Sacramento River Basin Geomorphic Classification

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

January, 2019

Executive Summary

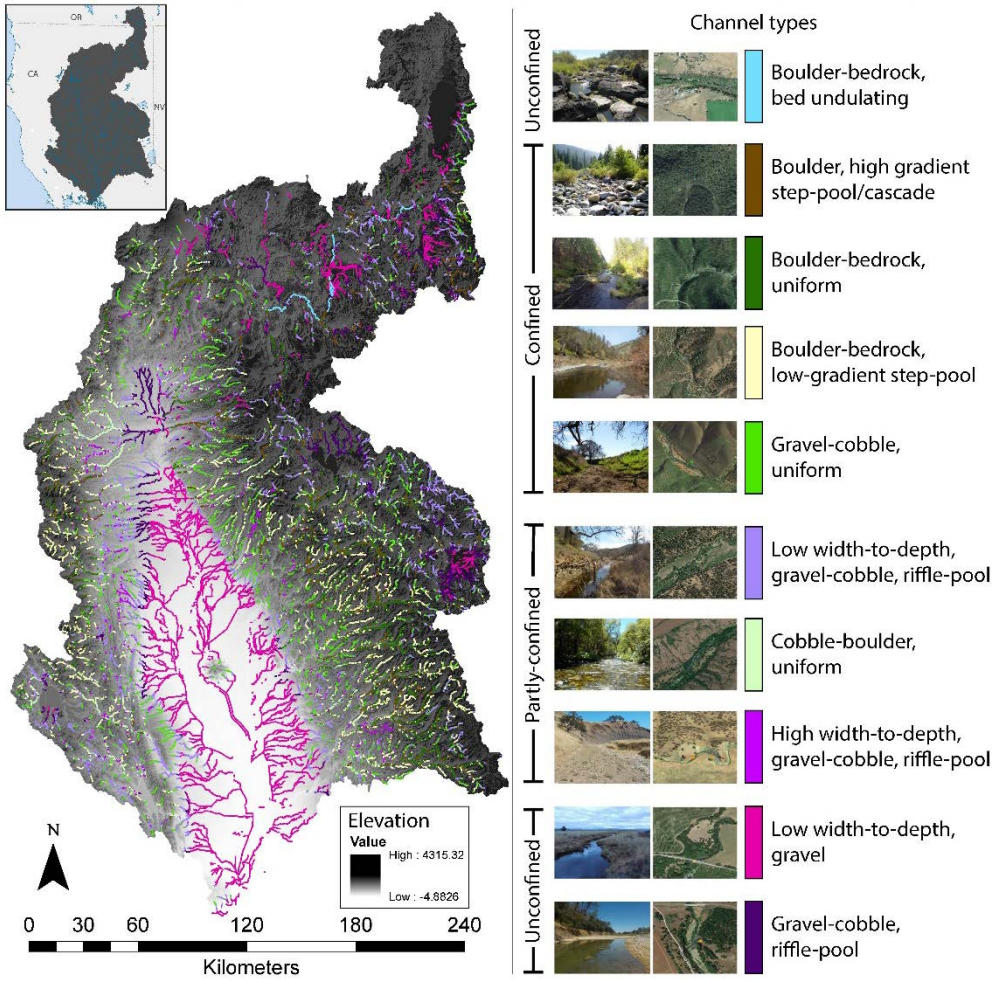
Colin F. Byrne, Hervé Guillon, Belize A. Lane, Gregory B. Pasternack and Samuel Sandoval-Solis

The objective of this research study was to determine a channel reach morphology classification and the spatial distribution of each geomorphic class throughout the Sacramento River Basin. A geomorphic class is defined as an archetypical stream form at the 10 – 20 channel width scale that has well-defined channel attributes (e.g. slope, bankfull width, etc.), topographic variability attributes (TVA) (e.g. coefficients of variation of width and depth), sediment composition (e.g. D50, D84, etc.) and landscape location (e.g. valley confined, partly confined or unconfined) that can be verified in the field. The classification was informed by 290 field-surveyed stream sites throughout different physiographic provinces, slope and contribution area bins. Surveyed streams in the Sacramento River basin were analyzed using multivariate statistical techniques to identify groups of reaches with similar stream forms. Two classifications were performed in this study: (1) an independent classification for the subset of sites in four hydrologic classes, and (2) a geomorphic classification for the entire Sacramento Basin.

This study focusses on the basin-wide geomorphic classification for the Sacramento Basin, ten channel types were identified within the Sacramento River basin (Figure ES-1). Sediment size and valley confinement were the most influential channel attributes in the clustering of channel types. In general, the classification progresses from confined mountainous upland to unconfined lowland streams, except for the first class, (1) *boulder-bedrock bed undulating*, located in the Modoc Plateau with specific volcanic terrain and non-fluvial legacy deposits. Four channel types were observed in confined settings, slope and TVAs were critical attributes in splitting these channel types after initial sediment size and confinement splits. Three partly-confined channel types provide key linkages between confined and unconfined river systems. Channel type 6 (*low width-to-depth ratio, gravel-cobble, riffle-pool,*) is the most common type of the survey, while bed undulation remains relatively absent, width variability increases significantly, likely due to the deposition and reworking of alluvial sediments in the partly-confined setting. Two final unconfined channel types were observed in the Sacramento River basin. Chanel type 9 (*low width-to-depth gravel*) occur in upland valleys and is the most common class within the Central Valley of California. These relatively deep, narrow streams display the smallest sediment sizes of all streams, they are laterally stable either due to well developed and heavily vegetated floodplains or anthropogenic land use and may exist as anastomosed streams. The ten channel types show the diversity of river-channels throughout the Sacramento Basin. These channel types will be used to evaluate the response aquatic and riparian ecosystems to different combinations of flow regimes and channel types.

The previous channel type classification was used as the training set to determine a statistical model using machine learning techniques that spatially predict channel types throughout the Sacramento Basin. The 290 field sites already classified are incorporated as labels into a large-scale supervised learning model. This model uses 300 coarse-scale features describing topography, lithology, soils, climate and land use. The spatial significance of the predictions from the best

models was assessed. In particular, the spatial pattern of channel types delineates distinct geomorphic areas with similar erosion processes. Figure ES-1 shows the spatial prediction of the ten channel types classified for the Sacramento Basin using a random forest model. 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.



ES-1 .- Geomorphic classification and spatial distribution of channel types in the Sacramento River Basin

Contents

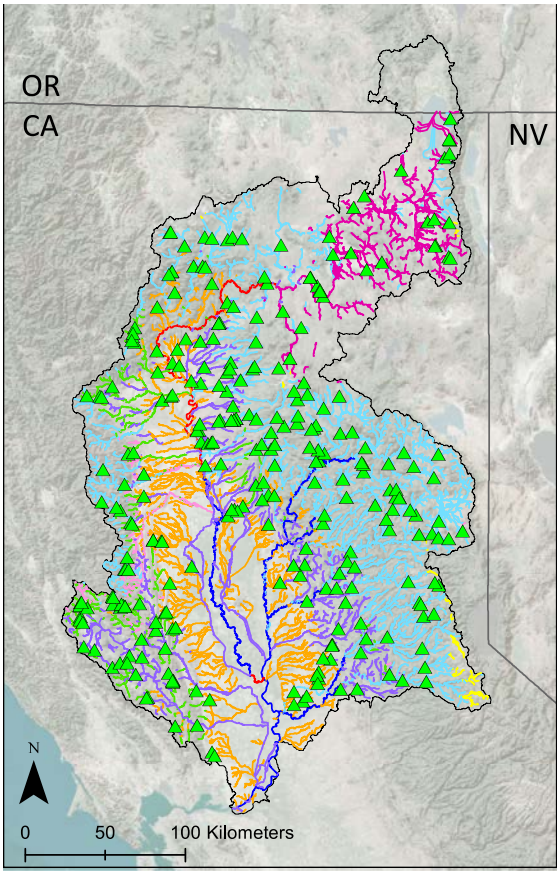
1 Classification of Channel-Reach Morphology

1.1 Objectives

The objective of this study was to organize field-surveyed stream sites into groups of similar channel form to generate a channel-reach morphology classification for California representing archetypical stream forms at the 10 – 20 channel width scale. Surveyed streams in the Sacramento River basin were analyzed using multivariate statistical techniques to identify groups of reaches with similar stream forms. Independent classifications were also performed for the subset of sites in four hydrologic classes within the Sacramento River basin: low-volume snowmelt and rain (LSR), perennial groundwater and rain (PGR), rain and seasonal groundwater (RGW), and winter storms (WS) (Lane et al., 2017a, 2018). The basin-wide geomorphic classification (i.e. independent of annual hydrologic regime) is referred to as the Sacramento classification, while independent geomorphic classifications are referred to as the LSR, PGR, RGW, and WS classifications. The archetypical channel-reach morphologies described in each classification are referred to as channel types.

1.2 Methodology

Channel types were developed based on 12 measured channel attributes: contributing area (Ac), slope (S), bankfull depth (bf.d), bankfull width (bf.w), bankfull width to depth ratio (bf.w.d), coefficient of variation of bankfull width (CV_bf.w), coefficient of variation of bankfull depth (CV_bf.d), sinuosity (sin), median grain size (D50), 84th percentile grain size (D84), channel roughness (bf.d.D50), and valley confinement distance (vc.dist.25pct). The *Sacramento* classification was developed based on 290 stream surveys including 141 sites surveyed by the University of California Davis (UCD) and 149 sites surveyed by the California State Water Resources Board Surface Water Ambient Monitoring Program (SWAMP) (Fig. 1). The *LSR* classification was developed from 147 sites (49 UCD, 98 SWAMP), the *PGR* classification from 54 sites (44 UCD, 10 SWAMP), the *RGW* classification from 51 sites (32 UCD, 19 SWAMP), and the *WS* classification from 38 sites (16 UCD, 22 SWAMP).



igh elevation, low precipitation

Figure 1. Site locations of 290 stream surveys used in the multivariate geomorphic classification. The Sacramento River basin stream network is represented by annual hydrologic regimes as defined by Lane et al. (2018b).

1.2.1 Site Selection

Field site locations for Sacramento basin surveys were selected using a random stratified sampling scheme based on slope and contributing area as documented in Lane et al., (2017b) (Fig 2). Slope bins were defined as <0.1%, 0.1-2%, 2-4%, 4-10%, and >10%. Contributing area bins differed based on physiographic province (i.e. Pacific Border or Cascade-Sierra Mountains). Pacific Border area bins were <50, 50-5,000, and >5,000 km^2, while Cascade-Sierra Mountains sites were <300, 300-9,000, and >9,000 km^2. The slope-area sampling protocol was designed to capture bins of transport capacity along the stream network. In addition, because some bins were likely to contain a substantially larger number of streams than others, bins were surveyed with an equal number of sites, therefore, placing a focus on all channel types, even rare forms, instead of the most dominant channel type.

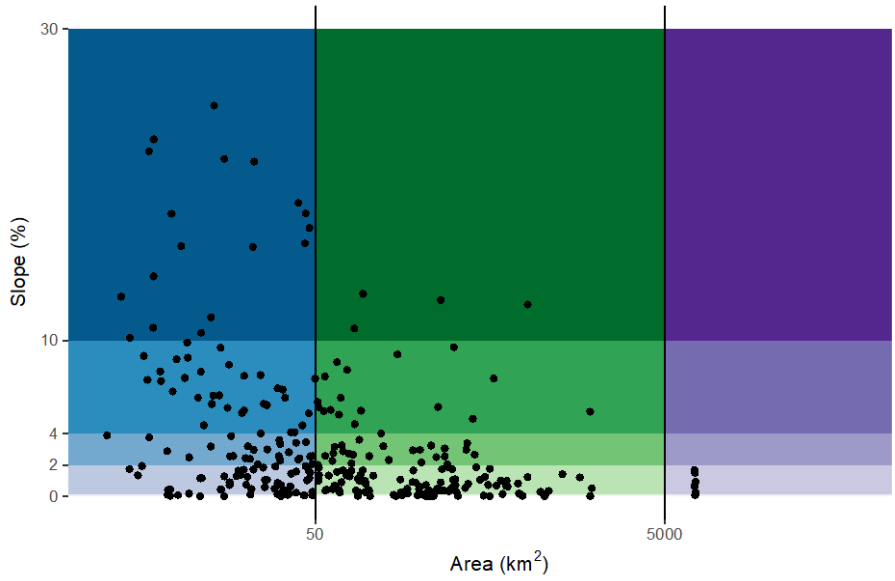


Figure 2. Site locations distributed across area-slope bins for Pacific Border binning protocol.

1.2.2 Data Acquisition and Processing

1.2.2.1 Field Surveying and Post-Processing

Field surveys were completed by UCD teams in the summers of 2015 through 2017. Surveying protocols were based upon SWAMP protocols (Ode, 2007). At each site, average wetted width was estimated to determine the length of channel to survey. Stream lengths of 150 and 250 m were surveyed based on wetted width estimations less than and greater than 10 m, respectively. This produced stream survey lengths of 10 to 100 channel widths. Eleven 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. Bankfull depth and water depth were recorded at the thalweg, Wolman pebble counts were conducted along each cross-section (Wolman, 1954), and a longitudinal survey was conducted along the thalweg at each cross-section.

Field data were processed in preparation for statistical analysis. In addition to the mean values of bankfull width, depth, and bankfull width-to-depth ratio, median and 84th percentile grain sizes of Wolman pebble counts were calculated. Slope was calculated from the best fit regression line of surveyed water surface elevations along longitudinal, thalweg transects. The roughness parameter was calculated as the ratio of bankfull depth to median grain size and 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 and depth are referred to as Topographic Variability Attributes (TVAs). Lane et al., (2017b) previously documented that TVAs displayed considerable importance in the identification of distinct channel types.

1.2.2.2 Geographic Information System 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 and slope were estimated for initial site selection based on sampling scheme protocols described in 2.1.1. 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 DEM 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.

Sinuosity has been used as a defining metric in previous classifications (Rosgen, 1994) and was calculated as the ratio of channel thalweg length to distance between upstream and downstream vertices within the GIS. Stream channels were digitized based upon aerial imagery, digital USGS topographic maps, and NHD layers for 1000 m. Because sinuosity will be greatly dependent on the scale at which it is calculated (Snow, 1989), 1000 m sinuosity represents the length of the channel at approximately 100 times the channel width.

Valley setting and valley confinement play both qualitative and quantitative roles in the majority of previous channel classification methodologies due to the influence of distinct valley setting processes in the creation of characteristic forms (Brierley and Fryirs, 2000; Fryirs et al., 2016; Rosgen, 1994). For example, a channel that abuts a valley wall may be defined by colluvial bed material, while streams in wide valleys would be more likely to be alluvial in nature. Because qualitative description of valley setting would not be applicable to a statistical cluster method of channel classification, a quantitative metric needed to be developed for analysis here. Various methodologies to determine valley confinement already exist, including estimation of valley bottom and channel-bounding margin assessment (Fryirs et al., 2016; Gilbert et al., 2016), the relationship between valley width and channel width (Bisson et al., 1996; Rosgen, 1994, 1996), and hydrologic estimation of bankfull width (Beechie and Imaki, 2014). As the SWAMP protocol does not call for measurement of floodprone width (i.e. the valley width at two times bankfull depth), entrenchment ratio could only be calculated with GIS measurement of floodprone width. This approach was deemed unsatisfactory as it was prone to substantial error in interpretation of floodprone width, especially at sites in more anthropogenically-influenced valley settings.

Instead, valley confinement was calculated as a distance from the channel thalweg to bounding valley wall. This provided a continuous geospatial metric applicable to the large number and variety of sites. The metric is reliant upon definition of a valley bottom slope. For the purposes of this study, 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). The 10 m DEM was converted to a slope raster and reclassified. Valley bottom polygons (i.e. areas with less than 25% slope) were created to clip cross-section polylines to site specific lengths. Four cross-sections per stream length were averaged to calculate a single valley confinement distance.

Confined, partly-confined, and unconfined nomenclature was defined by a logarithmic scale of <= 100 m, >100 and <= 1000 m, and > 1000 m, respectively.

1.2.3 Multivariate Statistical Classification

Statistical techniques used to define multi-dimensional clustering of field-surveyed geomorphic attributes were based upon Lane et al., (2017b). This iterative approach utilized non-metric multidimensional scaling (NMDS) (Anderson, 2001; Clarke, 1993; Kruskal, 1964), hierarchical clustering using Ward's algorithm (Murtagh and Legendre, 2014a, 2014b; Ward, 1963), classification and regression trees (CART) (De'ath and Fabricius, 2000), and Tukey's honestly significant differences to develop clustered groupings for each RGW, WS, and PGR hydrologic classes (Tukey, 1991). The `R` language was used for all statistical analysis (R Core Team, 2017). Initial correlations were conducted by calculating the Pearson's coefficient for zero to one rescaled attribute values using the `cor` function (`stats` package). The NMDS calculations were conducted with the use of the `metaMDS` function (`vegan` package) (Oksanen et al., 2018). Hierarchical clustering using Ward's algorithm utilized the `hclust` function with the method defined as `Ward.D2` (`stats` package) and the `NbClust` function to assess the suggested number of hierarchical clusters (`NbClust` package) (Murtagh and Legendre, 2014a). The classification trees were developed with the `rpart` function and pruned with the prune function (`rpart` package) (Therneau and Atkinson, 2018). Finally, Tukey's honestly significant differences were calculated with the `ghlt` function with inputs specified by an analysis of variance model using the `aov` function (`multcomp` and `stats` packages, respectively) (Hothorn et al., 2008).

Linear regressions between all geomorphic inputs were first conducted to remove highly correlated metrics. Remaining geomorphic attributes were rescaled from zero to one to remove the influence of large magnitude attributes. NMDS scaling in combination with principal component vectors allowed for the comparison between attributes and the plotting of multi-dimensional clusters in two-dimensional space. Hierarchical clustering with Ward's algorithm stratified the data by minimizing within-cluster variance and maximizing between-cluster variance. Ideally, this 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. Heuristic refinement of these groupings was also conducted based upon field reconnaissance and expert knowledge of specific field sites. Because branches within the hierarchical clustering do not necessarily have physical meaning, classification tree analysis was conducted as a method to understand the ability of geomorphic attributes to correctly define geomorphic classification. Pruning of the classification trees was conducted to ensure the number of final tree nodes matched the final number of channel types chosen. Tukey's honestly significant differences approach allowed for the comparison of attributes within each channel type as well. 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.

1.3 Results

1.3.1 Channel Classifications within Sacramento Basin Hydrologic Regimes

Statistical analysis using NMDS identified the most influential attributes in the multidimensional space. For LSR sites, D84, bf.w, and bf.d returned the greatest correlations (r^2 = 0.97, 0.80, and 0.54, respectively) (Fig. 3a). Sediment size and bankfull width were most influential along the primary and secondary axes, respectively. For PGR sites, D84, Ac, and bf.w.d returned the greatest correlations (r^2 = 0.94, 0.76, and 0.59, respectively) (Fig. 3b). Sediment size was most influential along the primary axis while contributing area was most influential along the secondary axis. NMDS of RGW sites indicated that D84, vc.dist, and Ac were the most highly correlated (r^2 = 0.89, 0.72, and 0.68, respectively) (Fig. 3c). Again, sediment size and contributing area were the most influential along the primary and secondary axes, respectively. Data from WS sites resulted in D84, Ac, and bf.w returning the greatest correlations (r^2 = 0.81, 0.71, and 0.66) (Fig. 3d). Similar to both PGR and RGW sites, the primary axis was most correlated with sediment size and the secondary axis with contributing area. Stress levels for LSR, PGR, RGW, and WS sites are displayed in Table 1. These values all correspond to stress levels that should likely lead to a usable ordination (Clarke, 1993). These results show consistency in the most influential attributes of clusters in multidimensional space, at least in terms of accounting for the most variance. However, statistically significant differences of attributes with less correlation to the primary and secondary axes were also critical in deciphering all channel types introduced later.

Table 1. Summary of statistical clustering results for LSR, PGR, RGW, and WS geomorphic lassification.

