South Fork Eel River Basin Geomorphic Classification




February 2019

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

Prepared for: California State Water Resources Control Board

Utah State University

UC DAVIS

Executive Summary

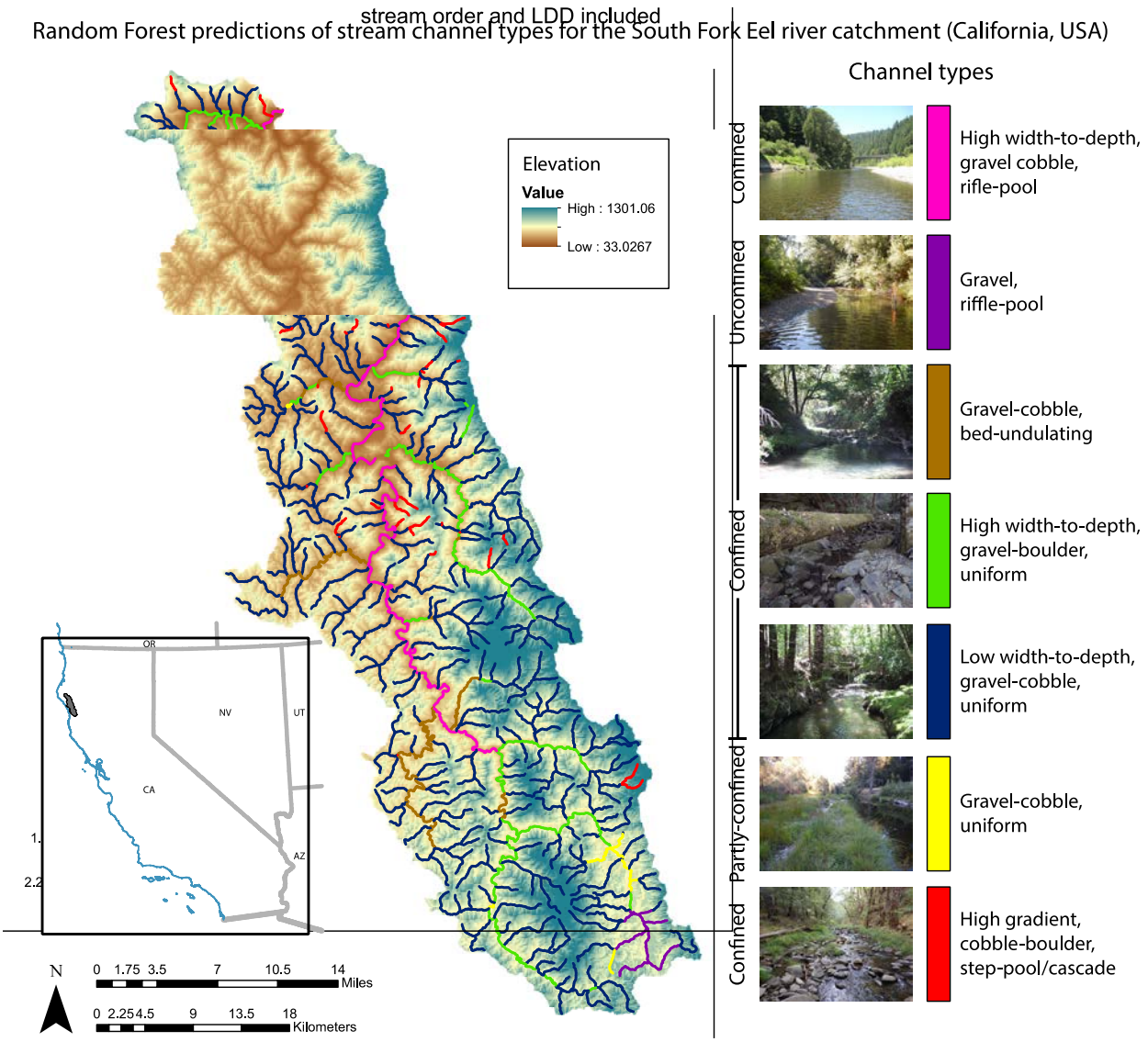
Hervé Guillon, Colin F. Byrne, Belize A. Lane, Gregory B. Pasternack and Samuel Sandoval-Solis

The objective of this research study was to determine a channel reach morphology classification and the spatial distribution of each geomorphic class throughout the South Fork of the Eel River (SFER) catchment. 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 97 field-surveyed streams. In the first year of the data collection campaign, a total of 54 field sites were surveyed based on their transport capacity (combination of slope and contributing area). In the second year of the field campaign, a total of 43 field sites were surveyed based on sediment supply characteristics (19 field sites) in addition to the transport capacity criteria used in the first year; and in river locations where ecological and habitat data has been collected historically (24 field sites). Surveyed streams in the SFER catchment were analyzed using multivariate statistical techniques to identify groups of reaches with similar stream forms.

This study focusses on the geomorphic classification for the SFER catchment. Seven channel types were identified in this region (Figure ES-1). Bankfull width-to-depth ratio and valley confinement were the most influential channel attributes in the clustering of channel types. The SFER catchment is dominated by confined settings throughout the region compared to other regions of California, and in general, the classification progresses from larger to smaller channel dimensions. As a result, five channel types occurred within confined settings and represent a wide variety of forms from headwater streams to the mainstem. Beginning in headwater streams, *gravel-cobble, low width-to-depth, uniform* and *high-gradient, cobble-boulder, step-pool/cascade* are channel types differentiated by slope and sediment grain size. *Gravel-cobble, low width-to-depth, uniform* streams are defined by low depth and width variability. *High-gradient, cobble-boulder, step-pool/cascade* streams are the steepest channels within the SFER catchment. Intermediate size, confined streams within the catchment are best represented by *gravel-cobble, bed-undulating* and *gravel-cobble, high width-to-depth ratio, uniform* channel types. The largest channel type occurs along the lower mainstem of the SFER. The *gravel-cobble, high width-to-depth ratio, riffle-pool* channel type is defined by a meandering, riffle-pool structure within a confined valley setting where valley walls may still influence channel form in addition to transverse, meandering flow patterns. The remaining two channel types were identified as *unconfined, gravel, riffle-pool*, a channel type that exists in wide valleys and its form is independent of hillslope processes predominantly located in the southeast portion of the SFER catchment, and *partly-confined, gravel-cobble, uniform* streams that exhibit low width-to-depth ratios but exist within a valley bottom that is wide enough to allow for floodplain development. These seven 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 SFER catchment. The 97 field sites already classified are incorporated as labels into a large-scale supervised learning model. This model uses 287 coarse-scale features describing topography,

