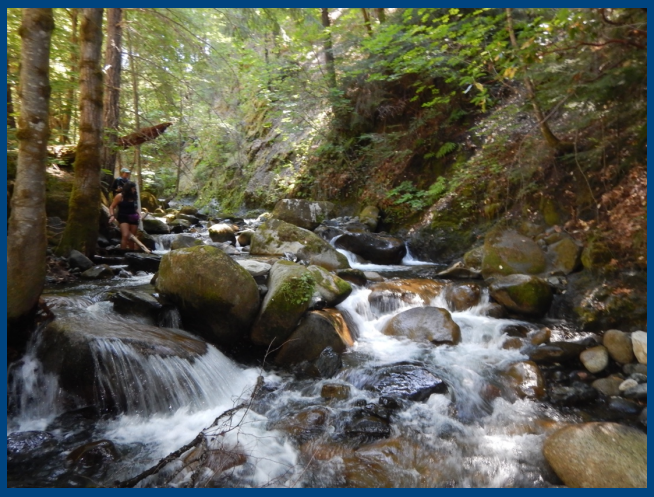
Coastal California Regional Geomorphic Classifications




February 2020

By: Colin F. Byrne, Hervé Guillon, Belize A. Lane, Gregory B. Pasternack and Samuel Sandoval-Solis

Prepared for: California State Water Resources Control Board

Utah State University

UC DAVIS

Executive Summary

Hervé Guillon, Colin F. Byrne, Belize A. Lane, Gregory B. Pasternack, and Samuel Sandoval-Solis

The objective of this research study was to determine a channel reach morphology classification and the spatial distribution of each geomorphic class for five coastal regions of the state of California: Klamath, North Coast, Central Coast (North and South) and South Coast. A geomorphic class is defined as an archetypical stream form at the 10 – 20 channel width scale (e.g. riffle-pool, plane bed, etc.) that has: (a) well-defined channel attributes (e.g. slope, bankfull width, etc.), (b) topographic variability attributes (TVA) (e.g. coefficients of variation of width and depth), (c) sediment composition (e.g. D50, D84, etc.) and (d) landscape location (e.g. valley confined, partly confined or unconfined) that can be verified in the field. The classification was informed by 596 field-surveyed streams in the coastal regions: Klamath (105), North Coast (201), Central Coast North (104) and South (119), and South Coast (67). Surveyed streams in the coastal region catchments were analyzed using multivariate statistical techniques to identify groups of reaches with similar stream forms. This study focusses on the geomorphic classification for the Coastal Region's catchments. A total of 34 channel types were identified in the coastal regions (Table ES-1). Figure ES-1 shows the location of 596 classified field-surveyed locations.

Table ES-1 - Channel types in the Coastal Regions of California

Klamath				North Coast				Central Coast - North				Central Coast - South				South Coast			
1	C	Bou-Bed	SP	1	P	Grv-Cob	RP	1	C	Cob-Bou	UN-BU	1	U	San-Grv	UN	1	U	San-Grv	UN
2	U	Grv	PB	2	C	Cob-Bou	CS-SP	2	C	Grv-Cob	RP	2	U	San-Grv	UN	2	P	Grv	BR
3	U	Grv-Cob	RP	3	U	Grv	RP	3	P	Grv	RP	3	P	Grv-Cob	UN	3	C	Bou	CS-SP
4	U	Grv	UN	4	C	Grv-Cob	UN-BU	4	U	Grv	UN	4	C	Grv-Cob	UN	4	C	Cob-Bou	UN
5	P	Grv-Cob	RP	5	C	Grv-Cob	PB	5	C	Cob-Bou	CS-SP	5	C	Cob-Bou	CS-SP	5	P	Grv-Cob	RP
6	C	Cob-Bou	CS-SP	6	C	Cob-Bou	UN-BU	6	C	Grv-Cob	UN	6	U	Grv-Cob	RP				
7	C	Cob-Bou	UN	7	C	Grv-Cob	RP					7	P	Grv-Cob	RP				
				8	C	Grv-Cob	UN					8	C	Cob-Bou	RP				

For each Region, the first column refers to *channel type number*; the second column refers to *confinement*: Confined (C), Unconfined (U), Partly Confined (P); the third column refers to *predominant sediment*: Boulder (Bou), Bedrock(Bed), Cobble (Cob), Gravel (Grv), Sand (San) and the fourth column refers to *channel form*: Step-Pool (SP), Riffle-Pool (RP), Uniform (UN), Plane Bed (PB), Bed Undulating (BU), Cascade (CS), and Braided (BR).

Channel types in confined (C) settings exhibit the greatest diversity, 18 channel types were identified in this setting, almost all channel forms (except braided) and predominant sediment (except sand) occur in this confinement setting and they occur in all regions. Channel types in partly confined (P) settings exhibit low diversity, 7 channel types were identified in this setting, channel form is predominantly riffle-pool (RP), the predominant sediment in these channel types is gravel and they occur mostly south of the San Francisco Bay. Channel types in unconfined (U) settings exhibit low diversity, 9 channel types were identified in this setting; two channel form are predominant in this setting, riffle-pool (RP) and Uniform (UN), the predominant sediment in these channel types are gravel and sand and they occur in all regions.

Within each region, the associated channel type classification was used as the training set to determine a statistical model using machine learning (ML) techniques that spatially predict channel types throughout each catchment. The field sites already classified are incorporated as labels into a large-scale supervised ML model. Three set of ML models were tested: Artificial Neural Networks (ANN), Support Vector Machine (SVM) and Random Forest (RF). ML models used from 287 to 147 coarse-scale as predictor variables describing topography, geology, soils, climate and land use. Across all regions, the distributions of the criterion Area Under the Curve

(AUC) and accuracy indicate that ANN is under-performing in comparison to RF and SVM. In addition, three performance metrics are used to assess the predictive modeling accuracy for SVM and RF: regional entropy, entropy rate and stream-interval entropy. In all regions, SVM and RF channel type predictions respect the general expected geomorphic organization of the landscape: confined high gradient channel types are dominantly predicted in mountainous areas where they are expected to occur while unconfined channel types are predicted in valley areas. For all regions, a random forest model was selected to spatially predict channel types for each catchment because it has the second best performance model (closet to SVM) but it provided consistent channel type predictions, which was not the case for SVM. In addition, the spatial significance of the predictions from the best models was assessed using expert-knowledge, with a focus on the general spatial organization of channel types across the different regions as well as their geomorphic relevance. Aerial imagery was used to confirm predictions at selected sample locations. In addition, this study also identified the uncertainty of a given reach to be classified correctly by using the Shannon-Weiner entropy index, that identifies at which location the prediction is more stable or not.

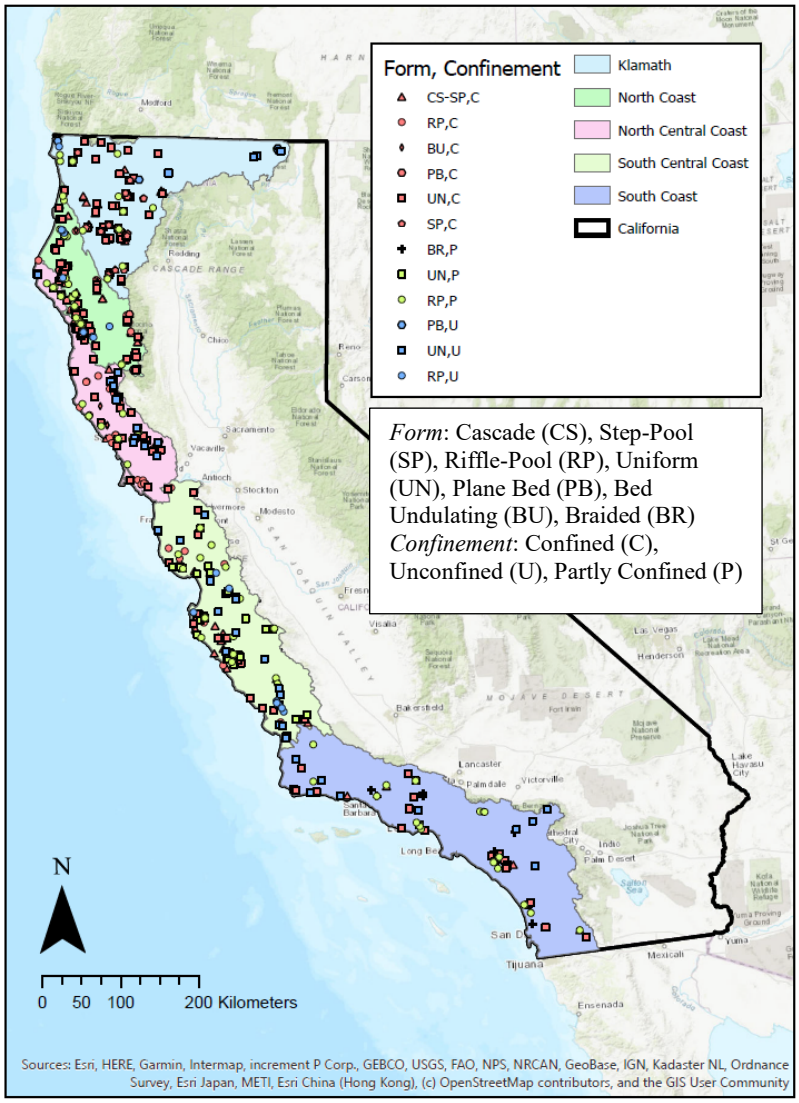


Figure ES-1 - Geomorphic classification of channel types in the Coastal Regions of California

Table of Contents

1. Objectives and Deliverables

To better understand how managed environmental flow prescriptions will impact native biota throughout the streams and rivers of California, knowledge of the types and distributions of physical habitat is critical. In management contexts, physical habitat is often defined at the reach-scale (i.e. 10-20 channel widths) and classification of reach-scale channel types serves as a method to simplify heterogeneous river conditions across entire stream networks. However, in order to understand these channel types and, thus, physical habitat throughout a network, two key steps are necessary: (1) a **classification** of the channel types that exist within a stream network must be defined, and (2) **predictions** of where those channel types occur must be developed. This report describes methods and results of addressing these two steps within five management regions of coastal California.

Five regions were the focus of this study: the Klamath (K), North Coast (NC), North Central Coast (NCC), South Central Coast (SCC), and South Coast (SC) (Figure 1). Region specific classifications and predictions are inherently more tailored to the existing stream conditions than previous classifications developed in other regions or broad universal classifications. Within each region, site-specific field data were first classified into unique channel types (Sections 2 and 3) and channel types were predicted across stream networks using remotely sensed databases (Sections 4 and 5). The goal of both approaches was to develop robust methodologies to understand differences between region-specific channel types and their existence throughout the regional stream network.

This report contains the following deliverables under contract *16-062-300-2 – Task 7 – Stream Sub Classification Developoment Guidance for Regions Outside of the Sacramento River Basin:*

- Methodology for channel type classification;
- Field site geomorphic classifications for contract regions: Klamath, North Coast, North Central Coast, South Central Coast and South Coast;
- Methodology for channel type predictions; and
- Channel type predictions for contract regions: Klamath, North Coast, North Central Coast, South Central Coast and South Coast.

2. Site Classification Methodology

2.1. Objectives

The objective of the channel-reach morphology classification was to organize field-surveyed sites into groups of similar channel form, thus representing archetypical stream forms within a region. These classifications approximate river form at the scale of 10 – 20 channel widths, otherwise defined as a streams' channel-reach morphology (Montgomery and Buffington, 1997). The objective was assessed independently in each of the five coastal management regions. The objective was achieved by surveying stream reaches and using multivariate statistical techniques to identify survey sites with more similar and dissimilar stream forms. The archetypical channel-reach morphologies described in each classification are referred to as channel types.

2.2. Data Acquisition and Processing

2.2.1. Site Selection

In an effort to survey stream sites representative of all existing channel types, a methodology was developed to determine the geomorphic settings of streams across the state based on two key geomorphic influences: sediment supply and transport capacity. Sites were selected at random from each stratified sampling bin within the stream network of each region. Bins were defined by estimates of sediment supply and transport capacity for each 200-m segment (Appendix A). Lateral valley confinement and hillslope sediment supply derived from the Revised Universal Soil Loss Equation (RUSLE) were used as indicators of sediment supply. Contributing drainage area and slope were used as indicators of transport capacity. Streams were first classified by sediment supply (confinement and RUSLE) and second by transport capacity (area and slope). Upper level bins were defined as confined, partly-confined, or unconfined and with high or low sediment supply, resulting in six upper level bins. Sites within each upper level bin were then grouped into five lower level contributing area-slope bins based on statistical distances. The resulting 30 bins (6 upper level bins by 5 lower level bins) were all sampled at least once. The remaining sites were then sampled across lower level bins within upper level bins with a greater proportion of sites. More detailed information on the estimation of binning metrics can be found in Section 2.2.3 and Appendix A.

2.2.2. Field Surveys and Post-Processing

A total of 596 surveys were included in the classification analysis across all regions (Figure 1). Survey methodologies are based on the California Water Board's Surface Water Ambient Monitoring Program (SWAMP) protocols (Ode, 2007). In the coastal regions, stream survey lengths were fifteen times mean bankfull width. Ten equally spaced cross-sectional surveys were completed along the channel length using rod and level techniques. The bankfull level was defined using geomorphic and vegetative indices as defined by Ode (2007) for SWAMP protocols, including slope breaks, change from annual to perennial vegetation, and changes in sediment size. These indices were used to calculate bankfull width (w) for each cross-section. Bankfull depth (d) and water depth were recorded at the thalweg. A Wolman pebble count was also conducted along each cross-section (Wolman, 1954). Finally, a longitudinal survey was conducted along the thalweg at each cross-section.

Field survey data was processed in preparation for statistical analysis to describe reach-average dimensional and sediment characteristics. In addition to the mean values of bankfull width (w), depth (d), and bankfull width-to-depth ratio (w/d), median (D_{50}) and 84th percentile (D_{84}) grain sizes of Wolman pebble counts were calculated. Slope (s) was calculated from the best fit regression line of surveyed water surface elevations along longitudinal, thalweg transects. The coefficients of variation of bankfull width and bankfull depth, or the ratio of standard deviation to mean, were calculated at each site using data from each surveyed cross-section. Here, coefficients of variation of width (CV_w) and depth (CV_d) are referred to as topographic variability attributes (TVAs, Lane et al., 2017b).

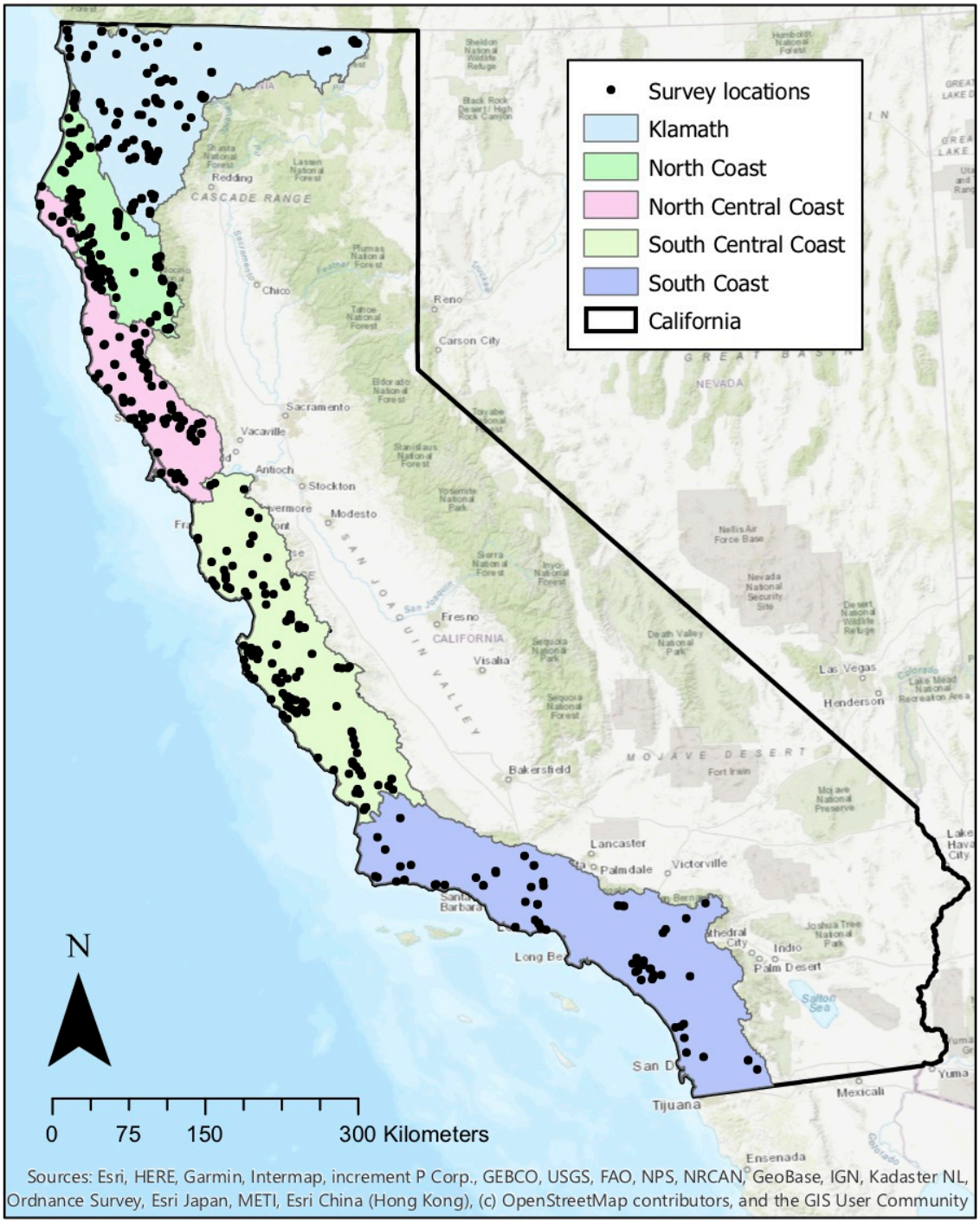


Figure 1. Regional boundaries and survey site locations used in geomorphic classification.

2.2.3. Geomorphic Information Systems Metrics

A geographic information system (GIS) [ESRI ArcGIS 10.4 (ESRI, 2016)] was used for geospatial analysis in both initial site selection and the estimation of geospatial attributes used in statistical methods.

Contributing area (A_c) and slope were estimated for initial site selection based on sampling scheme protocols described in 1.1.2. Contributing area was calculated based on the United States Geological Survey 10-m National Elevation Dataset (Gesch et al., 2002; NED) and streamlines defined by the National Hydrography Dataset version 2 (McKay et al., 2012; NHDPlusV2). Slope for pre-survey binning was calculated based on the ratio of the upstream and downstream differences in elevation and the length of the given stream segment. This technique provides a rough estimate of slope, but error is often associated with these estimates especially for short stream segments (Neeson et al., 2011) so slope was calculated from site surveys for statistical analysis.

Sediment supply bins were defined based on calculated valley confinement distances and RUSLE estimates. Valley confinement (C_v), which denotes the relative interaction of a stream and the nearest hillslope, was calculated as a distance from the channel thalweg to bounding valley wall. The metric is reliant upon definition of a valley wall, which was defined using a 10-m digital elevation model (DEM), where 25 percent slope was chosen as a threshold between valley bottom and valley wall capturing a medial value between clay and sand dominated hill footslopes (Carson, 1972). Confined, partly-confined, and unconfined nomenclature was defined by a logarithmic scale of <= 100 m, >100 and <= 1000 m, and > 1000 m, respectively. The RUSLE was used as a surrogate for hillslope erosion because it is a widely known methodology and calculation of the metric was achievable across the entire state of California. High and low sediment bins were defined as greater and less than 225 t/km^2/year, respectively. More detailed calculation methods can be found in Appendix A.

2.2.4. Multivariate Statistical Classification

Multivariate statistical clustering was used to develop regional channel types. Statistical techniques were based upon Lane et al. (2017a). The goal of this statistical methodological approach is to group sites with similar reach-averaged metrics The `R` language was used for all statistical analysis (`R` Core Team, 2017). Geomorphic attributes were rescaled from zero to one to remove the influence of large magnitude attributes. Linear correlations between all geomorphic inputs were first conducted to remove highly correlated metrics (`cor` function within the `stats` package). Euclidean distances between all site attributes were calculated for use in multidimensional clustering. Although ultimately not directly used in the classification of sites, NMDS scaling using Euclidean distances (`metaMDS` function within the `vegan` package) and calculation of principal component vectors (`prcomp` function within the `stats` package) were conducted to assess which attributes were most influential in distributing sites across multivariate space (Anderson, 2001; Clarke, 1993; Kruskal, 1964; Oksanen et al., 2019). Actual classification of sites was conducted using Ward's Hierarchical Clustering (WHC) with the Euclidean distance matrix (`hclust` function with the method defined as `Ward.D2` in the `stats` package) (Murtagh

and Legendre, 2014a, 2014b; Ward, 1963). This approach stratified the data by minimizing within-cluster variance and maximizing between-cluster variance. In geomorphic terms, this ideally means that more similar geomorphic settings are clustered together. Hubert Index values were used as one tool in selection of an appropriate number of channel types (`NbClust` function and package) (Charrad et al., 2012). However, heuristic refinement of these groupings was also conducted based upon field reconnaissance and expert knowledge of specific field sites. Classification tree analysis was conducted as a method to understand the stability of the hierarchical classification, which is presented here as the cross-validation percentage (`rpart` package) (De'ath and Fabricius, 2000; Therneau and Atkinson, 2018). The nonparametric Dunn's Test was used to quantify and compare clustering attributes across classified channel types (Kassambara, 2019).

The statistical methodology is iterative in the sense that both the hierarchical clustering and classification tree analysis can be altered with different combinations of input variables to better understand the geomorphic attributes most influential in clustering and classification. Ultimately, the number of channel types was selected based on the combination of hierarchical clustering, potential heuristic adjustment of those clusters based on expert opinion or field reconnaissance, and assessment of the stability of the resulting geomorphic classification using the classification tree approach.

3. Site Classification Results

3.1. Klamath

A total of 105 sites were used in the Klamath region classification. The final NMDS solution recorded two- and three-dimensional stress values of 0.119 and 0.072, respectively (Figure 2). The first and second principle component axes (PCAs) accounted for 43% and 22% of the variance in the data and were most greatly associated with D_{84} and C_v, respectively.

