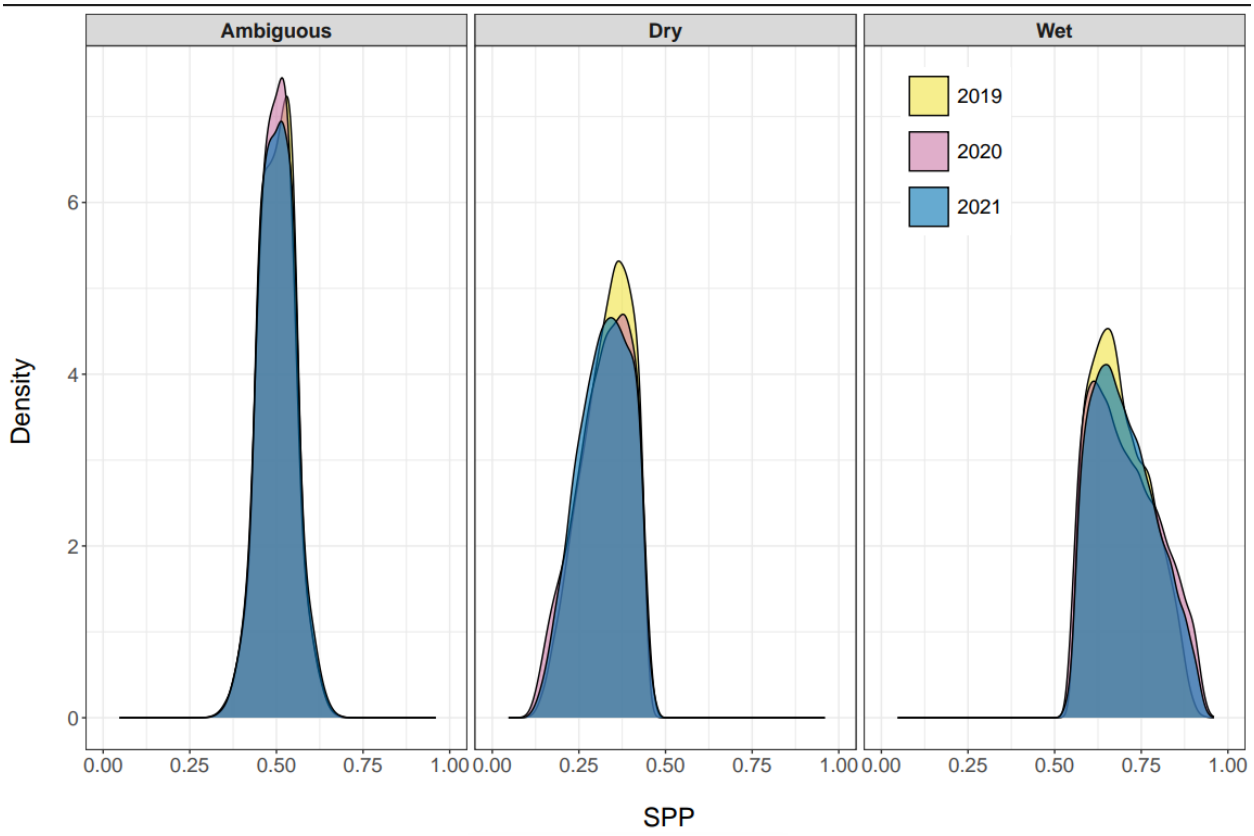

Figure S7.3. Distribution of streamflow permanence probability (SPP) and Ambiguous, Dry, and Wet classifications for subset prediction sub-reaches for years 2019, 2020, and 2021.