geology, soils, climate and land use. Figure ES-1 shows the spatial prediction of the seven channel types classified for the SFER catchment using a random forest (RF) model. Five variables are the main predictors for the RF model: drainage area, valley confinement, stream order, channel slope, and fractal dimension predictors (Hurst coefficients). The spatial significance of the predictions from the best models was assessed using expert-knowledge, with a focus on the general spatial organization of channel types across the SFER catchment as well as their geomorphic relevance. Aerial imagery was used to confirm predictions at selected sample locations. In addition, this study also identified the uncertainty of a given reach to be classified correctly by using the Shannon-Weiner entropy index, that identifies at which location the prediction is more stable or not.



ES-1. Geomorphic classification and spatial distribution of channel types in the South Fork Eel River catchment.

Contents

1 Classification of Channel-Reach Morphology

1.1 Objectives

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

1.2 Methodology

Channel types were developed based on measured channel attributes, including: 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), median grain size (D50), 84th percentile grain size (D84), channel roughness (bf.d.D50), and valley confinement distance (vc.dist). The SFER classification was determined from 97 surveys, all of which were conducted by Humboldt State University (HSU) personnel in summer of 2017 and 2018.

1.2.1 Site Selection

Field site locations for 2017 SFER catchment surveys were determined using a random stratified sampling scheme based on slope and contributing area as documented in Lane et al. (2017b). Slope bins were defined as <1%, 1-2%, 2-4%, 4-8%, and >8%. Contributing area bins were defined as <25, 25-200, 200-1,000, and >1,000 km^2. The slope-area sampling protocol was designed to capture stream reaches of differing transport capacity and thus expected differing channel forms. Because some bins were likely to contain more streams than others, an equal number of sites were surveyed in each slope-area bin. This places a focus on all channel types, even rare forms, instead of only the most dominant channel types. A total of 54 sites surveyed in 2017 were used in classification of SFER catchment streams (Fig. 1, green triangles).

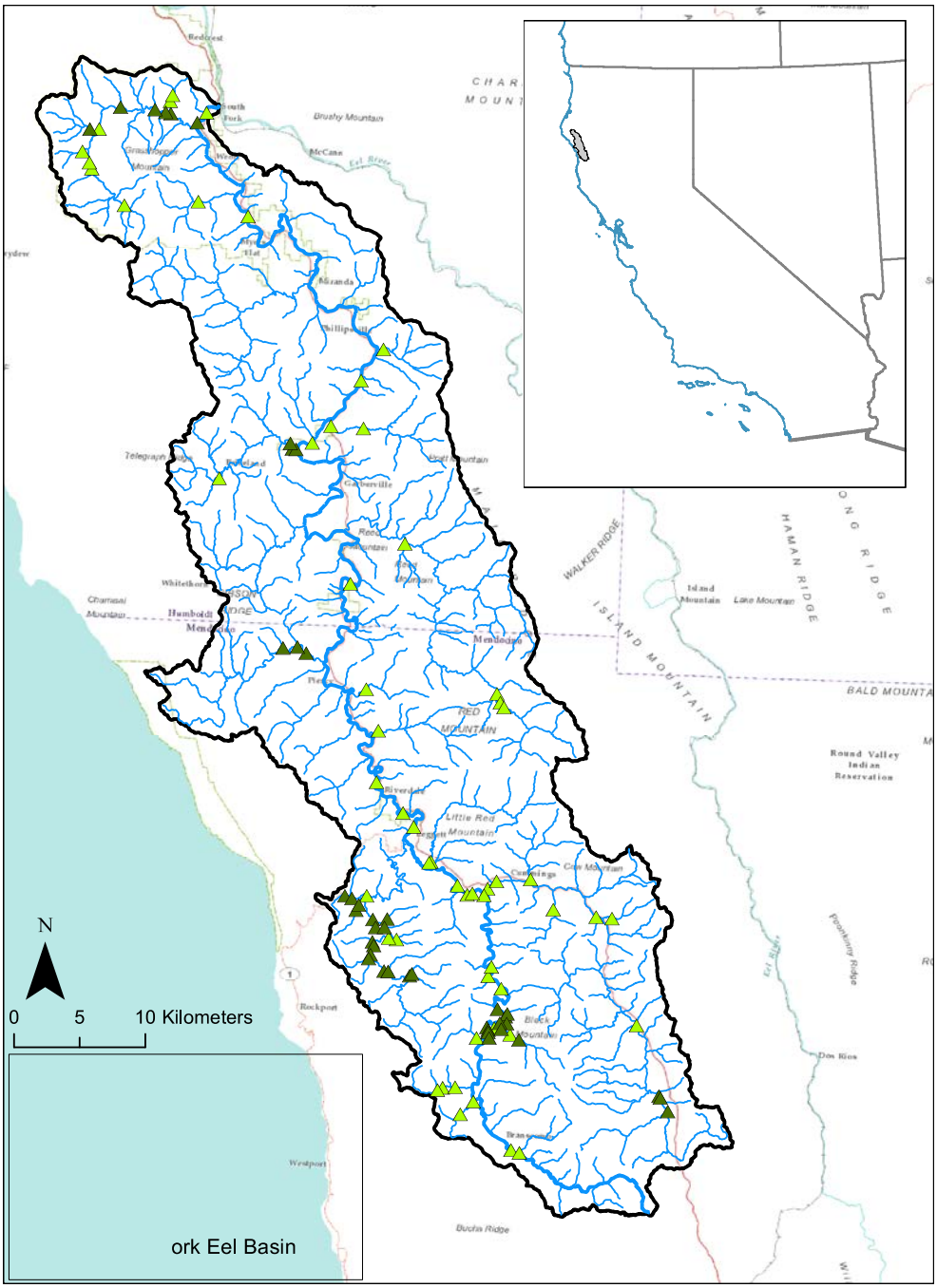


Figure 1. Site locations of 97 stream surveys used in the multivariate geomorphic classification.