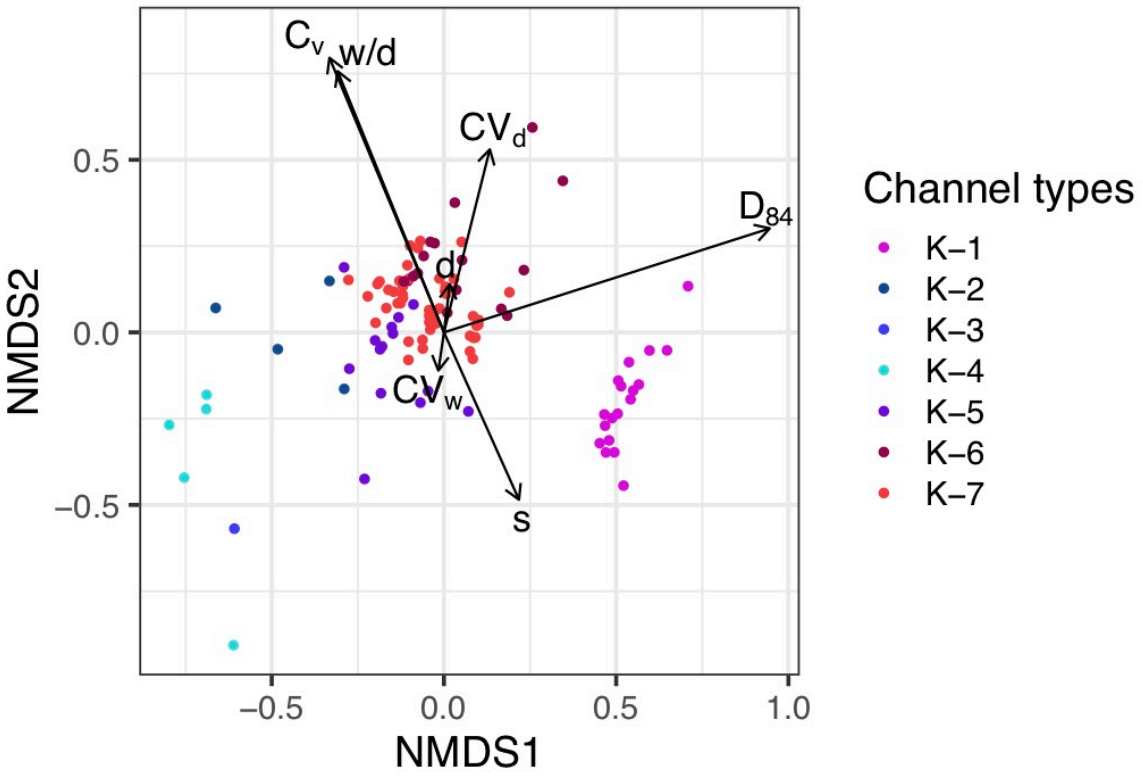


Figure 2. NMDS scatter plot with principal component vectors for attributes used in the classification of sites. Scattered sites are colored according to final channel type. Longer principal component vectors indicate greater influence on the scattering of sites.

Seven channel types were identified through the WHC with heuristic refinement (Figure 3). The final seven channel types were the result of CART analysis which resulted in a successful prediction rate of 96.2% (Figure 4). Ten-fold cross-validation, a measure of prediction on unseen data, was 82.9%. The final classification tree was defined by D_{84}, CV_d, s, and C_v. Final channel types were made up of between 1 and 47 sites. Median attributes of each channel type can be found in Table 1. The channel types are summarized in Figure 5. Distributions of channel type attributes and classified site locations can be found in Figure 6 and Figure 7, respectively.

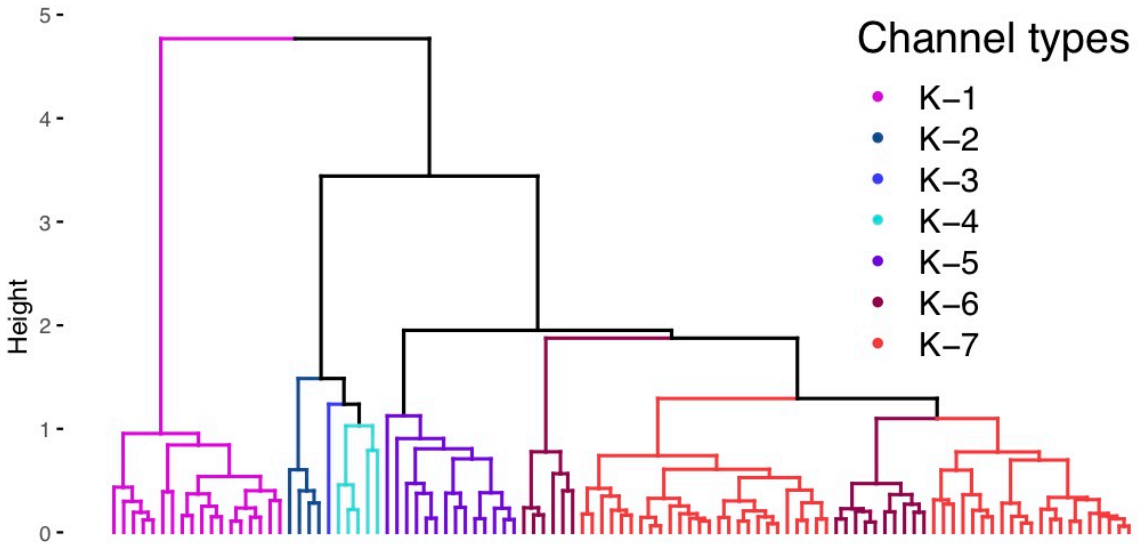


Figure 3. Ward's hierarchical clustering dendrogram with associated heuristic channel types.

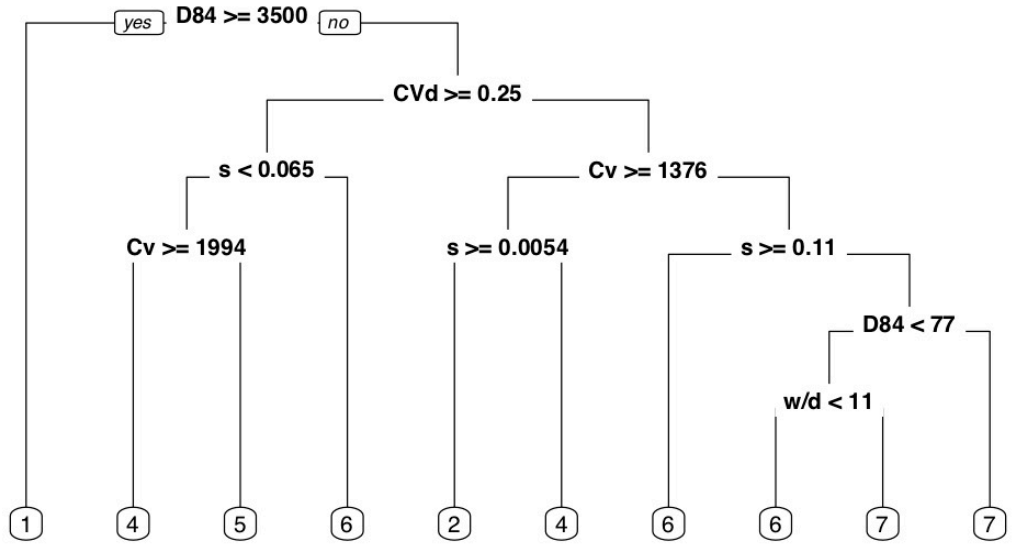


Figure 4. Final yes/no classification tree that produced suitable prediction and cross-validation percentages. Channel types are indicated by number at the bottom of the figure.

The Klamath classification resulted in three confined channel types, one partly-confined channel types, and three unconfined channel types. Confined channel types were dominated by the highest slopes and largest sediment sizes. *Confined, boulder-bedrock, step-pool* (K-1) streams were defined by moderate slopes compared to other confined classes, but a large proportion of the channel bottom at these sites was bedrock. This would suggest that these streams are able to

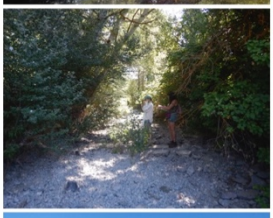
transport much of the sediment that is supplied to these channels. In contrast, *confined, cobble-boulder, cascade/step-pool* (K-6) streams were found to have smaller sediment sizes and steeper slopes. These streams are likely to be subjected to high sediment inputs from colluvial sources as the sites were also the steepest of all channel types in the Klamath region. The *confined, cobble-boulder, uniform* (K-7) channel type exhibited very low channel variability and the lowest slopes of the confined channel types. In the downstream continuum, *partly-confined, gravel-cobble, riffle-pool* (K-5) streams are likely to exist as the distance between valley walls increases, slope decreases, and alluvial floodplain formation begins. In the largest rivers within the Klamath region, valley width continues to expand and *unconfined, high-order, gravel-cobble, riffle-pool* (K-3) streams exist. These streams are characteristic of the mainstem rivers in the Klamath region. Other unconfined channel types were found to exist in localized stretches, including *unconfined, high width-to-depth, gravel uniform* (K-4) streams in the upland Scott and Shasta valleys, as well as *unconfined, low width-to-depth, gravel uniform* (K-2) streams in the more incised channel settings.

Table 1. Median channel attributes of seven channel types within the Klamath region.

Channel Type	A_c (km^2)	s	d (m)	w (m)	w/d	CV_d	CV_w	D50 (mm)	D84 (mm)	C_v (m)
K-1	23	0.0302	1.57	10.24	7.2	0.27	0.23	128	5000	30
K-2	38	0.0061	1.48	11.31	9.0	0.12	0.25	20	45	3173
K-3	1728	0.0010	5.31	152.38	31.9	0.29	0.18	45	90	2698
K-4	27	0.0053	1.14	25.55	28.6	0.28	0.23	11	45	4693
K-5	83	0.0134	1.70	21.39	13.7	0.37	0.26	45	159	131
K-6	7	0.1122	1.18	6.24	6.0	0.25	0.27	77	595	22
K-7	41	0.0230	1.45	12.50	9.1	0.17	0.22	90	1000	20

K-1 - Confined, boulder-bedrock, bed-undulating step-pool

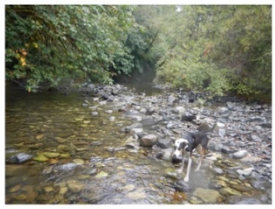
K-2 - Unconfined, low width-to-depth, gravel, plane bed



K-3 - Unconfined, high-order, gravel-cobble, riffle-pool

K-4 - Unconfined, high width-to-depth, gravel, uniform

K-5 - Partly-confined, gravel-cobble, riffle-pool



K-6 - Confined, cobble-boulder, cascade/step-pool

K-7 - Confined, cobble-boulder, uniform

Figure 5. The seven channel types within the Klamath region developed by multivariate statistical analysis with heuristic refinement.

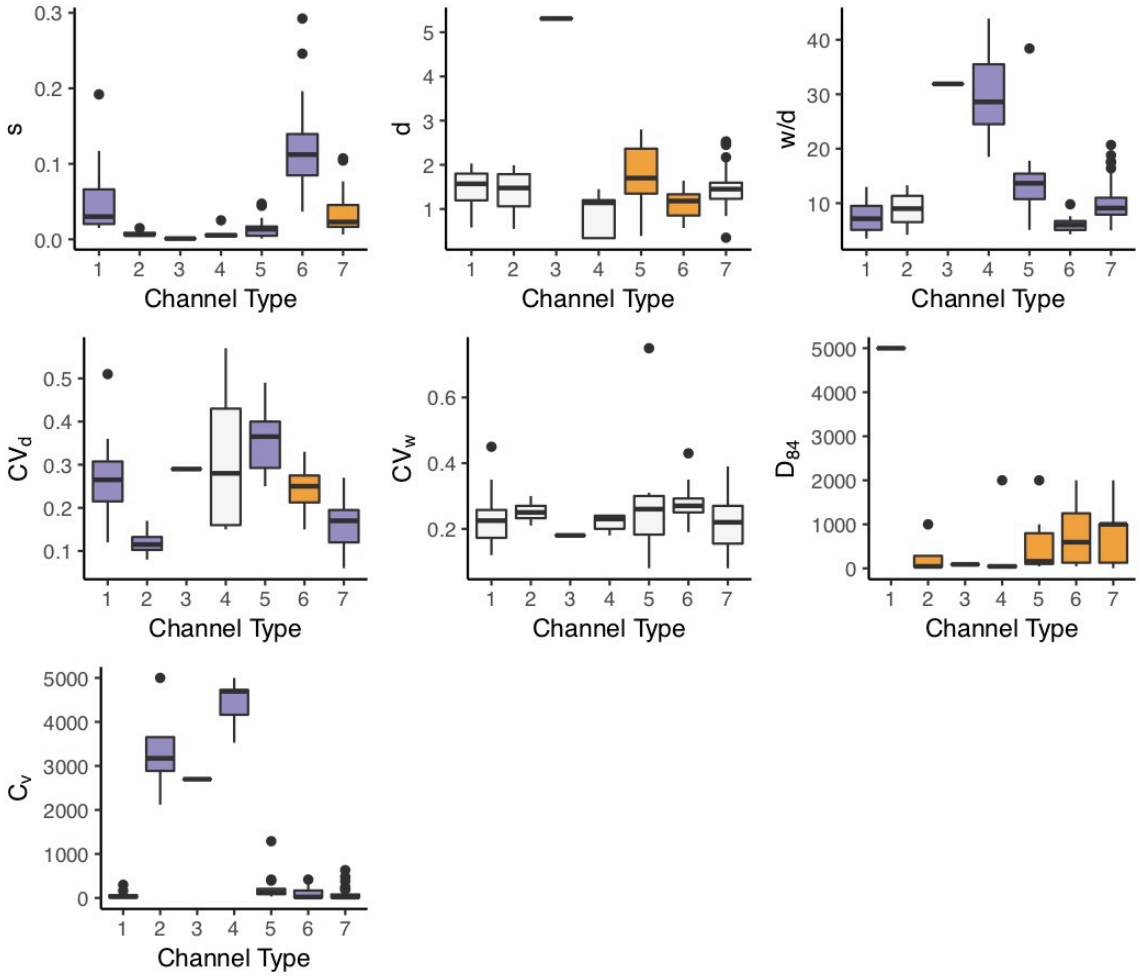


Figure 6. Box-and-whisker plots of classification metrics for each channel type in the Klamath region. Purple boxes represent channel types with significant differences from multiple other channel types based on Dunn's Test. Yellow boxes represent channel types significantly different than one other channel type and empty boxes represent a channel type with no significant differences.

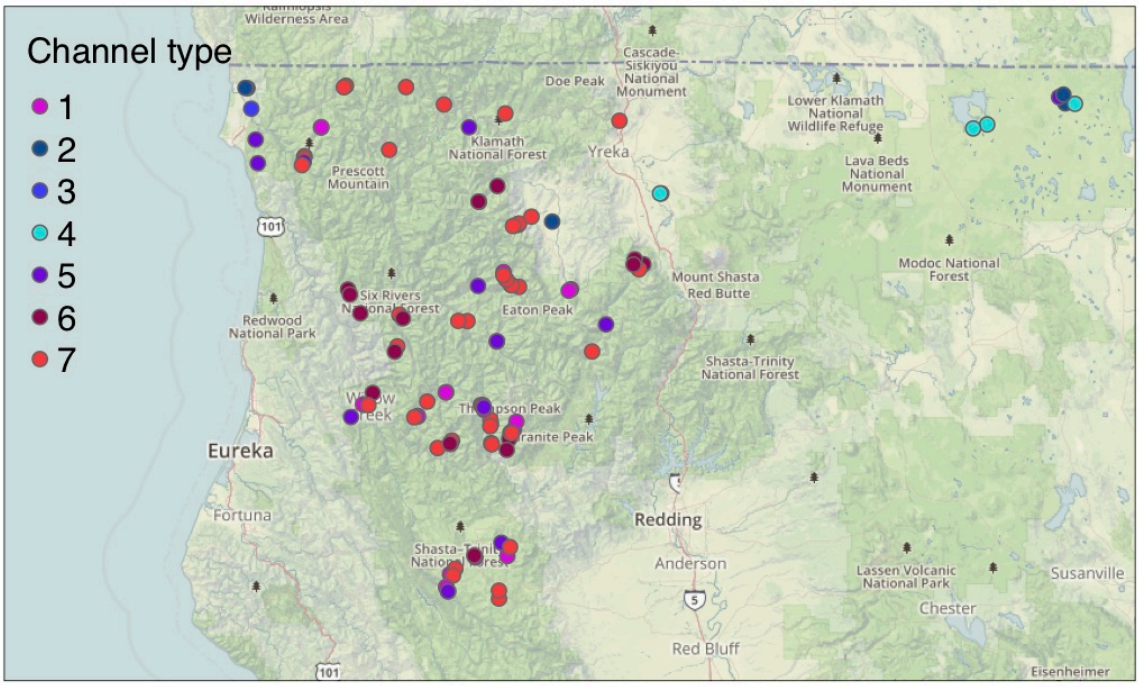


Figure 7. Location of site surveys colored by channel type in the Klamath region.

3.2. North Coast

A total of 201 sites were used in the North Coast region classification. The final NMDS solution recorded two- and three-dimensional stress values of 0.167 and 0.1, respectively (Figure 8). The first and second principle component axes (PCAs) accounted for 26% and 24% of the variance in the data and were most greatly associated with A_c and D_{84}, respectively.

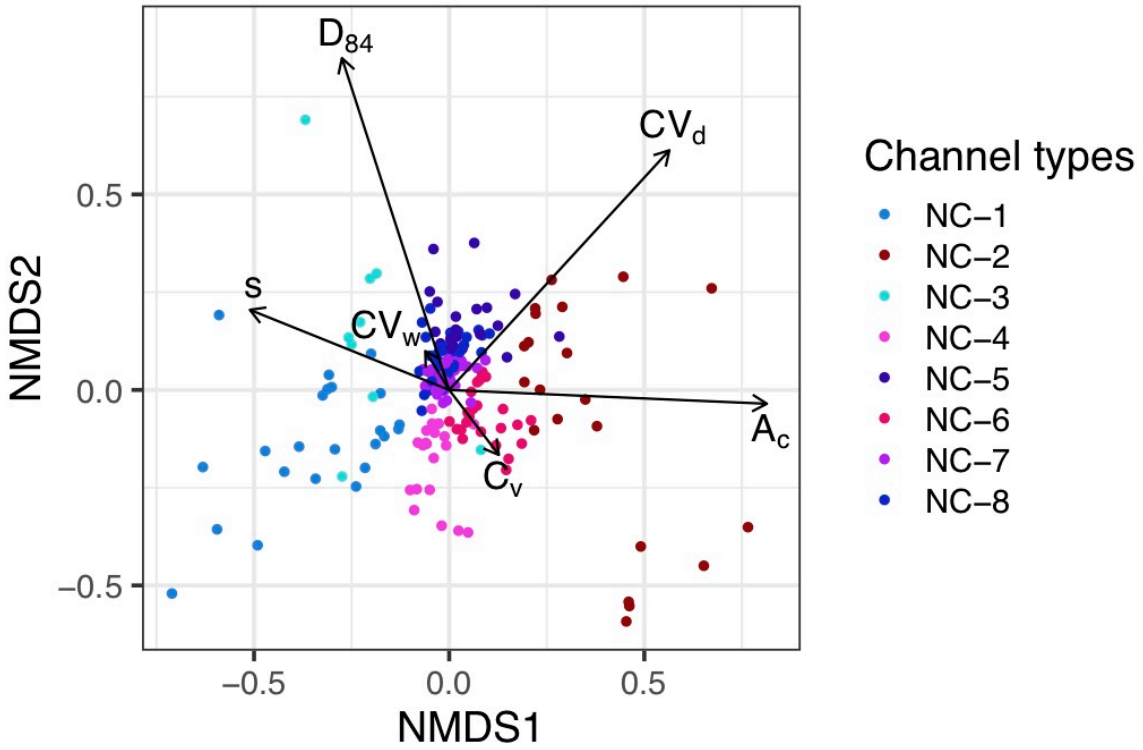


Figure 8. NMDS scatter plot with principal component vectors for attributes used in the classification of sites. Scattered sites are colored according to final channel type. Longer principal component vectors indicate greater influence on the scattering of sites.

Eight channel types were identified through the WHC with heuristic refinement (Figure 9). The final eight channel types were the result of CART analysis which resulted in a successful prediction rate of 93.5% (Figure 10). Ten-fold cross-validation, a measure of prediction on unseen data, was 85.6%. The final classification tree was defined by A_c, D_{84}, s, CV_d, CV_w, and C_v. Final channel types were made up of between 9 and 43 sites. Median attributes of each channel type can be found in Table 2. The channel types are summarized in Figure 11. Distributions of channel type attributes and classified site locations can be found in Figure 12 and Figure 13, respectively.

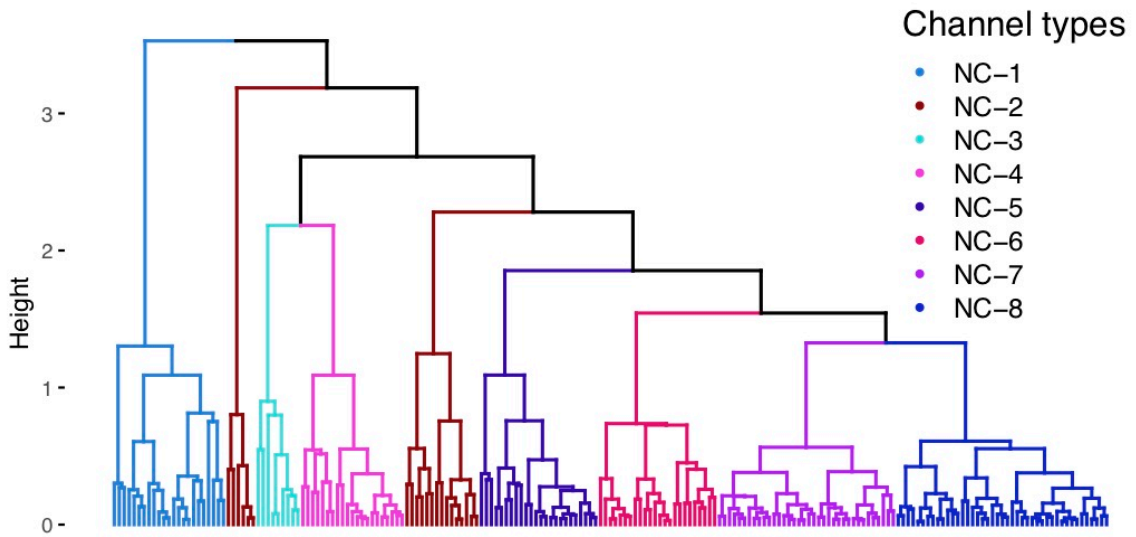


Figure 9. Ward's hierarchical clustering dendrogram with associated heuristic channel types.