	NMDS Stress	Hubert Index Suggested Cluster #	Chosen Cluster #	Heuristic Refinement (Y/N)	CART CP-value	CART Percent Classified Correctly	Cross-validation Percent Classified Correctly
LSR	0.163	4	7	Y	0.044	88.3	83.3
PGR	0.154	3	6	Y	0.036	94.4	85.2
RGW	0.138	4	6	Y	0.064	92.2	78.4
WS	0.144	2	5	Y	0.084	97.4	86.8

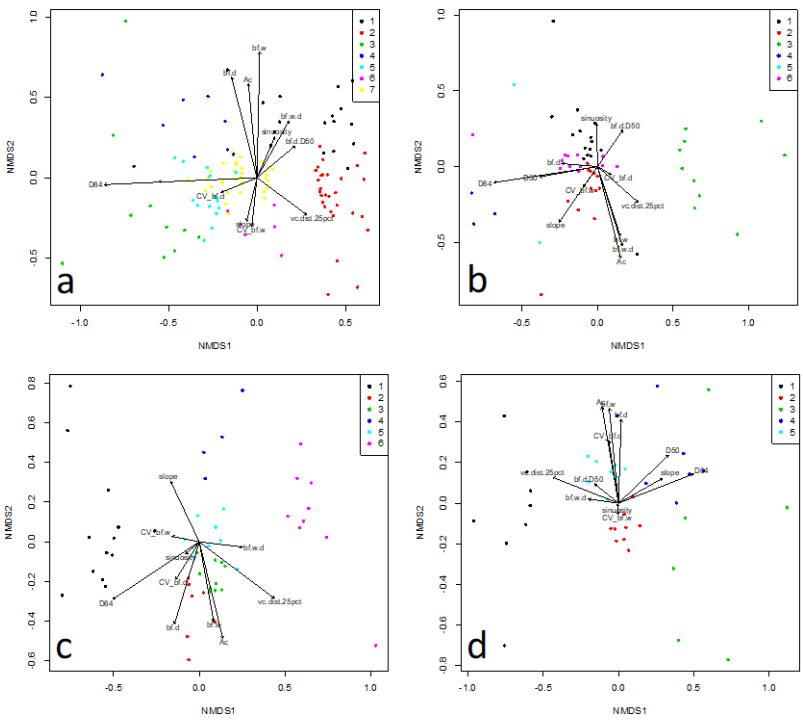


Figure 3. Non-metric multidimensional scaling with primary component vectors associated with rescaled geomorphic attributes of a) LSR, b) PGR, c) RGW, and d) WS sites. Colors represent final channel type classification independent of streams in other hydrologic classes.

Hierarchical clustering of LSR, PGR, RGW, and WS sites was conducted using the iterative methodology described above (Hubert Index and heuristic splitting information detailed in Table 1). LSR sites were statistically clustered into five groups, and two further groups were identified to be important based on high TVA and slope values (Fig. 4a). These sites exhibited clear differences in stream form from the other sites in the original statistical cluster. LSR clusters ranged in size from five to 33 sites. Hierarchical clustering of PGR sites led to five original statistical clusters and one heuristic cluster split from the largest statistical cluster (Fig. 4b). The statistical significance of this heuristic cluster was confirmed with Tukey's honestly significant differences. The number of sites in each cluster ranged from two to 14. Clusters with a small number of sites were typically avoided since sample size was likely to drive very site-specific differences to emerge rather than larger regional trends. However, in the PGR classification, clusters of two and three sites were retained due to clear physical geomorphic differences between neighboring clusters (e.g. most similar clusters) identified in the heuristic assessment. For RGW sites, eight statistical clusters were originally selected to incorporate key differences in channel types with more similar clustering (e.g. channel types 3 and 5 in Fig. 4c). Heuristically, two clusters with low sample size were recombined with larger, similar clusters using field knowledge and landscape setting information. This resulted in six final clusters with number of sites ranging from four to thirteen. Hierarchical methods for WS sites settled on four statistical clusters with a fifth cluster being extracted from the largest cluster group based upon channel area and size (Figs. 4d). Final WS clusters ranged in size from six to ten sites.

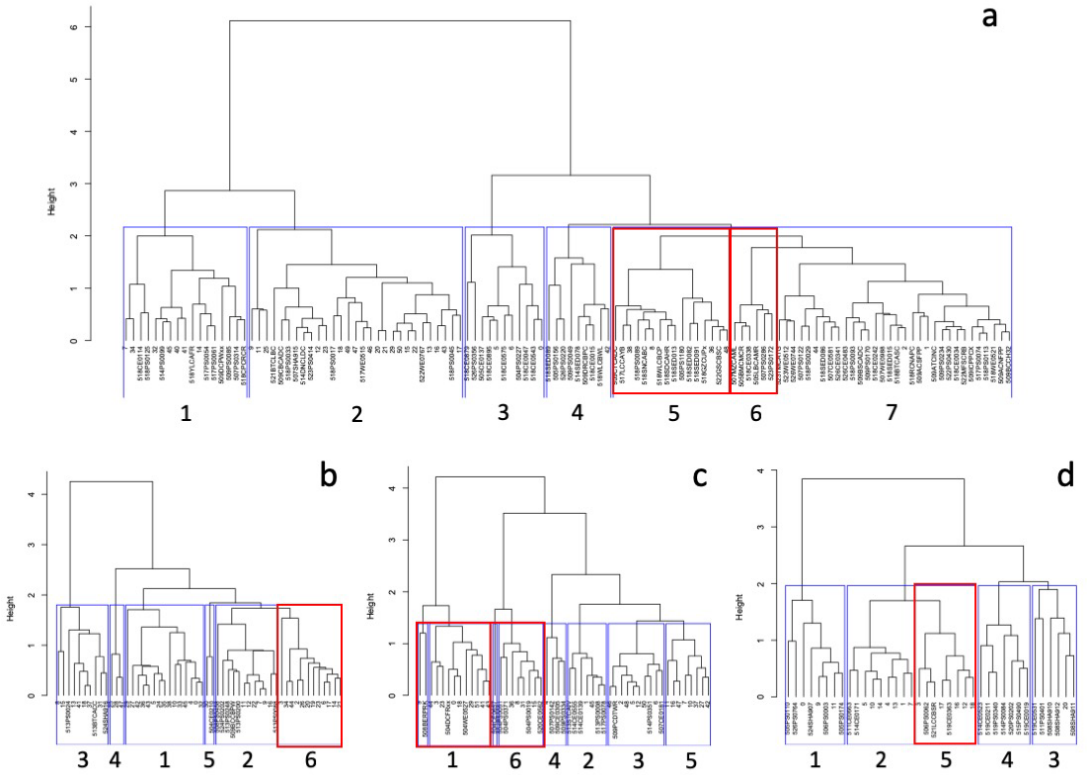


Figure 4. Hierarchical clustering using Ward's algorithm for rescaled geomorphic attributes of a) LSR, b) PGR, c) RGW, and d) WS sites. Blue boxes indicate groups based upon statistical clustering, while red boxes display heuristic refinement through combination or splitting of statistical groups.

Independent application of the classification tree analysis for LSR, PGR, RGW, and WS channel types indicated consistency in the most influential geomorphic attributes (Fig. 5). Sediment size (D84) and valley confinement distance (vc.dist.25pct) appear in each of the classification tree analyses for the four classifications. In addition, a cross-sectional geometry attribute (i.e. bankfull width or depth) appeared in each of the classification trees. Finally, a TVA and contributing area appeared in half of the classification trees. Complexity Parameter (CP) values applied in pruning of each classification tree are indicated in Table 1.

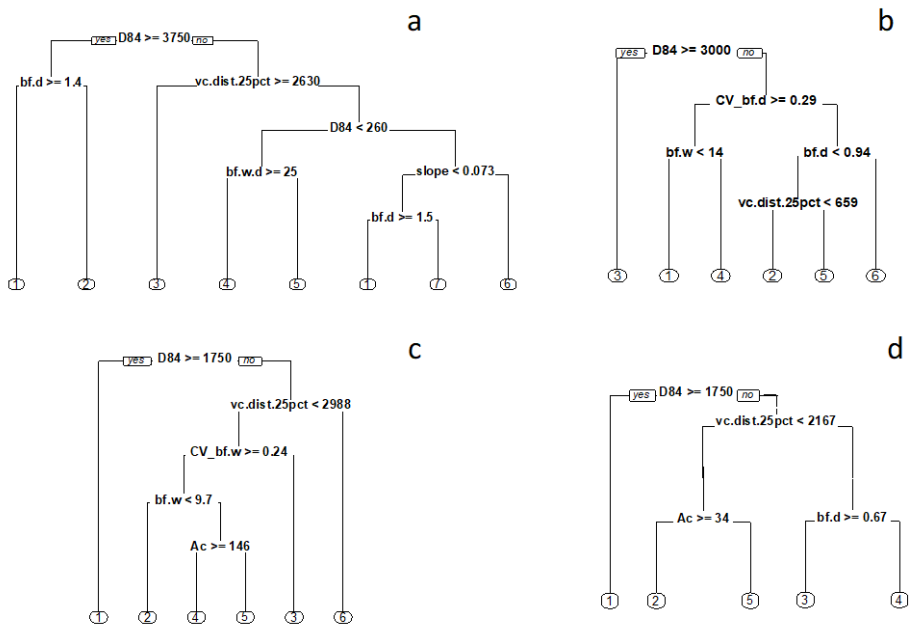


Figure 5. Classification and Regression Tree Analysis applied to channel types developed from hierarchical clustering with heuristic refinement for a) LSR, b) PGR, c) RGW and d) WS streams.

Seven channel types were identified in LSR streams: confined, boulder-bedrock, large area (LSR-1), confined, boulder-cobble, headwater (LSR-2), unconfined, gravel, low slope (LSR-3), partly-confined, cobble-gravel, uniform (LSR-4), partly-confined, gravel-cobble, riffle-pool (LSR-5), confined, cobble-boulder, cascade/step-pool (LSR-6), and confined, cobble-boulder, uniform (LSR-7) (Fig. 6). The classification tree was optimized using D84, bf.d, vc.dist.25pct, bf.w.d, and slope as attributes, which produced a classification rate of 88%. Cross-validation of classification trees are quantified in Table 1. Results of Tukey's honestly significant differences algorithm display the statistical differences between groups along the classification tree (Fig. 7).

While statistical differences in attributes drive the resulting channel types, it is important to place these channel types within the context of a watershed or river continuum. In the steepest landscapes, *confined, cobble-boulder, cascade/step-pool (LSR-6)* streams are present. With steep slopes, these streams exist in headwater settings with coarse sediment. In contrast, streams with lesser gradients and smaller grain sizes can also exist in headwater settings as *confined, boulder-cobble, headwater (LSR-2)* streams. If confined settings continue, headwater streams give way to larger *confined, boulder-bedrock, large area (LSR-1)* or *confined, cobble-boulder, uniform (LSR-7)* streams which exist at similar slopes but exhibit differences in TVA values. As the landscape becomes less confined, streams are likely to transition to *partly-confined, cobble-gravel, uniform (LSR-4)* or *partly-confined, gravel-cobble, riffle-pool (LSR-5)* streams, which are again notably different in the associated TVA values. Finally, in unconfined settings, *unconfined, gravel, low slope (LSR-3)* streams were found to exist and are likely to include single thread and anastomosed systems. Because LSR streams are likely to be found in mountainous regions of California, unconfined systems are likely limited to upland valleys where slopes decrease significantly.

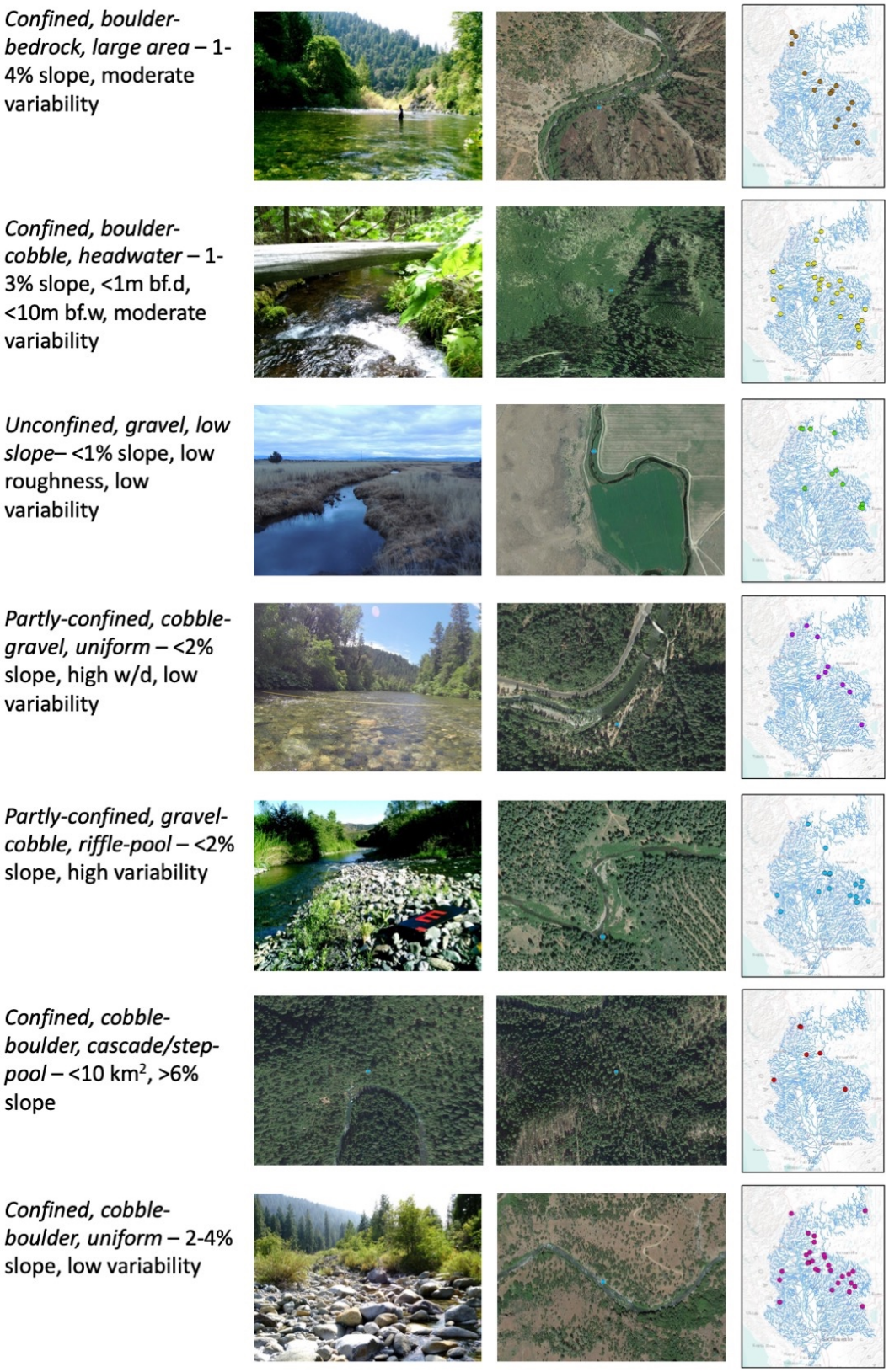


Figure 6. Channel types described by LSR statistical analysis with heuristic refinement.

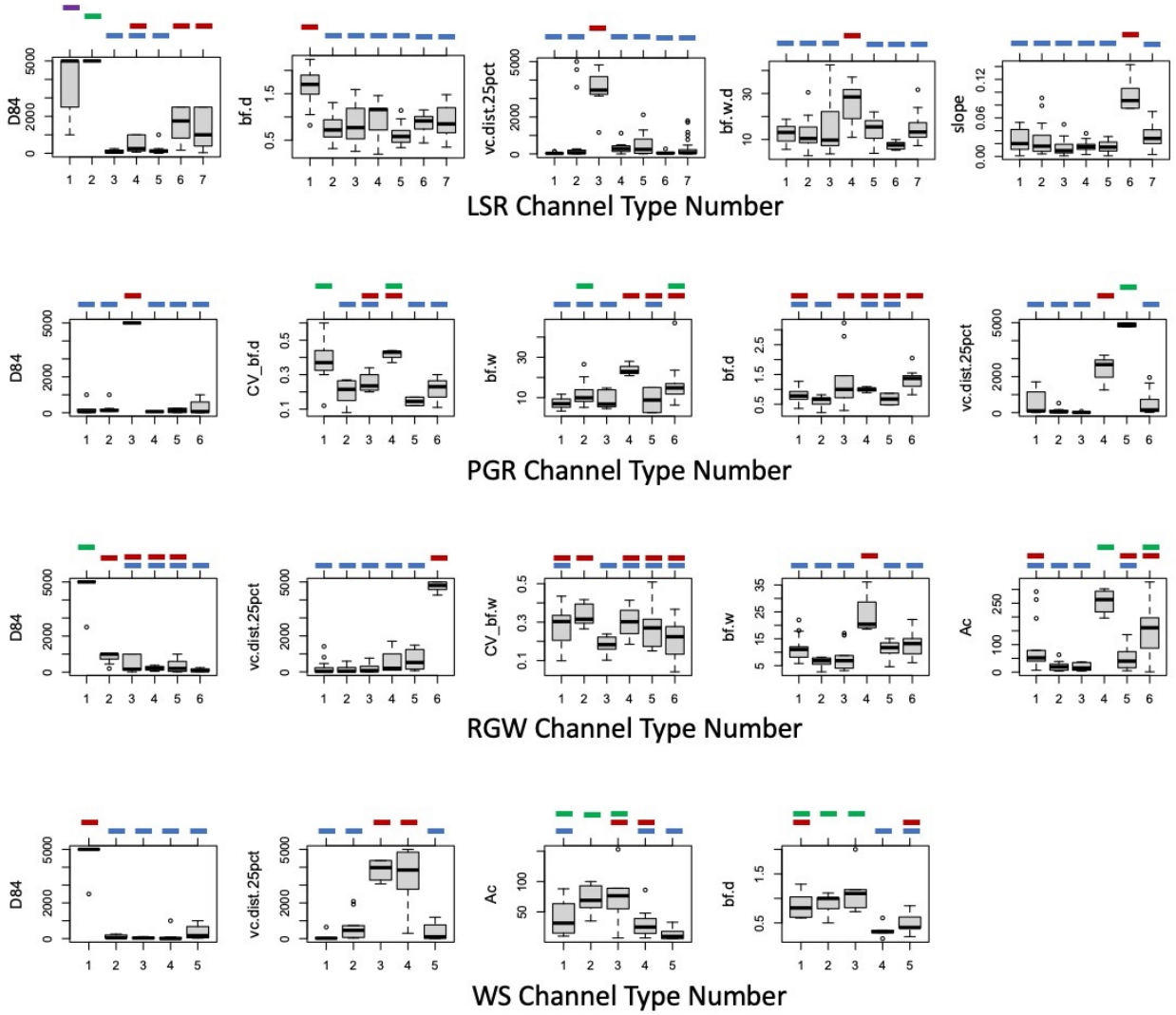


Figure 7. Tukey's honestly significant differences for PGR, RGW, and WS attributes found in the classification tree for each analysis. PGR, RGW, and WS attributes are represented by rows 1 through 3, respectively. Statistically significant differences between channel types for each attribute exist where color bars do not overlap. Colors are unique to each individual attribute for each hydrologic class.

Six channel types were identified in PGR streams: partly-confined headwater pool-riffle (PGR-1), confined uniform (PGR-2), confined bedrock (PGR-3), unconfined large area pool-riffle (PGR-4), unconfined uniform constrained (PGR-5), and partly-confined uniform (PGR-6) (Fig. 8). Based upon the hierarchical and heuristic cluster groups, 94% of all PGR sites were correctly classified based on D84, CV_bf.d, bf.w, bf.d, and vc.dist.25pct.

The *partly-confined headwater pool-riffle (PGR-1)* channel type is defined as a headwater stream due to small contributing areas. These sites often exist in upland valleys constrained by mountainous terrain. Downstream, these sites likely contribute to similar but larger and more uniform streams described here as *partly-confined uniform (PGR-6)*. These streams exist in a

similar valley setting described by confinement distance but have a significantly larger bankfull width (bf.w) and less depth variability (CV_bf.d). The *partly-confined uniform* streams have some degree of floodplain development in a partly-confined valley setting. Uniform channel type indicates low TVA values, an attribute shared by two other channel types in both confined and unconfined settings. *Confined uniform (PGR-2)* streams were identified in mountainous areas with steep valley walls defining the stream margin. These sites are likely influenced by colluvial processes in addition to longitudinal deposition and erosion of alluvial material within the stream channel. In comparison, *confined bedrock (PGR-3)* streams are defined by bedrock substrate with minimal alluvial material. The final two channel types identified in PGR streams are located within unconfined valley settings. *Unconfined, large area pool-riffle (PGR-4)* are identified by the large contributing area and distinct channel planform characteristics. These streams are of high stream order and take the identified characteristic form as the streams become unconfined in the Central Valley. The second channel type within unconfined valley settings are identified as *unconfined uniform, constrained (PGR-5).* These channel types are of variable contributing area and are substantially influenced by anthropogenic adjustments. The uniform nature of the channel dimensions (i.e. low TVA values) may be heavily influenced by human alteration in both agricultural and urban settings.