The 2018 binning of sites was expanded to account for estimated sediment supply in addition to transport capacity (Appendix A). Lateral valley confinement and hillslope sediment supply derived from the Revised Universal Soil Loss Equation (RUSLE) were used as indicators of sediment supply. Streams were first classified by sediment supply (confinement and RUSLE) and second by transport capacity (area and slope) (Fig. 2). Streams surveyed in 2017 were reassessed using the two-tiered binning approach. New bins that were under-surveyed in 2017 were identified for

2018 site surveys. These 19 sites are referred to as *data-gap sites* and were predominantly located in unconfined and low sediment supply locations. Upon identification of data-gap sites, the remainder of sites focused on locations within the SFER catchment identified as important for improved understanding of ecological processes and habitat conditions within the catchment (*eco-bio sites*). A total of 24 *eco-bio sites* were used in the classification with a predominant focus on streams within the Angelo Coast Range Reserve and Hollow Tree Creek watershed where access was less limited and more ecological data had previously been collected.

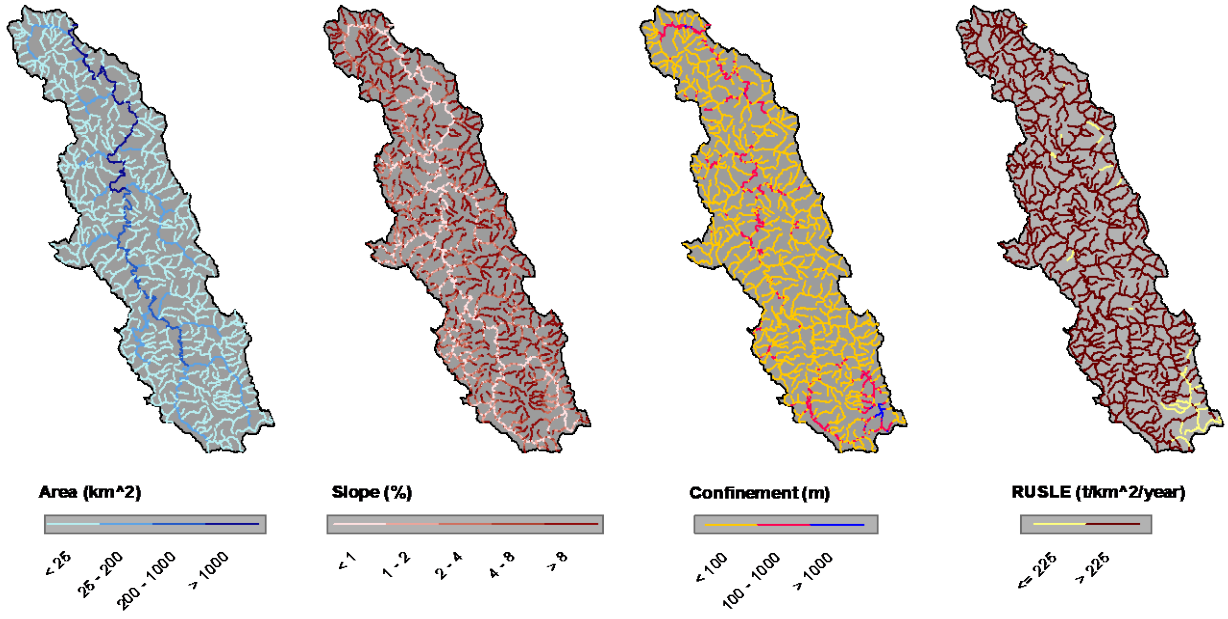


Figure 2. South Fork Eel River catchment streams binned by area, slope, confinement, and RUSLE values.

1.2.2 Data Acquisition and Processing

1.2.2.1 Field Surveying and Post-Processing

Survey methodologies are based on the California Water Board's Surface Water Ambient Monitoring Program (SWAMP) protocols (Ode, 2007). In the SFER catchment streams were surveyed along lengths equal to fifteen times mean bankfull width. Eight 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. A Wolman pebble count was also conducted along each cross-section (Wolman, 1954). Finally, a longitudinal survey was conducted along the thalweg at each cross-section.

Field survey data was processed in preparation for statistical analysis. 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 while 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 were significant in the identification of distinct channel types and process interpretation.

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.

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 of channel types (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 South Fork Eel River Catchment Classification

1.3.1.1 Multivariate Statistical Clustering

Bankfull width-to-depth ratio and valley confinement were identified as the most influential attributes in multivariate clustering. The final NMDS solution recorded a stress value of 0.156 with a non-metric coefficient of determination of 0.976 between observed dissimilarity and ordination distance (Fig. 3). The first and second principle component axes (PCAs) resulting from the NMDS ordination explained 52% of the variance in the data with loadings of 0.62 for bf.w.d and 0.67 for vc.dist for PCA-1 and PCA-2, respectively. Inclusion of PCA-3 and -4 explained 76% of variance, for which CV_bf.d and bf.d are the attributes with the greatest loading values. Within the SFER catchment, bankfull depth was highly correlated with bankfull width, indicating that bankfull depth is representative of bankfull width as well.