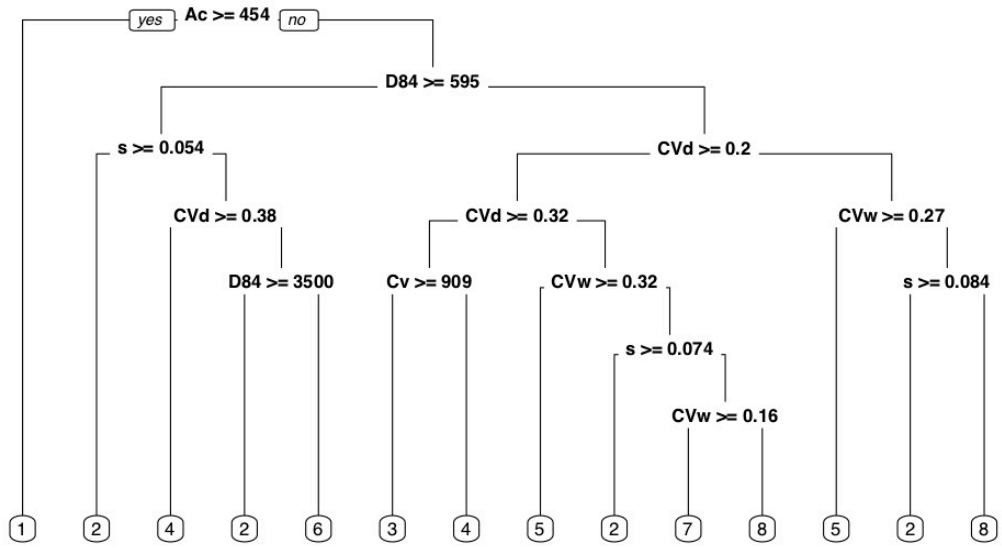


Figure 10. Final yes/no classification tree that produced suitable prediction and cross-validation percentages. Channel types are indicated by number at the bottom of the figure.

The North Coast classification resulted in six confined channel types, one partly-confined channel type, and one unconfined channel type. Confined channel types were most significantly influenced by slope, sediment size, and TVAs. The steepest streams in the region are categorized as *confined, cobble-boulder, cascade/step-pool* (NC-2) systems. These streams have grain roughness characteristics due to colluvial inputs and large sediment sizes. Two confined channel types exhibit uniform reach-scale channel dimensions but differ significantly in grain size: *confined, cobble-boulder, uniform* (NC-6) and *confined, gravel-cobble, uniform* (NC-8) streams. The final three confined channel types all exhibit the same sediment size but differ significantly

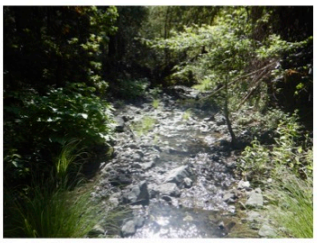
from each other and uniform streams in a combination of stream variability. *Confined, gravel-cobble, uniform width-bed undulating streams* (NC-4) exhibit uniform widths but bed undulation due to depositional and scour patterns. *Confined, cobble-boulder, width variable-plane bed* (NC-5) streams do not exhibit the channel bed scouring of the bed undulating streams but do show high variability in width. Finally, *confined, cobble-boulder, riffle-pool* (NC-7) streams show relatively high variability in both width and depth, thus exhibiting a more classic riffle-pool morphology. The partly-confined streams within the North Coast region are characteristic of the larger mainstem rivers in the region and are defined as *partly-confined, high order, gravel-cobble, riffle-pool* (NC-1) streams. Finally, *unconfined, gravel-cobble, riffle-pool* (NC-3) streams exist within the relatively few valley bottoms within the region and exhibit variability associated with riffle-pool formation.

Table 2. Median channel attributes of eight channel types within the North Coast region.

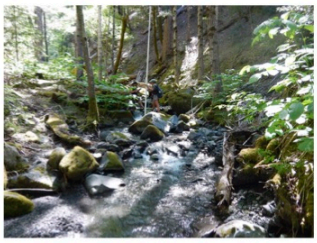
Channel Type	A_c (km^2)	s	d (m)	w (m)	w/d	CV_d	CV_w	D50 (mm)	D84 (mm)	C_v (m)
NC-1	959	0.0013	2.56	61.35	24.0	0.31	0.22	22.6	64	132
NC-2	6	0.0974	1.16	7.24	6.6	0.28	0.20	64	1000	3
NC-3	46	0.0026	1.50	14.42	10.4	0.30	0.22	16	32	1749
NC-4	67	0.0062	1.28	15.65	12.1	0.43	0.19	32	128	30
NC-5	12	0.0171	1.20	10.45	9.9	0.18	0.33	32	64	43
NC-6	37	0.0126	1.22	13.94	11.4	0.22	0.19	77	1000	21
NC-7	12	0.0098	1.22	10.53	9.6	0.25	0.26	32	77	31
NC-8	18	0.0138	1.14	8.18	8	0.17	0.17	32	64	48

NC-1 – Partly-confined, high order, gravel-cobble, riffle-pool

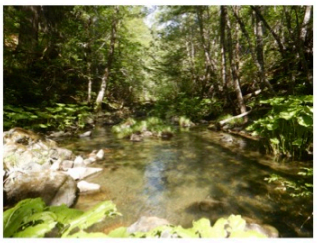
NC-5 – Confined, gravel-cobble, width variable-plane bed



NC-2 – Confined, cobble-boulder, cascade/step-pool

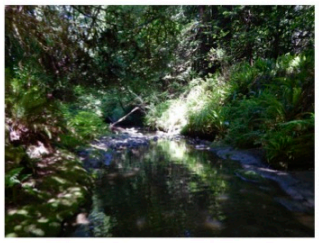


NC-6 – Confined, cobble-boulder, uniform

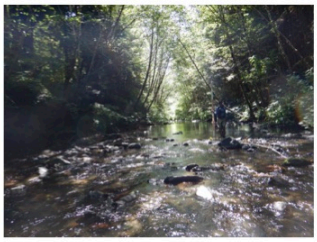


NC-3 – Unconfined, gravel, riffle-pool

NC-7 – Confined, gravel-cobble, riffle-pool



NC-4 – Confined, gravel-cobble, uniform width-bed undulating



NC-8 – Confined, gravel-cobble, uniform

Figure 11. The eight channel types within the North Coast region developed by multivariate statistical analysis with heuristic refinement.

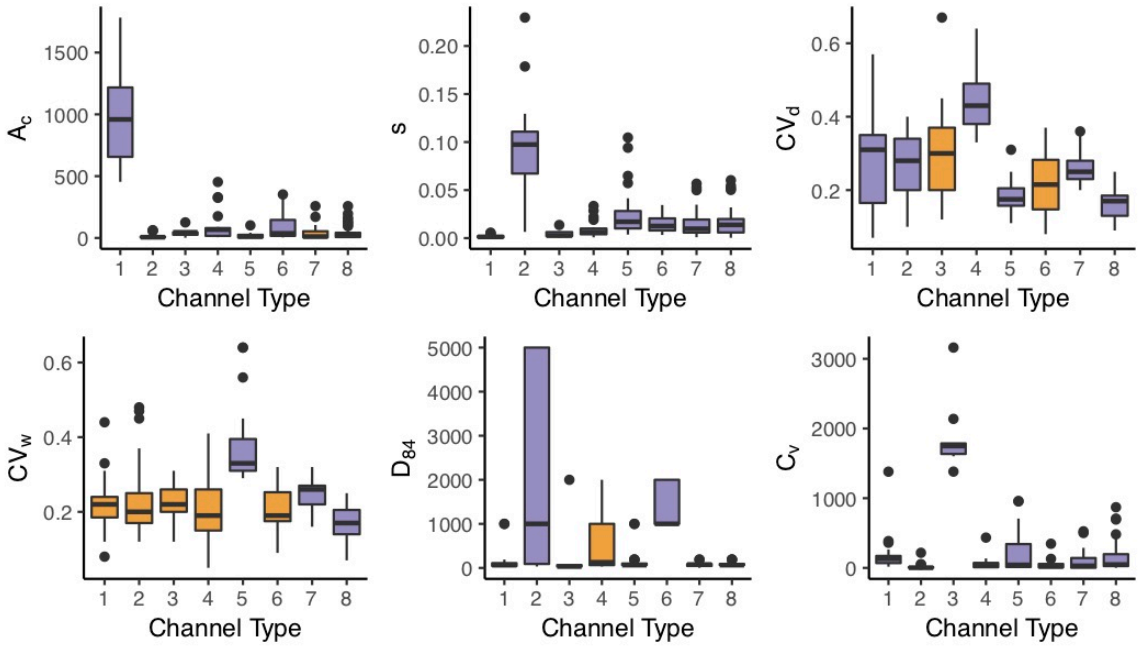


Figure 12. Box-and-whisker plots of classification metrics for each channel type in the North Coast region. Purple boxes represent channel types with significant differences from multiple other channel types based on Dunn's Test. Yellow boxes represent channel types significantly different than one other channel type and empty boxes represent a channel type with no significant differences.

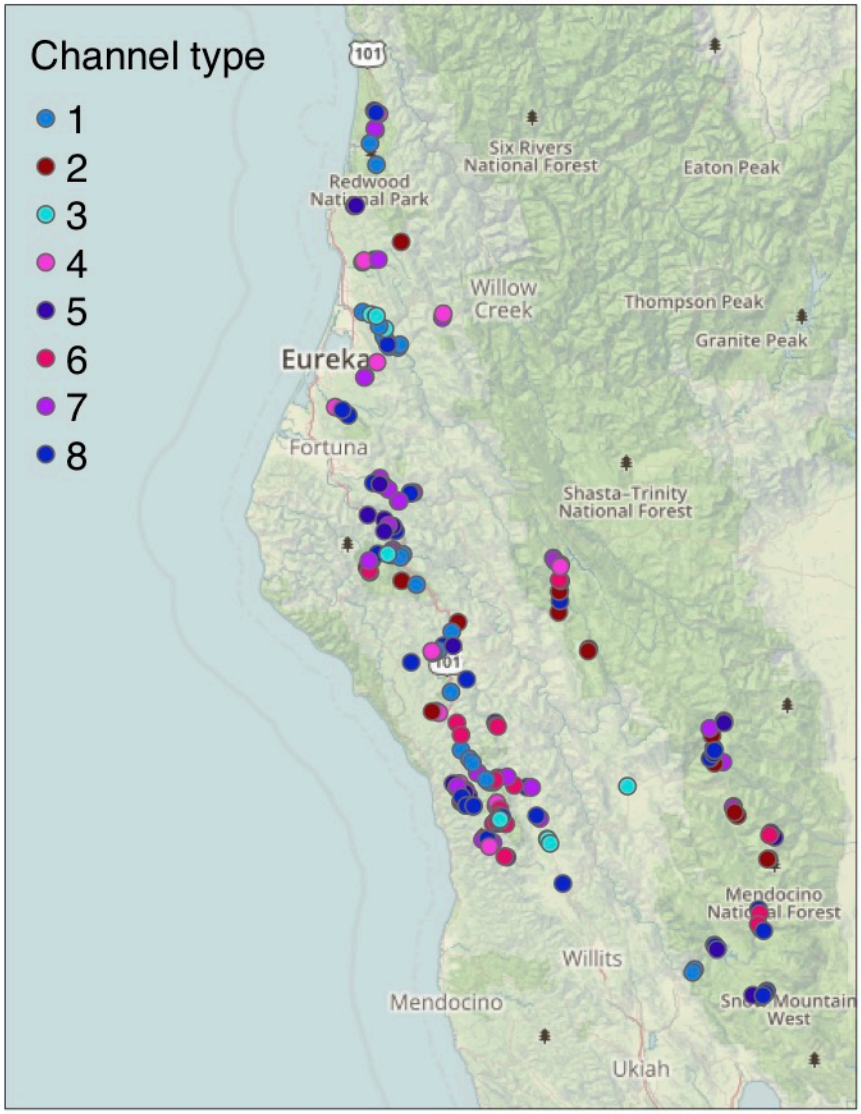


Figure 13. Location of site surveys colored by channel type in the North Coast region.

3.3. North Central Coast

A total of 104 sites were used in the North Central Coast region classification. The final NMDS solution recorded two- and three-dimensional stress values of 0.163 and 0.094, respectively (Figure 14). The first and second principle component axes (PCAs) accounted for 33% and 22% of the variance in the data, respectively. Both were most greatly associated with D_{84}.

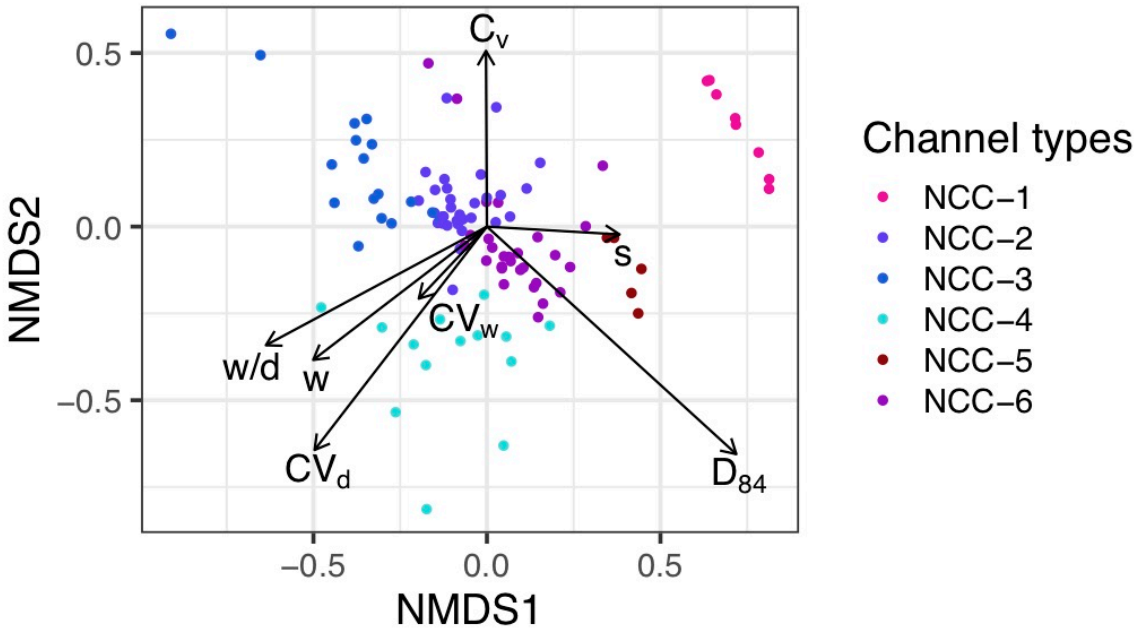


Figure 14. NMDS scatter plot with principal component vectors for attributes used in the classification of sites. Scattered sites are colored according to final channel type. Longer principal component vectors indicate greater influence on the scattering of sites.

Six channel types were identified through the WHC with heuristic refinement (Figure 15). The final six channel types were the result of CART analysis which resulted in a successful prediction rate of 89.4% (Figure 16). Ten-fold cross-validation, a measure of prediction on unseen data, was 85.6%. The final classification tree was defined by CV_d, *w/d,* C_v, D_{84}, *w*, and *s*. Final channel types were made up of between 5 and 32 sites. Median attributes of each channel type can be found in Table 3. The channel types are summarized in Figure 17. Distributions of channel type attributes and classified site locations can be found in Figure 18 and Figure 19, respectively.

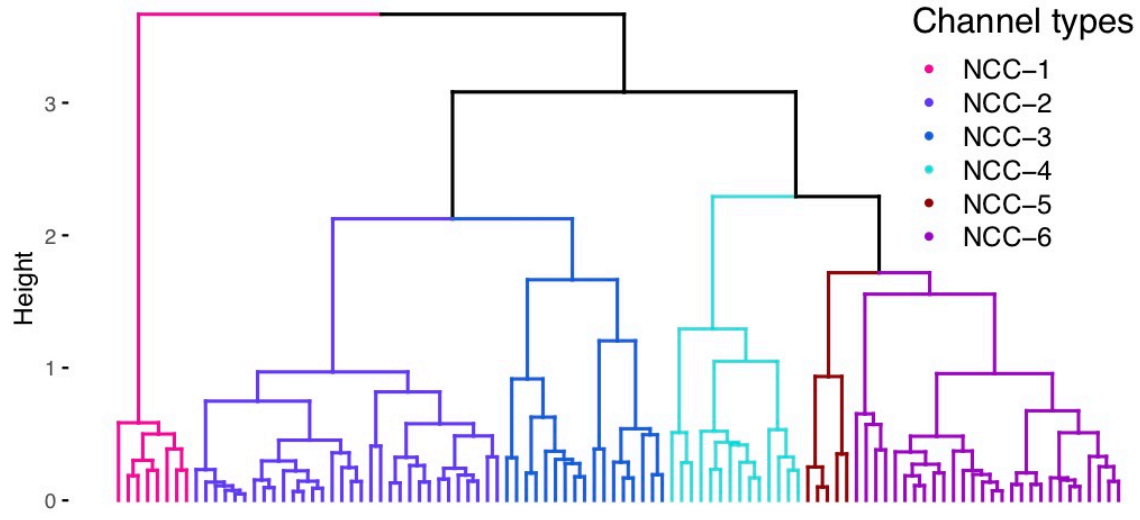


Figure 15. Ward's hierarchical clustering dendrogram with associated heuristic channel types.

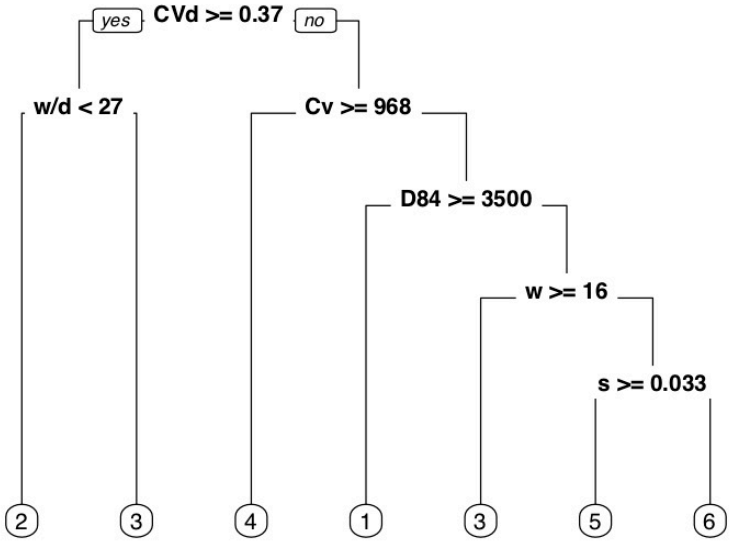


Figure 16. Final yes/no classification tree that produced suitable prediction and cross-validation percentages. Channel types are indicated by number at the bottom of the figure.

Streams in the North Central Coast region were classified into four confined channel types, one partly-confined channel type, and one unconfined channel type. Streams in steep, mountainous settings are classified as *confined, cobble-boulder, cascade/step-pool* (NCC-5) streams. At moderate slopes and smaller confined, sediments sizes, TVA values decrease significantly from other confined channel types and result in *confined, gravel-cobble, uniform* (NCC-6) streams. Confined streams with the lowest slopes are differentiated by both sediment size and TVA values. *Confined, cobble-boulder, uniform width-bed undulating* (NCC-1) streams show uniform width characteristics but high depth variability. These streams are likely subjected to a greater colluvial inputs at the valley margin. The last confined channel type, *confined, gravel-cobble, riffle-pool* (NCC-2), exists at the lowest slopes and has smaller sediment sizes than NCC-1. In addition, these streams show both high width and depth variability and exhibit a characteristic riffle-pool morphology. The largest rivers, both in size and contributing area, in the North Central Coast region are *partly-confined, gravel, riffle-pool* (NCC-3) streams. Finally, the *unconfined, gravel, uniform* (NCC-4) streams show low TVA values and exist in valley settings throughout the region

Table 3. Median channel attributes of six channel types within the North Central Coast region.