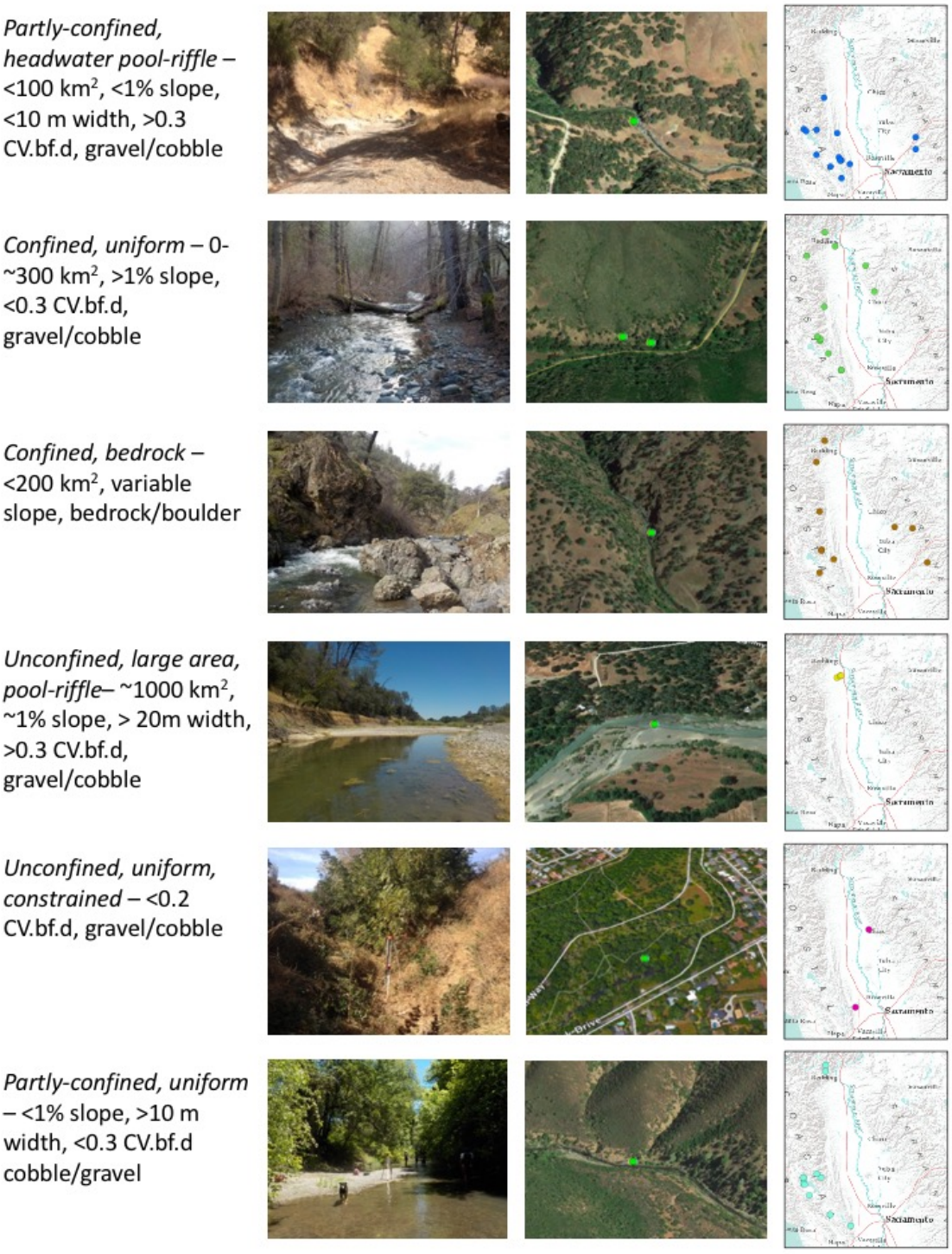


Figure 8. Channel types described by PGR statistical analysis with heuristic refinement.

Six channel types were also identified in streams defined by RGW hydrology, although many types were distinct from those identified in PGR streams. These channel types were described as: *confined bedrock (RGW-1), confined step pool (RGW-2), confined uniform (RGW-3), partly-confined lowland pool-riffle (RGW-4), partly-confined upland pool-riffle (RGW-5),* and *unconfined constrained (RGW-6)* (Fig. 9). Classification trees correctly placed 92% of statistically and heuristically defined channel types. Attributes which define statistical splits within the RGW dataset include D84, vc.dist.25pct, CV_bf.w, bf.w, and Ac (Fig. 5c). Figure 7 displays the significant differences between channel types which drive the classification tree for RGW sites.

As nearly 80% of RGW streams occur in the eastern region of the Sacramento River basin, RGW streams predominantly flow out of the Sierra Nevada Mountain Range. This is in contrast to PGR streams which mainly exist or originate within the western Coast Ranges of the Sacramento basin. Differences in elevation and aspect between the Coast and Sierra Nevada Ranges drive differences in hydrologic regime. Differences in hydrology, geology, and ecology all appear in RGW streams when qualitatively compared to PGR channel types. Headwater streams in RGW streams originate predominantly in the Sierra Nevada Range and to a lesser degree in the Coast Range. These channel types are identified as *confined, step pool* and *confined, uniform* and differentiated by slope. Steeper streams take the form of *step pool* channels with greater TVA values while lower slope streams have considerably more uniform channel dimensions, especially width. *Confined, bedrock* channel types also exist within RGW streams. These streams also have relatively small areas and remain confined by steep valley walls. In portions of the hydrologic class where confinement distance is greater, *partly confined, lowland pool-riffle* and *partly-confined, upland pool-riffle* streams exist. Statistically significant differences in width and contributing area create independent groupings of these channel types, even with similar planform descriptions. *Partly-confined, lowland pool-riffle* exist in both wide valleys of the Coast Range and as valley walls dissipate on the western side of the Sierra Nevada Range. Likely downstream of these streams in the Central Valley are *unconfined, constrained* channels. These streams are likely altered or constrained by urban or agricultural influences, thus limiting the natural lateral processes of the streams. Ultimately, these streams flow from alluvial fans near the foothills of the Sierra Nevada Range to the main-stem Sacramento River.

Five channel types were identified for streams defined by WS hydrology: confined step pool (WS-1), partly-confined gravel-cobble (WS-2), unconfined pool-riffle (WS-3), unconfined uniform headwater (WS-4), and partly-confined headwater gravel-cobble (WS-5) (Fig. 10). Classification trees correctly placed 97% of streams into the correct group based upon statistical splits in D84, vc.dist.25pct, Ac, and bf.d (Fig 5d). The statistical differences in channel types are plotted in Figure 7.

While WS streams are more distributed across the state, they are defined by lower total contributing areas than PGR or RGW streams. In addition, several channel types associated with WS hydrology appear to depend on their location within the watershed. For example, *confined step pool* channels exist to a large extent in the northern mountainous regions defined by WS hydrology. These steep slope channels have similar morphology when qualitatively compared to the RGW step pool channel type. *Confined step pool* streams have small contributing areas and are the most confined of streams in this hydrologic class. *Partly-confined, headwater gravel-cobble* and *unconfined, uniform headwater* streams also exhibit small contributing areas. However, the spatial context of these streams is considerably different than the step pool channels. While

partly-confined, headwater gravel-cobble streams have high slopes similar to the *confined step pool* channel type, these streams exist in predominantly foothill, unconsolidated terrain. This spatial location likely creates more horizontal slope distribution instead of vertical adjustment as in the *confined step pool* streams. While *unconfined, uniform headwater* streams also have limited contributing area, these streams originate in the Central Valley region. Therefore, these small streams have low slope. Because the streams are within the Central Valley, when flow becomes concentrated to the point of channel formation, these channels are likely constrained by urban and agricultural influences. *Partly-confined gravel-cobble* and *unconfined pool-riffle* have greater contributing areas than the other three channel types in this hydrologic class but still not as large as some streams in PGR or RGW hydrologic classifications. *Partly-confined gravel-cobble* streams likely exist predominantly in foothill regions, but are of higher stream order than *partly-confined, headwater gravel-cobble* streams. The two types of streams are similar in their existence in highly erodible, unconsolidated landscapes, but *partly-confined gravel-cobble* are statistically larger in size. Finally, *unconfined pool-riffle* streams exist as larger rivers within the Central Valley. In addition to the classification tree differences, these streams also have significantly greater variability in depth due to their pool-riffle form. Even though anthropogenic influences likely constrain these rivers to some degree, it may be that discharges characteristic of *unconfined pool-riffle* require more lateral area, thus leading to more observable river form compared with smaller streams in the Central Valley.

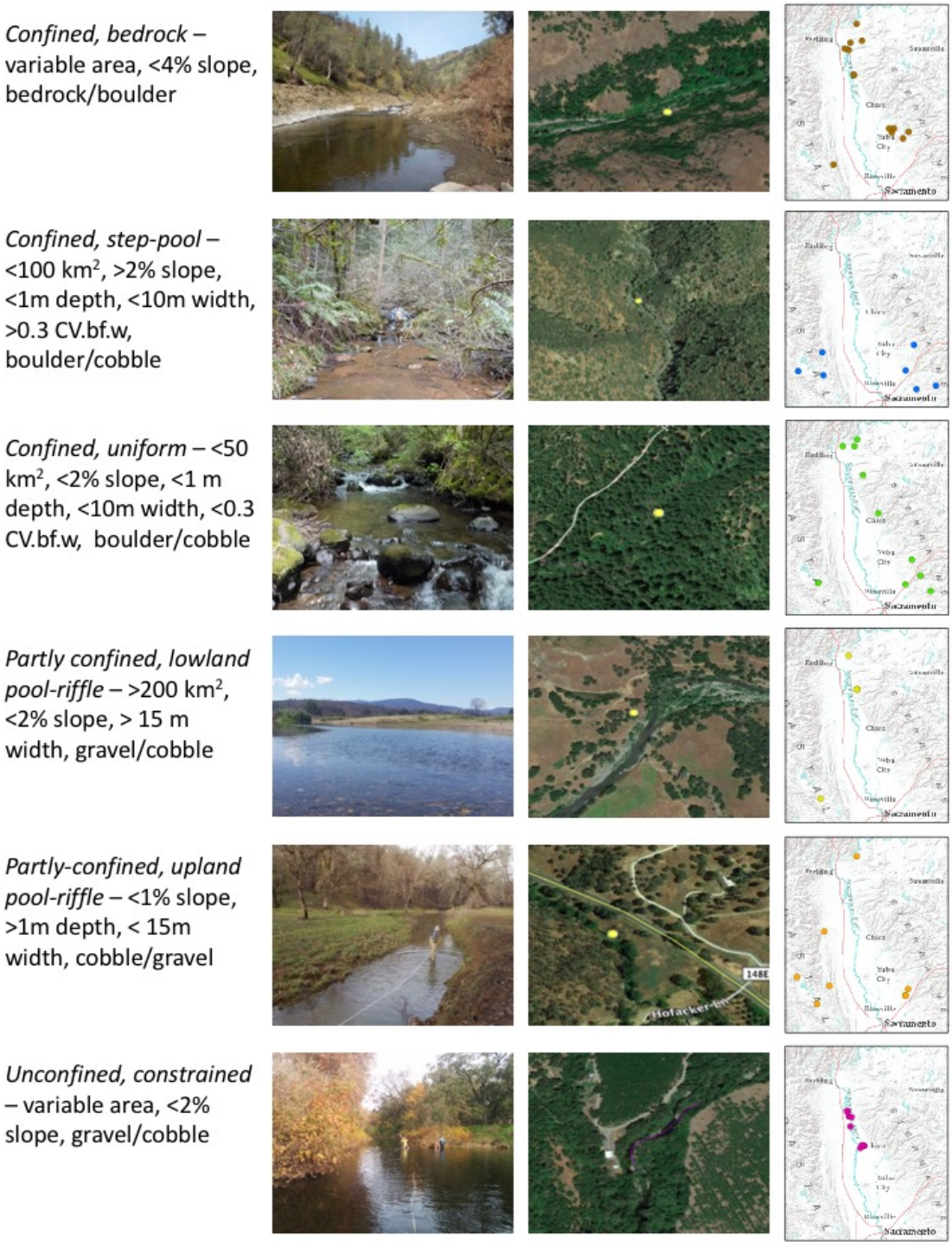


Figure 9. Channel types described by RGW statistical analysis with heuristic refinement.

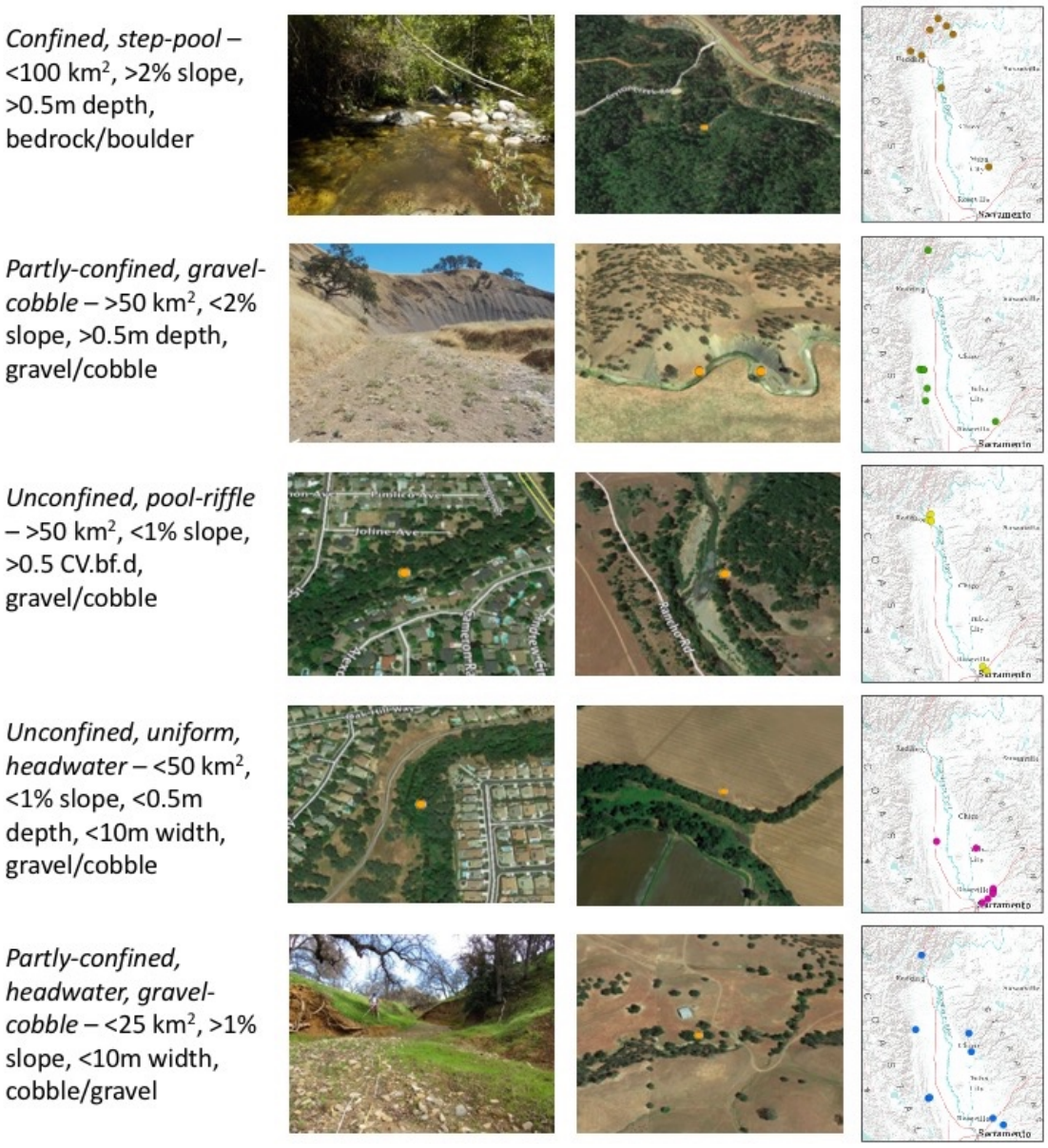


Figure 10. Channel types described by WS statistical analysis with heuristic refinement.

1.3.2 Unified Sacramento River Basin Classification

1.3.2.1 Multivariate Statistical Clustering

Sediment size and valley confinement were identified as the most influential attributes in multivariate clustering. The final NMDS solution recorded a stress value of 0.139 with a non-metric coefficient of determination of 0.981 between observed dissimilarity and ordination distance (Fig. 11a). The first and second principle component axes (PCAs) resulting from the NMDS ordination explained 63% of the variance in the data with loadings of 0.94 for D84 and 0.91 for vc.dist for PCA-1 and PCA-2, respectively. These strong loading values indicate a large

amount of clustering in the NMDS analysis can be attributed to these two variables alone. Inclusion of PCA-3 and -4 explained 80% of variance, for which bf.w and CV_bf.d are the attributes with the greatest loading values.

Ten channel types were identified through the Ward's clustering with heuristic refinement and CART (Figs. 11 and 12). The Hubert Index suggested three Ward's clusters as the optimal number of groupings, however this was due to strong breaks in sediment size and valley confinement. As three groups was insufficient to describe the full number of channel types within the basin, secondary indications by Hubert Index values at 10 and 14 groups were the focus of heuristic refinement. The final ten channel types were the result of CART analysis which resulted in a successful prediction rate of 84%. Ten-fold cross-validation of this prediction was 75%.

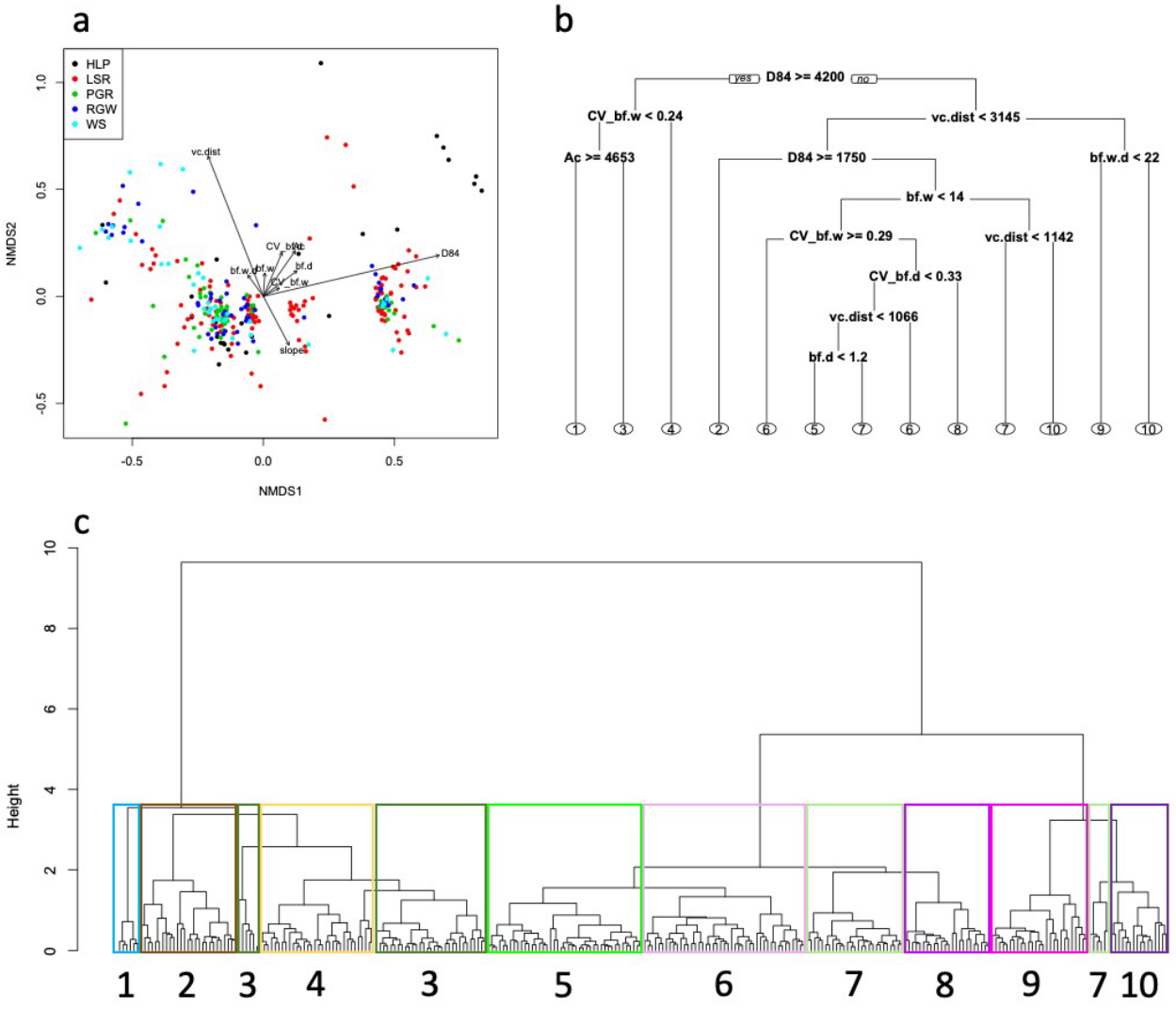


Figure 11. Results from a) non-metric multidimensional scaling, b) classification and regression tree, and c) hierarchical clustering by Ward's algorithm analyses resulting in ten geomorphic channel types.