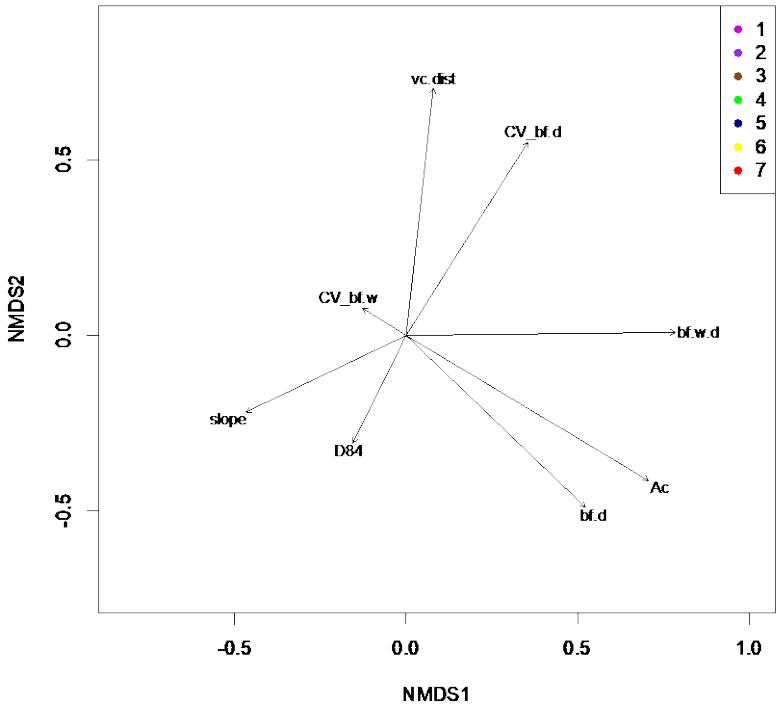


Figure 3. NMDS scatter plot with principal component vectors for attributes used in the classification of sites. Scattered sites are colored according to final channel type. Longer principal component vectors indicate greater influence on the scattering of sites.

Seven channel types were identified through the WHC with heuristic refinement and CART (Figs. 4 and 5). The Hubert Index suggested three Ward's clusters as the optimal number of groupings, however three channel types did not sufficiently describe the variability in stream form across the

catchment. Further clusters were analyzed for geomorphic significance and the final seven channel types were the result of CART analysis which resulted in a successful prediction rate of 89%. Ten-fold cross-validation of this prediction was 75%. Width-to-depth ratio, contributing area, sediment size, depth variability, and valley confinement were the channel attributes that defined the final classification tree

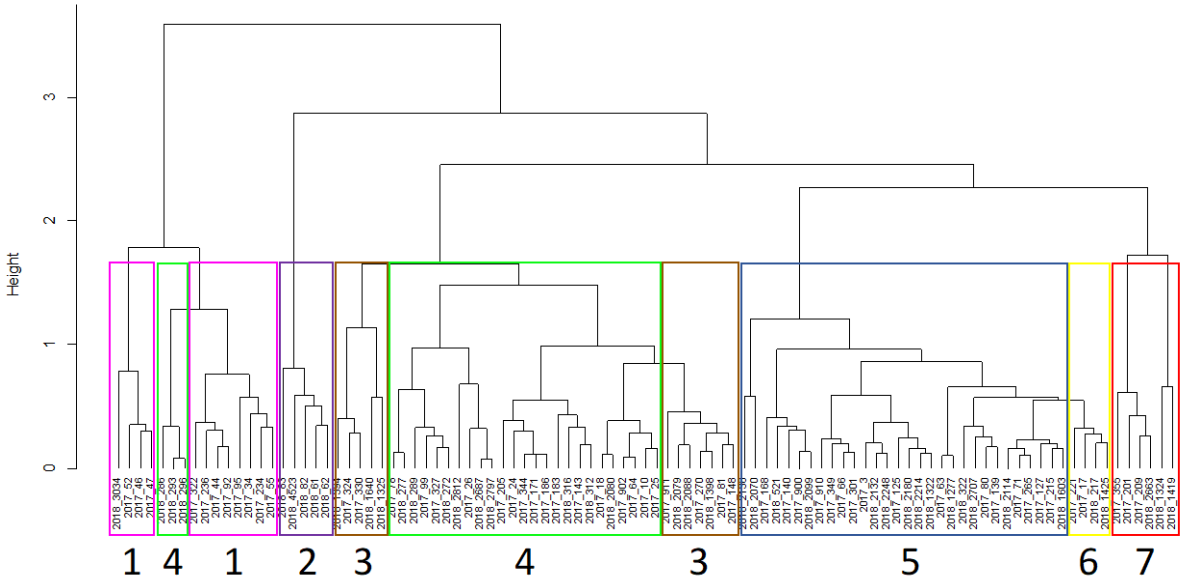


Figure 4. Ward's hierarchical clustering dendrogram with associated heuristic channel types in colored and numbered boxes.

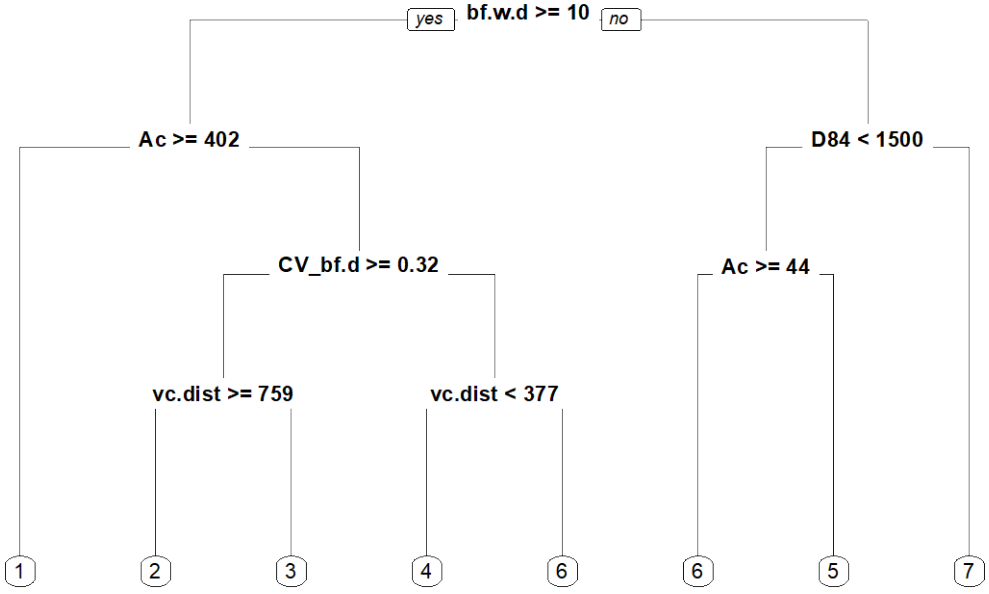


Figure 5. Final yes/no classification tree that produced suitable prediction and cross-validation percentages. Channel types are indicated by number at the bottom of the figure.