Channel Type	A_c (km^2)	s	d (m)	w (m)	w/d	CV_d	CV_w	D50 (mm)	D84 (mm)	C_v (m)
NCC-1	24	0.0070	0.65	7.99	12.3	0.33	0.18	64	5000	24
NCC-2	37	0.0054	0.56	8.01	17.9	0.44	0.23	22.6	64	73
NCC-3	184	0.0016	0.90	20.50	30.1	0.41	0.26	22.6	45	155
NCC-4	21	0.0047	0.36	4.87	15.2	0.25	0.18	22.6	54.5	1947
NCC-5	5	0.0622	0.44	4.25	9.7	0.25	0.26	64	2000	15
NCC-6	17	0.0105	0.53	6.37	13.0	0.23	0.22	32	90	60

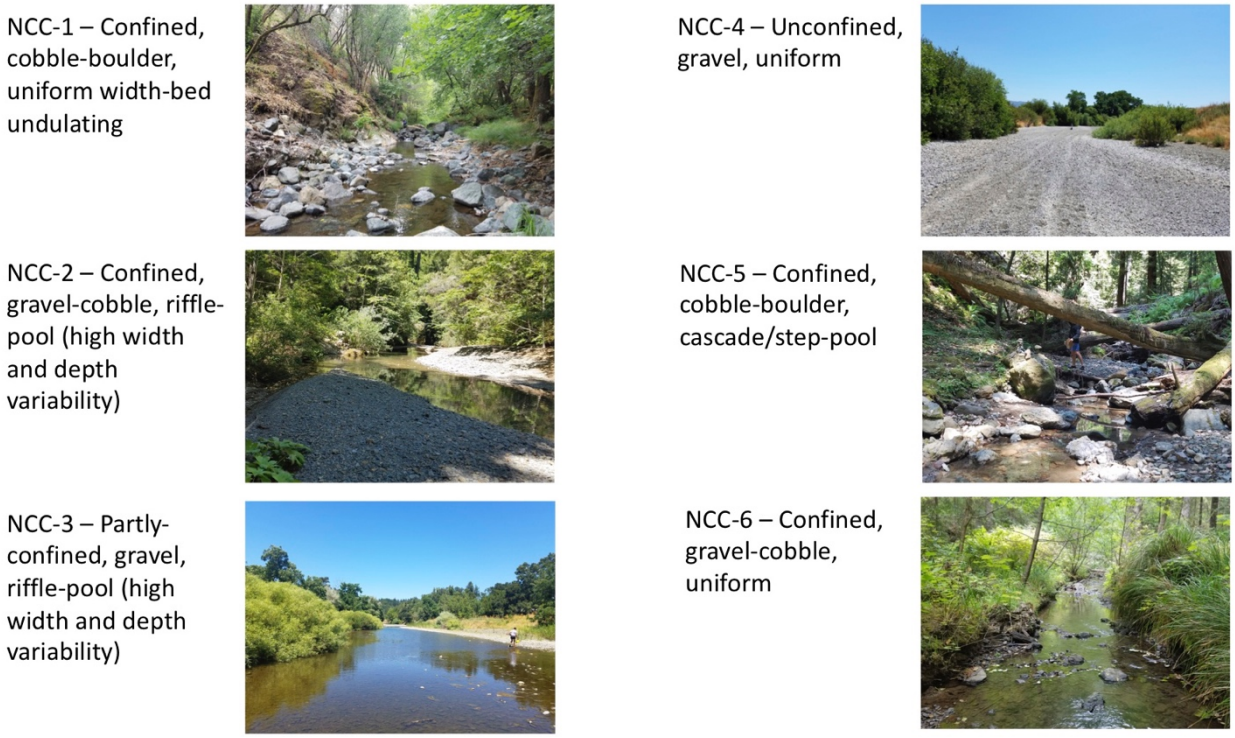


Figure 17. The six channel types within the North Central Coast region developed by multivariate statistical analysis with heuristic refinement.

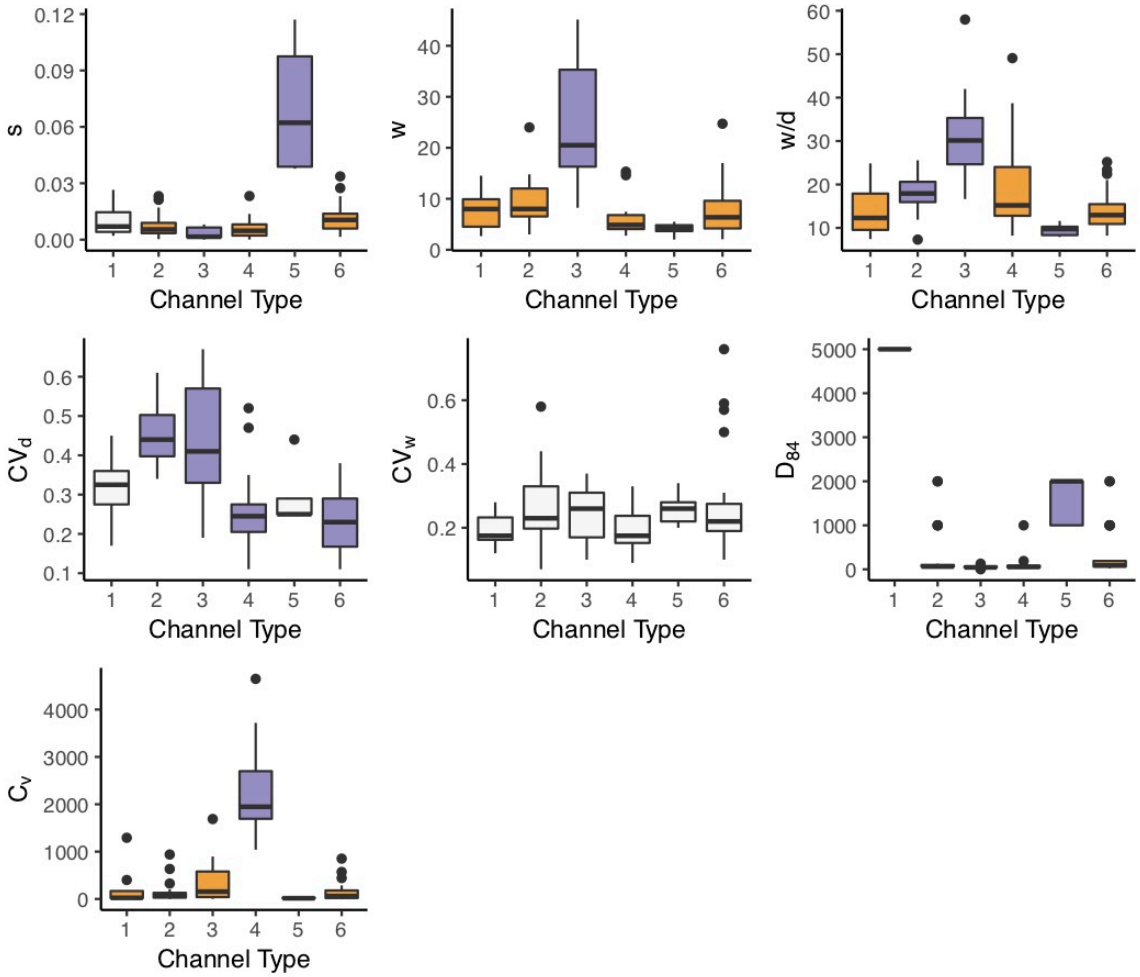


Figure 18. Box-and-whisker plots of classification metrics for each channel type in the North Central Coast region. Purple boxes represent channel types with significant differences from multiple other channel types based on Dunn's Test. Yellow boxes represent channel types significantly different than one other channel type and empty boxes represent a channel type with no significant differences.

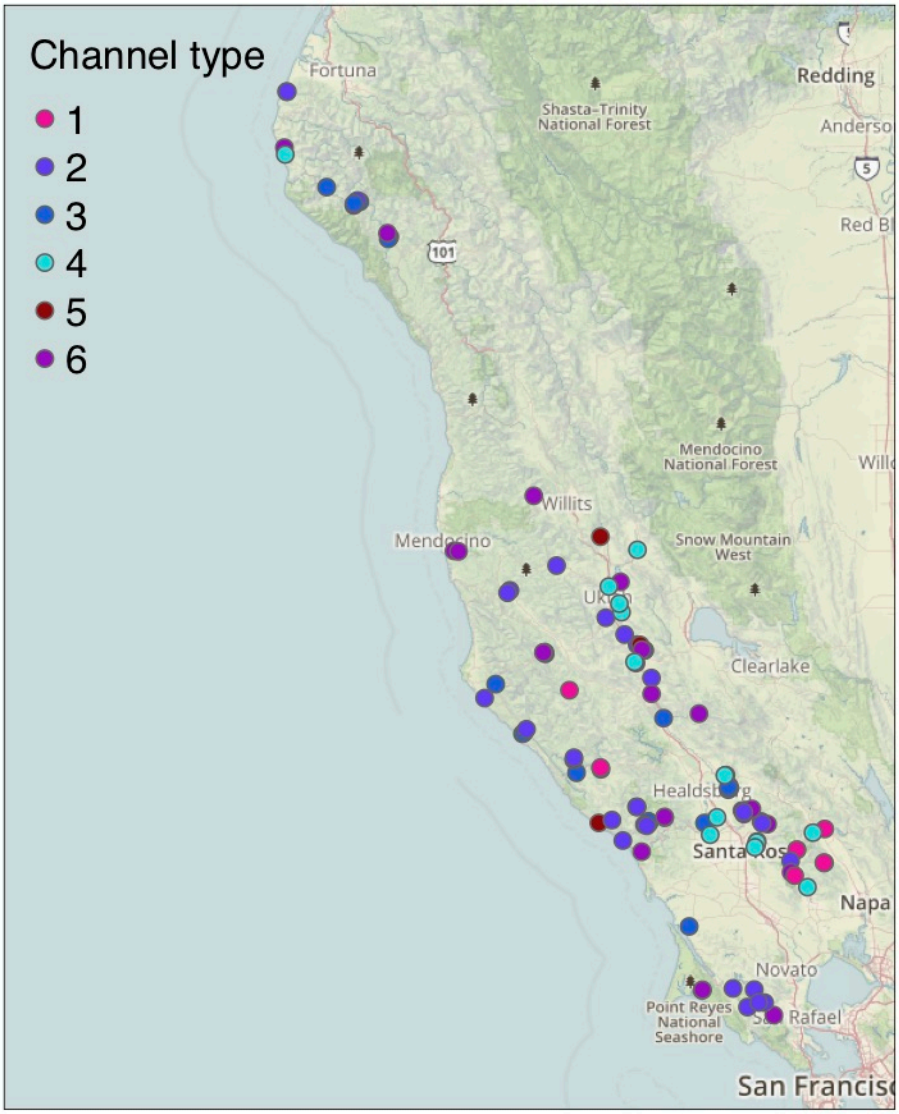


Figure 19. Location of site surveys colored by channel type in the North Central Coast region.

3.4. South Central Coast

A total of 119 sites were used in the South Central Coast region classification. The final NMDS solution recorded two- and three-dimensional stress values of 0.141 and 0.095, respectively (Figure 20). The first and second principle component axes (PCAs) accounted for 28% and 23% of the variance in the data and were most greatly associated with D_{84} and CV_d, respectively.

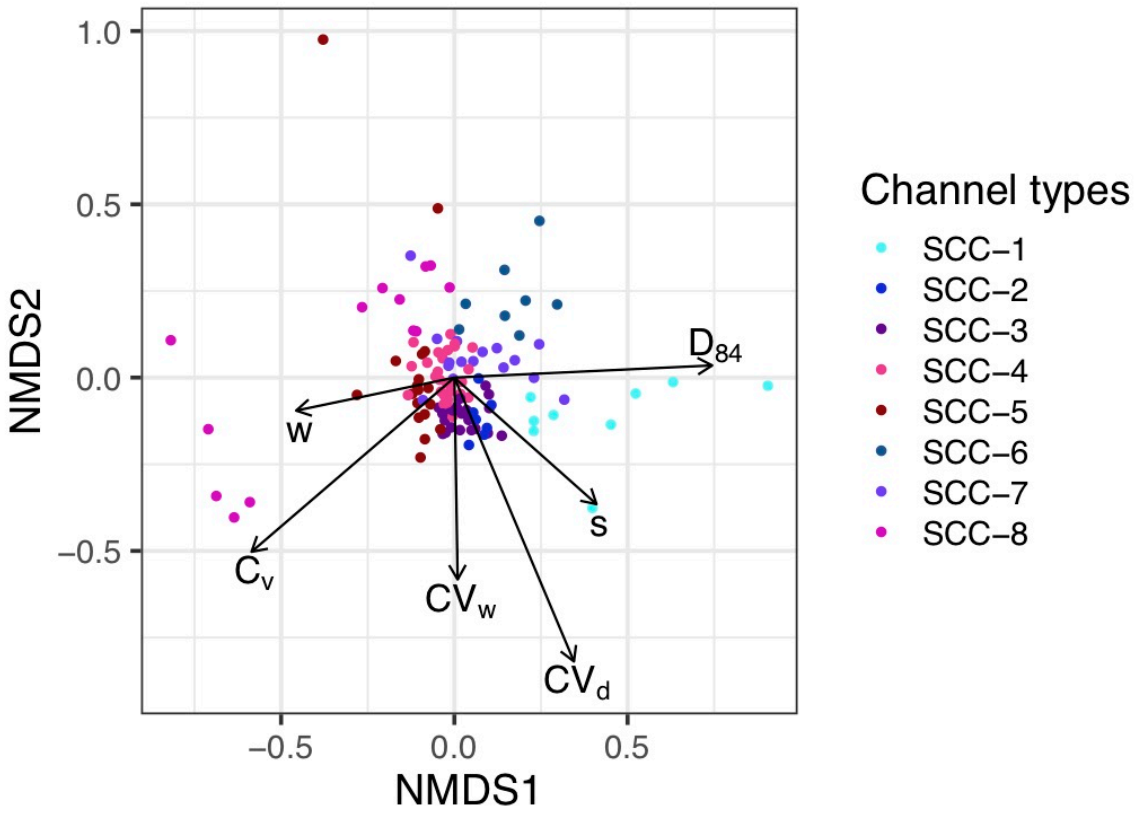


Figure 20. NMDS scatter plot with principal component vectors for attributes used in the classification of sites. Scattered sites are colored according to final channel type. Longer principal component vectors indicate greater influence on the scattering of sites.

Eight channel types were identified through the WHC with heuristic refinement (Figure 21). The final eight channel types were the result of CART analysis which resulted in a successful prediction rate of 89.9% (Figure 22). Ten-fold cross-validation, a measure of prediction on unseen data, was 74.8%. The final classification tree was defined by D_{84}, CV_d, C_v, and w. Final channel types were made up of between 7 and 27 sites. Median attributes of each channel type can be found in Table 4. The channel types are summarized in Figure 23. Distributions of channel type attributes and classified site locations can be found in Figure 24 and Figure 25, respectively.

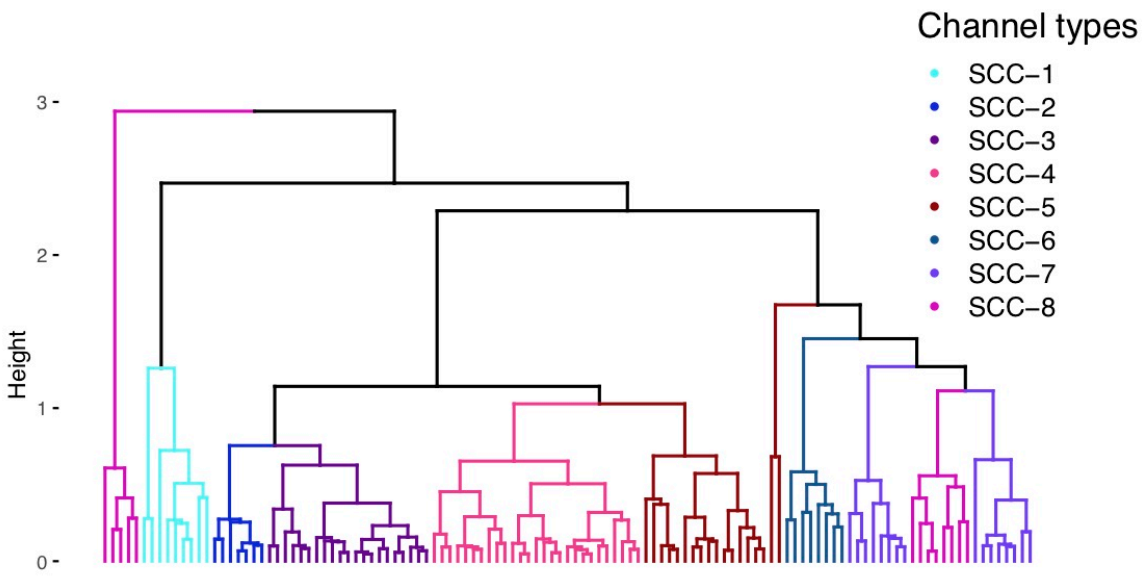


Figure 21. Ward's hierarchical clustering dendrogram with associated heuristic channel types.

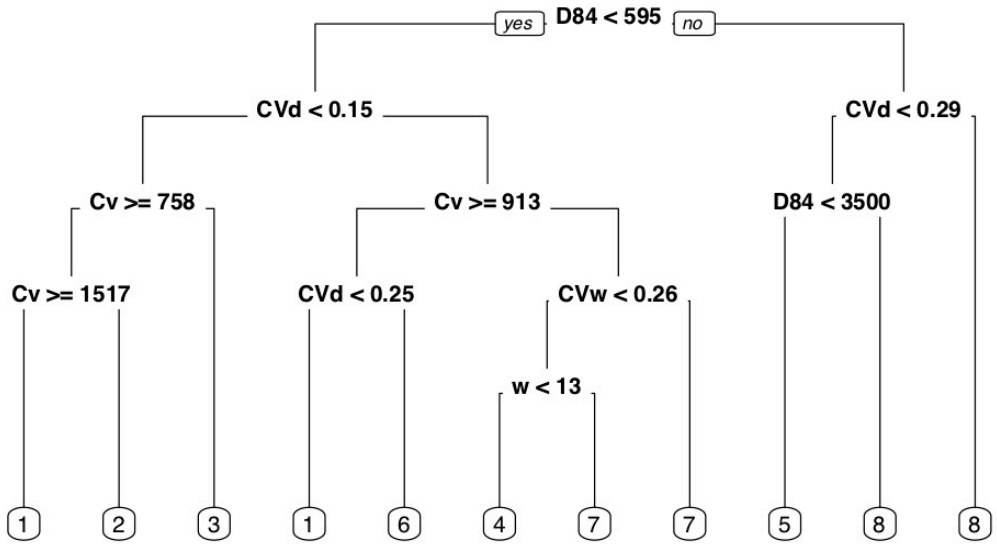


Figure 22. Final yes/no classification tree that produced suitable prediction and cross-validation percentages. Channel types are indicated by number at the bottom of the figure.

Streams in the South Central Coast region were classified into three confined channel types, two partly-confined channel types, and three unconfined channel types. Confined channel types were most clearly differentiated by slope, sediment size, and TVA values. *Confined, cobble-boulder, cascade/step-pool* (SCC-5) streams are found at the greatest slopes and also have the largest sediment sizes. At lower slopes, but of similar grain size, *confined, cobble-boulder, riffle-pool* (SCC-8) streams exhibit high TVA values. This is in contrast to *confined, gravel-cobble, uniform* (SCC-4) streams that exhibit low TVA values and also have smaller sediment sizes. Two types of

partly-confined streams, *partly-confined, gravel-cobble, uniform* (SCC-3) and *partly-confined, gravel-cobble, riffle-pool* (SCC-7) are differentiated by low and high TVA values, respectively. These differences also exist in unconfined settings. *Unconfined, gravel-cobble, riffle-pool* (SCC-6) streams show high width and depth variability, while both *unconfined, high-order, sand-gravel, uniform* (SCC-1) and *unconfined, low-order, sand-gravel, uniform* (SCC-2) streams show low width and depth variability. The SCC-1 channel type is differentiated from SCC-2 channel type by greater contributing areas and width-to-depth ratios.

Table 4. Median channel attributes of seven channel types within the South Central Coast region.

Channel Type	A_c (km^2)	s	d (m)	w (m)	w/d	CV_d	CV_w	D50 (mm)	D84 (mm)	C_v (m)
SCC-1	157	0.0035	0.87	15.71	16.3	0.13	0.19	2.8	16	2520
SCC-2	11	0.0048	0.55	3.85	6.5	0.12	0.20	5.6	32	1055
SCC-3	22	0.0088	0.70	7.14	9.6	0.11	0.19	16	64	122
SCC-4	14	0.0123	0.61	5.00	7.9	0.22	0.16	16	64	73
SCC-5	28	0.0256	0.73	6.22	8.2	0.17	0.17	77	1000	13
SCC-6	51	0.0075	0.87	6.50	10.4	0.36	0.25	22.6	38.5	1954
SCC-7	97	0.0088	0.97	11.87	13.2	0.22	0.29	27.3	109	213
SCC-8	20	0.0147	0.87	5.48	8.9	0.38	0.24	64	1000	79

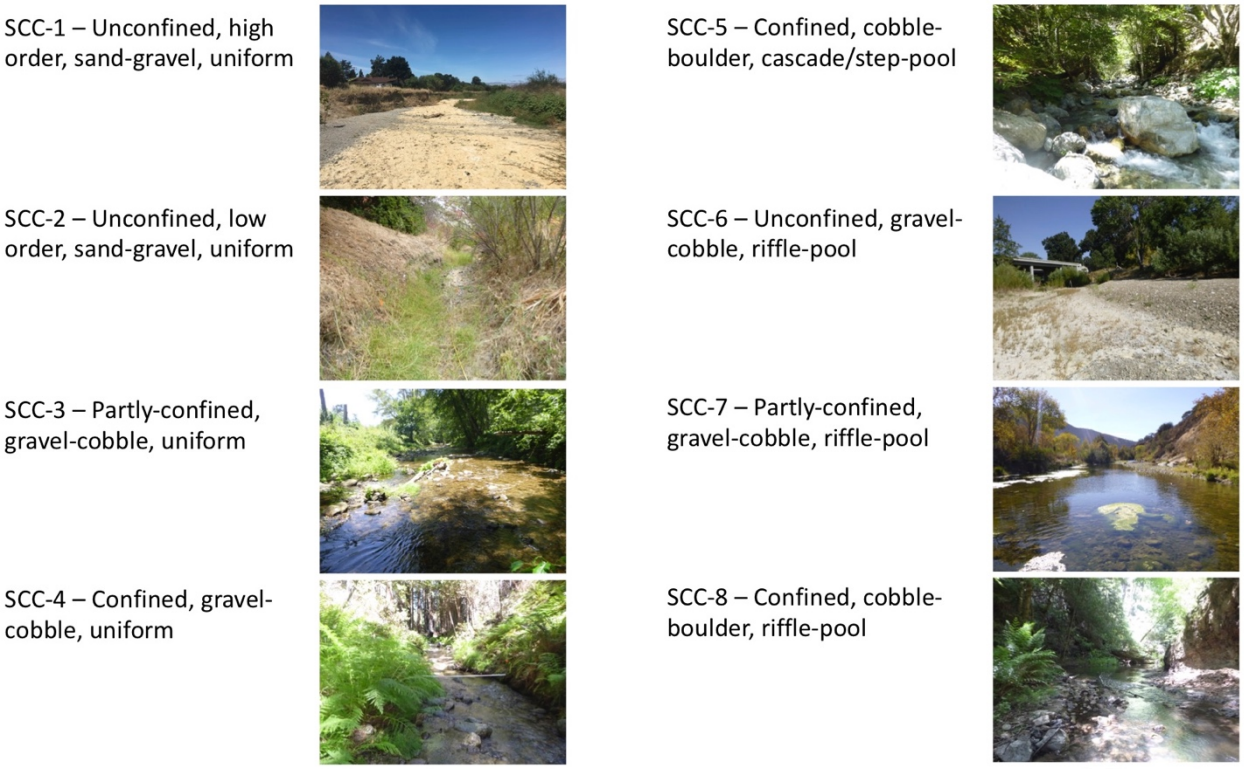


Figure 23. The seven channel types within the South Central Coast region developed by multivariate statistical analysis with heuristic refinement.