Final channel type clusters were made up of between 6 and 45 sites. Clusters with a small number of sites were typically avoided, as outliers were likely to drive site-specific differences rather than larger basin trends. However, it was ultimately the uniqueness of cluster attributes that drove final channel types. Because sediment size and valley confinement played such an important role in site clustering, the classification is broadly numerically-organized from large to small clast size (Fig. 13). Channel types were then organized by median valley confinement (Fig. 12). While there was not a high correlation between sediment size and confinement distance using individual site data (R^2 = -0.27), there is a general inverse relationship between channel type sediment size and valley confinement distance. Median attributes of each channel type can be found in Table 2.

Table 2. Median channel attributes of ten channel types within the Sacramento River basin.

Channel Type	Ac	slope	bf.d	bf.w	bf.w.d	bf.d.D50	CV_bf.d	CV_bf.w	sinuosity	D50	D84	vc.dist.25pct
1	7441	0.004	1	15.2	14.5	8.59	0.49	0.23	1.15	128	5000	1090
2	84	0.042	0.97	11	11.02	5.4	0.2	0.26	1.2	250	2500	28
3	100.5	0.014	1.085	10.935	10.755	5.5	0.23	0.205	1.2	190	5000	45.5
4	31	0.0185	0.94	6.66	7.3	6.11	0.34	0.32	1.19	128	5000	23
5	30	0.02	0.66	6.58	9.4	9.62	0.2	0.18	1.12	57	200	62
6	32	0.012	0.7	6.83	8.6	23.09	0.23	0.32	1.2	40	95	598
7	164	0.0142	1.31	16.17	13.42	15.87	0.19	0.17	1.23	87	380	114
8	54	0.00625	0.735	11.575	16.79	27.875	0.415	0.25	1.19	27.3	130	104
9	74	0.007	1.03	8.12	8.51	64.6	0.23	0.24	1.15	11	45	4688
10	170	0.00915	1.095	17.84	20.515	35.15	0.295	0.26	1.135	28	64	2868.5

Given the general relationships between confinement and sediment size, the classification progresses from confined mountainous upland to unconfined lowland streams. A notable exception is the first channel type, which fits within the conceptual framework of large to small sediment size rivers, but the sites are classified as unconfined. This lack of confinement indicates colluvial and mass movement processes are unlikely in these settings. Therefore, the large sediment clasts and unique Modoc Plateau volcanic terrain at these locations are likely non-fluvial legacy deposits (Hauer and Pulg, 2018). The remaining channel types generally follow the pattern of large sediment size-high confinement to small sediment size-low confinement progression.

Four channel types were observed in confined settings: *boulder, high-gradient, step-pool/cascade*; *boulder-bedrock, uniform*; *boulder-bedrock, low-gradient, step-pool*; and *gravel-cobble, uniform*. Slope and TVAs were critical attributes in splitting these channel types after initial sediment size and confinement splits. *Boulder, high-gradient, step-pool/cascade* (n = 27) streams are defined by a significantly high slope value compared to all other groupings and less bedrock influence than other large-grained channel types (Fig. 13). These streams have relatively low depth variability, but moderate width variability. While step-pool streams are thought to have high depth variability, the statistical conflation of cascade and step-pool systems likely resulted in a dampened depth variability. A second partial explanation for low depth variability may be that depth variability in steep streams is lower relative to lower gradient step-pool and riffle pool styles (Wohl et al., 1993). As slope decreases, two channel types with boulder-bedrock sediment sizes and confined settings exist: *boulder-bedrock, uniform* and *boulder-bedrock, low-gradient, step-pool*. *Boulder-bedrock, uniform* streams (n = 36) exhibit low TVA values. Similar to plane bed streams in other literature (Grant et al., 1990; Montgomery and Buffington, 1997), uniform terminology is used here to signify smaller changes in both depth and width along a reach. However, it should be noted that

uniform streams do still exhibit some variability in channel dimensions. At slopes between the previous two channel types, *boulder-bedrock, low-gradient, step-pool* reaches (n = 33) exhibit high TVA values. Although width and depth undulation is substantial, the bf.w.d values are small compared to other channel types. The final confined channel type exhibits significantly smaller sediment sizes as compared to the other three confined channel types. Similar to *boulder-bedrock, uniform* streams, *gravel-cobble, uniform* streams (n = 43) display low TVA. This channel types also exhibited the smallest median channel dimensions in the basin.

Three partly-confined channel types provide key linkages between confined and unconfined river systems: low width-to-depth ratio, gravel-cobble, riffle-pool, cobble-boulder, uniform, and high width-to-depth ratio, gravel-cobble, riffle-pool. Partly-confined low width-to-depth, gravel-cobble, riffle-pool (n = 45) streams exhibit similar grain sizes and dimensions to confined gravel-cobble, uniform streams. While bed undulation remains relatively absent, width variability increases significantly, likely due to the deposition and reworking of alluvial sediments in the partly-confined setting. These streams are the most common channel type of the sites surveyed. Due to their small size, bed undulations may be difficult to measure. Similar to confined gravel-cobble, uniform streams are partly-confined cobble-boulder, uniform streams (n = 33). While TVA values remain low for these uniform streams, bf.w.d is significantly higher than both the confined gravel-cobble, uniform and partly-confined, low width-to-depth ratio, gravel-cobble, riffle-pool streams. In addition, larger sediment sizes exist compared to other partly- and unconfined channel types. While partly-confined cobble-boulder, uniform streams exhibit the lowest median depth variability, partly-confined, high width-to-depth ratio, gravel-cobble, riffle-pool streams exhibit the greatest depth variability, with the exception of the unique bed undulating, boulder-bedrock channel type. In combination with moderate width variability, the smaller sediment sizes likely lead to bedform resistance in compare to grain resistance in the cobble-boulder, uniform channel type.

Two final unconfined channel types were observed in the Sacramento River basin: *low width-to-depth gravel* and *gravel-cobble, riffle-pool*. *Low width-to-depth, gravel* streams (n = 27) occur in the most unconfined settings. These relatively deep, narrow streams exist both in upland valleys and within the Central Valley of California and display the smallest sediment sizes of all streams. These streams are laterally stable either due to well developed and heavily vegetated floodplains or anthropogenic land use and may exist as anastomosed streams. The *low width-to-depth, gravel* channel type exhibits low depth variability and only moderate width variability. In comparison, the unconfined *gravel-cobble, riffle-pool* channel type has relatively high depth variability and moderate width variability. While depth variability is lower than partly-confined *high width-to-depth ratio, gravel-cobble, riffle-pool* streams, higher width-to-depth ratios in these unconfined *gravel-cobble, riffle-pool* streams may be an indication of systems near the threshold of a braided morphology.

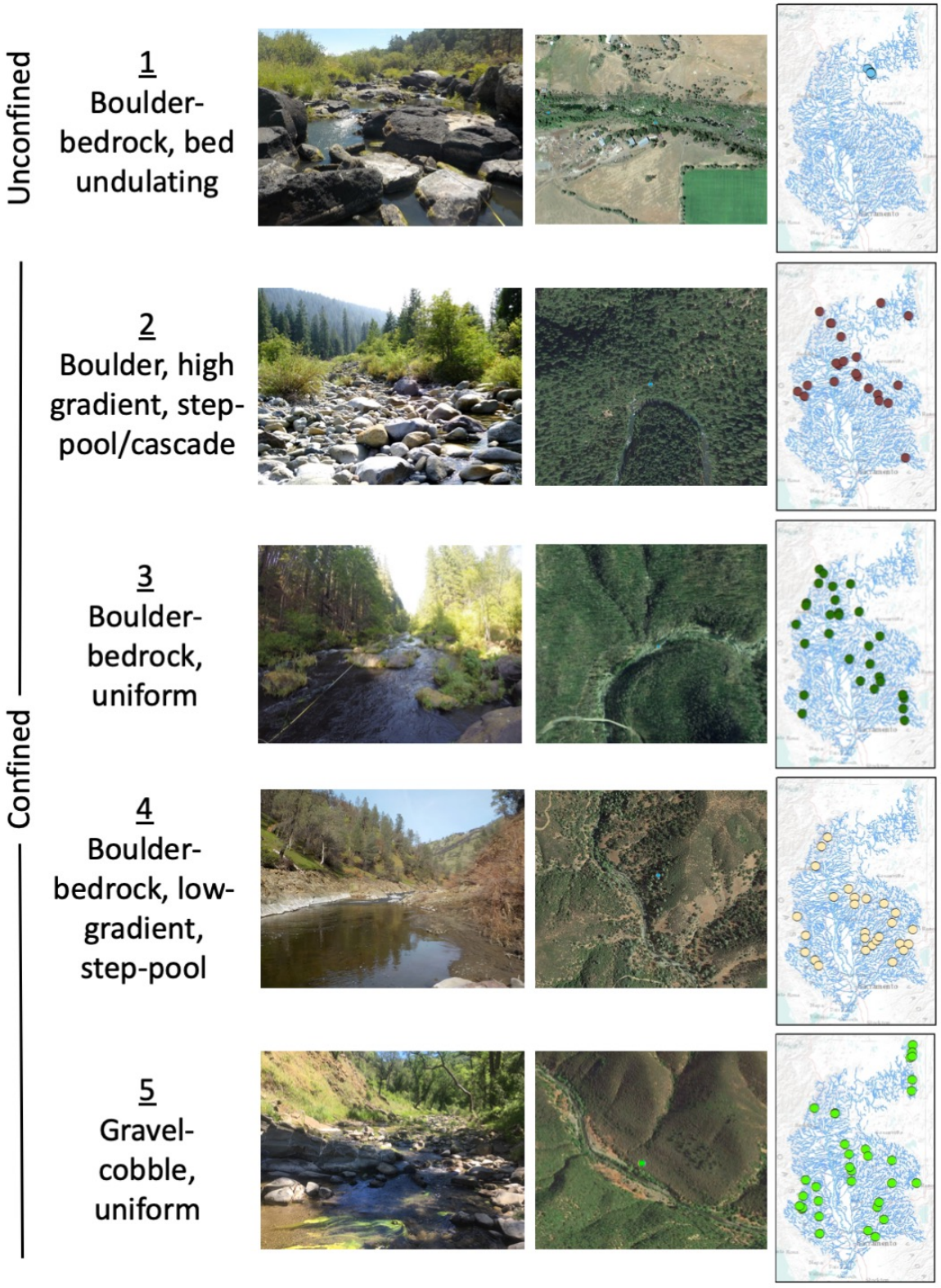


Figure 12. The ten channel types within the Sacramento River basin developed by multivariate statistical analysis with heuristic refinement.

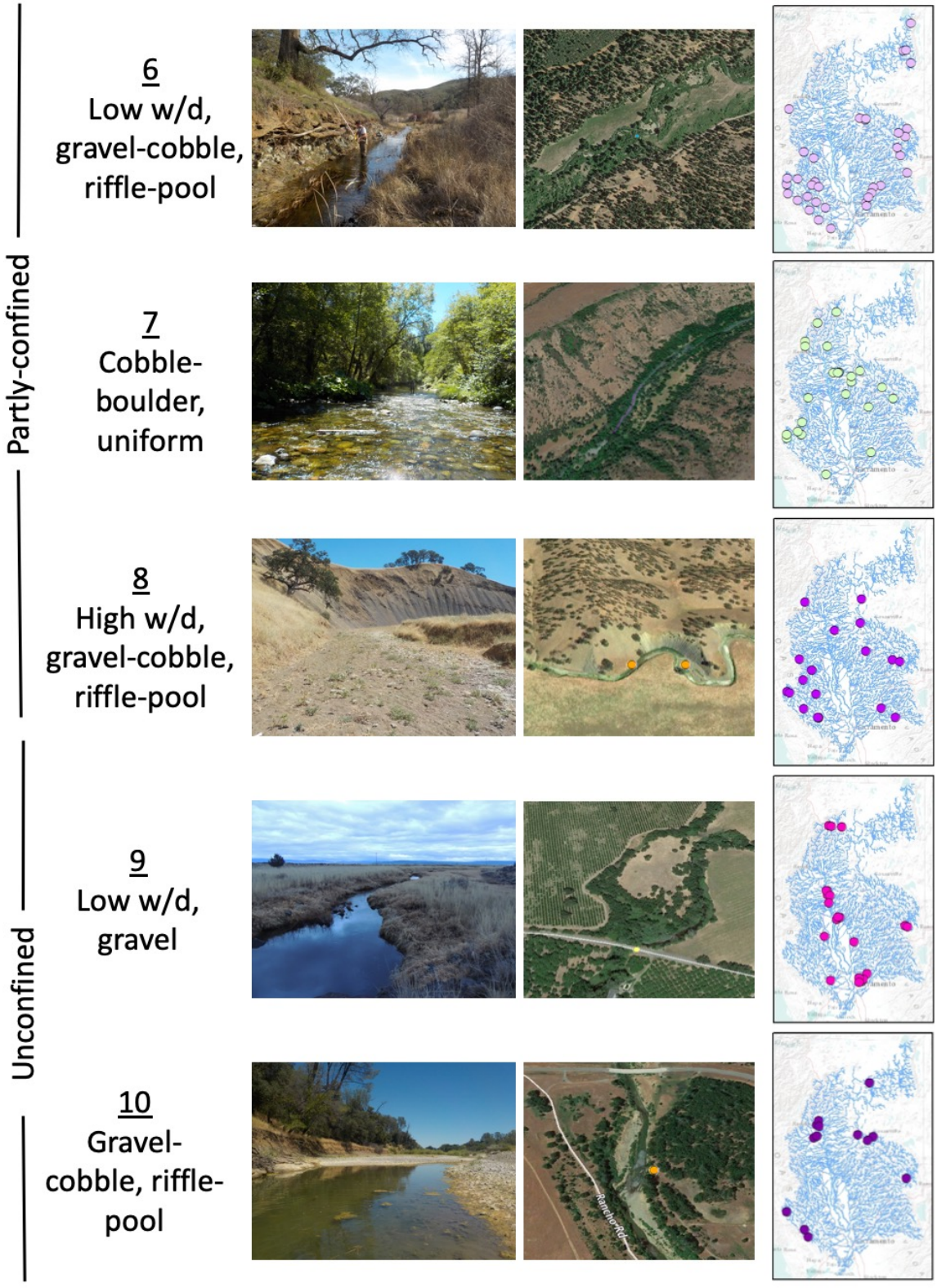


Figure 12 (cont'd). The ten channel types within the Sacramento River basin developed by multivariate statistical analysis with heuristic refinement.

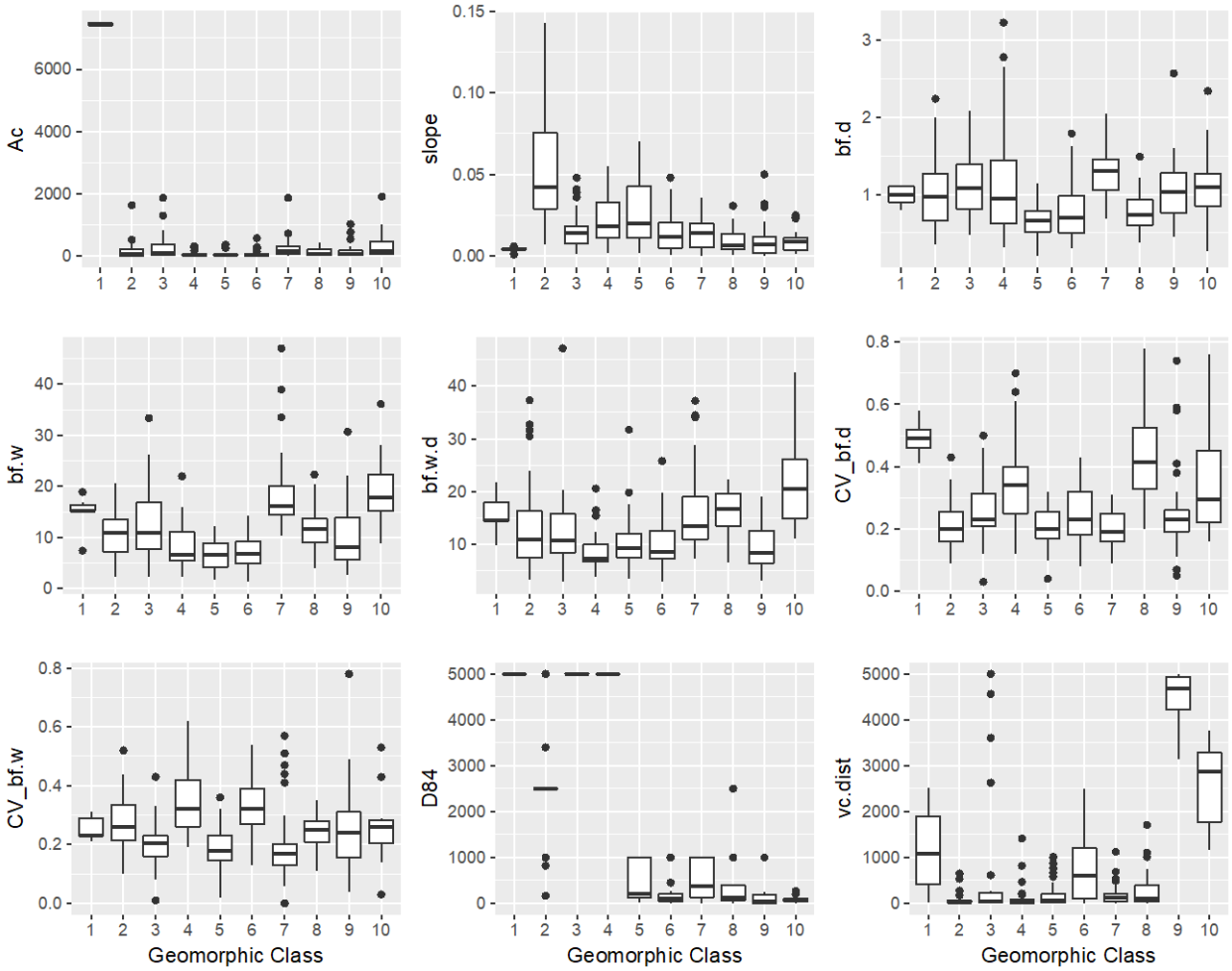


Figure 13. Tukey's Honestly Significant Differences between attributes of each channel type in the Sacramento River basin.

2 Predictions of Channel Types over the Sacramento Basin

2.1 Methods

The channel type of each stream segment in the region was predicted using a supervised classification approach: given a set of predictors (variables) what label (channel type) should be assigned to each stream segment? Ten hydro-geomorphic channel types were previously derived using an unsupervised learning approach (see Section 1.3.1). Since each channel type was expert inspected in the unsupervised learning phase, the labels are treated as noiseless (e.g. Borut, Dragan, and Nada 2010; Garcia, Lorena, and Carvalho 2012; Sluban, Gamberger, and Lavrač 2013; Garcia, Carvalho, and Lorena 2015). The frequency of stream reaches in each classes is unbalanced (Table 3), with the fewest streams in class 1 and the majority in classes 5 and 6.

A multi-tiered machine-learning framework was developed to select the best set of predictors, pre-processing and classifier – or classifying algorithm – to perform the classification task. The four stages of this framework are described below (Figure 14): (1) define a tractable problem for reducing predictor noise; (2) assess the statistical performance of classifiers; (3) assess the stability of the spatial predictions; and (4) evaluate the geomorphic relevance of the predictions.