Final channel types were made up of between 4 and 30 sites. Clusters with a small number of sites were avoided, as outliers were likely to drive site-specific differences rather than larger catchment trends. However, it was ultimately the uniqueness of cluster attributes that drove final channel types, especially in valley confinement settings where influences from hillslope processes are less likely. Median attributes of each channel type can be found in Table 1.

Table 1. Median channel attributes of seven channel types within the South Fork Eel River catchment.

Channel Type	Ac	slope	bf.d	bf.w	bf.w.d	bf.d.D50	CV_bf.d	CV_bf.w	D50	D84	vc.dist
1	789	0.0012	2.47	53.21	20.37	93.20	0.33	0.22	27.3	77	85
2	26	0.0026	1.54	15.49	10.86	108.20	0.37	0.25	11	32	1749
3	69	0.0064	1.43	17.03	13.39	39.47	0.48	0.20	27.3	140	11
4	68	0.0076	1.19	19.41	13.49	24.80	0.25	0.21	45	1000	42
5	8	0.0139	1.02	7.29	6.94	44.03	0.22	0.26	27.3	90	15
6	73	0.0060	1.24	9.37	9.57	52.97	0.21	0.22	30.5	86.5	444
7	13	0.0801	1.14	8.32	6.16	13.07	0.28	0.27	96	2000	9

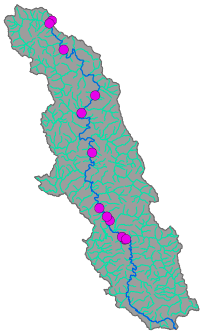
The SFER catchment is dominated by confined settings throughout the region compared to other regions of California. As a result, five of the seven channel types described by multivariate statistical classification occurred within confined settings. The remaining two channel types were split between partly-confined and unconfined settings. In general, the classification progresses from larger to smaller channel dimensions. Although bankfull depth and bankfull width-to-depth ratio were not highly correlated across all sites (R^2=0.33), the classification also generally transitions from larger to smaller bankfull width-to-depth ratio.

Five channel types are observed within confined valley settings and represent a wide variety of forms from headwater streams to the mainstem SFER (Fig. 6). The SFER catchment channel types are described as follows: *cobble, high width-to-depth ratio, riffle-pool* (cluster 1); *gravel-cobble, bed-undulating* (cluster 2); *gravel-cobble, high width-to-depth ratio, uniform* (cluster 4); *gravel-cobble, low width-to-depth ratio, uniform* (cluster 5); and *high-gradient, cobble-boulder, step-pool/cascade* (cluster 7). Beginning in headwater streams, streams with the smallest contributing area include *gravel-cobble, low width-to-depth, uniform* and *high-gradient, cobble-boulder, step-pool/cascade* channel types. These channel types are most clearly differentiated by slope and sediment grain size (Fig. 7), which will also drive differences in associated form. *Gravel-cobble, low width-to-depth, uniform* streams (n = 30) are defined by low depth and width variability, which results in the uniform channel description. The uniform channel description does not preclude periodic bed structures or bed structures created by local hillslope settings, but the variability in these streams are statistically less than the same attributes in other observed channel types. *High-gradient, cobble-boulder, step-pool/cascade* streams (n = 6) are the steepest channels within the SFER catchment. Intermediate size, confined streams within the catchment are best represented by *gravel-cobble, bed-undulating* and *gravel-cobble, high width-to-depth ratio, uniform* channel types. *Gravel-cobble, bed-undulating* streams (n = 12) are characterized by large values in depth variability. These streams likely display well-defined, periodic bed structures commonly referred to as riffles and pools but may also include forced bed structures as well. In comparison, *gravel-cobble, high width-to-depth ratio, uniform* (n = 28) streams lack the statistically high depth

variability values. This suggests that these streams display a more planar bed structure compared to bed-undulating streams. The largest channel type in the SFER catchment occurs along the lower mainstem of the SFER. The *gravel-cobble, high width-to-depth ratio, riffle-pool* channel type (n = 12) is defined by a meandering, riffle-pool structure within a confined valley setting where valley walls may still influence channel form in addition to transverse, meandering flow patterns.

The remaining two channel types were identified as *unconfined, gravel, riffle-pool* (cluster 2) and *partly-confined, gravel-cobble, uniform* (cluster 6) streams. *Unconfined, gravel, riffle-pool* (n = 5) streams exist in wide valleys and exist independently of hillslope processes. The riffle-pool morphology is indicative of point bar and floodplain development. The unconfined streams are predominantly located in the southeast portion of the SFER catchment. *Partly-confined, gravel-cobble, uniform* (n = 4) streams exhibit low width-to-depth ratios but exist within a valley bottom that is wide enough to allow for floodplain development. However, hillslope processes are likely to influence the channel type where the channel abuts local hillslopes. The channel dimensions are lower in variability than the unconfined channel type and are more similar in value to the uniform streams in confined SFER catchment settings.

1 – Confined, gravel-cobble, high width-to-depth ratio, riffle-pool



2 – Unconfined, gravel, riffle-pool

3 – Confined, gravel-cobble, bed-undulating

4 – Confined, gravel-boulder, high width-to-depth, uniform

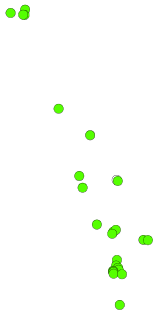


Figure 6. The seven channel types within the South Fork Eel River catchment developed by multivariate statistical analysis with heuristic refinement.

5 – Confined, gravel-cobble, low width-to-depth,

6 – Partly-confined, gravel-cobble, uniform

7 – Confined, high gradient, cobble-boulder, step-pool/cascade

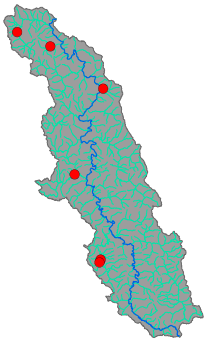


Figure 6 (cont'd). The seven channel types within the South Fork Eel River catchment developed by multivariate statistical analysis with heuristic refinement.