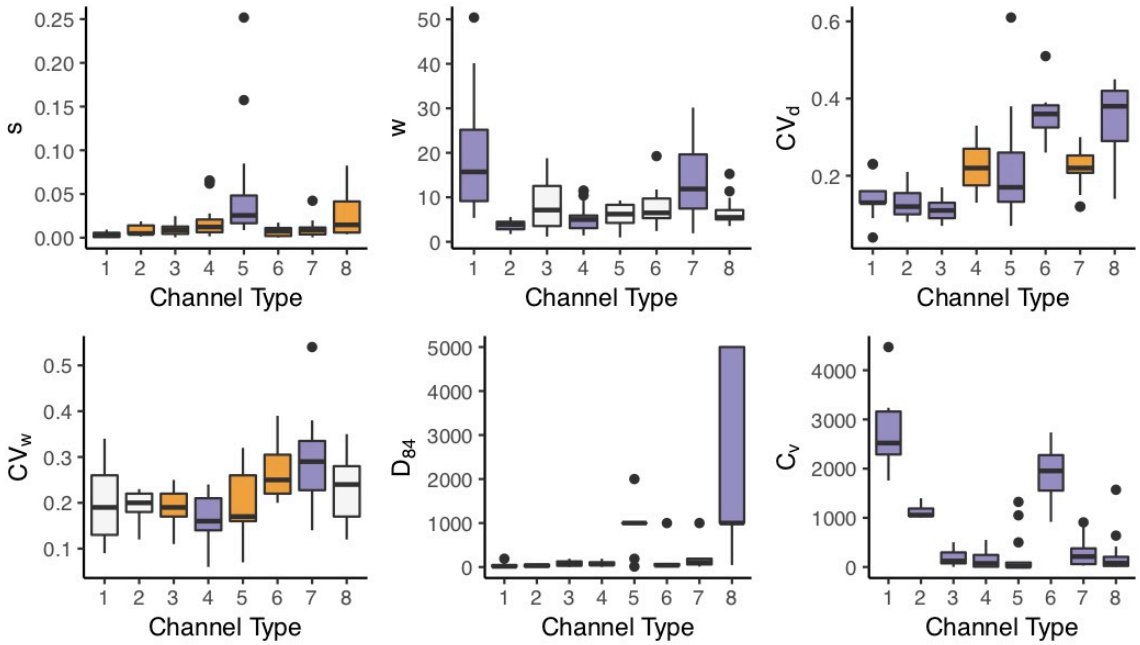


Figure 24. Box-and-whisker plots of classification metrics for each channel type in the South Central Coast region. Purple boxes represent channel types with significant differences from multiple other channel types based on Dunn's Test. Yellow boxes represent channel types significantly different than one other channel type and empty boxes represent a channel type with no significant differences.

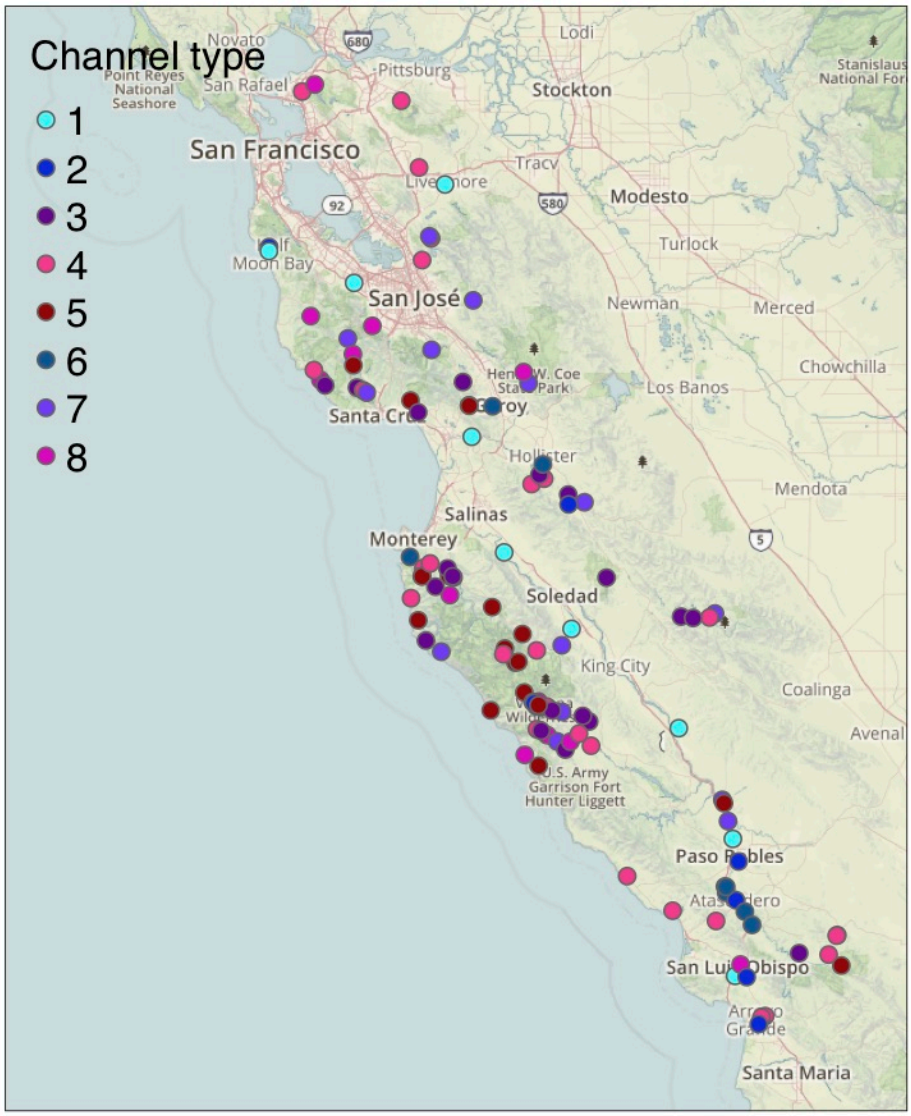


Figure 25. Location of site surveys colored by channel type in the South Central Coast region.

3.5. South Coast

A total of 67 sites were used in the South Coast region classification. The final NMDS solution recorded two- and three-dimensional stress values of 0.164 and 0.103, respectively (Figure 26). The first and second principle component axes (PCAs) accounted for 30% and 20% of the variance in the data and were most greatly associated with C_v and *w/d*, respectively.

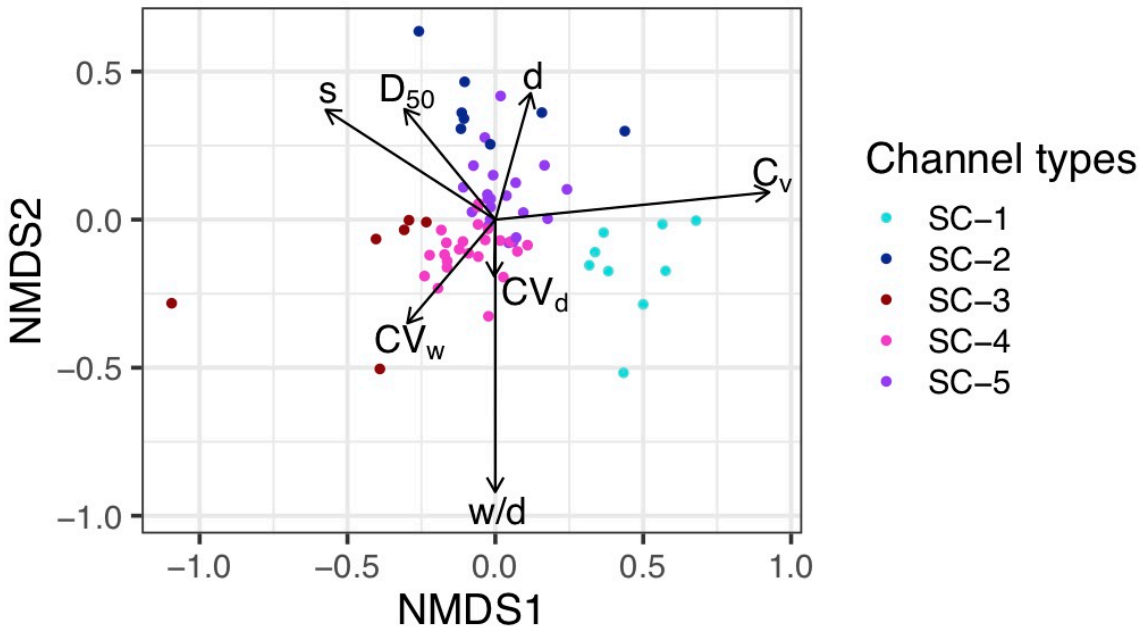


Figure 26. NMDS scatter plot with principal component vectors for attributes used in the classification of sites. Scattered sites are colored according to final channel type. Longer principal component vectors indicate greater influence on the scattering of sites.

Five channel types were identified through the WHC with heuristic refinement (Figure 27). The final five channel types were the result of CART analysis which resulted in a successful prediction rate of 91% (Figure 28). Ten-fold cross-validation, a measure of prediction on unseen data, was 82.1%. The final classification tree was defined by CV_d, *w/d*, *s*, and C_v. Final channel types were made up of between 6 and 23 sites. Median attributes of each channel type can be found in Table 5. The channel types are summarized in Figure 29. Distributions of channel type attributes and classified site locations can be found in Figure 30 and Figure 31, respectively.

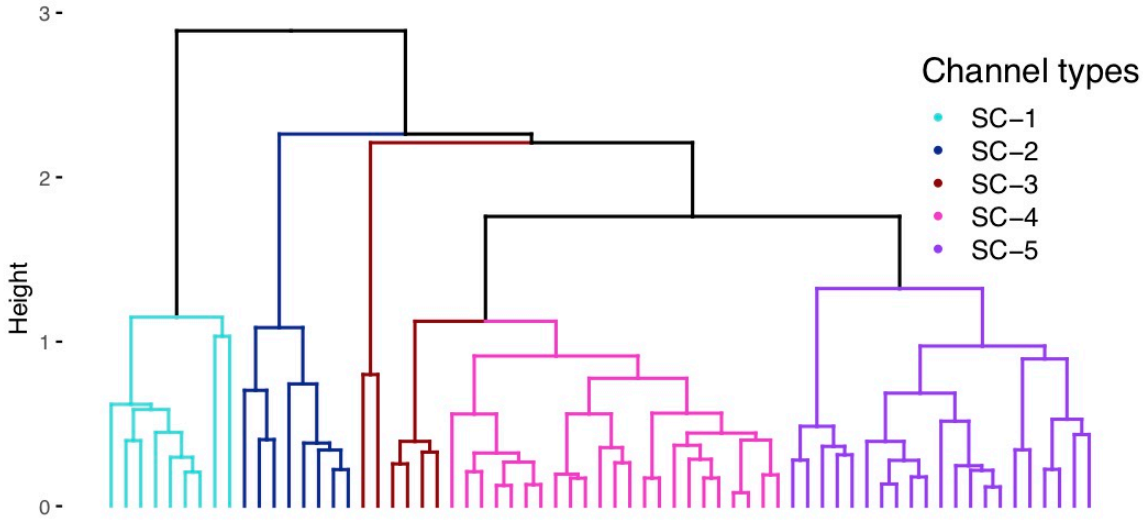


Figure 27. Ward's hierarchical clustering dendrogram with associated heuristic channel types.

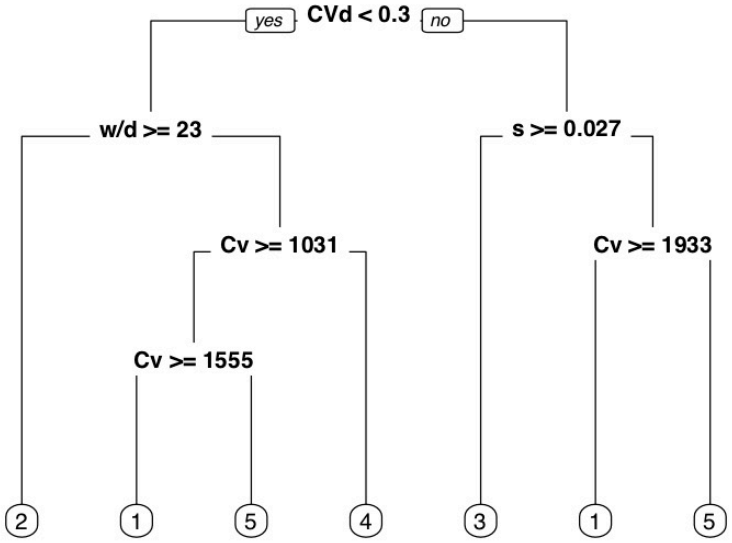


Figure 28. Final yes/no classification tree that produced suitable prediction and cross-validation percentages. Channel types are indicated by number at the bottom of the figure.

The South Coast region was defined by two confined channel types, two partly-confined channel types, and one unconfined channel type. Confined channel types were defined most prominently by slope. *Confined, cobble-boulder, cascade/step-pool* (SC-3) streams exist at steeper slopes and have greater variability than *confined, gravel-cobble, uniform* (SC-4) streams. Partly-confined streams show large differences in width-to-depth ratio due to the multi-threaded characteristics of *partly-confined, gravel, braided* (SC-2) streams. The SC-2 streams are in contrast to the single-threaded *partly-confined, gravel-cobble, riffle-pool* (SC-5) streams, which also exhibit high depth variability. Finally, streams in unconfined settings generally show small sediment sizes and low TVA values and are classified as *unconfined, sand-gravel, uniform* (SC-1) streams.

Table 5. Median channel attributes of five channel types within the South Coast region.

Channel Type	A_c (km^2)	s	d (m)	w (m)	w/d	CV_d	CV_w	D50 (mm)	D84 (mm)	C_v (m)
SC-1	52	0.0053	0.63	6.67	10.0	0.24	0.20	5.6	22.6	2480
SC-2	124	0.0143	0.42	17.92	36.7	0.25	0.30	10.8	48	202
SC-3	28	0.0397	0.54	4.94	9.7	0.32	0.28	90	2000	31
SC-4	25	0.0182	0.64	5.68	9.6	0.21	0.21	11	190	94
SC-5	182	0.0054	0.60	7.78	12.8	0.35	0.28	16	128	126

SC-1 – Unconfined, sand-gravel, uniform

SC-2 – Partly-confined, gravel, braided

SC-3 – Confined, boulder, cascade/step-pool

SC-4 – Confined, cobble-boulder, uniform

SC-5 – Partly-confined, gravel-cobble, riffle-pool

Figure 29. The five channel types within the South Coast region developed by multivariate statistical analysis with heuristic refinement.

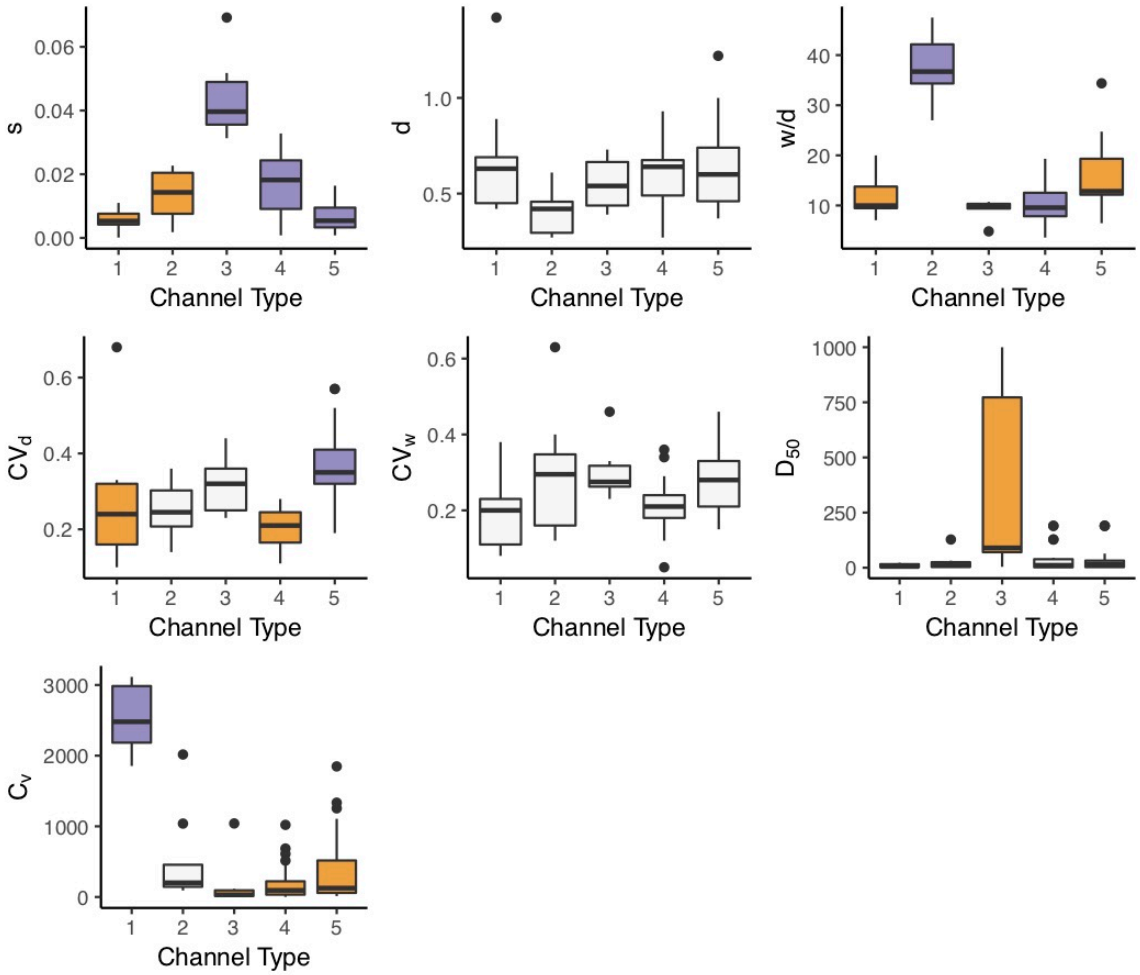


Figure 30. Box-and-whisker plots of classification metrics for each channel type in the South Coast region. Purple boxes represent channel types with significant differences from multiple other channel types based on Dunn's Test. Yellow boxes represent channel types significantly different than one other channel type and empty boxes represent a channel type with no significant differences.

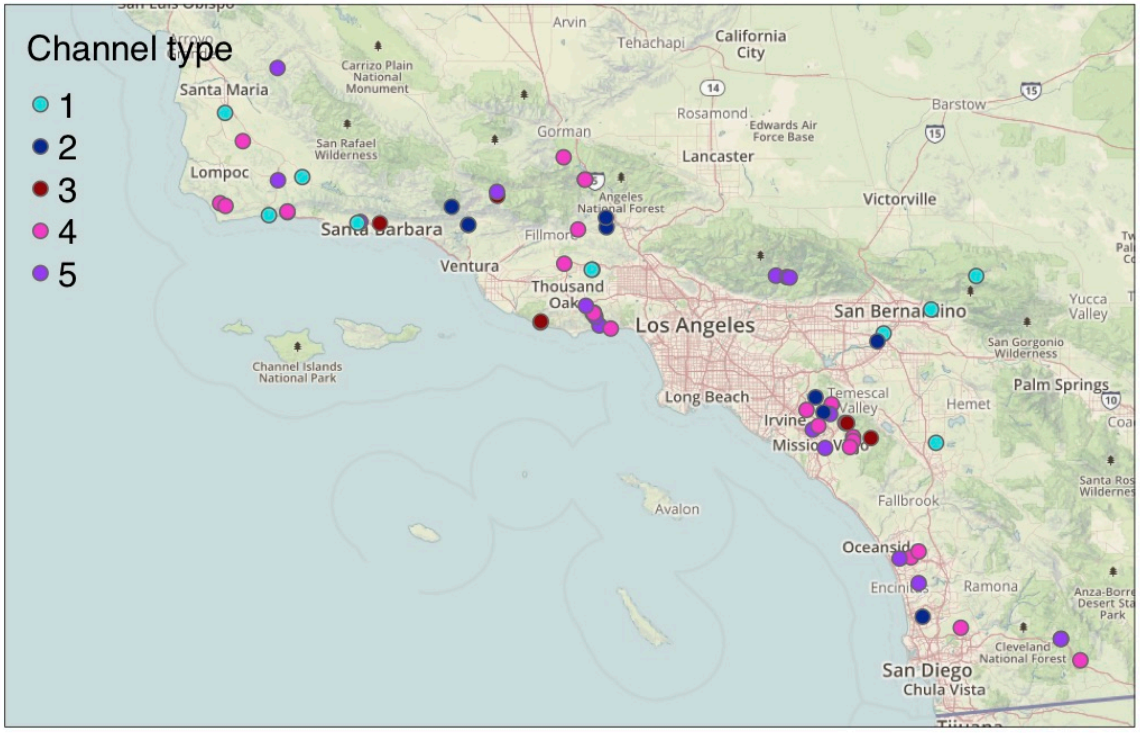


Figure 31. Location of site surveys colored by channel type in the South Coast region.

4. Stream Network Prediction Methodology

4.1. Background on Machine Learning

In the cognition process defining machine learning (ML), model parameters are *internally* optimized against some performance metrics: the model *trains* and self-improves. *Hyper*-parameters set prior to training define the architecture of the model and *tuning* is used to select hyper-parameters leading to best performance. For a polynomial model, the order of the polynomial is the hyper-parameter, its coefficients are parameters optimized by minimizing the sum of the squared residuals and tuning identifies the order yielding the highest performance. Importantly, what separates ML from a curve fitting exercise is ML's ability to generalize patterns. Such generalization is achieved by *resampling* the initial dataset into a training set and a test set such that the robustness of the learned pattern can be assessed against data unseen during training.

Learning tasks are usually separated into unsupervised and supervised ML approaches. Both approaches use input data, predictor variables or predictors to extract or predict some information about a dataset. Unsupervised learning identifies patterns in input data and uses these patterns to cluster observations. Supervised learning uses known information to approximate the relationship between input data and output which can then be used to predict output from new input. When the distribution of such output is continuous, the supervised learning task is a regression problem. When the output distribution is discrete, it corresponds to labels or classes and the supervised

learning task is a classification problem. For a regression problem, the relationship learned between output and input corresponds to a mathematical mapping. For a classification problem, what is learned pertains to *class boundaries*, divisions of the multi-dimensional predictor space separating subsets where observations of a given class are dominantly located.