Figure 14. Flowchart of the four-fold machine learning framework.

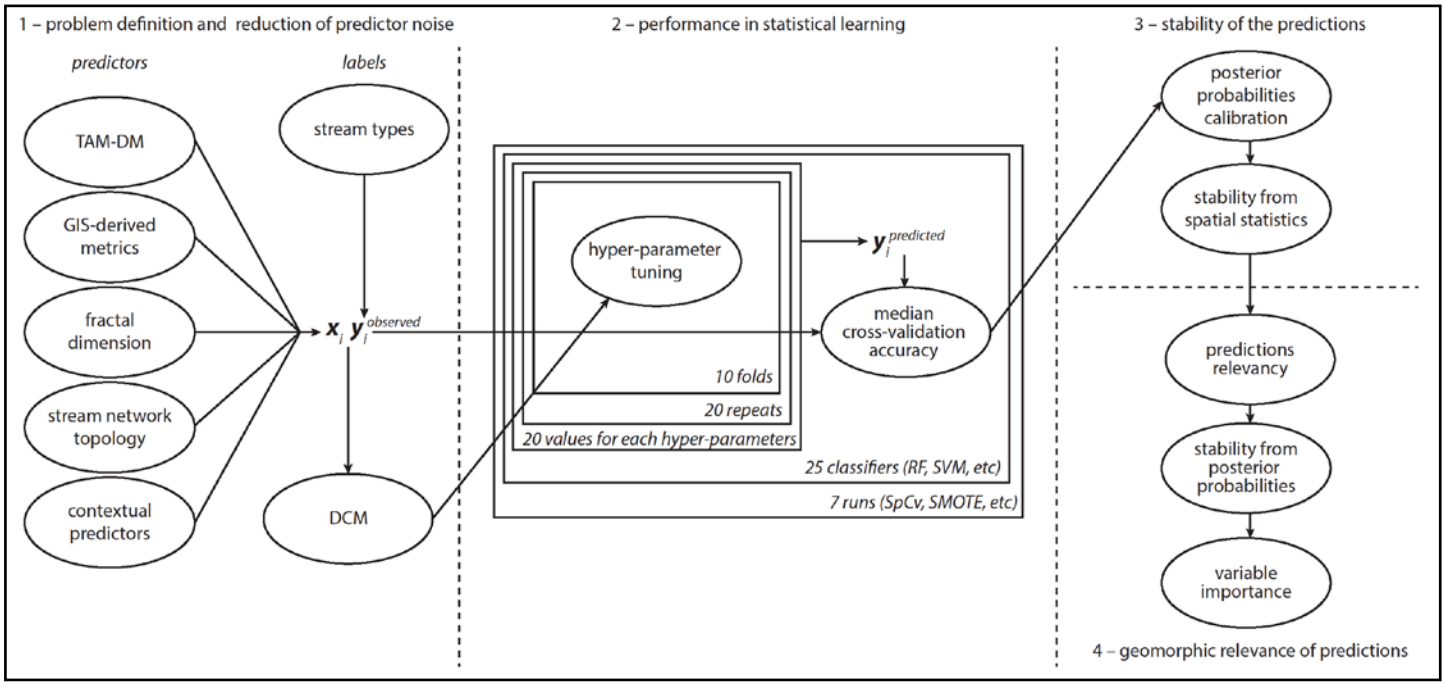


Table 3. Number of observations for each class of channel type and prevalence (i.e. relative frequency).

HG class	1	2	3	4	5	6	7	8	9	10
number	6	27	36	33	43	45	33	24	27	16
prevalence	0.021	0.093	0.12	0.11	0.15	0.16	0.11	0.083	0.093	0.06

2.1.1 Defining a Tractable Problem and Reducing Predictor Noise

As most data-driven approach, the set of predictors used in input influences the output. In the following paragraphs, the input variables to the machine learning algorithms are described. However, as irrelevant predictors might induce noise and deteriorate the predicting performance of the classifiers, an avenue to filter predictor noise is detailed.

Following a data-driven approach, a large set of predictors is initially included in the machine learning framework. This initial set of predictors corresponds to a set of 287 metrics commonly understood to influence channel morphology (e.g. Rosgen 1994; Montgomery and Buffington 1998; Buffington and Montgomery 2013; Teutschbein et al., 2018): topography, geology, soils, land cover and climate. Several predictors are derived from two core datasets: (i) the 10-m National Elevation Dataset (Gesch et al., 2002; NED); and (ii) the stream segments from the National Hydrology Dataset (McKay et al., 2012; NHDPlusV2) with their associated hydrologic class (Lane et al., 2018). The initial set of predictors included 108 Terrain Analysis Metrics, 3 GIS-derived metrics, 32 fractal dimension metrics, 4 network topology metrics and 140 contextual variables.

Several Terrain Analysis Metrics (TAM) were calculated from the NED 10-m Digital Elevation Model (DEM),: elevation z, slope, aspect, roughness ($max[z_i] - min[z_i]$), flow direction, planform curvature, profile curvature, Topographic Position Index (TPI, $\langle z_i - \lfloor z_j \rfloor \rangle_j$) and Terrain Ruggedness Index (TRI, $\langle |z_i - z_j| \rangle_j$), where z_i is the elevation of the current DEM cell; z_j are the 8 neighboring cells and brackets indicate averaging all j. This calculation was performed using the `R` package raster (Hijmans et al., 2018). In addition, as curvature is linked to erosion (Hurst et al., 2012), planform and profile curvatures estimations following Evans method (Florinsky 1998) were added to the C code of the raster package. Since the distribution of Terrain Analysis Metrics differentiates different stages of landscape maturity (Bonetti and Porporato 2017) and channel types (Lane, Pasternack, et al., 2017), the following Distribution Metrics (DM) were estimated in addition to the mean value of each TAM: mean, median, minimum, maximum, standard-deviation and skewness. The TAM-DM metrics were estimated over two spatial scales: a 512×512-m tile centered at the midpoint of each stream segment and along a 100-m wide near-channel buffer. The calculation of the TAM-DM over the ~ 109k stream segments was performed using UC Davis High Performance Computing (HPC) FARM cluster.

Three predictors were derived using ArcGIS (ESRI 2016): channel confinement (Fryirs, Wheaton, and Brierley 2016), channel slope and sediment supply. Confinement was defined as the distance between the channel and the closest high slope as described in Section 1.2.2.2. Sediment supply was estimated using the Revised Universal Soil Loss Equation (Renard et al., 1997; RUSLE) from

data publicly available from the California Waterboards and from NLCD 2011 data combined with Haan, Barfield, and Hayes (1994) lookup table.

Metrics describing the stream network topology were extracted from the NHDPlusV2 dataset: drainage areas, Strahler's stream order and Local Drainage Density (LDD). Strahler's stream order (Strahler 1957) captures the hierarchy of streams in a drainage basin with low-order streams combining to form higher order streams. For a given stream segment, LDD corresponds to the drainage area divided by the total length of the network upstream from the segment and has been recently shown to identify areas with distinct geomorphic processes (Danesh-Yazdi, Tejedor, and Foufoula-Georgiou 2017).

The fractal dimension (Mandelbrot 1967), which corresponds to the slope of the relationship between the average standard deviation and the spatial scale, has been shown to reflect the influence of tectonics, lithology and erosion processes (e.g. Xu, Moore, and Gallant 1993; Carr 1997; Sung, Chen, and Chao 1998; Sung and Chen 2004; Faghih and Nourbakhsh 2015; Liucci and Melelli 2017). Here, fractal dimension was calculated following the methods from Liucci and Melelli (2017) in 3 steps over 32 sets of 5 consecutive scales: (i) the initial elevation raster was aggregated to a standard deviation raster; (ii) the standard deviation rasters were aggregated to an average standard deviation raster; and (iii) a regression over the log-transformed stack of average standard deviation rasters provided an estimation of the fractal dimension at each cell.

The fractal dimension is obtained by performing a linear regression over the log-transformed stack of average standard deviation rasters. The relationship between fractal dimension and scale is known to vary with the scale of observation with a crossover between shorter and longer length-scales (Pastor-Satorras and Rothman 1998; Dodds and Rothman 2000; Duclut and Delamotte 2017), such that fractal dimension represents tectonism at large scale (Wilson and Dominic 1998; Sung and Chen 2004; Faghih and Nourbakhsh 2015; Liucci and Melelli 2017) and erosion processes at short scale (Lifton and Chase 1992; Sung and Chen 2004; Faghih and Nourbakhsh 2015; Liucci and Melelli 2017). To address this issue, the fractal dimension was computed over a sliding window of five scales with starting scales of 20, 30, 50, 70 and 90-m. Fractal dimension was also computed at 32 different upper tile scales ranging from ~0.6 km to ~82 km as an alternative to using multi-scale decomposition of the topography (Cazenave et al., 2012; Buscombe 2016; Agarwal et al., 2017; Newman, Lindsay, and Cockburn 2018). The calculation of the fractal dimension over the state of California was performed using UC Davis High Performance Computing (HPC) FARM cluster.

The initial resolution of the NED raster was 10-m, meaning each cell is 10×10-m. The initial elevation raster was aggregated to five standard deviation rasters with pixel sizes of 20, 40, 80, 160, and 320-m. Each cell of the 20-m standard deviation raster corresponds to the standard deviation of the elevation computed over $2 \times 2 = 4$ cells of the initial elevation raster ($32 \times 32 = 1024$ cells for the 320-m raster). Then, the standard deviation rasters were aggregated to an average standard deviation raster. The pixel size of such a raster corresponds to the smallest scale at which a meaningful averaging can be computed for the largest scale raster (320-m): 640-m. In consequence, each cell of the five 640-m rasters corresponds to the average standard deviation over 32×32, 16×16, 8×8, 4×4 and 2×2 pixels for 20-m, 40-m, 80-m, 160-m, 320-m rasters, respectively.

Contextual predictors were provided by the publicly available Stream-Catchment Dataset (StreamCat; Hill et al., 2015). These predictors are integrated over the entire upstream watershed and at the limited part of the catchment draining to a stream segment. In particular, the following set of contextual predictors were selected: lithology (Cress et al., 2010; initial resolution >1 km); soil characteristics (Schwarz and Alexander 1995; STATSGO, 1 km initial resolution); land cover (Homer et al., 2015; NLCD 2011, 30-m initial resolution); runoff and air temperature climatologies between 1981 and 2010 (PRISM Climate Group 2004, 800–m initial resolution); mines (count); Indices of Catchment and Watershed Integrity (Thornbrugh et al., 2018). In addition, some of the StreamCat predictors were computed within a 100-m riparian buffer.

One key challenge of this project is the prediction of channel reaches defined from data acquired from 10^0 to 10^2m scale with predictors typically available at the 10^2 to 10^5-m scale. In addition, the definition of relevant scales is often difficult (e.g. Archfield et al., 2015). The complexity of adding coarse predictors to the classification can be assessed using Data Complexity Measures (Ho and Basu 2002; Lorena et al., 2018 DCM). DCMs inform on the linearity of the problem, the complexity of the class boundaries as well as the underlying structure of the observations within the predictor space. DCM can be used to filter the predictors (e.g. Garcia, Carvalho, and Lorena 2015) that ultimately make the problem more complex by including noise from irrelevant or coarse predictors. After assessing the problem complexity with the initial complete set of all predictors, complexity was evaluated when removing the fractal dimension predictors, the stream network topology metrics, and the contextual variables.

2.1.2 Assessing the Performance of Classifiers in Statistical Learning

The classification is performed by a classifier, an algorithm that inputs predictor variables and outputs classified stream segment observations. The accuracy of a classifier corresponds to the number of predicted labels that match the observed labels. The set of parameters that produces the most accurate predictions for a given classifier is determined in the training phase using the predictors extracted at the location of the field surveys and their associated channel type.

The classifier with the best performance is often unknown at the start, so Luengo and Herrera (2013) proposed a range of DCMs to assess effective classifiers. DCMs computed to filter the predictors were used to determine best classifiers, and 11 different machine learning algorithms were tested: Partial Least Squares (PLS); Naive Bayes (NB); Multivariate Adaptive Regression Splines (MARS); SVM with linear (L-SVM) and radial kernel (R-SVM); k-Nearest Neighbors (k-NN); Classification And Regression Tree (CART); bagged-trees (BaT); Random Forest (RF); Linear Discriminant Analysis (LDA); Flexible Discriminant Analysis (FDA); Regularized Linear Discriminant Analysis (RLDA); Artificial Neural Network (ANN).

Two key algorithms considered, SVM and RF, are described below (for additional information see e.g. Shen 2018; Shen et al., 2018; Rahmati et al., 2017). Implementation of the algorithms was performed using the R packages caret (Kuhn 2008; Kuhn and others 2018) and h2o (H2O.ai 2018):

A linear SVM finds the linear boundary between two distinct classes so that it maximizes the margin between the boundary and each class closest point(s). Those points are the support vectors for the boundary. More flexible decision boundaries (i.e. non-linear) can be obtained by a non-linear kernel version of the SVM, that is transforming the predictor space in such a way that the

problem becomes linearly solvable in the transformed predictor space. The most common kernel used to perform such a "kernel trick" is the radial basis function. Both for L-SVM and R-SVM, solving a multi-class problem is achieved by breaking it down in a set of binary problems.

Random Forest is an ensemble model combining regression trees (i.e. the forest) and is widely used in natural sciences, especially ecology. Each individual regression tree is built from a random subset of the predictors. Tuning is usually performed on the number of predictors included in each tree, m_{try} but tuning the number of trees and sample size has been found to be valuable (Probst, Wright, and Boulesteix 2018). In addition, RF appears to be robust to predictors noise (e.g. Fox et al., 2017).

Most classifiers are defined so that some pre-processing steps are required. One common step is to deal with class imbalance (see below) which requires the following steps to be taken: centering and scaling predictors; and dealing with missing values with k-NN imputation. A combination of additional preprocessing steps were also tested: Box-Cox transformations; removing near-zero variance predictors; removing correlated predictors; and transforming predictors with Principal Component Analysis (PCA) or Independent Component Analysis (ICA). In addition to centering and scaling, Box-Cox transformations attempt to collapse the distribution of each predictor to a normal distribution – an assumption behind numerous classifiers (e.g. Csillik, Evans, and Drăguț 2015).

The datasets used in machine learning applications are often divided into a training set and a testing set; the training set is used to tune the hyper-parameters and the training set to assess the accuracy. In the case of smaller dataset, resampling allows all data to be used both in training and in testing. The most common resampling is the v-fold cross-validation (Burman 1989) with $v = 10$. In such 10-fold cross-validation, the data are randomly separated in 10 parts or folds. Successively, 1 fold is held out and the 9 other folds are used to train the classifier. The performance of the classifiers is assessed against the hold-out fold. The reported accuracy is then often the median over 10 cross-validation accuracies and yields an estimate of the performance of the classifier against unseen data. In our case, 20 repeats of 10-fold cross-validation were used to address the bias that might be introduced by the initial random selection of the folds. The median cross-validation accuracies were estimated over the accuracy from 200 different folds.

Spatial cross-validation (Schratz et al., 2018) is a variant of the standard v-fold cross-validation which addresses the issue of spatial correlation between the training data points. Spatial cross-validation ensures that the hold-out folds are spatially disjointed from the training folds. The folds are created from the coordinates of the training set points using the k-means algorithm (Hartigan and Wong 1979). In a multiclass problem such as ours, with an expected significant spatial variability, ensuring that each fold contains examples of all classes is complex. Regular cross-validation with stratified folds and spatial cross-validation were compared and the heterogeneity of the spatial cross-validation folds was assessed.

To address the unbalance of the training set (Table 3), the Synthetic Minority Oversampling Technique (SMOTE) was used (Chawla et al., 2002). SMOTE relies on assigning predictors along the edges connecting the k-nearest neighbours from randomly selected observations. As synthetic data points need to be included in the spatial cross-validation scheme, spatial cooridnates must be reliably assigned. SMOTE was used to generate the synthetic predictors and the required

geographic information was derived using a Gaussian noise with perturbation of 10% on the position of the observation selected by the SMOTE.

For each run and for each classifier, the set of best hyper-parameters was selected with a grid search across 20 different values per parameter. The one-standard-error rule was applied, selecting the simplest set of hyper-parameters within one-standard-deviation of the most accurate set of hyper-parameters. The best model was selected from these runs based on its median cross-validation accuracy. A paired t-test between the distribution of cross-validation accuracies from the best classifier and other classifiers was performed to assess their similarity. These additional classifiers are also reported.

In total, the 7 following runs were performed:

- base-a: imbalanced classes
- base-b: imbalanced classes; without StreamCat predictors
- SpCV-b: imbalanced classes; without StreamCat predictors; spatial-cross-validation;
- SMOTE-a: balanced classes
- SMOTE-b: balanced classes; without StreamCat predictors
- SpCV-SMOTE-a: balanced classes; spatial-cross-validation;
- SpCV-SMOTE-b: balanced classes; without StreamCat predictors; spatial-cross-validation;

In a base run, the class imbalance was left untouched. In SMOTE runs, the class imbalance was dealt with using the k-NN-based SMOTE algorithm which creates duplicate observations with a non-parametrized random noise. In SpCV runs, a spatial cross-validation was performed.

2.1.3 Estimating the Stability of the Spatial Predictions of Channel Types

The statistical learning selects either one or multiple classifiers as the most accurate. Before assessing the stabilitiy of their predictions, the posterior probabilities of each classifier, that is the probability of an observation to be of a given class, are calibrated. This calibration improves classifiers performance (Zadrozny 2002; Niculescu-Mizil and Caruana 2005). In that regards, few approaches are common practice: acknowledging the sigmoid-shape of most reliability plots, Platt and others (1999) proposed a sigmoid calibration to correct for this effect. Other useful approaches include Bayesian calibration and isotonic scaling (Zadrozny and Elkan 2002). Hereafter, we present results from classifiers which posterior calibration was performed using a multinomial regression. Such approach is a straightforward extension of the binomial case corresponding to the logistic Platt's scaling (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.

Calibrated classifiers provide more accurate predictions but the stability of their predictions needs to be assessed. Importantly, estimating spatial stability is built upon the premise that channel types are organized hierarchically within the landscape. In other words, it is expected that combination of channel types are restricted to certain specific areas.

Such a stability assessment was performed using spatial statistics. First, all stream segments were converted to points using their midpoint as reference. Then, 10 rasters of the relative risk were

computed. Relative risks correspond here to the spatially varying estimate of the probability of each of the 10 classes to occur. As this is a non-parametric estimate using kernel smoothing, one key point is the selection of the bandwidth with which the relative risk is computed. The bandwidth value determines the amount of smoothing introduced by the Gaussian kernel: large bandwidth values correspond to a high degree of smoothing and vice-versa.

There is no general rule for selecting the appropriate bandwidth so bandwidth was selected as the mean value between: (i) the median length of the stream segments as defined in the NHDPlusV2; (ii) the mean length of the stream segments as defined in the NHDPlusV2; (iii) the inflexion point of the Ripley's K and Besag's L function; (iv) the inflexion point of the pair correlation function $g(r)$. The Ripley's K, Besag's L and the pair correlation $g(r)$ functions characterize a point pattern (e.g. Illian et al., 2008). In particular, they identify how the clustering evolves with increasing spatial scale. An inflexion point marks the scale at which the clustering starts to slow down with increasing scale and was estimated using a segmented linear regression. After the kernel computation at the right bandwidth, the domain outside from the concave boundaries of the set of each class points was masked from the relative risk rasters. This constrains the probability estimates to the spatial domain within which each class was observed.

Evenness, richness and entropy raster were then computed (Thoms et al. 2018). Richness is the number of classes with significant occurrence of probability p_i at a given location. A significant probability was defined as being higher than the no-information rate of the problem (i.e. $1/10 = .1$). Evenness or Simpson's evenness index is defined as $E = \frac{D}{D_{max}}$ with $D = \frac{1}{\sum {p_i}^2}$. Shannon-Weiner's entropy or diversity is defined as $H' = -\sum p_i ln\, p_i$. In statistical mechanics, entropy corresponds to the notion of disorder with higher entropy linked to a higher number of possible state for a given particle in a gas (e.g. Gibbs 1902). These metrics were used by Thoms, Scown, and Flotemersch (2018) to characterize the diversity of physical typology of river networks and defined functional process zones that is areas with similar hydro-geomorphic characteristics. Interestingly, the concept of entropy also connects back to earlier studies from Leopold and Langbein (1962); Scheidegger (1964, 1967, 1968a,b) linking statistical mechanics and geomorphology. In our case, the entropy of the predictions was used to select the classifier that produced the most stable predictions that is the predictions with the lowest entropy.