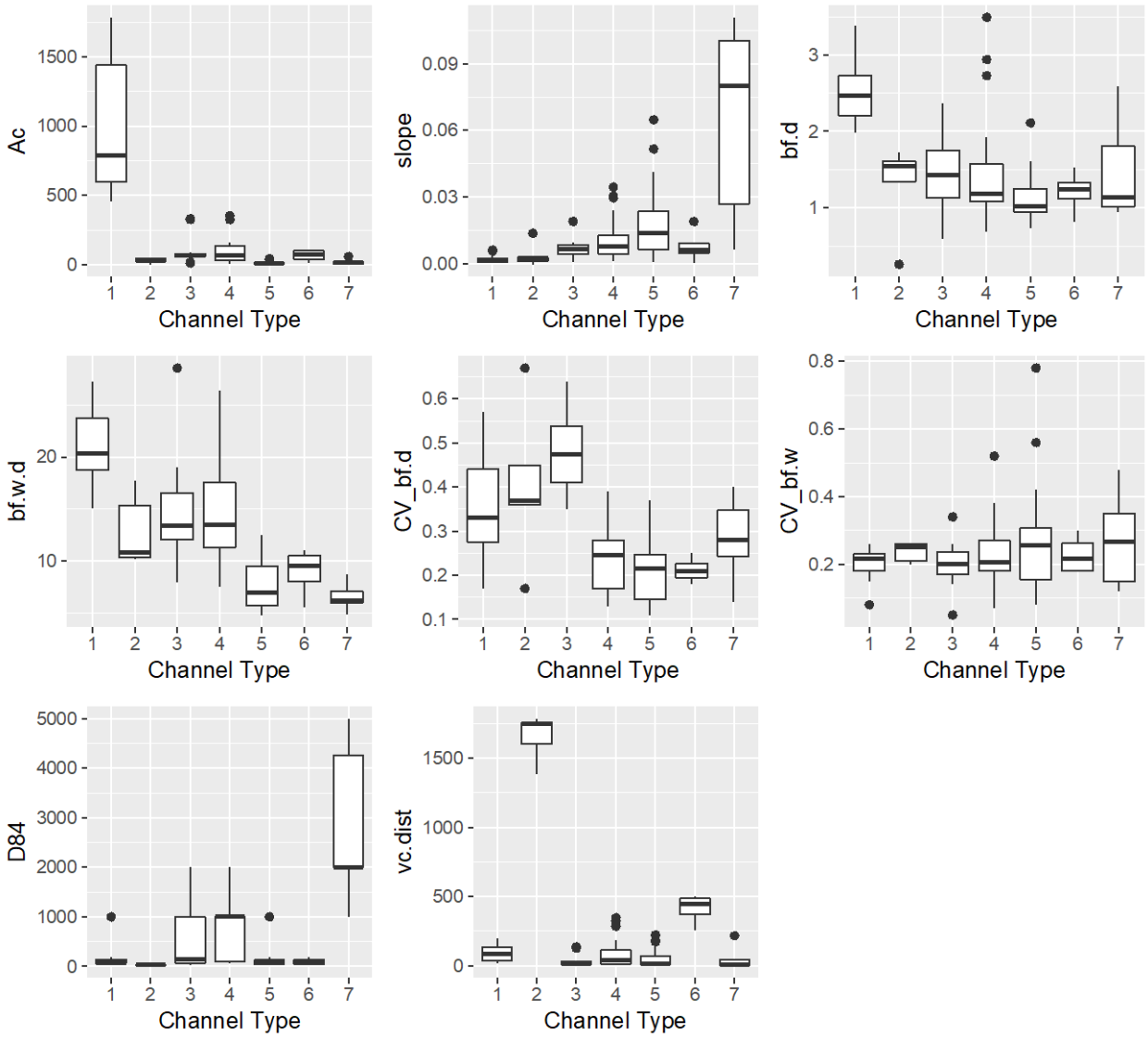


Figure 7. Tukey's Honestly Significant Differences between attributes used in the multivariate clustering of each channel type in the SFER catchment.

2 Predictions of Channel Types in the South Fork Eel River Catchment

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? Seven 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 2), with the fewest streams in classes 2 and 6 and the majority in classes 4 and 5.

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 8): (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 8. Flowchart of the four-fold machine learning framework.