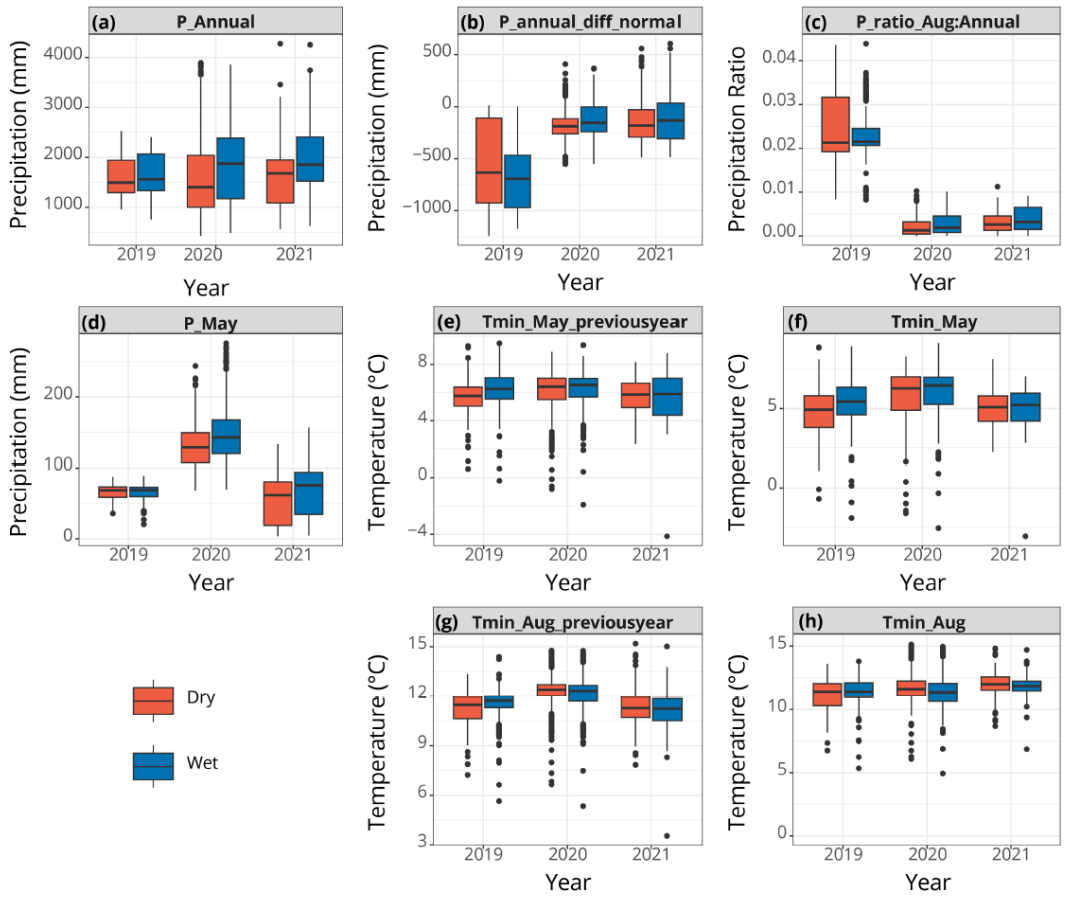

Figure S7.4. Boxplots of climate covariates for 2019, 2020, and 2021 for wet and dry observations used in training data.

Supplemental References

Abedi, R., Costache, R., Shafizadeh-Moghadam, H., Pham, Q.B., 2022. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto International 37, 5479–5496. https://doi.org/10.1080/10106049.2021.1920636

Al Farizi, W. S., Hidayah, I., & Rizal, M. N., 2021. Isolation Forest Based Anomaly Detection: A Systematic Literature Review. *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, 118–122. https://doi.org/10.1109/ICITACEE53184.2021.9617498

Barnhart, T.B., Schultz, A.R., Siefken, S.A., Thompson, F.E., Welborn, T., Sando, T.R., McCarthy, P.M., 2020. Flow Conditioned Parameter Grids for Mechanistic, Statistical, and Machine Learning Hydrologic Models: A Seamless Basin Characteristic Dataset for the Contiguous United States 2020, H028-06.

Bischl, B., Mersmann, O., Trautmann, H., & Weihs, C., 2012. Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation. *Evolutionary Computation*, *20*(2), 249–275. Evolutionary Computation. https://doi.org/10.1162/EVCO_a_00069

Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., Lang, M., 2020. Benchmark for filter methods for feature selection in high-dimensional classification data. Computational Statistics & Data Analysis 143, 106839. https://doi.org/10.1016/j.csda.2019.106839

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., 2022. xgboost: Extreme Gradient Boosting.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

O'Connor, J. E., Mangano, J. F., Anderson, S. W., Wallick, J. R., Jones, K. L., & Keith, M. K, 2014. Geologic and physiographic controls on bed-material yield, transport, and channel morphology for alluvial and bedrock rivers, western Oregon. *Geological Society of America Bulletin*, *126*(3–4), 377–397. https://doi.org/10.1130/B30831.1

Omernik, J. M., & Griffith, G. E., 2014. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environmental Management*, *54*(6), 1249–1266. https://doi.org/10.1007/s00267-014-0364-1

Oregon Department of Geology and Mineral Industries (DOGAMI), [2025] LiDAR Digital Terrain Model Mosaic.

Esri, Redlands, CA, 2021. ArcGIS Professional: Version 2.8.

Esri, Redlands, CA, 2019. Arc Hydro: Overview of Terrain Preprocessing Workflows. Esri White Paper.

Friedman, J. H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Friedman, J. H., 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Natekin, A., & Knoll, A., 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*. https://doi.org/10.3389/fnbot.2013.00021

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33, 1–22. https://doi.org/10.18637/jss.v033.i01

Gendaszek, A.S., Dunham, J.B., Torgersen, C.E., Hockman-Wert, D.P., Heck, M.P., Thorson, J., Mintz, J., Allai, T., 2020. Land-Cover and Climatic Controls on Water Temperature, Flow Permanence, and Fragmentation of Great Basin Stream Networks. Water 12, 1962. https://doi.org/10.3390/w12071962

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach Learn 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1

Harar, P., Elbrächter, D., Dörfler, M., Johnson, K.D., 2022. Redistributor: Transforming Empirical Data Distributions. https://doi.org/10.48550/arXiv.2210.14219

Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li. 2008. "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning." In 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 1322–28. Hong Kong, China: IEEE. https://doi.org/10.1109/IJCNN.2008.4633969.

Henderson, S.G., Nelson, B.L. (Eds.), 2006. Handbooks in Operations Research and Management Science: Simulation. Vol. 13. Elsevier.