This study aims to predict channel types in five study regions: the Klamath, North Coast, North Central Coast, South Central Coast, and South Coast. In each region, a supervised ML classification approach answers the following question: Given a set of predictors, which label (channel type) should be assigned to each stream-reach? To achieve this study aim, we developed a two-tiered ML framework to: (i) define a tractable problem; (ii) assess the performance of models in statistical learning; and (iii) assess the performance of models in predictive modeling (Figure 32).

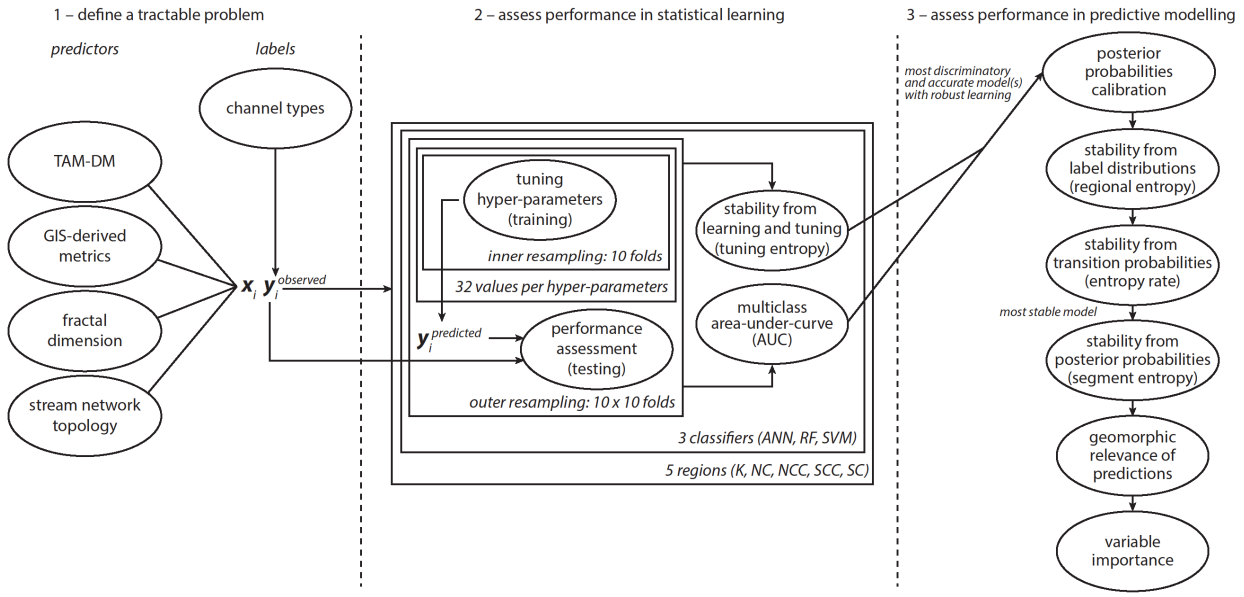


Figure 32: Schematic of the three-tiered machine-learning framework.

4.2. Problem Definition

Our ML approach derives the relationship between labels corresponding to regional channel types defined in Section 3 and 147 predictors (Table 6). Several categories of predictors with a documented influence on landscape and channel morphology were considered: a) channel confinement ; b) stream network topology ; c) statistical roughness or fractal dimension of the topography (Pastor-Satorras and Rothman 1998; Dodds and Rothman 2000; Wilson and Dominic 1998; Lifton and Chase 1992; Sung and Chen 2004; Faghih and Nourbakhsh 2015; Liucci and Melelli 2017; Duclut and Delamotte 2017); and d) nine Terrain Analysis Metrics (e.g. curvature, (Rosgen 1994; Hurst et al. 2012; Prancevic and Kirchner 2019)). Previous research in the Sacramento Basin and the South Fork Eel River catchment highlighted that contextual variables like geology, soils, land cover and climate did not contribute to the classification and made the problem noisier and more complex.

Since the distribution of Terrain Analysis Metrics across the landscape differentiates stages of landscape maturity (Bonetti and Porporato 2017) and channel types (Lane et al. 2017), five Distribution Metrics were estimated: mean, median, minimum, maximum, standard-deviation and skewness. Most predictors were derived from three core datasets: (i) the 10-m National Elevation Dataset (Gesch et al. 2002) NED; (ii) the stream network from the National Hydrology Dataset (McKay et al. 2012) NHDPlusV2; and (iii) the Stream-Catchment Dataset (StreamCat; Hill et al. 2015) which aggregates previous data sources at the level of the NHD stream lines.

The essence of the problem addressed in this study is the prediction of channel types, derived from data at the 10^0 to 10^2 m scale, by predictors available at coarser spatial scales, typically 10^2 to 10^5 m (Table predictors). In addition, defining the relevant scales to distinguish controlling physical processes is often difficult. To address this challenge, we considered 32 sets of 5 nested spatial scales with upper scales from ~ 0.6 km to ~ 82 km when calculating the fractal dimension following the box-counting method from Liucci and Melelli (2017). This can be seen as an alternative to using multiscale decomposition of the topography. In addition, the Terrain Analysis Metrics and Distribution Metrics were estimated over two spatial coverages to represent hillslope and near-channel processes, respectively: a $512 - \mathrm{m}$ square tile centered at the midpoint of each stream-interval and along a 100-m wide near-channel buffer. These two spatial coverages combined with nine Terrain Analysis Metrics and five Distribution Metrics result in 108 TAM-DM predictors (Table 6).

Table 6. Predictors used in the Machine Learning framework. The 10-m National Elevation Dataset (Gesch et al., 2002) and the Stream-Catchment Dataset (Hill et al., 2015) are publicly available on download platform from the United States Geological Survey and the United States Environmental Protection Agency, respectively. The stream network from the National Hydrology Dataset (McKay et al., 2012) is publicly available on both platforms.

predictors group	predictor name	spatial scale	original data	methodology
TAM-DM (108)	elevation	512 m; 100-m buffer	Gesch et al. (2002)	Hijmans et al. (2018)
	slope	512 m; 100-m buffer	Gesch et al. (2002)	Hijmans et al. (2018)
	aspect	512 m; 100-m buffer	Gesch et al. (2002)	Hijmans et al. (2018)
	roughness	512 m; 100-m buffer	Gesch et al. (2002)	Hijmans et al. (2018)
	flow direction	512 m; 100-m buffer	Gesch et al. (2002)	Hijmans et al. (2018)
	planform curvature	512 m; 100-m buffer	Gesch et al. (2002)	Florinsky (1998)
	profile curvature	512 m; 100-m buffer	Gesch et al. (2002)	Florinsky (1998)
	Topographic Position Index	512 m; 100-m buffer	Gesch et al. (2002)	Hijmans et al. (2018)
	Terrain Ruggedness Index	512 m; 100-m buffer	Gesch et al. (2002)	Hijmans et al. (2018)
GIS-metrics (3)	channel slope	200 m	Gesch et al. (2002)	ESRI (2016)
	confinement	-	Gesch et al. (2002)	Byrne et al. (2019)
	sediment supply	-	Haan, Barfield, and Hayes (1994)	Renard et al. (1997)
network topology (4)	drainage area	-	McKay et al. (2012)	Hill et al. (2015)
	Strahler's stream order	-	McKay et al. (2012)	Strahler (1957)
	Local Drainage Density	-	McKay et al. (2012)	Danesh-Yazdi, Tejedor, and Foufoula-Georgiou (2017)
fractal dimension (32)	Hurst coefficients	640 m to 82 km	Gesch et al. (2002)	Liucci and Melelli (2017)

4.3. Assess Performance in Statistical Learning

Assessment of the statistical performance quantifies the stability of the learning process. While numerous ML models exist, the best model for a given task is often unknown at the start (Luengo and Herrera 2013), leading to the common practice of training multiple models on the same task. A machine learning model is defined by its underlying algorithm, pre-processing, resampling, tuning strategy and performance metric. We describe these model components in the following sections.

4.3.1. Algorithms

Based on previous results in the Sacramento Basin and the South Fork Eel River catchment, three prominent ML algorithms were trained: Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Network (ANN). These three key algorithms are described below and implementation of the algorithms was performed using the R packages mlr (Bernd Bischl et al. 2016).

Linear SVM finds the linear boundary between two distinct classes by maximizing the margin between the class boundary and each class's closest point(s) (Cortes and Vapnik 1995). Those points are the support vectors for the boundary. Non-linear boundaries are obtained by a non-linear kernel version of SVM, transforming predictor space so that the problem becomes linearly solvable. The most common kernel used to perform this so-called kernel trick is the radial basis function. SVMs solve multi-class problems by transforming them into a set of two-classes problems for which multiple binarization strategies exist.

RF is an ensemble of classification and regression trees built from random subsets of predictors (Breiman 1984). At each split of each tree, a predictor is chosen as splitting variable based on an information selection process (e.g. Gini coefficient) which ultimately provides a measure of variable importance. In addition, the ensemble decision process from uncorrelated (random) trees lead to great performance when the training dataset is reduced, noisy or both (Fox et al. 2017).

ANN are formed by successive layers of connected neurons each characterized by a weight and an activation. The weight describes the strength of the connection of the neuron to neurons in the next layer. The activation results from the combination, through an activation function, of the inputs that a neuron receives from the previous layer. The first and last layers of such networks correspond to input and output while middle layers are termed hidden. A network with a large number of hidden layers is called a deep ANN with emerging applications in hydrologic sciences (Shen 2018), geophysics (Bergen et al. 2019) and earth system science (Reichstein et al. 2019).

4.3.2. Pre-processing

Prior to statistical learning, a set of transformations was applied to predictors. Such a pre-processing includes estimating missing values with median imputation, removing predictors with near-zero variance and applying centering and scaling transformations to collapse each predictor distribution to a normal distribution. The labels in this study suffer from the common ML challenge of unequal representation.The Synthetic Minority Oversampling Technique (SMOTE, (Chawla et al. 2002)) assigns predictors along the edges connecting the 5-Nearest Neighbors from randomly selected observations and was used in SMOTE runs to address the imbalance of the training set.

4.3.3. Resampling

The main underlying resampling technique used here is 10-fold cross-validation, which allows all data to be used in both training and testing (Burman 1989): the data are randomly separated in ten parts or folds and successively, while one fold is held out, the nine other folds are used for training while performance is assessed against the hold-out fold. Repeated cross-validation addresses the potential bias introduced by the initial random selection of the folds. Because the structure of the spatial correlation is different across each region, we refrain to use spatial cross-validation (Schratz et al. 2018) for comparing results between regions. In addition, the benchmark was performed using nested resampling which estimates the robustness of the tuning process and limits over-fitting by using two nested loops: an inner loop for model tuning and an outer loop for model selection (B. Bischl et al. 2012). While traditional resampling leads to a distribution of model performance, nested resampling additionally provides a distribution of best-tuned hyper-parameters. Here, the outer resampling is a 10-fold stratified cross-validation repeated 10 times, the inner resampling is a 10-fold stratified cross-validation.

4.3.4. Tuning Strategy

For each model and for each region, the set of best hyper-parameters was tuned by a discrete grid search across 32 different values per parameter for RF and SVM, and with a random discrete search with 32 iterations for ANN. Coupled with the 10 repeated 10-fold outer cross-validation and with the inner 10-fold cross-validation, this tuning strategy leads to the training of 96,000 models per region.

4.3.5. Performance Metric

The performance in statistical learning is assessed by benchmarking machine-learning models across the five regions of study using area-under-curve (AUC) and hyper-parameter tuning entropy. AUC is preferred here to accuracy for its higher discrimination performance, its clear relation to class-separability and its suitability for limited dataset (Rosset 2004; Huang and Ling 2005; Ferri, Hernández-Orallo, and Modroiu 2009). In addition to AUC, the performance of the model is assessed by estimating the hyper-parameter tuning entropy from the distribution of their best-tuned hyper-parameters resulting from the nested resampling. Increasing entropy H corresponds to increasing unpredictability and here to an unstable learning process (Shannon 1948):

$$H = -\sum_i p_i \log p_i$$

with p_i probabilities.

4.3.6. Posterior Calibration

ML models output posterior probabilities that often require calibration. Such calibration corrects the potential distortion of the posterior probabilities when compared to empirical probabilities and improves model performance (DeGroot and Fienberg 1983; Zadrozny 2002; Niculescu-Mizil and Caruana 2005). Given the sigmoid-shape of most distortions, J. Platt and others (1999) proposed a sigmoid calibration to address this effect. Other useful approaches include Bayesian

calibration and isotonic scaling (Zadrozny and Elkan 2002), which both require binarizing the problem (J. C. Platt, Cristianini, and Shawe-Taylor 2000; Kijsirikul and Ussivakul 2002; Dong, Frank, and Kramer 2005; A. C. Lorena and Carvalho 2010; Quiterio and Lorena 2016; Adnan and Islam 2015; Melnikov and Hüllermeier 2018). In this study, posterior calibration was performed using a multinomial regression; a straightforward extension of the binomial case corresponding to the logistic Platt's scaling (J. Platt and others 1999). The R package glmnet was used to fit a generalized linear model with an elastic net penalty and with a 10-fold cross-validation.

4.4. Assess Performance in Predictive Modeling

Assessing predictive modeling performance corresponds to evaluating the stability of the generalization process with a focus on providing a measure of uncertainty. In probabilistic predictive modeling, such as weather forecasting, prediction skill measures the difference between a predicted value and a reference forecast (Gneiting and Raftery 2007). If a reference forecast is unavailable, as in the current study, one strategy is to estimate the entropy H from predicted posterior probabilities which prognosticates the prediction skill of a model (Stephenson and Dolas-Reyes 2000; Roulston and Smith 2002; Daley and Vere-Jones 2004; Nearing and Gupta 2015). In addition, the entropy rate $H(\chi)$ defines (to the limit) the variability of entropy in a Markovian sequence of prediction and the predictability of the sequence.

In this study, the predictive performance of the selected most accurate ML models was quantitatively assessed using regional entropy, entropy rate and stream-interval entropy. First, the regional entropy is the entropy of the distribution of the most probable channel types across the entire area of study. This regional entropy of predictions is to be compared to the regional entropy of training labels that is the distribution of the channel types observed from field data. These field data have been collected following a stratified random sampling. In consequence, a ML model with good ability to generalize will yield predictions with a regional entropy close to the one derived from the training set. Second, at the network scale, the stability of the predictions was derived by computing the entropy rate from the transition probabilities between each channel types. Third, stream-interval entropy was derived from the posterior probabilities of the model with the highest median AUC, the lowest tuning entropy and the lowest entropy rate. Such a stream-interval entropy represents the stability of the predictions at the stream-interval scale and is a measure of prediction uncertainty.

The map of predicted channel types was investigated using expert-knowledge, with a focus on the general spatial organization of channel types across study area as well as their geomorphic relevance. Aerial imagery was used to qualitatively confirm predictions at selected sample locations. Finally, the variable importance of predictors was investigated to support geomorphic interpretability of model predictions.

5. Stream Network Prediction Results

5.1. Performance in Statistical Learning

5.1.1. Performance Metrics

In all regions, the performance metrics in statistical learning of all models are high (Figure 33, Figure 34). The median cross-validated multiclass AUC is above 0. 90 and the median cross-validated accuracy is above 70%. In terms of AUC, the best performing region is the Klamath region while the worst performing is the South Coast. Across all regions, the distributions of AUC and accuracy indicate that ANN is under-performing in comparison to RF and SVM. Combined with the complexity of the ANN model, this leads us to only consider RF and SVM in the remaining results.

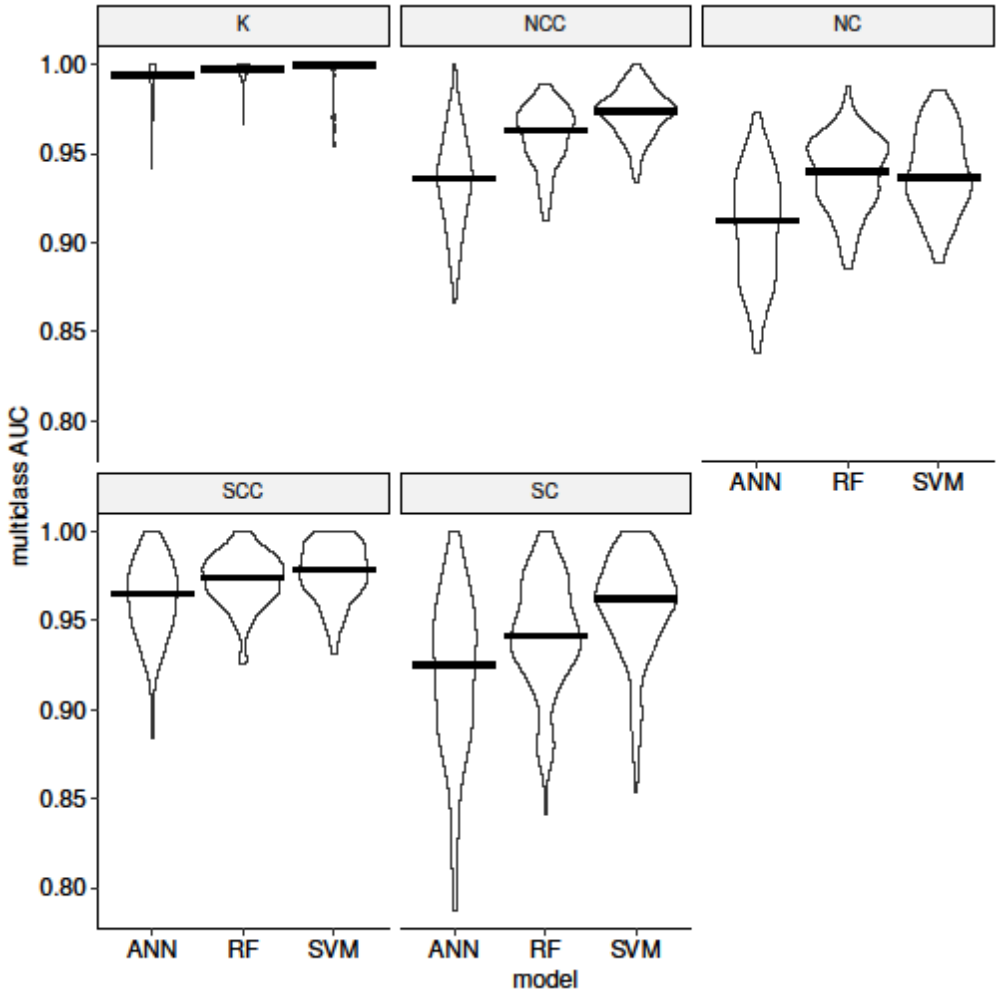


Figure 33: Distribution of cross-validated multiclass AUC for the ANN, RF and SVM models across the five regions of study.

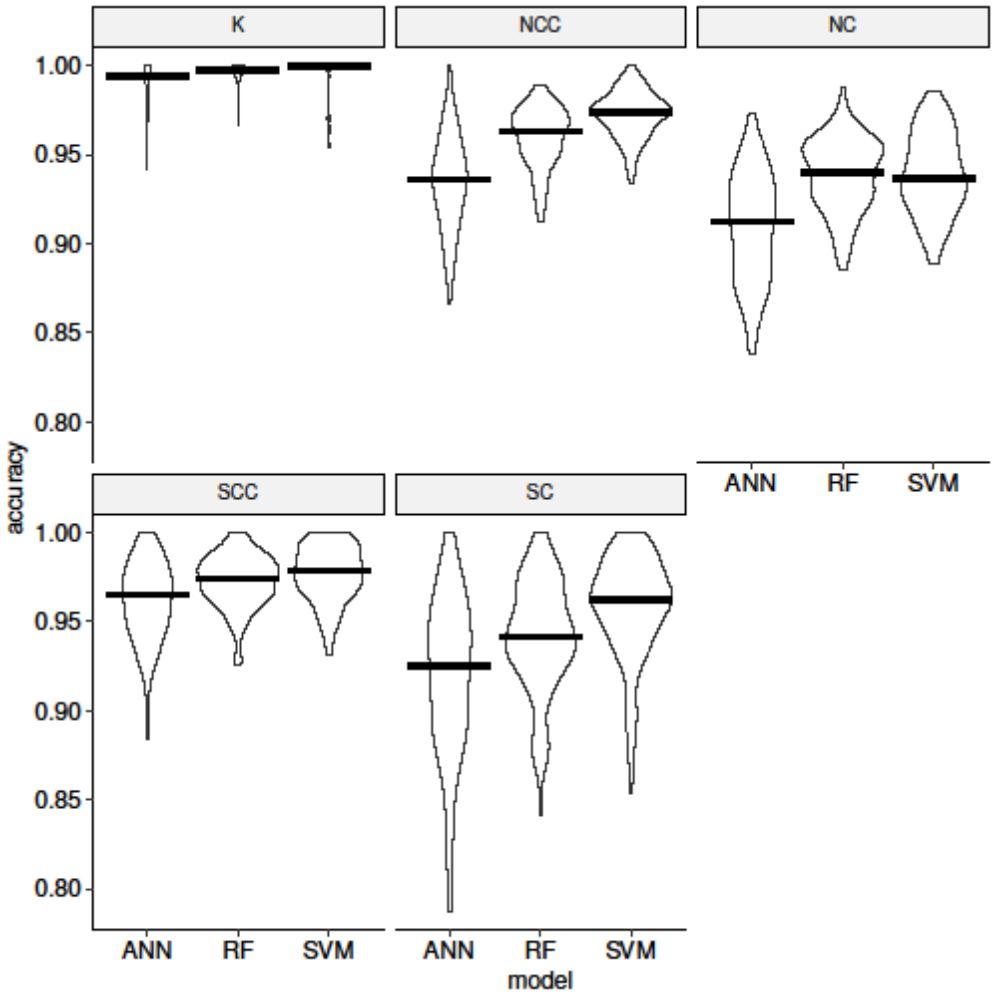


Figure 34: Distribution of cross-validated accuracy AUC for the ANN, RF and SVM models across the five regions of study.

5.1.2. Nested Resampling Results

Like regular resampling, nested resampling diagnoses the ability to generalize while also providing an evaluation of the stability of the tuning process. This stability is represented by a distribution of the hyper-parameters selected as best performing. Across all regions, the learning process of the RF model is more consistent across the training data set with distributions more defined than the ones corresponding to the SVM model (Figure 35). This visual inspection is supported by the entropy values associated with the hyper-parameter values distribution (Figure 36): RF scores consistently lower than SVM. However, RF stability is higher in Klamath, North Coast and North Central Coast than in South Central Coast and South Coast regions.