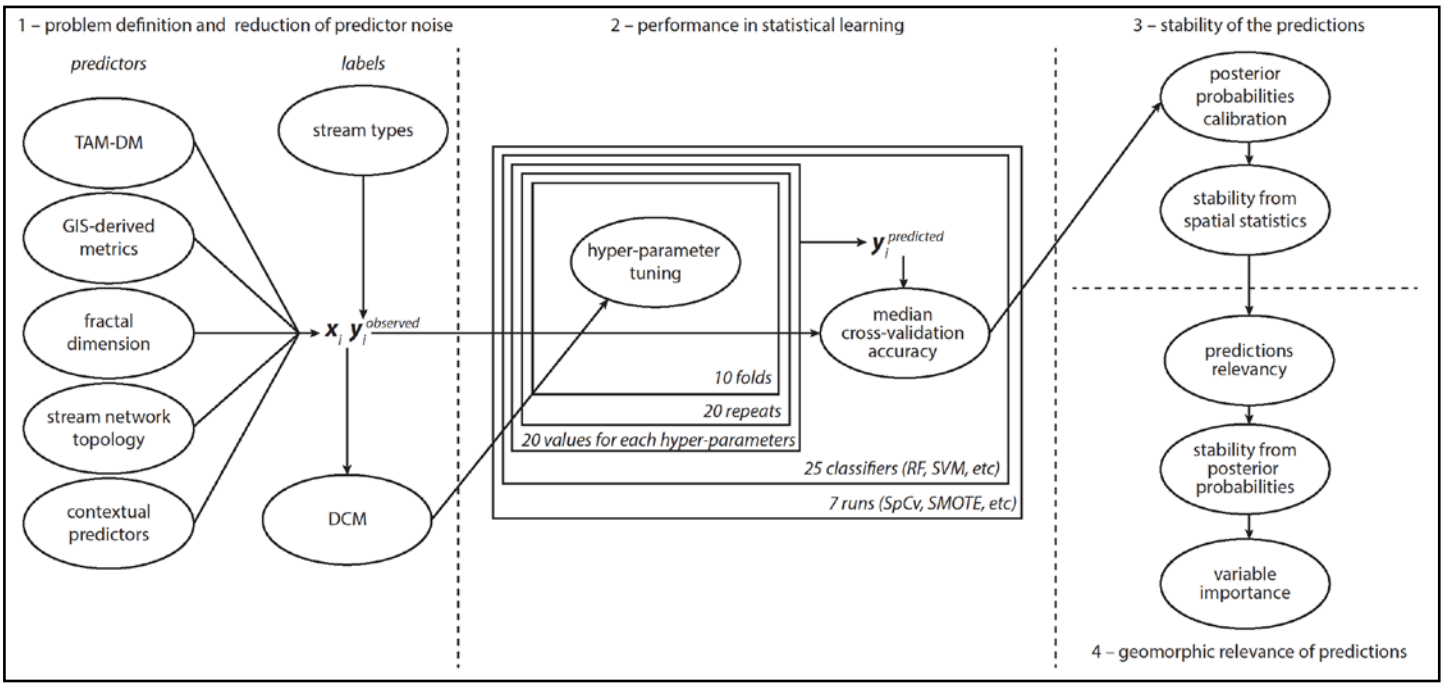


Table 2. Number of observations for each class of channel type and prevalence (i.e. relative frequency).

HG class	1	2	3	4	5	6	7
number	12	4	12	28	30	4	6
prevalence	0.13	0.042	0.13	0.29	0.31	0.042	0.06

2.1.1 Defining a Tractable Problem and Reducing Predictor Noise

As most data-driven approaches, the set of predictors used as 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.

A large set of predictors is initially included in the machine learning framework. This initial set includes 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 $\sim$ 8k 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 (see Appendix 2: Fig. 1).

Contextual predictors were obtained from the publicly available Stream-Catchment Dataset (StreamCat; Hill et al., 2015). These predictors are integrated over the entire upstream watershed draining to each stream segment. Contextual predictors included: 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); long-term runoff and air temperature 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 Drăguţ 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 balance of the training set (Table 2), the Synthetic Minority Oversampling Technique (SMOTE) was used (Chawla et al., 2002). SMOTE relies on assigning predictors along the edges connecting the k-nearest neighbors from randomly selected observations. As synthetic data points need to be included in the spatial cross-validation scheme, spatial coordinates 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 3 following runs were performed:

- base-a: imbalanced classes
- SMOTE-b: balanced classes; without StreamCat predictors
- 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 SFER catchment 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) multiple classifiers show good performance in statistical learning; (iii) spatial statistics highlight the higher stability of Random Forest predictions; and (iv) Random Forest predictions capture large-scale and fine-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. 9). 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 2 is 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 uniform classes 4 vs 5.

Figure 9. 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.

Our previous results in the Sacramento Basin showed that 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.

Here, we show that removing the coarse contextual variable does not impede overall run accuracy (Table 3). 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 3).

Table 3. 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	SMOTE-b	SpCV-SMOTE-b†
maximum median cross-validation accuracy	0.4	0.79	0.77
best classifier(s)	k-NN PLS NB	RF R-SVM	L-SVM, R-SVM RF, BaT, ANN MARS, FDA

a: with StreamCat predictors
b: without StreamCat predictors
base: imbalanced classes
SMOTE: classes balanced with SMOTE algorithm
SpCV: Spatial Cross Validation
†: run selected

2.2.2 Multiple Classifiers Show Good Performance in Statistical Learning

From the statistical learning step of our four-fold framework, seven classifiers emerge as the most accurate: Linear SVM, Radial SVM, Random Forest, Bagged Trees, Artificial Neural Network, Multiple Adaptative Regression Spline and Flexible Discriminant Analysis (Fig. 10). 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 L-SVM, RF, BaT, ANN, MARS and FDA spatial predictions.