2.1.4 Assessing the Geomorphic Relevance of the Predictions

The most accurate and stable classifier was selected to perform the final classification. The resulting map of predicted channel types was investigated using expert-knowledge, with a focus on the general spatial organization of channel types across the Sacramento Basin landscapes as well as their geomorphic relevance. Aerial imagery was used to confirm predictions at selected sample locations. In addition, for each stream segment, the richness, evenness and entropy of the posterior probabilities were calculated. A map of the entropy of each stream segment was then produced to estimate the stability of the predictions at the stream segment scale. Finally, the variable importance of predictors was investigated to provide a gray box model rather than a non-interpretable black box.

2.2 Results

In this section, the results from our four-fold machine learning framework are presented. In particular, we describe in the following how: (i) removing coarse-scale contextual predictors leads to a simpler classification problem; (ii) Support Vector Machine and Random Forest outperform other classifiers (iii) spatial statistics highlight the higher stability of Random Forest predictions; and (iv) Random Forest predictions capture the large-scale organization of the landscape.

2.2.1 Removing Coarse-scale Contextual Predictors Leads to a Simpler Classification Problem

Data Complexity Measures (DCM) were computed for each pair of classes and for a set of DCM categories (Fig. 15). This detailed account of complexity informs that the problem should be tractable. In particular, the overall low value of the linearity measures indicates that linear class boundaries should be found. Conversely, neighborhood based methods (e.g. k-NN) should perform worse than linear classifiers. The overlapping of classes in the predictor space highlights which classes are expected to be harder to separate. Here, the unconfined classes 1 and 10 are shown to be more easily discriminated from other classes. In contrast, the pair of classes that are expected to be the hardest to separate is classes 4 vs 3.

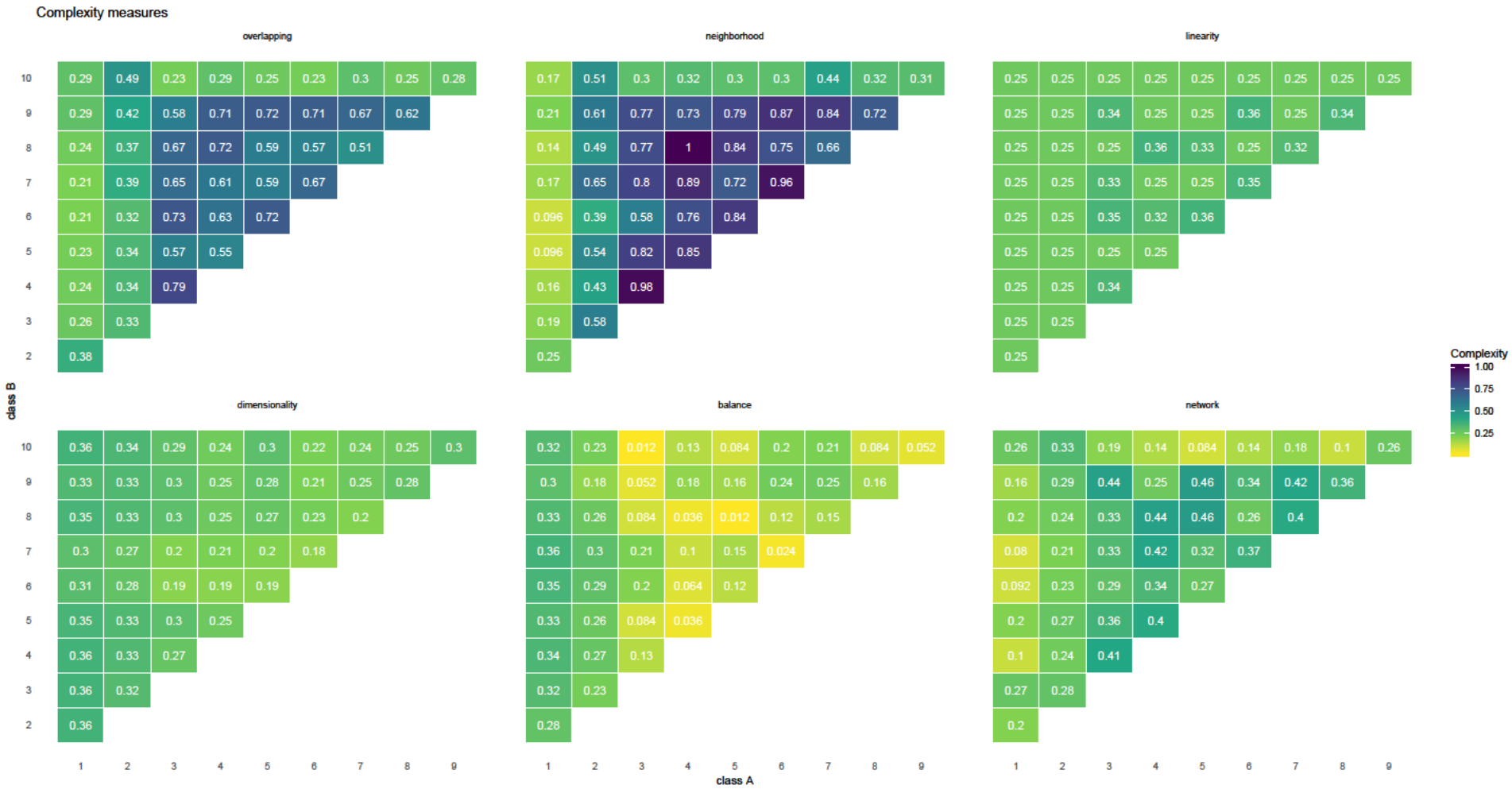


Figure 15. Summmary of complexity measures for each pair of classes. The Data Complexity Measures (DCM) were computed following (Lorena et al., 2018) and were grouped by the following categories: overlapping, neighborhood, linearity, dimensionality, balance and network DCMs. As each categories pertain to a different number of DCM, the score for each category and class pairs is a normalized sum with signed unitary weights to account for complexity increase or decrease with each specific DCM. For each pair of classes, values for between 0 and 1, with 0 indicating a simple problem and 1 a complex one.

The specific values of the DCM were compared to the set of rules from Luengo and Herrera (2013). In that regard, both SVM-based and tree-based classifiers were moderately well-behaved and torelably ill-behaved, meeting 2/9 and 4/11 rules, respectively. Furthermore, removing the StreamCat predictors decreased the overall complexity of the problem. The problem has then better dimensionality, features with a better discriminative power and the same performance in terms of linear separation between classes but higher degree of non-linearity. Removing the contextual variables also led to an increase in the neighborhood DCMs indicating a slightly more complex decision boundary. Nonetheless, network DCMs showed that data points are more clustered and with more hubs in the predictor space. In contrast, removing stream order or fractal dimension predictors increased the complexity of the classification problem across all DCMs. In particular, removing the fractal dimension predictors worsened the network DCMs indicating a more disconnected and less clustered network of data points in the predictor space. Nonetheless, removing the fractal dimension predictions makes some of the predictors more significantly more prevalent at separating classes.

The removal of the coarse contextual variable does not impede overall run accuracy (Table 4). Balancing the classes with SMOTE significantly improves the accuracy of the predictions. Conversely, using the spatial cross-validation (SpCV) leads to decreased accuracy for all runs. Therefore, results are described from the run using spatial cross-validation, SMOTE and without coarse contextual predictors (SpCV-SMOTE-b, Table 4).

Table 4. Results from the different runs performed. The maximum value for the median cross-validation is reported for each run as well as the best classifier(s). If more than one classifier is reported, this means that the distributions of the cross-validation of these classifiers are indistinguishable from a statistical point of view.

run	base-a	base-b	SpCV-b	SMOTE-a	SMOTE-b	SpCV-SMOTE-a	SpCV-SMOTE-b†
max median x-val acc	0.39	0.38	0.34	0.7	0.7	0.6	0.61
best classifier(s)	MARS FDA	MARS FDA RF	MARS FDA CART (RF*)	R-SVM	R-SVM	L-SVM R-SVM	L-SVM RF R-SVM

a: with StreamCat predictors

b: without StreamCat predictors

base: imbalanced classes

SMOTE: classes balanced with SMOTE algorithm

SpCV: Spatial Cross Validation

†: run selected

*: RF achieved a high accuracy but was not statistically similar to MARS

2.2.2 Support Vector Machine and Random Forest Outperform other Classifiers

From the statistical learning step of our four-fold framework, three classifiers emerge as the most accurate: Linear SVM, Radial SVM and Random Forest (Fig. 16). Linear SVM had the lowest computational cost (<1 minute), radial SVM and RF were the most costly (>2 hours). Since linear SVM is a simpler model than radial, only L-SVM results were considered and we compared RF and L-SVM spatial predictions.

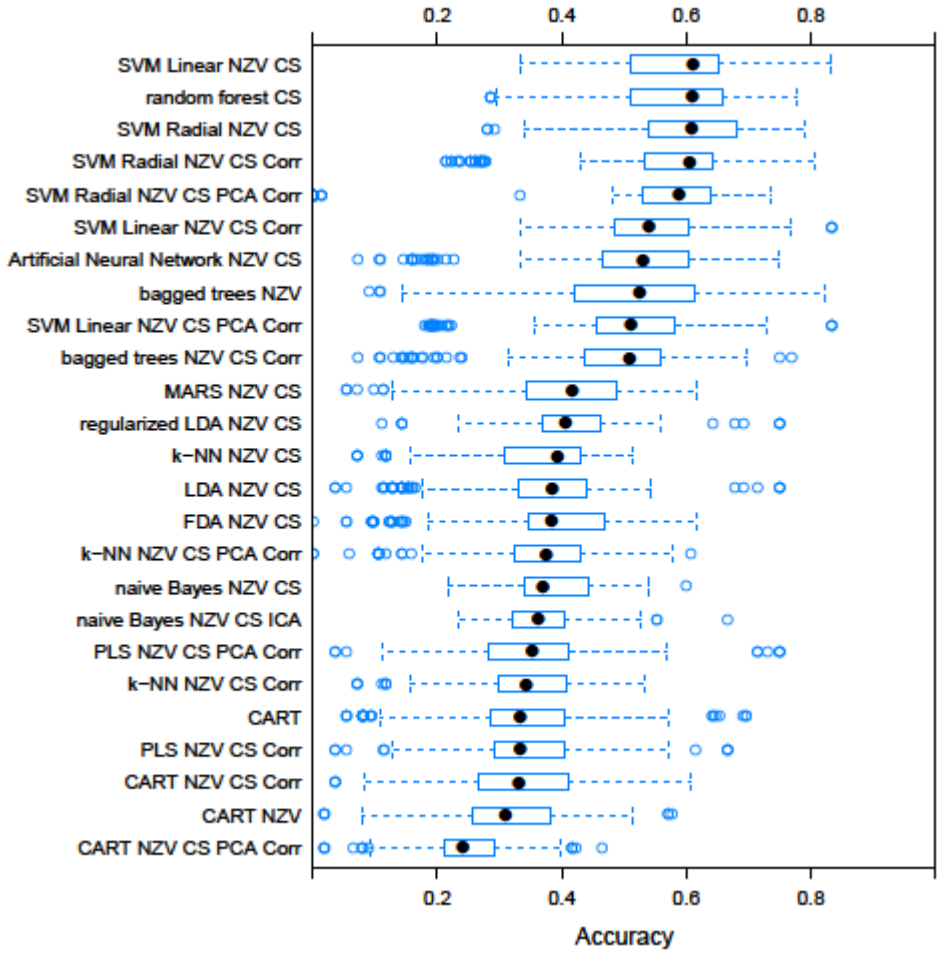


Figure 16. Distributions of cross-validation accuracies for all classifiers for the SpCV-SMOTE-b run (Table 3). Corr indicates that Box-Cox transformations were applied when required; NZV means that the predictors with Near-Zero-Variance were filtered; ICA and PCA mean that Independant Component Analysis and Principal Component Analysis were performed on the predictors, respectively.

2.2.3 Spatial Statistics Highlight the Higher Stability of Random Forest Predictions

Posterior probability scaling was performed to improve the predictive power of L-SVM and RF model. The results of that process are shown in Fig. 17 for RF; similar results were obtained for L-SVM. Calibration pushes the probability towards the diagonal so that the observed probabilities more closely match the true classes. In addition, this highlights where the RF model seems highly confident. Classes with posterior probabilities further from 1 are harder to discriminate (Fig. 15).

Spatial statistics were used to assess the prediction stability using RF and LSVM classifiers. These methods rely on a characteristic length-scale or bandwidth computed as the average of four measures: (i) the median length of the stream segments from NHDv2Plus, 967.2 m; (ii) the mean length of the stream segments from NHDv2Plus, 1464 m; (iii) the inflexion point of the Besag's L function, 1163 ± 16.04 m; (iv) the inflexion point of the pair correlation function $g(r)$, 653.4 ± 8.104 m. Averaging yielded a bandwidth of $1062 \simeq 1000$ m to derive the relative risk rasters. Maps also computed the bandwidth derived from Stoyan's rule of thumb of 537.7 m. Nonetheless, these maps showed very little difference with the maps produced according our procedure for selecting the appropriate bandwidth.

Comparison of the spatial statistics of the SVM and RF predictions shows similar spatial patterns in the probability of occurrence of each of the 10 classes (Fig. 18). Two notable exceptions are the distributions of hotspots for class 5 and 10. In addition, RF predictions achieve higher probability as shown by the generally darker and bluer shades (Fig. 18). These higher probabilities translate into an overall entropy over the prediction domain that is lower for RF than for L-SVM (Fig. 19). Hence, the predictions from the RF classifier are more stable than the ones from the L-SVM classifier. This difference might be explained by the evidenced performance of RF for classification and spatial extrapolation. In addition, RF hyper-parameter m_{try} best value is 17; a value low enough so that the trees of the forest are not significantly correlated making the ensemble process more stable (Probst et al., 2018). In contrast, the best value of the cost parameter of the LSVM is $C = 0.25 < 1$ meaning that the accuracy of the LSVM comes at the cost of some under-fitting translating into poorer predictions for data unseen in training as the class boundaries may not be well constrained.

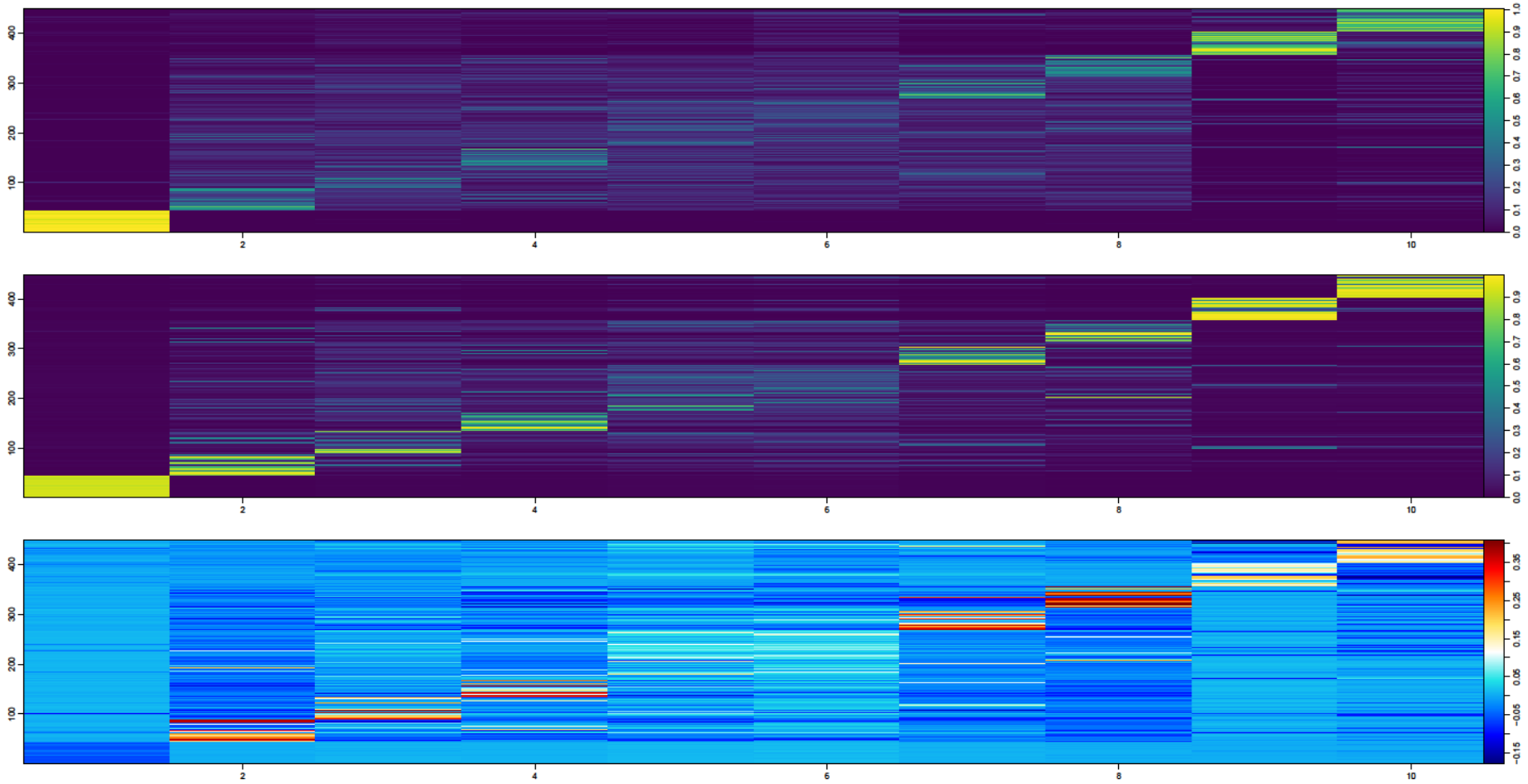


Figure 17. Results of the posterior probability scaling for Random Forest. This figure is analogous to a reliability plot, with the vertical axis corresponding to training observations. These points are sorted by classes so that points that have the label 1 are at the bottom of the vertical axis whereas points with class 10 are at the top. a-b) Each row displays the posterior probabilities that the observation belongs to each of the 10 classes (horizontal axis). An ideal model would yield a yellow diagonal line showing a very high probability of belonging to the true class of the observation. a) corresponds to probabilities before calibration, b) is the output from the calibration and c) represents the difference between a) and b).

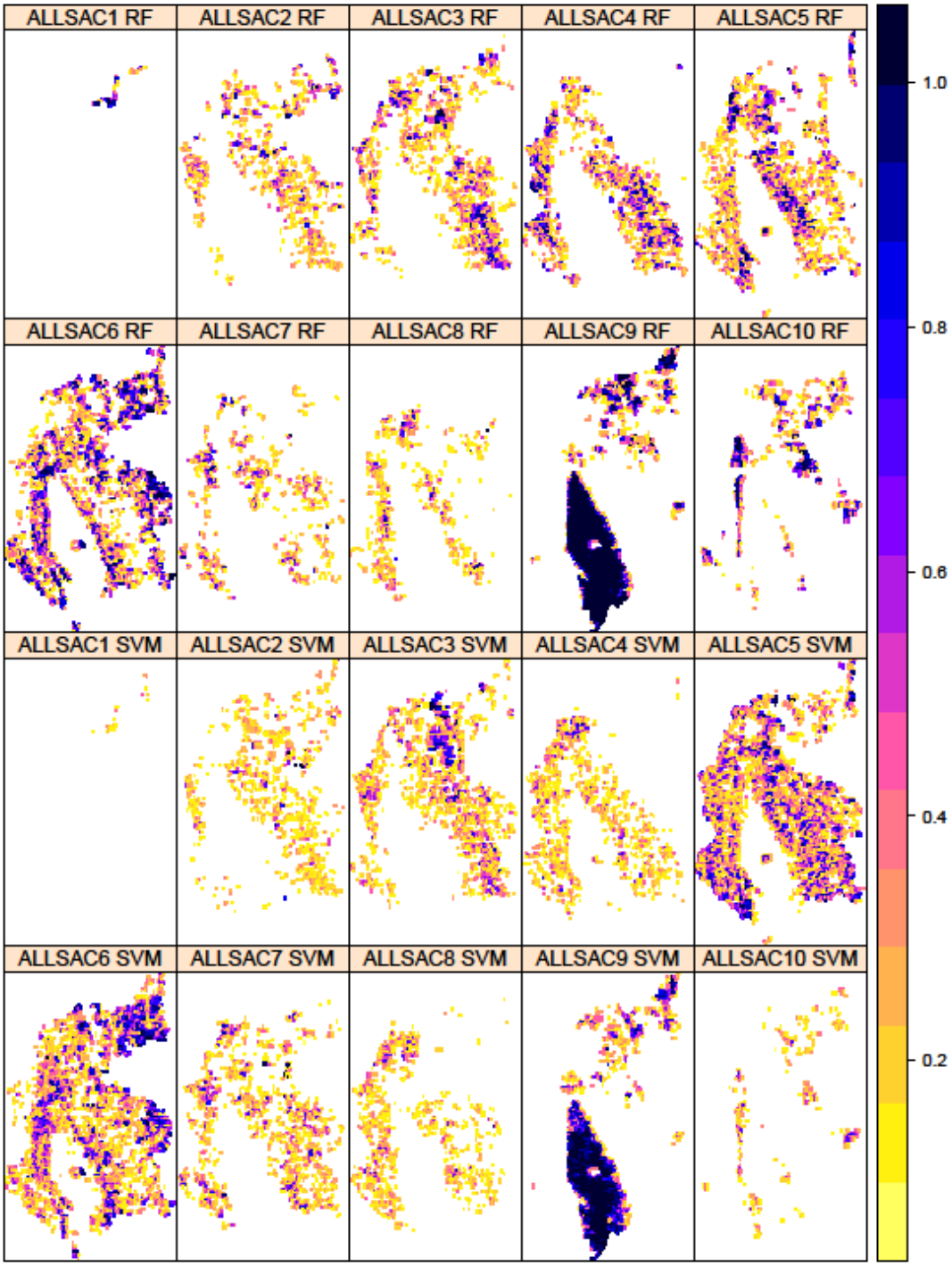


Figure 18. Relative risk maps for the 10 hydro-geomorphic classes identified in the Sacramento Basin. First and second row pertain to the RF model predictions; third and fourth rows

correspond to the predictions from L-SVM. The appropriate bandwidth of the smoothing kernel was determined to be ~1000 m.