Hosmer, D.W., Lemeshow, S., 1992. Confidence Interval Estimation of Interaction. Epidemiology 3, 452–456.

Jaeger, K. L., Sando, R., Dunn, S. B., & Gendaszek, A. S. (2023). Predicting probabilities of late summer surface flow presence in a glaciated mountainous headwater region. *Hydrological Processes*, *37*(2), e14813. https://doi.org/10.1002/hyp.14813

Jaeger, K.L., Sando, R., McShane, R.R., Dunham, J.B., Hockman-Wert, D.P., Kaiser, K.E., Hafen, K., Risley, J.C., Blasch, K.W., 2019. Probability of Streamflow Permanence Model (PROSPER): A spatially continuous model of annual streamflow permanence throughout the Pacific Northwest. Journal of Hydrology X 2, 100005. https://doi.org/10.1016/j.hydroa.2018.100005

Jensen, C.K., McGuire, K.J., Shao, Y., Andrew Dolloff, C., 2018. Modeling wet headwater stream networks across multiple flow conditions in the Appalachian Highlands. Earth Surface Processes and Landforms 43, 2762–2778. https://doi.org/10.1002/esp.4431

Kuhn, M., Johnson, K., others, 2013. Applied predictive modeling. Springer.

Kursa, M.B., Rudnicki, W.R., 2010. Feature Selection with the Boruta Package. Journal of Statistical Software 36, 1–13.

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L., Bischl, B., 2019. mlr3: A modern object-oriented machine learning framework in R. Journal of Open Source Software. https://doi.org/10.21105/joss.01903

Malley, J.D., Kruppa, J., Dasgupta, A., Malley, K.G., Ziegler, A., 2012. Probability Machines. Methods Inf Med 51, 74–81. https://doi.org/10.3414/ME00-01-0052

Moidu, H., Obedzinski, M., Carlson, S.M., Grantham, T.E., 2021. Spatial Patterns and Sensitivity of Intermittent Stream Drying to Climate Variability. Water Resources Research 57, e2021WR030314. https://doi.org/10.1029/2021WR030314

Naghibi, S.A., Hashemi, H., Berndtsson, R., Lee, S., 2020. Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. Journal of Hydrology 589, 125197. https://doi.org/10.1016/j.jhydrol.2020.125197

Penaluna, B.E., Burnett, J.D., Christiansen, K., Arismendi, I., Johnson, S.L., Griswold, K., Holycross, B., Kolstoe, S.H., 2022. UPRLIMET: UPstream Regional LiDAR Model for Extent of Trout in stream networks. Sci Rep 12, 20266. https://doi.org/10.1038/s41598-022-23754-0

Python Software Foundation, 2018. Python Language Reference, version 2.7. Available at http://www.python.org

Pihur, V., Datta, Susmita, Datta, Somnath, 2009. RankAggreg, an R package for weighted rank aggregation. BMC Bioinformatics 10, 62. https://doi.org/10.1186/1471-2105-10-62

R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Sahin, E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Appl. Sci. 2, 1308. https://doi.org/10.1007/s42452-020-3060-1

Sando, R., Jaeger, K.L., Farmer, W.H., Barnhart, T.B., McShane, R.R., Welborn, T.L., Kaiser, K.E., Hafen, K.C., Blasch, K., York, B., Shallcross, A., 2022. Predictions and drivers of sub-reach-scale annual streamflow permanence for the upper Missouri River basin: 1989–2018. Journal of Hydrology X 17, 100138. https://doi.org/10.1016/j.hydroa.2022.100138

Signorell, A., 2023. *RobScale function—RDocumentation* (Version 0.99) [Computer software]. https://www.rdocumentation.org/packages/DescTools/versions/0.99.54/topics/RobScale

Speiser, J.L., Miller, M.E., Tooze, J., Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert Systems with Applications 134, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Tsamardinos, I., Rakhshani, A., Lagani, V., 2015. Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization. Int. J. Artif. Intell. Tools 24, 1540023. https://doi.org/10.1142/S0218213015400230

U.S. Geological Survey (USGS), 2021. National Hydrography Dataset. https://prd-tnm.s3.amazonaws.com/index.html?prefix=StagedProducts/Hydrography/NHD/HU4/GPKG/

Vinayak, R.K., Gilad-Bachrach, R., 2015. DART: Dropouts meet Multiple Additive Regression Trees, in: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics. Presented at the Artificial Intelligence and Statistics, PMLR, pp. 489–497.

Wager, S., Hastie, T., Efron, B., 2014. Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. The Journal of Machine Learning Research 15, 1625–1651.

Wright, M. N., & Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*, 1–17. https://doi.org/10.18637/jss.v077.i01

Yu, J.-W., Yoon, Y.-W., Baek, W.-K., Jung, H.-S., 2021. Forest Vertical Structure Mapping Using Two-Seasonal Optic Images and LiDAR DSM Acquired from UAV Platform through Random Forest, XGBoost, and Support Vector Machine Approaches. Remote Sensing 13, 4282. https://doi.org/10.3390/rs13214282