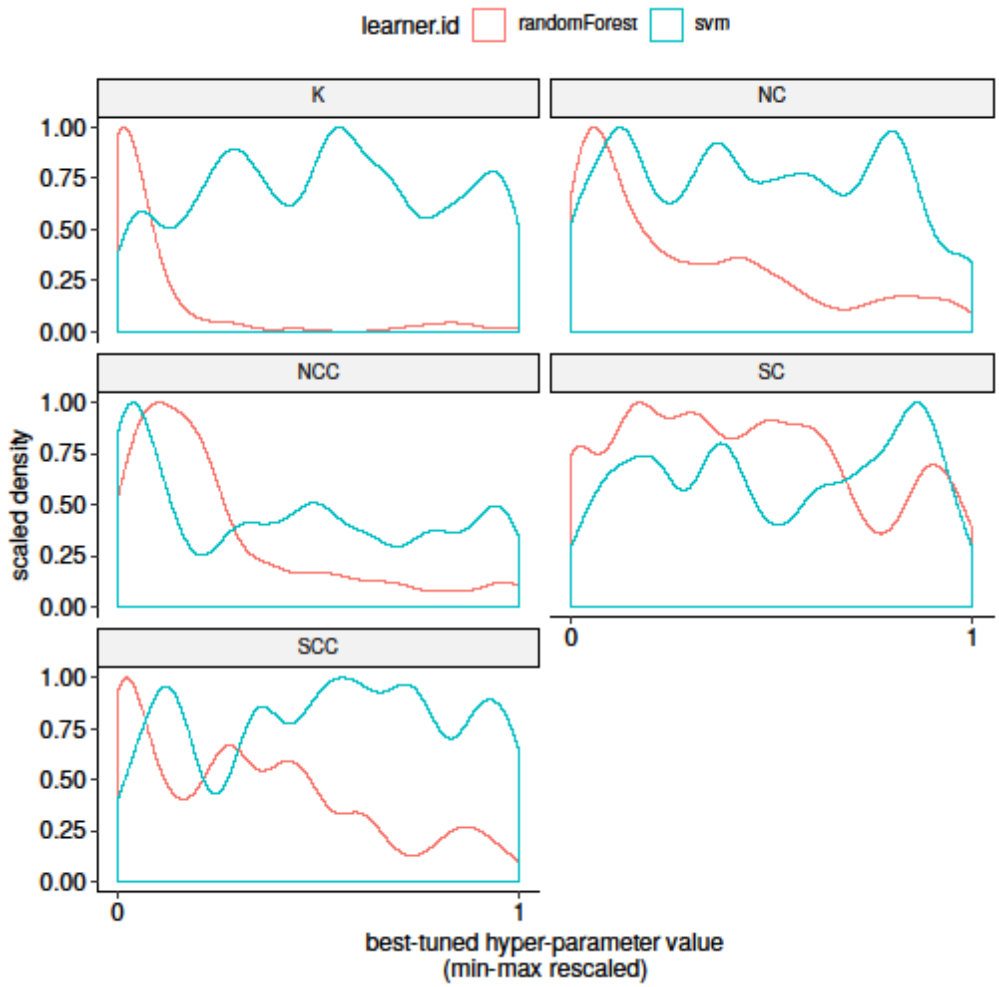


Figure 35: Distributions of best-tuned hyper-parameters resulting from nested resampling.

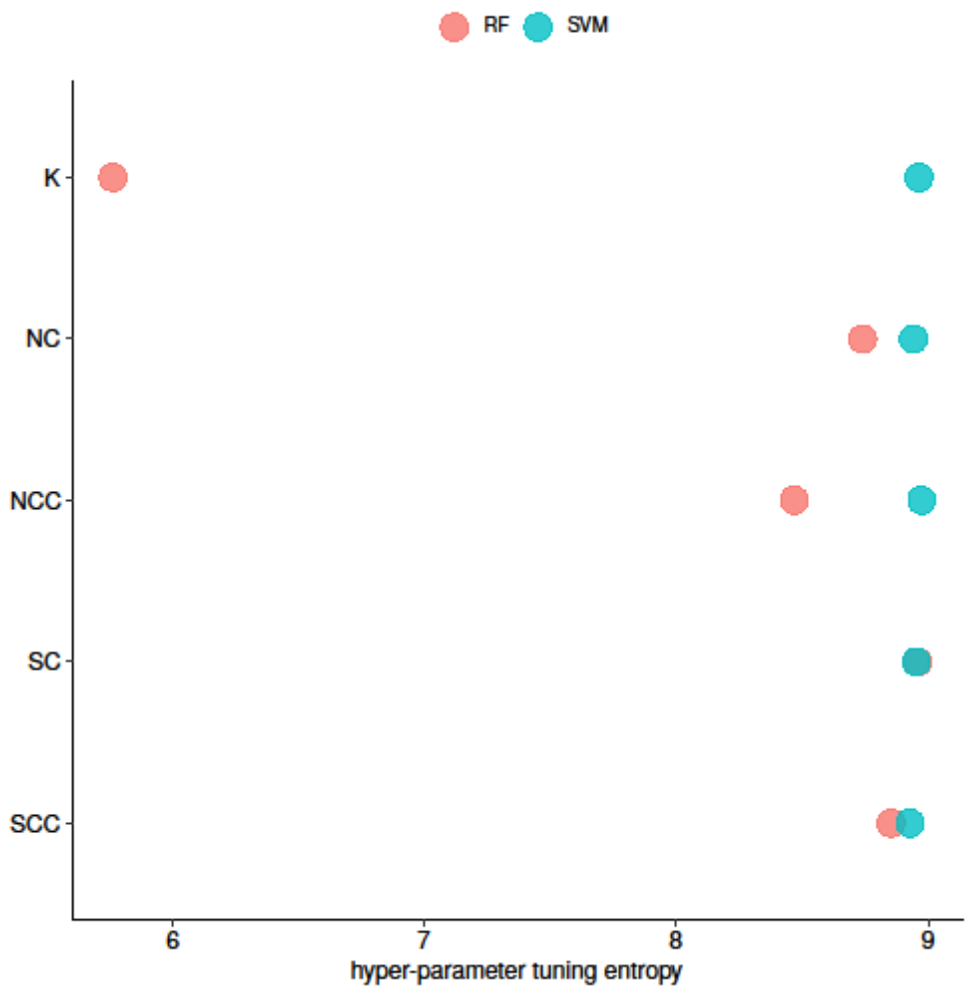


Figure 36: Tuning entropy corresponding to the distributions of best-tuned hyper-parameters (Figure 35).

5.2. Performance in Predictive Modeling

Performance in predictive modeling is assessed from the ML model predictions with three metrics: regional entropy, entropy rate and stream-interval entropy. Regional entropy values fall the closest to the regional entropy for training for RF, underlining that the distribution of RF channel type predictions are closer than SVM predictions distributions to the distribution of the training set (Figure 37). The gap between regional entropy for training and for SVM diagnoses a poor generalization: SVM is not consistently predicting some channel types over the entire region.

While entropy rate values are higher for RF than for SVM, this is linked to SVM's poor generalization (Figure 37). SVM is not consistently predicting all channel types, with some channel types entirely missing from its predictions. This results in a lower number of channel types to transition from one stream interval to another and mechanically lower entropy rate. Such finding highlights that entropy rate should be considered in combination with other metrics for model selection. RF entropy rate is the lower in the Klamath region where statistical learning was the best performing and the most stable.

In all regions, channel type predictions respect the general expected geomorphic organization of the landscape (Figure 38-42a). For example, confined high gradient channel types are dominantly predicted in mountainous areas where they are expected to occur (dark red and pink colors; Figure 38-42a). Similarly, unconfined channel types are predicted in valley areas (teal and blue colors; Figure 38-42a). Two caveats are rule-based post-hoc heuristic refinements performed in the Klamath and South Central Coast regions. In the Klamath region, using stream order achieved better discrimination for the *unconfined, high order, gravel-cobble riffle-pool* (K-3). In the South Central Coastal, using valley confinement improved separation of the three unconfined classes. These two cases illustrate the limit of using coarse-scale geospatial databases to predict reach-scale channel types, which might only be discriminated with near-channel or in-channel predictors (e.g. width, variability). Investigations are ongoing using topographic lidar datasets in order to better channel types predictions. Finally, the associated stream-interval entropy provides a measure of the uncertainty associated with the predictions and should guide future investigations and applications (Figure 38-42b). For example, high entropy values (purple) indicate a lower degree of confidence in a channel type prediction.

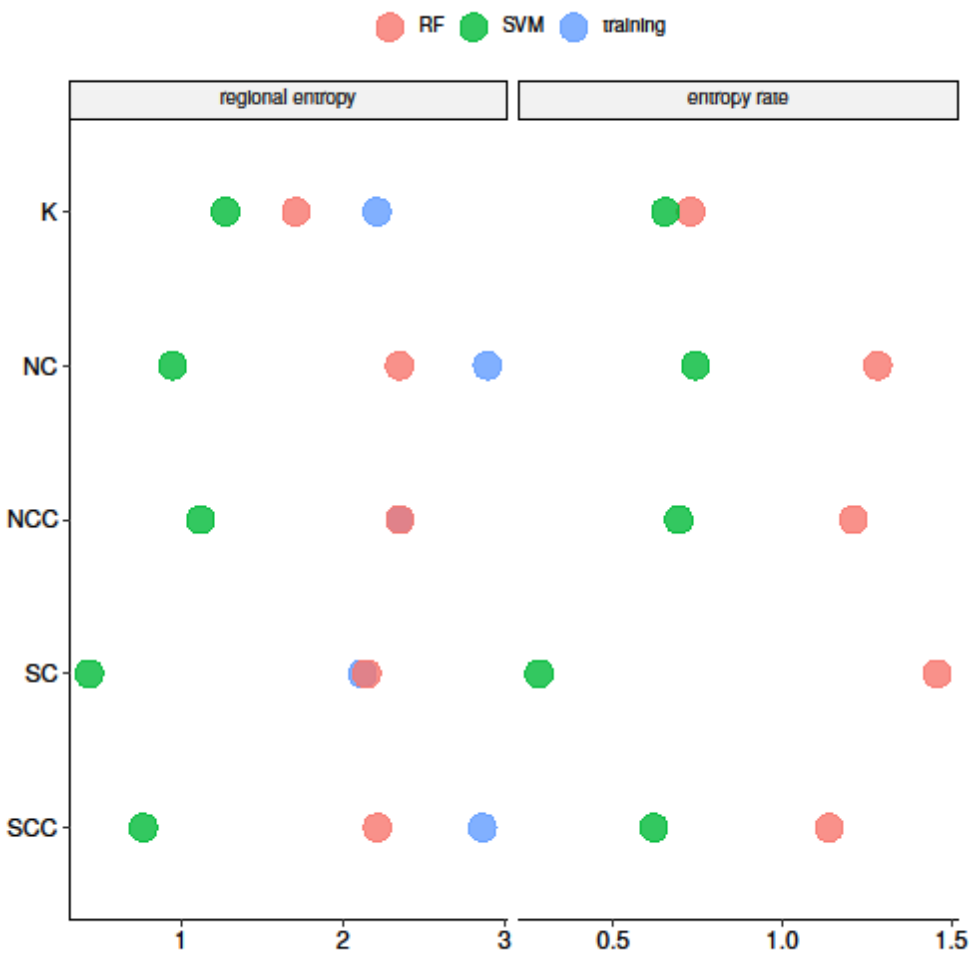


Figure 37: Regional entropy and entropy rate across all regions.

5.3. Channel Type Predictions

5.3.1. Klamath

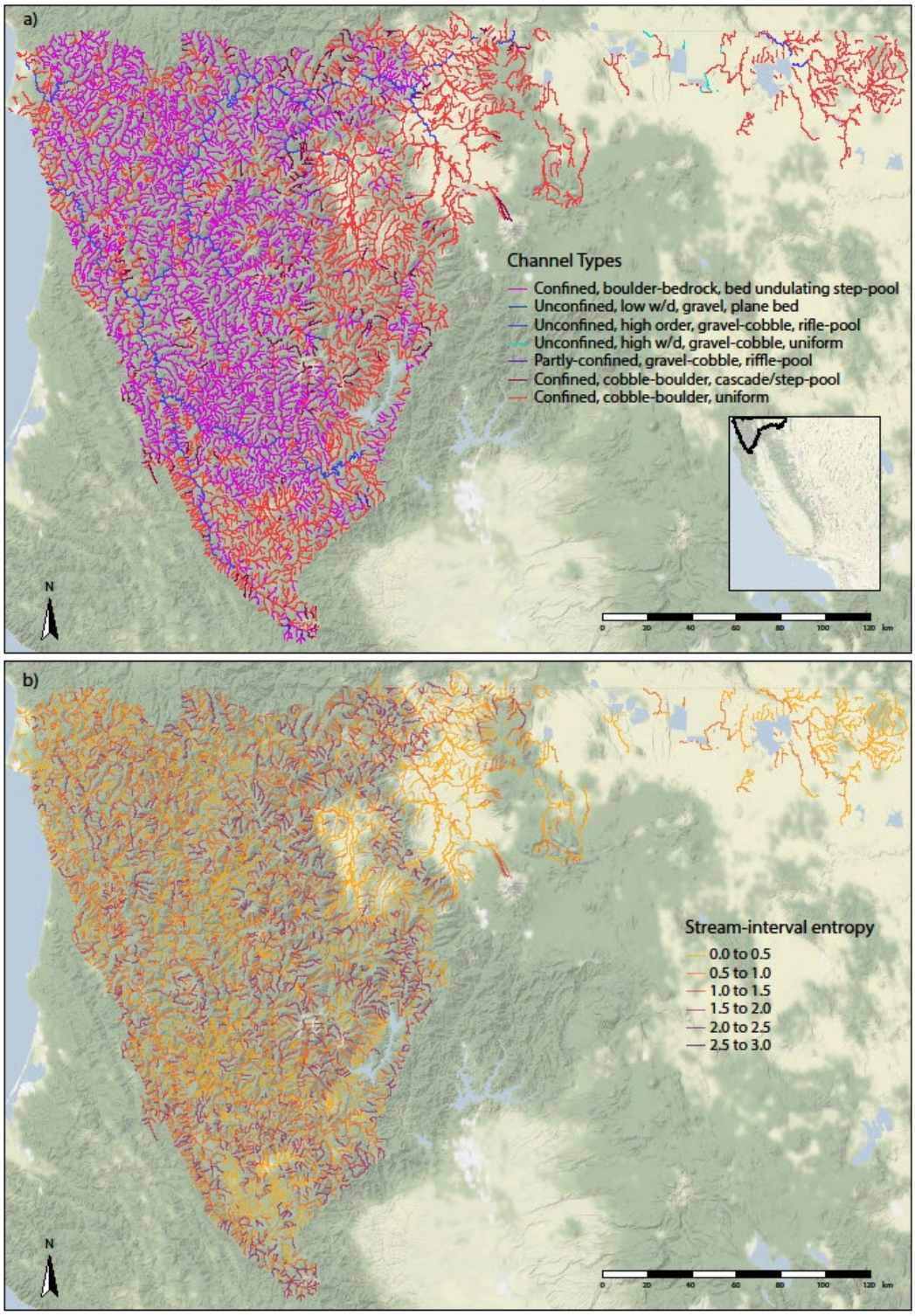


Figure 38: Results for the Klamath region: a) map of the spatial predictions of channel types; inset shows general location in California; b) stream-interval posterior entropy.

5.3.2. North Coast

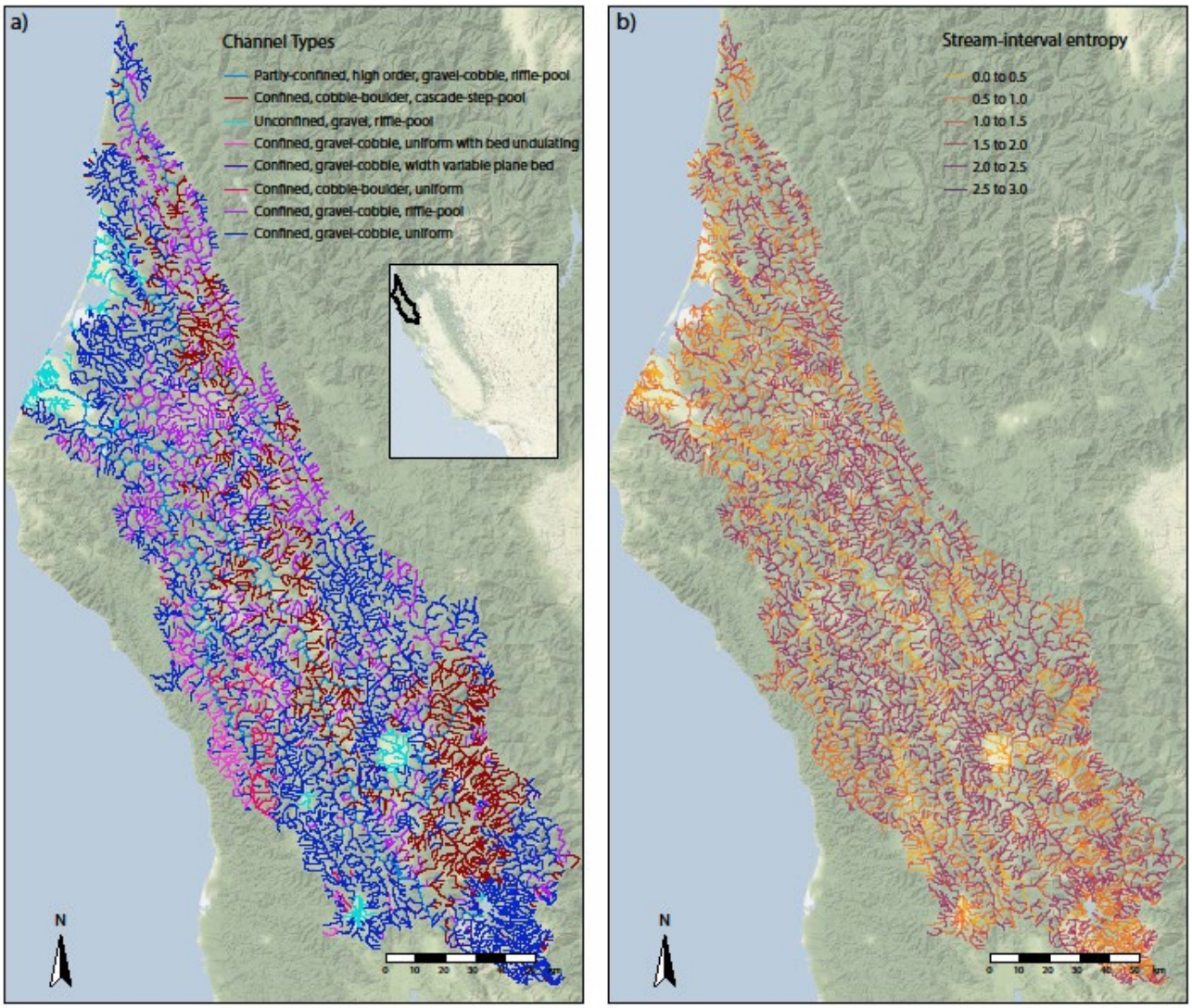


Figure 39: Results for the North Coast region: a) map of the spatial predictions of channel types; inset shows general location in California; b) stream-interval posterior entropy.

5.3.3. North Central Coast

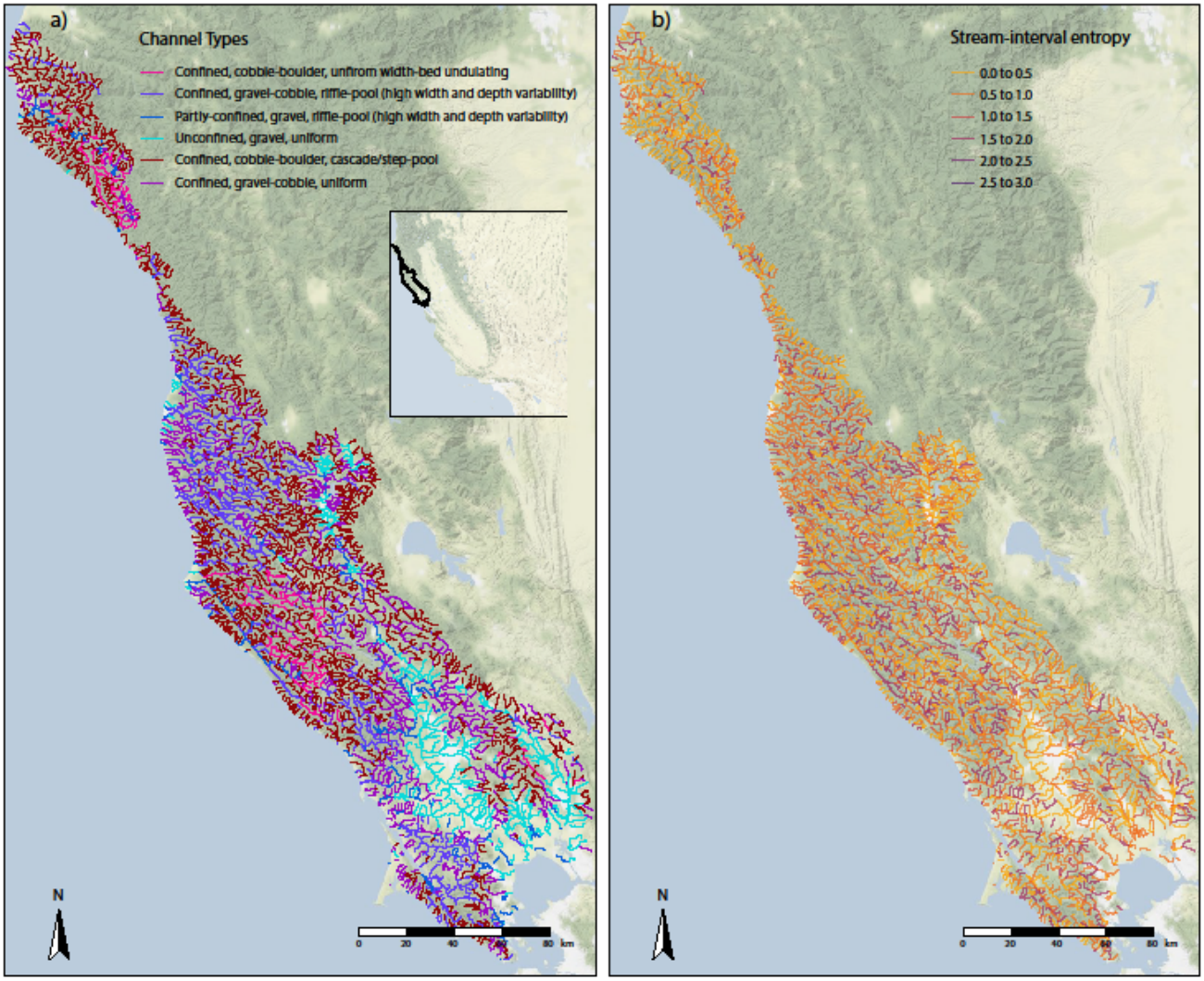


Figure 40: Results for the North Central Coast region: a) map of the spatial predictions of channel types; inset shows general location in California; b) stream-interval posterior entropy.