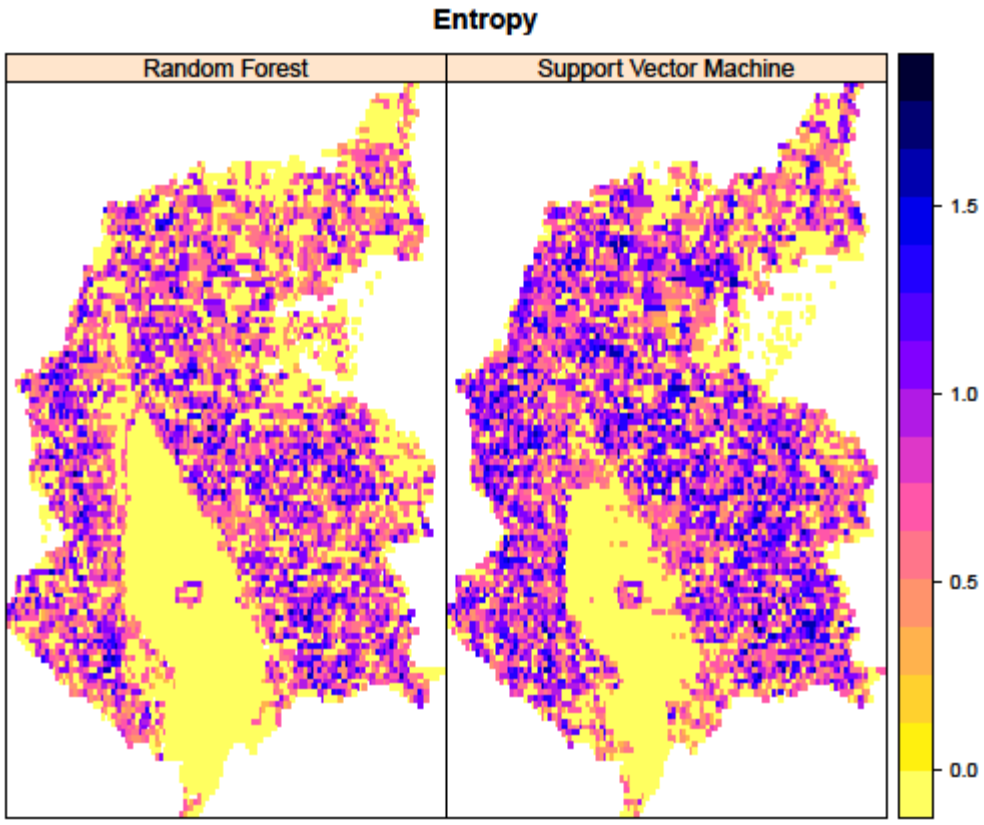


Figure 19. Map of the stability of predictions at the basin scale from the Shannon-Weiner's entropy of the RF and L-SVM classifiers. The entropy value is computed from the relative risks maps (Fig. 18). A low entropy indicates the number of significantly probable channel types is lower and thus that the classifier prediction are stable. A higher entropy means that the number of significantly probable channel types is higher and thus that the classifier prediction are more unstable.

2.2.4 Random Forest Predictions Capture Large-scale Landscape Organization

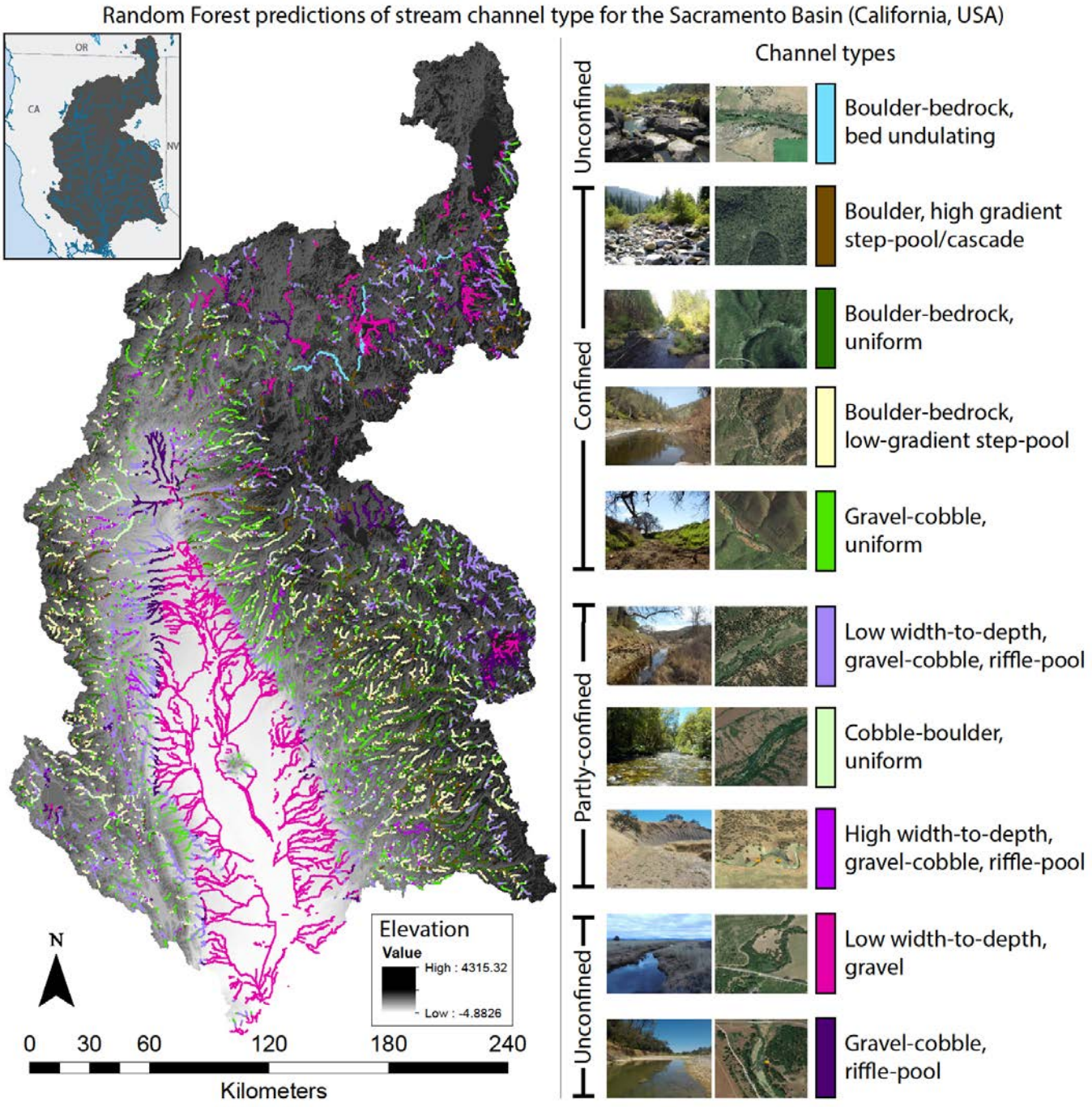
The geomorphic relevance of the predictions from the RF classifier was investigated. An overview of RF predictions of channel type (Fig. 20) shows that the large-scale organization of the landscape with channel types occurring where expected e.g. unconfined meandering stream in the Central

Valley, boulder heavy step-pools in the mountainous areas (see also Section 1.3.1). Additional investigations of the predictions were done in combination with aerial imagery and showed a general good agreement. However, in some areas the predictions did not correspond to what an expert would predict with the *caveat* that vegetation often impedes a highly confident expert judgment. Hence, at a smaller scale, predictions appear noisier with a level of misclassification that could be expected both from the pairwise DCMs analysis (Fig. 15) and from the median cross-validation accuracy (61%).

To provide a quantitative assessment of the uncertainty of the predictions, the entropy from the posterior probabilities associated to each stream segments was derived (Fig. 21). The predictions from the RF appear highly stable in the Central Valley, with more uncertainties in the mountainous area and a high level of instability in the Modoc Plateau. Interestingly, the entropy has a non-linear relationship with elevation throughout the Sierra Nevada towards the Central Valley as highlighted by the profile through the Feather River catchment (Fig. 22). In particular, this highlights the Sierra foothills as an area where the number of significantly probable channel types increases markedly. This pivotal area corresponds to the boundary between the orogen of the Sierra Nevada and the foreland basin of the Central Valley. Significant concentration both in discharge and sediment load is expected to occur at that location. The combination of these factors coupled with the remaining potential energy from the elevation and with the shift in confinement explains the higher number of channel types statistically probable for a stream segment.

The variable importance plot for the RF classifier clarifies in part the black box nature of some machine learning approaches (Fig. 23). Three variables appear significantly more important than the other predictors: valley confinement, drainage area and stream order. Apart from this 3 variables, channel slope and LDD, the most important predictors are dominated by fractal dimension predictors (Hurst coefficients) underlining their relevance. This finding was already suggested by the DCM analysis and is supported by similar variable importance from the Deep Artificial Neural Network (which underperformed in statistical learning).

Figure 20. Map of the Sacramento Basin with the spatial predictions of the 10 types of channel.



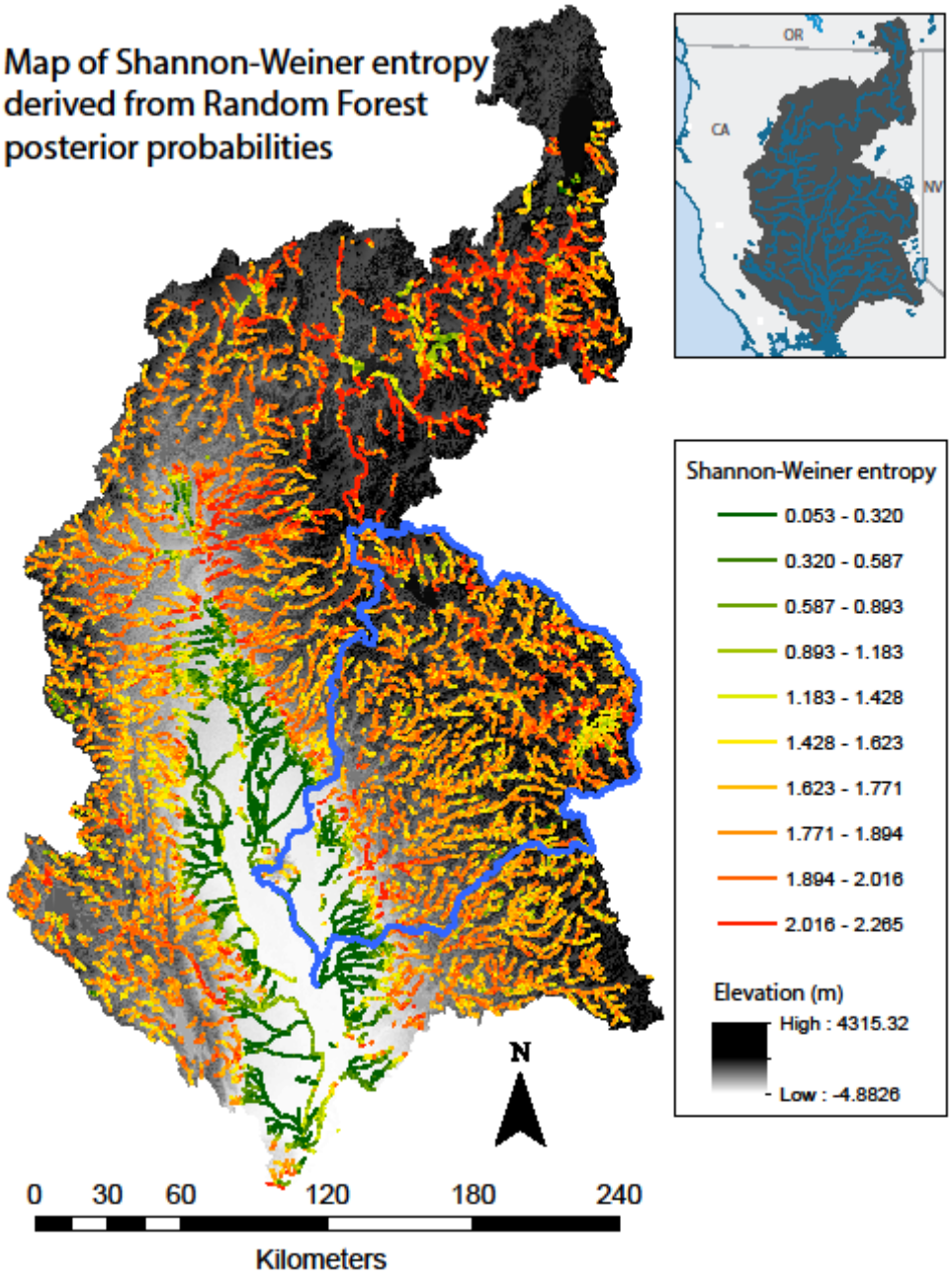


Figure 21. Map of the Sacramento Basin with the entropy of each spatial predictions. The Feather River catchment is highlighted in blue (Fig. 22).

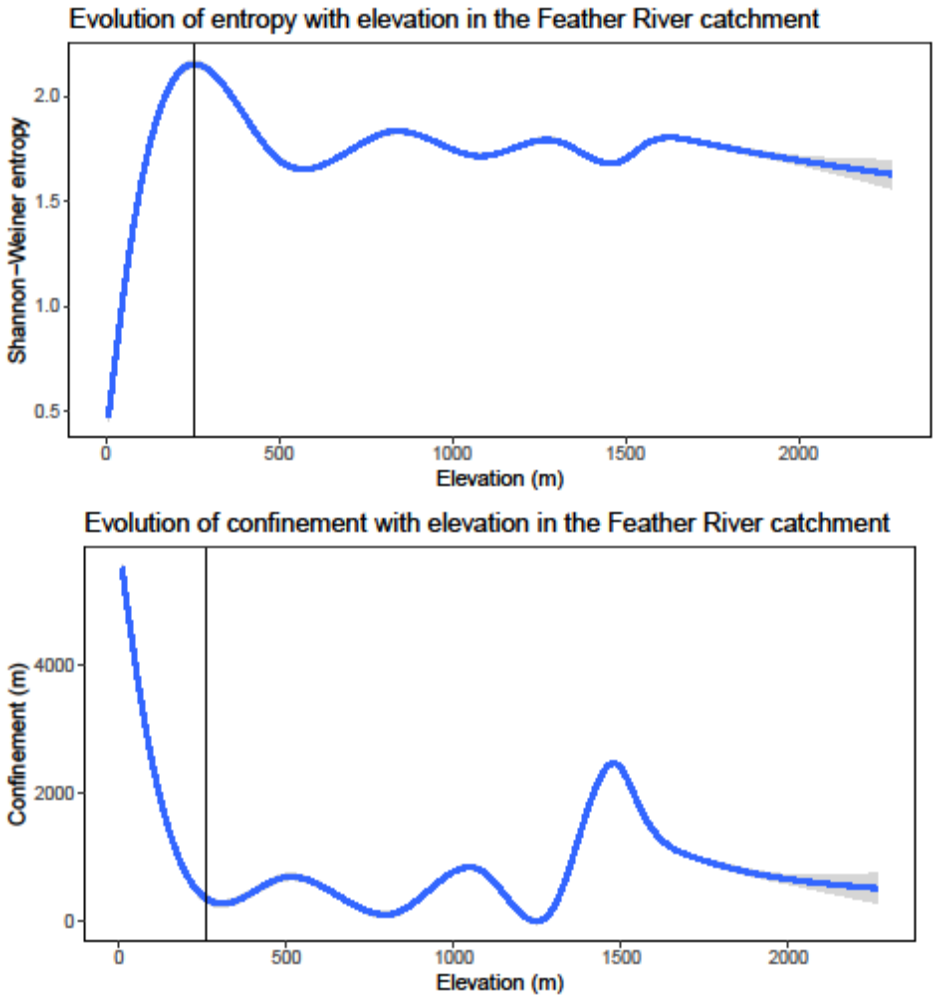


Figure 22. Evolution of entropy and confinement with elevation in the Feather River catchment (Fig. 21).

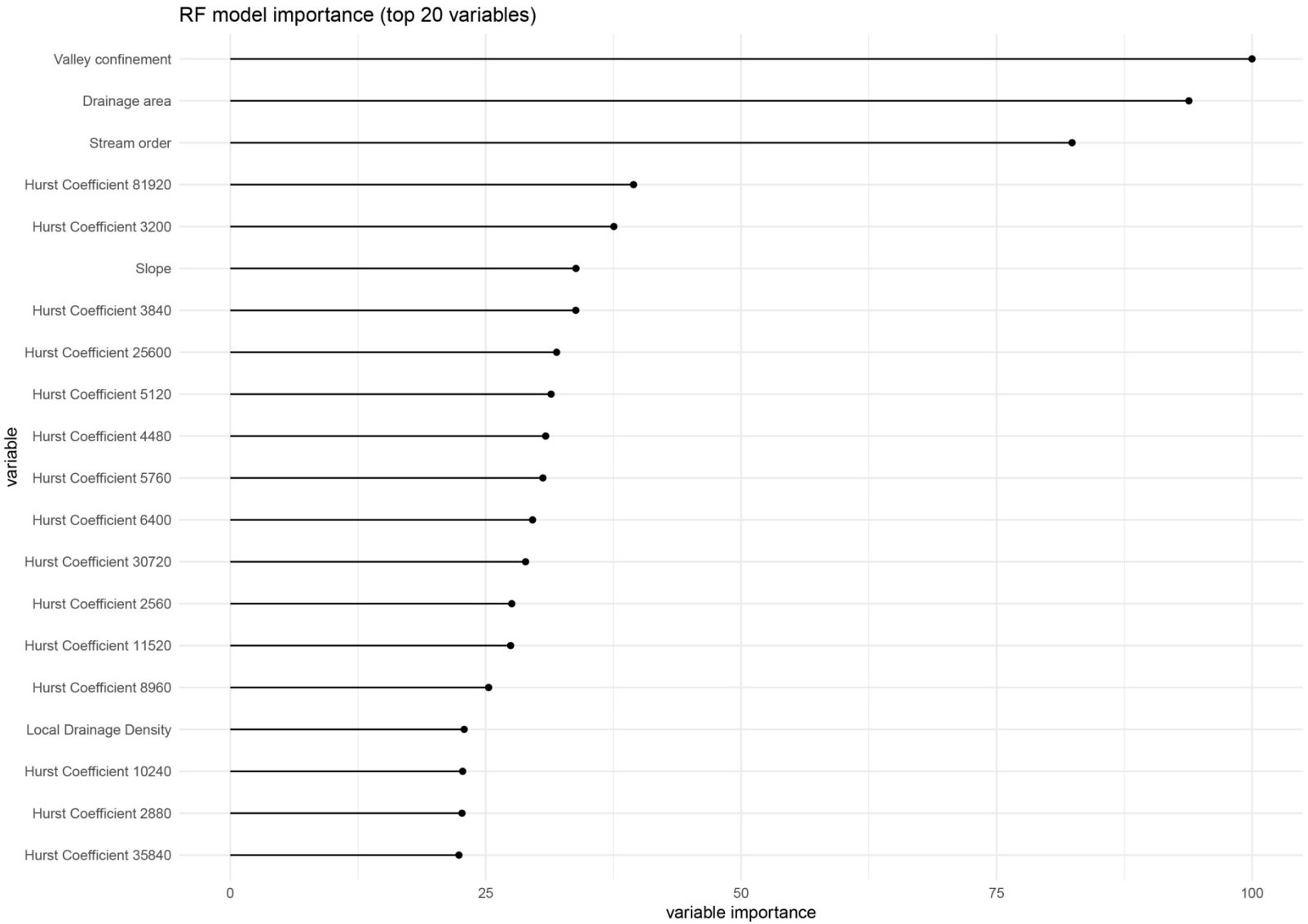


Figure 23. Variable importance of the RF model.

3 References

Agarwal, Ankit, Norbert Marwan, Maheswaran Rathinasamy, Bruno Merz, and Jürgen Kurths. 2017. "Multi-Scale Event Synchronization Analysis for Unravelling Climate Processes: A Wavelet-Based Approach." Nonlinear Processes in Geophysics 24 (4). Copernicus GmbH: 599–611. doi:10.5194/npg-24-599-2017.

Alexander, J.S., R.B. Zelt, and N.J. Schaepe. 2009. "Geomorphic Segmentation, Hydraulic Geometry, and Hydraulic Microhabitats of the Niobrara River, Nebraska—Methods and Initial Results." U.S. Geological Survey Scientific Investigations Report.

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26: 32–46. DOI: 10.1111/j.1442-9993.2001.01070.pp.x

Archfield, Stacey A., Martyn Clark, Berit Arheimer, Lauren E. Hay, Hilary McMillan, Julie E. Kiang, Jan Seibert, et al., 2015. "Accelerating Advances in Continental Domain Hydrologic Modeling." Water Resources Research 51 (12). American Geophysical Union (AGU): 10078–91. doi:10.1002/2015wr017498.

Beechie T, Imaki H. 2014. Predicting natural channel patterns based on landscape and geomorphic controls in the Columbia River basin, USA. Water Resources Research 50: 39–57. DOI: 10.1002/2013WR013629

Bisson PA, Montgomery DR, Buffington JM. 1996. Valley segments, stream reaches, and channel units. Methods in stream ecology: 23–52.

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.

Borut, Sluban, Gamberger Dragan, and Lavra Nada. 2010. "Advances in Class Noise Detection." Frontiers in Artificial Intelligence and Applications 215 (ECAI 2010). IOS Press: 1105–6. doi:10.3233/978-1-60750-606-5-1105.

Brierley GJ, Fryirs K. 2000. River Styles, a Geomorphic Approach to Catchment Characterization: Implications for River Rehabilitation in Bega Catchment, New South Wales, Australia. Environmental Management 25: 661–679. DOI: 10.1007/s002670010052

Buffington, J.M., and D.R. Montgomery. 2013. "Geomorphic Classification of Rivers." In Treatise on Geomorphology, 730–67. Elsevier. doi:10.1016/b978-0-12-374739-6.00263-3.

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.

Buscombe, Daniel. 2016. "Spatially Explicit Spectral Analysis of Point Clouds and Geospatial Data." Computers & Geosciences 86 (January). Elsevier BV: 92–108. doi:10.1016/j.cageo.2015.10.004.

Carr, James R. 1997. "Statistical Self-Affinity, Fractal Dimension, and Geologic Interpretation." Engineering Geology 48 (3-4). Elsevier: 269–82.

Carson MA. 1972. Hillslope form and process, . University Press: Cambridge [Eng.

Cazenave, Pierre W., Justin K. Dix, David O. Lambkin, and Lisa C. McNeill. 2012. "A Method for Semi-Automated Objective Quantification of Linear Bedforms from Multi-Scale Digital Elevation Models." Earth Surface Processes and Landforms 38 (3). Wiley-Blackwell: 221–36. doi:10.1002/esp.3269.

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 KR. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117–143. DOI: 10.1111/j.1442-9993.1993.tb00438.x

Cress, Jill, David Soller, Roger Sayre, Patrick Comer, and Harumi Warner. 2010. Terrestrial Ecosystems—Surficial Lithology of the Conterminous United States. https://pubs.usgs.gov/sim/3126/.

Csillik, O., I.S. Evans, and L. Drăguţ. 2015. "Transformation (Normalization) of Slope Gradient and Surface Curvatures, Automated for Statistical Analyses from DEMs." Geomorphology 232 (Mar). Elsevier BV: 65–77. doi:10.1016/j.geomorph.2014.12.038.

Danesh-Yazdi, Mohammad, Alejandro Tejedor, and Efi Foufoula-Georgiou. 2017. "Self-Dissimilar Landscapes: Revealing the Signature of Geologic Constraints on Landscape Dissection via Topologic and Multi-Scale Analysis." Geomorphology 295 (October). Elsevier BV: 16–27. doi:10.1016/j.geomorph.2017.06.009.

De'ath G, Fabricius KE. 2000. Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis. Ecology 81: 3178–3192. DOI: 10.1890/0012-

9658(2000)081[3178:CARTAP]2.0.CO;2

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.

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

ESRI. 2016. ArcGIS Desktop . Environmental Systems Research Institute: Redlands, CA

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.

Florinsky, Igor V. 1998. “Accuracy of Local Topographic Variables Derived from Digital Elevation Models.” International Journal of Geographical Information Science 12 (1). Informa UK Limited: 47–62. doi:10.1080/136588198242003.

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.

Fryirs, Kirstie A., Joseph M. Wheaton, and Gary J. Brierley. 2016. “An Approach for Measuring Confinement and Assessing the Influence of Valley Setting on River Forms and Processes.” Earth Surface Processes and Landforms 41 (5). Wiley-Blackwell: 701–10. doi:10.1002/esp.3893.

Garcia, Luis Paulo F., Ana Carolina Lorena, and Andre C.P.L.F. Carvalho. 2012. “A Study on Class Noise Detection and Elimination.” In 2012 Brazilian Symposium on Neural Networks. IEEE. doi:10.1109/sbrn.2012.49.

Garcia, Luís PF, André CPLF de Carvalho, and Ana C Lorena. 2015. “Effect of Label Noise in the Complexity of Classification Problems.” Neurocomputing 160. Elsevier: 108–19.

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.

Gibbs, J. W. 1902. Elementary Principles in Statistical Mechanics. Univ. Press, New Haven, Conn, Yale.

Gilbert JT, Macfarlane WW, Wheaton JM. 2016. The Valley Bottom Extraction Tool (V-BET): A GIS tool for delineating valley bottoms across entire drainage networks. Computers & Geosciences 97: 1–14. DOI: 10.1016/j.cageo.2016.07.014

Grant GE, Swanson FJ, Wolman MG. 1990. Pattern and origin of stepped-bed morphology in high-gradient streams, Western Cascades, Oregon. GSA Bulletin 102: 340–352. DOI: 10.1130/0016-7606(1990)102<0340:PAOOSB>2.3.CO;2

H2O.ai. 2018. R Interface for H2O. https://github.com/h2oai/h2o-3.

Haan, Charles Thomas, Billy J Barfield, and Julie Candler Hayes. 1994. Design Hydrology and Sedimentology for Small Catchments. Elsevier.

Hartigan, J. A., and M. A. Wong. 1979. “A K-means Clustering Algorithm.” Journal of the Royal Statistical Society. Series C (Applied Statistics) 28 (1): 100–108. http://www.jstor.org/stable/2346830?origin=JSTOR-pdf.

Hauer C, Pulg U. 2018. The non-fluvial nature of Western Norwegian rivers and the implications for channel patterns and sediment composition. CATENA 171: 83–98. DOI: 10.1016/j.catena.2018.06.025

Hijmans, Robert J, Jacob van Etten, Joe Cheng, Jonathan A Greenberg, Oscar Perpinan Lamigueiro, Andrew Bevan, and others. 2018. Package “Raster”.

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.

Ho, Tin Kam, and M. Basu. 2002. “Complexity Measures of Supervised Classification Problems.” IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (3). Institute of Electrical; Electronics Engineers (IEEE): 289–300. doi:10.1109/34.990132.

Homer, Collin, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Xian, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. 2015. “Completion of the 2011 National Land Cover Database for the Conterminous United States–representing a Decade of Land Cover Change Information.” Photogrammetric Engineering & Remote Sensing 81 (5). American Society for Photogrammetry; Remote Sensing: 345–54.

Torsten Hothorn, Frank Bretz and Peter Westfall. 2008. Simultaneous Inference in General Parametric Models. Biometrical Journal 50(3), 346--363.

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.

Illian, Janine, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. 2008. Statistical Analysis and Modelling of Spatial Point Patterns. Vol. 70. John Wiley & Sons.

Jha, Rajan, and Panayiotis Diplas. 2017. “Elevation: A Consistent and Physically-Based Framework for Classifying Streams.” Journal of Hydraulic Research 56 (3). Informa UK Limited: 299–312. doi:10.1080/00221686.2017.1354928.

Kruskal JB. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29: 1–27. DOI: 10.1007/BF02289565

Kuhn, Max, and others. 2018. Package “Caret”.

Kuhn, Max. 2008. “Building Predictive Models in R Using the Caret Package.” Journal of Statistical Software, Articles 28 (5): 1–26. doi:10.18637/jss.v028.i05.

Lane BA, Pasternack GB, Dahlke HE, Sandoval-Solis S. 2017b. The role of topographic variability in river channel classification. Progress in Physical Geography: 0309133317718133. DOI: 10.1177/0309133317718133

Lane, B. A., E. Dahlke Helen, G. Pasternack, and S. Sandoval-Solis. 2017. “Revealing the diversity of natural hydrologic regimes in California with relevance for environmental flows applications.” Journal of the American Water Resources Association.

Lane, Belize A., Gregory B. Pasternack, and Samuel Sandoval Solis. 2018. “Integrated Analysis of Flow, Form, and Function for River Management and Design Testing.” Ecohydrology 11 (5). Wiley: e1969. doi:10.1002/eco.1969.

Lane, Belize A., Samuel Sandoval-Solis, Eric D. Stein, Sarah M. Yarnell, Gregory B. Pasternack, and Helen E. Dahlke. 2018. “Beyond Metrics? The Role of Hydrologic Baseline Archetypes in Environmental Water Management.” Environmental Management, June. Springer Nature. doi:10.1007/s00267-018-1077-7.

Leopold, Luna B., and Walter B. Langbein. 1962. “Th Concept of Entropy in Landscape Evolution.” Geological Survey Professional Paper.

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.

Lin, Henry W., Max Tegmark, and David Rolnick. 2017. “Why Does Deep and Cheap Learning Work so Well?” Journal of Statistical Physics 168 (6). Springer Nature: 1223–47. doi:10.1007/s10955-017-1836-5.

Lisenby, Peyton E., and Kirstie A. Fryirs. 2017. “’Out with the Old?’ Why Coarse Spatial Datasets Are Still Useful for Catchment-Scale Investigations of Sediment (Dis)connectivity.” Earth Surface Processes and Landforms 42 (10). Wiley: 1588–96. doi:10.1002/esp.4131.

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 C., Luís P. F. Garcia, Jens Lehmann, Marcilio C. P. Souto, and Tin K. Ho. 2018. “How Complex Is Your Classification Problem? A Survey on Measuring Classification Complexity.”

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.

Mandelbrot, B. 1967. “How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension.” Science 156 (3775). American Association for the Advancement of Science (AAAS): 636–38. doi:10.1126/science.156.3775.636.

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).

Montgomery DR, Buffington JM. 1997. Channel-reach morphology in mountain drainage basins. Geological Society of America Bulletin 109: 596–611.

Montgomery, D. R., and J. M. Buffington. 1998. “River Ecology and Management: Lessons from the Pacific Coastal Ecoregion.” In, edited by R. Naiman and R Bilby, 13–42. New York, Springer-Verlag.

Murtagh F, Legendre P. 2014a. Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? Journal of Classification 31: 274–295. DOI: 10.1007/s00357-014-9161-z

Murtagh F, Legendre P. 2014b. Ward’s Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. Journal of Classification 31: 274–295. DOI: 10.1007/s00357-014-9161-z

Neeson TM, Gorman AM, Whiting PJ, Koonce JF. 2011. Factors Affecting Accuracy of Stream

Channel Slope Estimates Derived from Geographical Information Systems. North American Journal of Fisheries Management 28: 722–732. DOI: 10.1577/M05-127.1

Newman, D.R., J.B. Lindsay, and J.M.H. Cockburn. 2018. “Evaluating Metrics of Local Topographic Position for Multiscale Geomorphometric Analysis.” Geomorphology 312 (July). Elsevier BV: 40–50. doi:10.1016/j.geomorph.2018.04.003.

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 PR. 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

O’Brien, G.O., and J. Wheaton. 2014. “River Styles Report for the Middle Fork John Day Watershed, Oregon.” Ecogeomorphology; Topographic Analysis Lab, Utah State University.

O’Brien, Gary R., Joseph Wheaton, Kirstie Fryirs, Peter McHugh, Nicolaas Bouwes, Gary Brierley, and Chris Jordan. 2017. “A Geomorphic Assessment to Inform Strategic Stream Restoration Planning in the Middle Fork John Day Watershed, Oregon, USA.” Journal of Maps 13 (2). Informa UK Limited: 369–81. doi:10.1080/17445647.2017.1313787.

Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner. 2018. vegan: Community Ecology Package. R package version 2.4-6. 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.

Perry, George L. W., and Mark E. Dickson. 2018. “Using Machine Learning to Predict Geomorphic Disturbance: The Effects of Sample Size, Sample Prevalence, and Sampling Strategy.” Journal of Geophysical Research: Earth Surface 123 (11). American Geophysical Union (AGU): 2954–70. doi:10.1029/2018jf004640.

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.

PRISM Climate Group. 2004. PRISM Gridded Climate Data. Oregon State University. http://prism.oregonstate.edu.

Probst, Philipp, Marvin Wright, and Anne-Laure Boulesteix. 2018. “Hyperparameters and Tuning Strategies for Random Forest.” ArXiv Preprint ArXiv:1804.03515.

Pyne, Matthew I., Daren M. Carlisle, Cristopher P. Konrad, and Eric D. Stein. 2017. “Classification of California Streams Using Combined Deductive and Inductive Approaches: Setting the Foundation for Analysis of Hydrologic Alteration.” Ecohydrology 10 (3). Wiley: e1802. doi:10.1002/eco.1802.

R Core Team. 2017. R: A Language and Environment for Statistical Computing . R Foundation for Statistical Computing: Vienna, Austria [online] Available from: https://www.R-project.org/

Rahmati, Omid, Nasser Tahmasebipour, Ali Haghizadeh, Hamid Reza Pourghasemi, and Bakhtiar Feizizadeh. 2017. “Evaluation of Different Machine Learning Models for Predicting and Mapping the Susceptibility of Gully Erosion.” Geomorphology 298 (December). Elsevier BV: 118–37. doi:10.1016/j.geomorph.2017.09.006.

Renard, Kenneth G, George R Foster, GA Weesies, DK McCool, DC Yoder, and others. 1997. Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE). Vol. 703. United States Department of Agriculture Washington, DC.

Rinaldi, M., B. Belletti, F. Comiti, L. Nardi, M. Bussettini, L. Mao, and A.M. Gurnell. 2015. “The Geomorphic Units Survey and Classification System (GUS), Deliverable 6.2, Part 4, of REFORM (REstoring Rivers fOR Effective Catchment Management), a Collaborative Project (Large-Scale Integrating Project) Funded by the European Commission Within the 7 Th Framework Programme Under Grant Agreement 282656.”

Rosgen DL. 1996. Applied river morphology . Wildland Hydrology

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.

Scheidegger, A. 1964. “Some Implications of Statistical Mechanics in Geomorphology.” International Association of Scientific Hydrology. Bulletin 9 (1). Informa UK Limited: 12–16. doi:10.1080/02626666409493650.

Scheidegger, A.E. 1968. “Horton’s Law of Stream Order Numbers and a Temperature-Analog in River Nets.” Water Resources Research.

Scheidegger. 1967. “A Complete Thermodynamic Analogy for Landscape Evolution.” International Association of Scientific Hydrology. Bulletin 12 (4). Informa UK Limited: 57–62. doi:10.1080/02626666709493550.

Scheidegger. 1968. “Microcanonical Ensembles of River Nets.” International Association of Scientific Hydrology. Bulletin 13 (4). Informa UK Limited: 87–90. doi:10.1080/02626666809493629.

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.

Schwarz, Gregory E, and RB Alexander. 1995. “State Soil Geographic (STATSGO) Data Base for the Conterminous United States.”

Shen, Chaopeng, Eric Laloy, Amin Elshorbagy, Adrian Albert, Jerad Bales, Fi-John Chang, Sangram Ganguly, et al., 2018. “HESS Opinions: Incubating Deep-Learning-Powered Hydrologic Science Advances as a Community.” Hydrology and Earth System Sciences 22 (11). Copernicus GmbH: 5639–56. doi:10.5194/hess-22-5639-2018.

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.

Sluban, Borut, Dragan Gamberger, and Nada Lavrač. 2013. “Ensemble-Based Noise Detection: Noise Ranking and Visual Performance Evaluation.” Data Mining and Knowledge Discovery 28 (2). Springer Nature: 265–303. doi:10.1007/s10618-012-0299-1.

Smith, Mark W. 2014. “Roughness in the Earth Sciences.” Earth-Science Reviews 136 (September). Elsevier BV: 202–25. doi:10.1016/j.earscirev.2014.05.016.

Snow RS. 1989. Fractal sinuosity of stream channels. Pure and applied geophysics 131: 99–109.

Strahler, Arthur N. 1957. “Quantitative Analysis of Watershed Geomorphology.” Transactions, American Geophysical Union 38 (6). American Geophysical Union (AGU): 913. doi:10.1029/tr038i006p00913.

Sung, Q.-C., Y.-C. Chen, and P.C. Chao. 1998. “Spatial Variation of Fractal Parameters and Its Geological Implications.” Terrestrial, Atmospheric and Oceanic Sciences Journal 9 (4): 655–72.

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.

Teutschbein, Claudia, Thomas Grabs, Hjalmar Laudon, Reinert H. Karlsen, and Kevin Bishop. 2018. “Simulating Streamflow in Ungauged Basins Under a Changing Climate: The Importance of Landscape Characteristics.” Journal of Hydrology 561 (June). Elsevier BV: 160–78. doi:10.1016/j.jhydrol.2018.03.060.

Therneau, Terry and Beth Atkinson. 2018. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-13. https://CRAN.R-project.org/package=rpart

Thoms, Martin, Murray Scown, and Joseph Flotemersch. 2018. “Characterization of River Networks: A GIS Approach and Its Applications.” JAWRA Journal of the American Water Resources Association, April. Wiley. doi:10.1111/1752-1688.12649.

Thornbrugh, Darren J., Scott G. Leibowitz, Ryan A. Hill, Marc H. Weber, Zachary C. Johnson, Anthony R. Olsen, Joseph E. Flotemersch, John L. Stoddard, and David V. Peck. 2018. “Mapping Watershed Integrity for the Conterminous United States.” Ecological Indicators 85 (February). Elsevier BV: 1133–48. doi:10.1016/j.ecolind.2017.10.070.

Tukey JW. 1991. The Philosophy of Multiple Comparisons. Statistical Science 6: 100–116.

Ward JHJ. 1963. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58: 236–244. 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.

Wohl EE, Vincent KR, Merritts DJ. 1993. Pool and riffle characteristics in relation to channel gradient. Geomorphology 6: 99–110. DOI: 10.1016/0169-555X(93)90041-Y

Wolman MG. 1954. A method of sampling coarse river-bed material. Eos, Transactions American Geophysical Union 35: 951–956. DOI: 10.1029/TR035i006p00951

Xu, Tingbao, Ian D. Moore, and John C. Gallant. 1993. “Fractals, Fractal Dimensions and Landscapes — a Review.” Geomorphology 8 (4): 245–62. doi:https://doi.org/10.1016/0169-555X(93)90022-T.

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.

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