5.3.4. South Central Coast

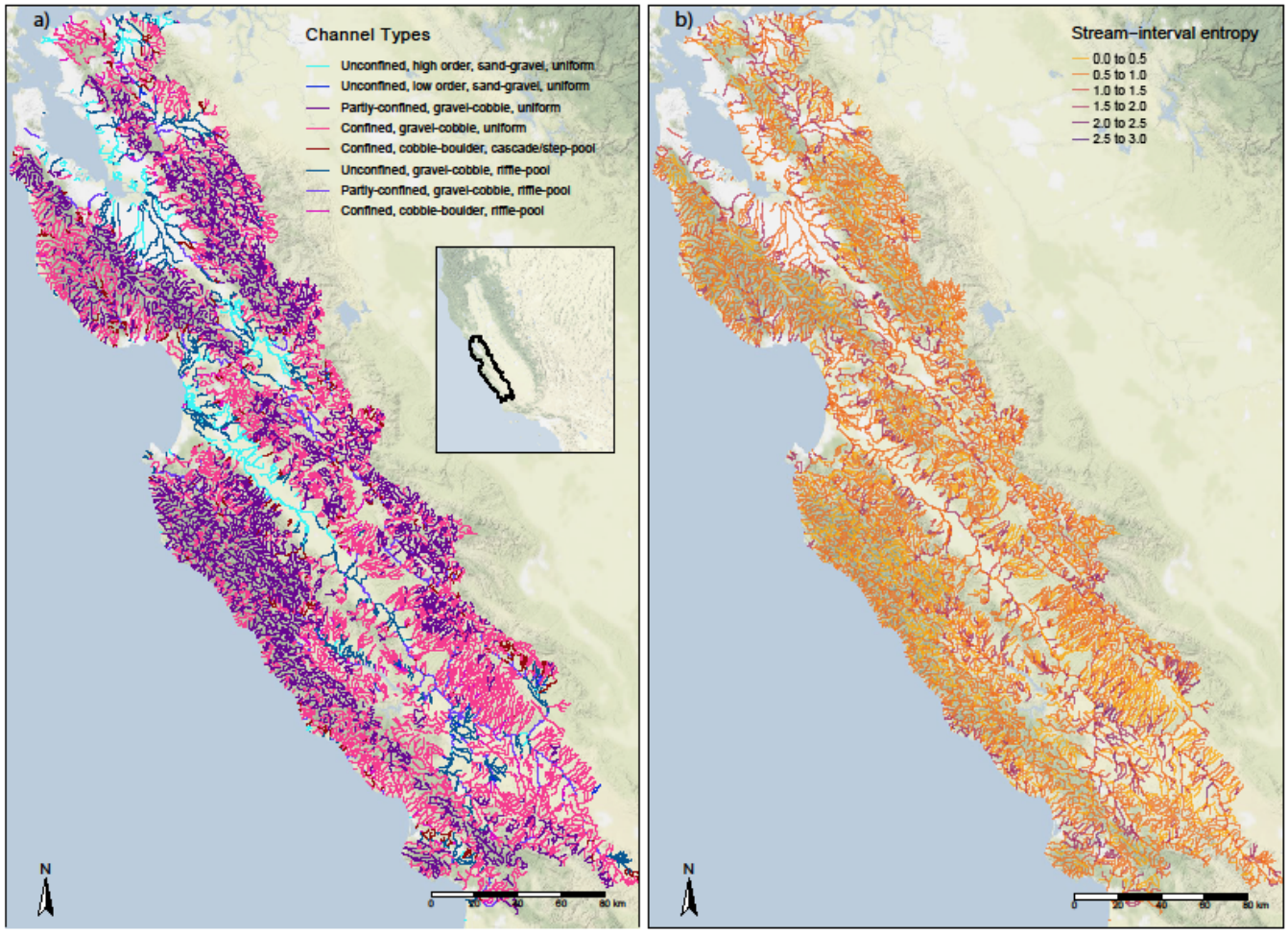


Figure 41: Results for the South Central Coast region: a) map of the spatial predictions of channel types; inset shows general location in California; b) stream-interval posterior entropy.

5.3.5. South Coast

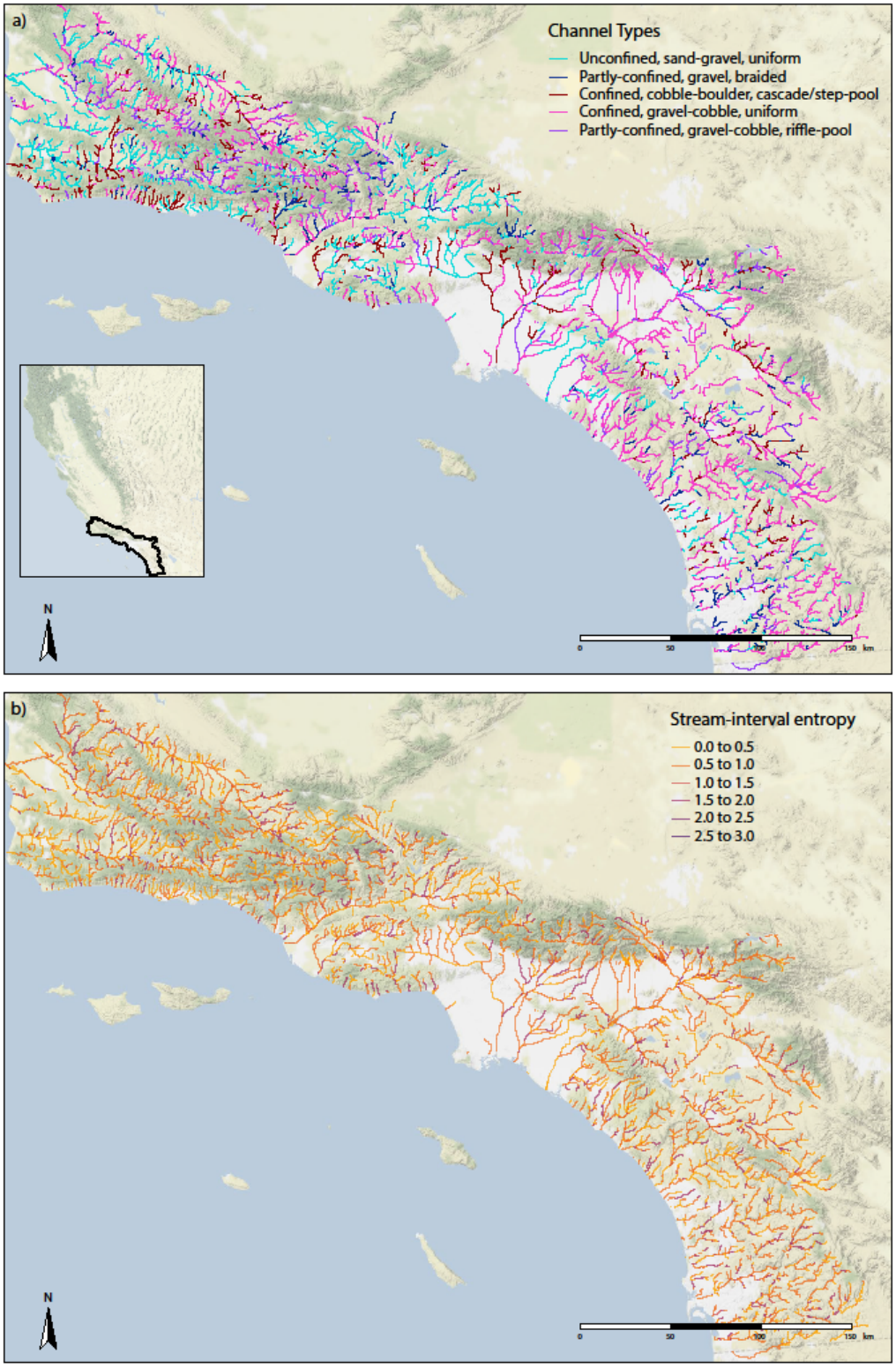


Figure 42: Results for the South Coast region: a) map of the spatial predictions of channel types; inset shows general location in California; b) stream-interval posterior entropy.

6. References

Adnan, Md Nasim, and Md Zahidul Islam. 2015. "One-Vs-All Binarization Technique in the Context of Random Forest." In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 385–90.

Anderson, Marti J. 2001. "A New Method for Non-Parametric Multivariate Analysis of Variance." *Austral Ecology* 26 (1): 32–46. doi:10.1111/j.1442-9993.2001.01070.pp.x.

Bergen, Karianne J, Paul A Johnson, V Maarten, and Gregory C Beroza. 2019. "Machine Learning for Data-Driven Discovery in Solid Earth Geoscience." *Science* 363 (6433). American Association for the Advancement of Science: eaau0323.

Bischl, B., O. Mersmann, H. Trautmann, and C. Weihs. 2012. "Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation." *Evolutionary Computation* 20 (2). MIT Press - Journals: 249–75. doi:10.1162/evco_a_00069.

Bischl, Bernd, Michel Lang, Lars Kotthoff, Julia Schiffner, Jakob Richter, Erich Studerus, Giuseppe Casalicchio, and Zachary M. Jones. 2016. "mlr: Machine Learning in R." *Journal of Machine Learning Research* 17 (170): 1–5. http://jmlr.org/papers/v17/15-066.html.

Bonetti, S., and A. Porporato. 2017. "On the Dynamic Smoothing of Mountains." *Geophysical Research Letters* 44 (11). Wiley-Blackwell: 5531–9. doi:10.1002/2017gl073095.

Breiman, Leo. 1984. *Classification and Regression Trees*. Chapman; Hall/CRC.

Burman, Prabir. 1989. "A Comparative Study of Ordinary Cross-Validation, V-Fold Cross-Validation and the Repeated Learning-Testing Methods<." *Biometrika* 76 (3). JSTOR: 503. doi:10.2307/2336116.

Carson, Michael A. 1972. *Hillslope Form and Process*. Cambridge Geographical Studies ; 3. Cambridge: University Press.

Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. 2012. "NbClust Package: Finding the Relevant Number of Clusters in a Dataset." *J. Stat. Softw*.

Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. "SMOTE: Synthetic Minority over-Sampling Technique." *Journal of Artificial Intelligence Research* 16: 321–57.

Clarke, K. R. 1993. "Non-Parametric Multivariate Analyses of Changes in Community Structure." *Australian Journal of Ecology* 18 (1): 117–43. doi:10.1111/j.1442-9993.1993.tb00438.x.

Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20 (3). Springer Nature: 273–97. doi:10.1007/bf00994018.

Daley, Daryl J, and David Vere-Jones. 2004. "Scoring Probability Forecasts for Point Processes: The Entropy Score and Information Gain." *Journal of Applied Probability* 41 (A). Cambridge University Press: 297–312.

De'ath, Glenn, and Katharina E. Fabricius. 2000. "Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis." *Ecology* 81 (11): 3178–92. doi:10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.

DeGroot, Morris H, and Stephen E Fienberg. 1983. "The Comparison and Evaluation of Forecasters." *Journal of the Royal Statistical Society: Series D (the Statistician)* 32 (1-2). Wiley Online Library: 12–22.

Dodds, Peter Sheridan, and Daniel H. Rothman. 2000. "Scaling, Universality, and Geomorphology." *Annual Review of Earth and Planetary Sciences* 28 (1). Annual Reviews: 571–610. doi:10.1146/annurev.earth.28.1.571.

Dong, Lin, Eibe Frank, and Stefan Kramer. 2005. "Ensembles of Balanced Nested Dichotomies for Multi-Class Problems." In *European Conference on Principles of Data Mining and Knowledge Discovery*, 84–95. Springer.

Duclut, C., and B. Delamotte. 2017. "Nonuniversality in the Erosion of Tilted Landscapes." *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics* 96 (012149).

Faghih, Ali, and Ahmad Nourbakhsh. 2015. "Implication of Surface Fractal Analysis to Evaluate the Relative Sensitivity of Topography to Active Tectonics, Zagros Mountains, Iran." *Journal of Mountain Science* 12 (1). Springer Nature: 177–85. doi:10.1007/s11629-014-3005-5.

Ferri, César, José Hernández-Orallo, and R Modroiu. 2009. "An Experimental Comparison of Performance Measures for Classification." *Pattern Recognition Letters* 30 (1). Elsevier: 27–38.

Fox, Eric W., Ryan A. Hill, Scott G. Leibowitz, Anthony R. Olsen, Darren J. Thornbrugh, and Marc H. Weber. 2017. "Assessing the Accuracy and Stability of Variable Selection Methods for Random Forest Modeling in Ecology." *Environmental Monitoring and Assessment* 189 (7). Springer Nature. doi:10.1007/s10661-017-6025-0.

Gesch, Dean, Michael Oimoen, Susan Greenlee, Charles Nelson, Michael Steuck, and Dean Tyler. 2002. "The National Elevation Dataset." *Photogrammetric Engineering and Remote Sensing* 68 (1). ASPRS AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND: 5–32.

Gneiting, Tilmann, and Adrian E Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477). Taylor & Francis: 359–78.

Hill, Ryan A., Marc H. Weber, Scott G. Leibowitz, Anthony R. Olsen, and Darren J. Thornbrugh. 2015. "The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States." *JAWRA Journal of the American Water Resources Association* 52 (1). Wiley-Blackwell: 120–28. doi:10.1111/1752-1688.12372.

Huang, Jin, and Charles X Ling. 2005. "Using Auc and Accuracy in Evaluating Learning Algorithms." *IEEE Transactions on Knowledge and Data Engineering* 17 (3). IEEE: 299–310.

Hurst, Martin D., Simon M. Mudd, Rachel Walcott, Mikael Attal, and Kyungsoo Yoo. 2012. "Using Hilltop Curvature to Derive the Spatial Distribution of Erosion Rates." *Journal of*

Geophysical Research: Earth Surface 117 (F2). Wiley-Blackwell: n/a–n/a. doi:10.1029/2011jf002057.

Kassambara, Alboukadel. 2019. "Rstatix: Pipe-Friendly Framework for Basic Statistical Tests." https://CRAN.R-project.org/package=rstatix.

Kijsirikul, B., and N. Ussivakul. 2002. "Multiclass Support Vector Machines Using Adaptive Directed Acyclic Graph." In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN02 (Cat. No.02ch37290)*. IEEE. doi:10.1109/ijcnn.2002.1005608.

Kruskal, J. B. 1964. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika* 29 (1): 1–27. doi:10.1007/BF02289565.

Lane, B. A., G. Pasternack, E. Dahlke Helen, and S. Sandoval-Solis. 2017. "The Role of Topographic Variability in River Channel Classification." *Progress in Physical Geography*.

Lifton, Nathaniel A., and Clement G. Chase. 1992. "Tectonic, Climatic and Lithologic Influences on Landscape Fractal Dimension and Hypsometry: Implications for Landscape Evolution in the San Gabriel Mountains, California." *Geomorphology* 5 (1): 77–114. doi:https://doi.org/10.1016/0169-555X(92)90059-W.

Liucci, Luisa, and Laura Melelli. 2017. "The Fractal Properties of Topography as Controlled by the Interactions of Tectonic, Lithological, and Geomorphological Processes." *Earth Surface Processes and Landforms*, August. Wiley-Blackwell. doi:10.1002/esp.4206.

Lorena, Ana Carolina, and André C.P.L.F. de Carvalho. 2010. "Building Binary-Tree-Based Multiclass Classifiers Using Separability Measures." *Neurocomputing* 73 (16-18). Elsevier BV: 2837–45. doi:10.1016/j.neucom.2010.03.027.

Luengo, Julián, and Francisco Herrera. 2013. "An Automatic Extraction Method of the Domains of Competence for Learning Classifiers Using Data Complexity Measures." *Knowledge and Information Systems* 42 (1). Springer Nature: 147–80. doi:10.1007/s10115-013-0700-4.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. and Rea. 2012. *NHDPlus Version 2: User Guide*. United States Environmental Protection Agency (EPA).

Melnikov, Vitalik, and Eyke Hüllermeier. 2018. "On the Effectiveness of Heuristics for Learning Nested Dichotomies: An Empirical Analysis." *Machine Learning*, June. Springer Nature. doi:10.1007/s10994-018-5733-1.

Murtagh, Fionn, and Pierre Legendre. 2014. "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?" *Journal of Classification* 31 (3): 274–95. doi:10.1007/s00357-014-9161-z.

Nearing, Grey S., and Hoshin V. Gupta. 2015. "The Quantity and Quality of Information in Hydrologic Models." *Water Resources Research* 51 (1). American Geophysical Union (AGU): 524–38. doi:10.1002/2014wr015895.

Neeson, Thomas M., Ann Marie Gorman, Peter J. Whiting, and Joseph F. Koonce. 2008. "Factors Affecting Accuracy of Stream Channel Slope Estimates Derived from Geographical

Information Systems." *North American Journal of Fisheries Management* 28 (3): 722–32. doi:10.1577/M05-127.1.

Niculescu-Mizil, Alexandru, and Rich Caruana. 2005. "Predicting Good Probabilities with Supervised Learning." In *Proceedings of the 22nd International Conference on Machine Learning - ICML 05*. ACM Press. doi:10.1145/1102351.1102430.

Ode, P. R. 2007. "Standard Operating Procedures for Collecting Benthic Macroinvertebrate Samples and Associated Physical and Chemical Data for Ambient Bioassessments in California." *California State Water Resources Control Board. Surface Water Ambient Monitoring Program (SWAMP) Bioassessment SOP* 1.

Oksanen, Jari, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. 2019. "Vegan: Community Ecology Package." https://CRAN.R-project.org/package=vegan.

Pastor-Satorras, Romualdo, and Daniel H. Rothman. 1998. "Scaling of a Slope: The Erosion of Tilted Landscapes." *Journal of Statistical Physics* 93 (3/4). Springer Nature: 477–500. doi:10.1023/b:joss.0000033160.59155.c6.

Platt, John C, Nello Cristianini, and John Shawe-Taylor. 2000. "Large Margin Dags for Multiclass Classification." In *Advances in Neural Information Processing Systems*, 547–53.

Platt, John, and others. 1999. "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods." *Advances in Large Margin Classifiers* 10 (3). Cambridge, MA: 61–74.

Prancevic, Jeff P., and James W. Kirchner. 2019. "Topographic Controls on the Extension and Retraction of Flowing Streams." *Geophysical Research Letters* 46 (4). American Geophysical Union (AGU): 2084–92. doi:10.1029/2018gl081799.

Quiterio, Thaise M, and Ana C Lorena. 2016. "Determining the Structure of Decision Directed Acyclic Graphs for Multiclass Classification Problems." In *Intelligent Systems (Bracis), 2016 5th Brazilian Conference on*, 115–20. IEEE.

Reichstein, Markus, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566 (7743). Springer Nature: 195–204. doi:10.1038/s41586-019-0912-1.

Rosgen, David L. 1994. "A Classification of Natural Rivers." *CATENA* 22 (3): 169–99. doi:https://doi.org/10.1016/0341-8162(94)90001-9.

Rosset, Saharon. 2004. "Model Selection via the Auc." In *Proceedings of the Twenty-First International Conference on Machine Learning*, 89. ACM.

Roulston, Mark S, and Leonard A Smith. 2002. "Evaluating Probabilistic Forecasts Using Information Theory." *Monthly Weather Review* 130 (6): 1653–60.

Schratz, Patrick, Jannes Muenchow, Jakob Richter, and Alexander Brenning. 2018. "Performance Evaluation and Hyperparameter Tuning of Statistical and Machine-Learning Models Using Spatial Data." *arXiv Preprint arXiv:1803.11266*.

Shannon, Claude Elwood. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (3). Wiley Online Library: 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Shen, Chaopeng. 2018. "A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists." *Water Resources Research*, November. American Geophysical Union (AGU). doi:10.1029/2018wr022643.

Stephenson, David B, and Francisco J Dolas-Reyes. 2000. "Statistical Methods for Interpreting Monte Carlo Ensemble Forecasts." *Tellus A: Dynamic Meteorology and Oceanography* 52 (3). Taylor & Francis: 300–322.

Sung, Quo-Cheng, and Yen-Chieh Chen. 2004. "Self-Affinity Dimensions of Topography and Its Implications in Morphotectonics: An Example from Taiwan." *Geomorphology* 62 (3-4). Elsevier BV: 181–98. doi:10.1016/j.geomorph.2004.02.012.

Team, R Core. 2017. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Therneau, Terry M, and Elizabeth J Atkinson. 2018. "Rpart: Recursive Partitioning and Regression Trees." Mayo Foundation. https://CRAN.R-project.org/package=rpart.

Ward, Joe H. Jr. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58 (301): 236–44. doi:10.1080/01621459.1963.10500845.

Wilson, Thomas H., and Jovita Dominic. 1998. "Fractal Interrelationships Between Topography and Structure." *Earth Surface Processes and Landforms* 23 (6). John Wiley & Sons, Ltd: 509–25. doi:10.1002/(SICI)1096-9837(199806)23:6<509::AID-ESP864>3.0.CO;2-D.

Wolman, M. Gordon. 1954. "A Method of Sampling Coarse River-bed Material." *Eos, Transactions American Geophysical Union* 35 (6): 951–56. doi:10.1029/TR035i006p00951.

Zadrozny, Bianca. 2002. "Reducing Multiclass to Binary by Coupling Probability Estimates." In *Advances in Neural Information Processing Systems*, 1041–8.

Zadrozny, Bianca, and Charles Elkan. 2002. "Transforming Classifier Scores into Accurate Multiclass Probability Estimates." In *Proceedings of the Eighth Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 694–99. ACM.