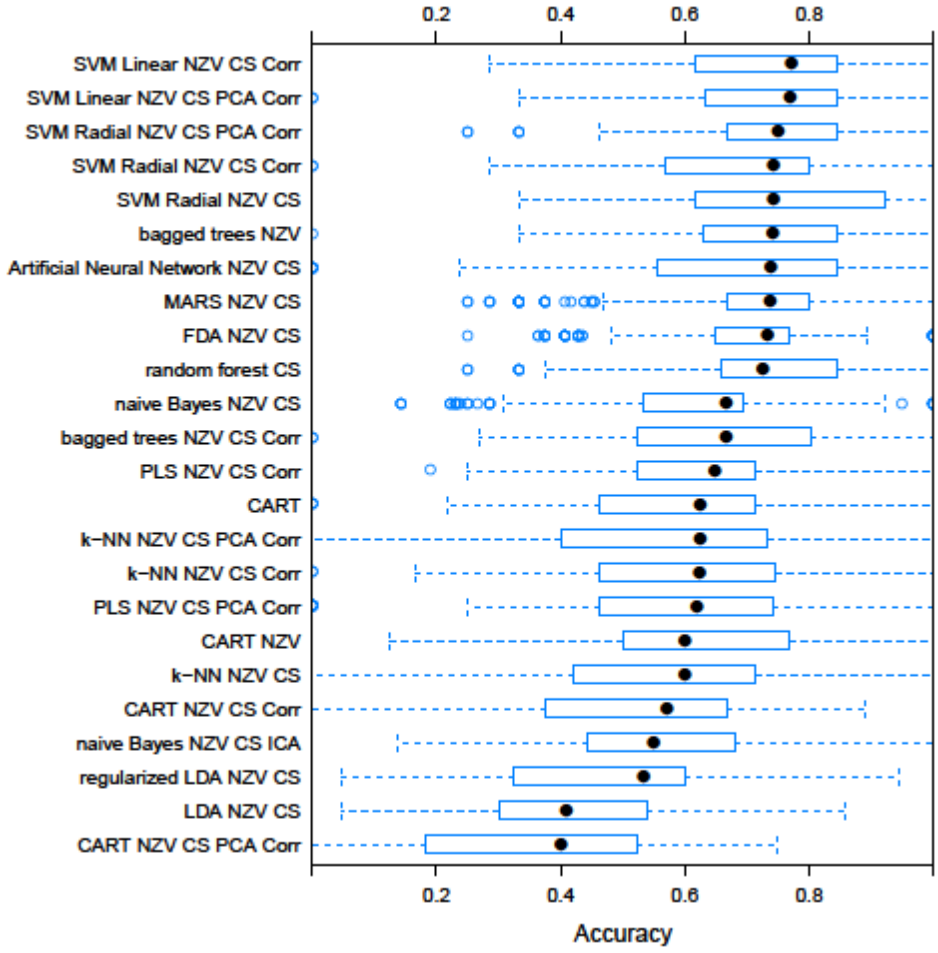


Figure 10. Distributions of cross-validation accuracies for all classifiers for the SpCV-SMOTE-b run (Table 3). Corr indicates that Box-Cox transformations were applied when required; NZV means that the predictors with Near-Zero-Variance were filtered; ICA and PCA mean that Independant Component Analysis 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 all classifiers. The results of that process are shown in Fig. 11 for RF. 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. 11).

Spatial statistics were used to assess the prediction stability using L-SVM, RF, BaT, ANN, MARS and FDA 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 six classifiers predictions highlights different spatial patterns in the probability of occurrence of each of the 7 classes (Fig. 12). Class 1, the SFE main steam is predicted with a high probability of occurrence by RF, FDA, MARS and BaT. In constrast, the spatial pattern from L-SVM and ANN is less clear. Those two classifiers display a more scattered pattern for class 2 and 6. Examining the entropy map (Fig. 13) and entropy histograms (Fig. 14) underlines that RF, MARS and FDA outperform L-SVM, ANN and BaT. Comparing the stacked distributions of entropy for RF, MARS and FDA classifiers shows that RF displays a more skewed left-tail and a lighter right-tail (Fig. 15). These considerations lead us to selection RF as the classifier displaying the most stable spatial predictions. 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).

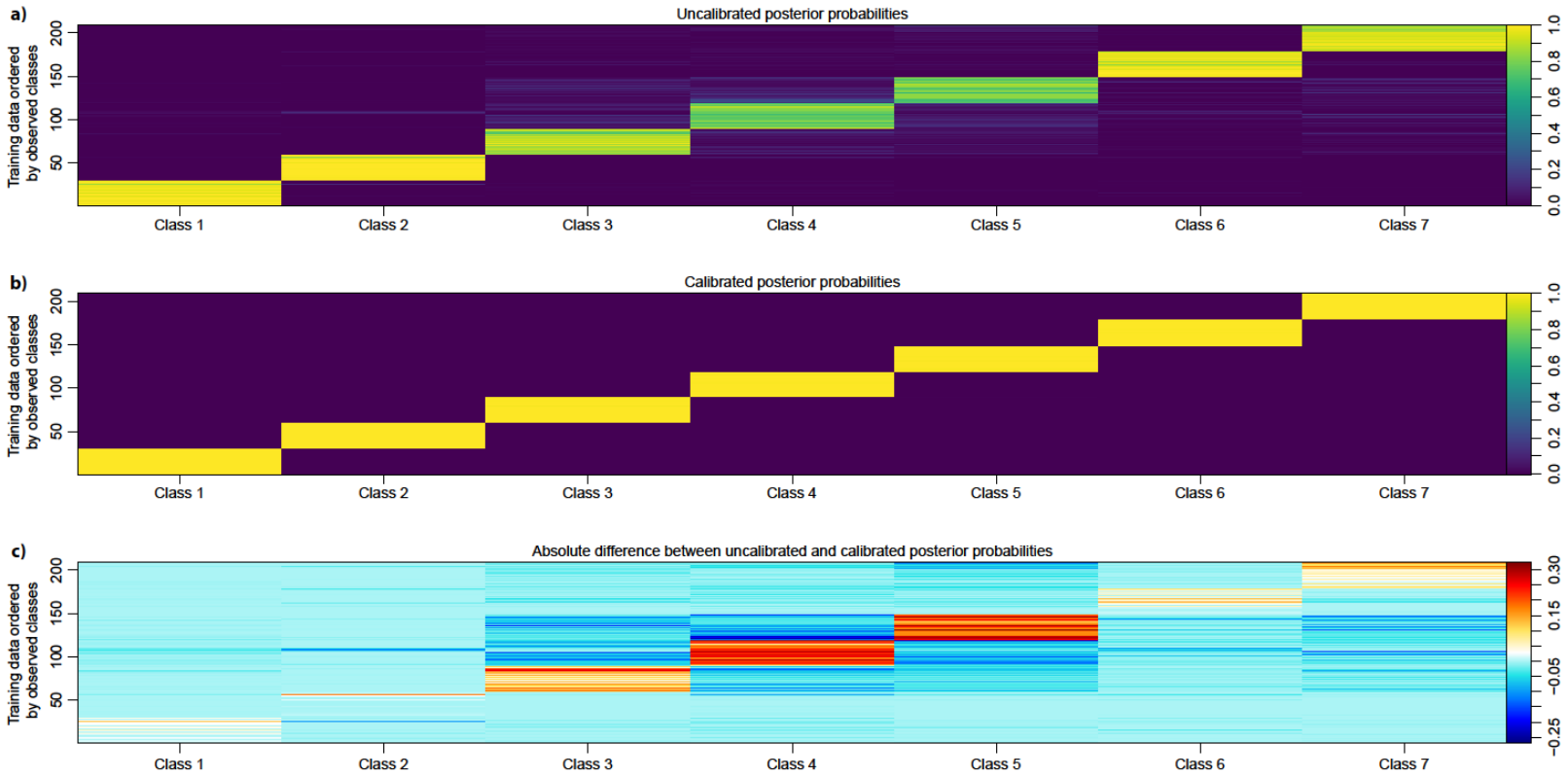


Figure 11. Results of the posterior probability scaling for Random Forest. This figure is analogous to a reliability plot, with the vertical axis corresponding to training observations. These points are sorted by classes so that points that have the label 1 are at the bottom of the vertical axis whereas points with class 7 are at the top. a-b) Each row displays the posterior probabilities that the observation belongs to each of the 7 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).

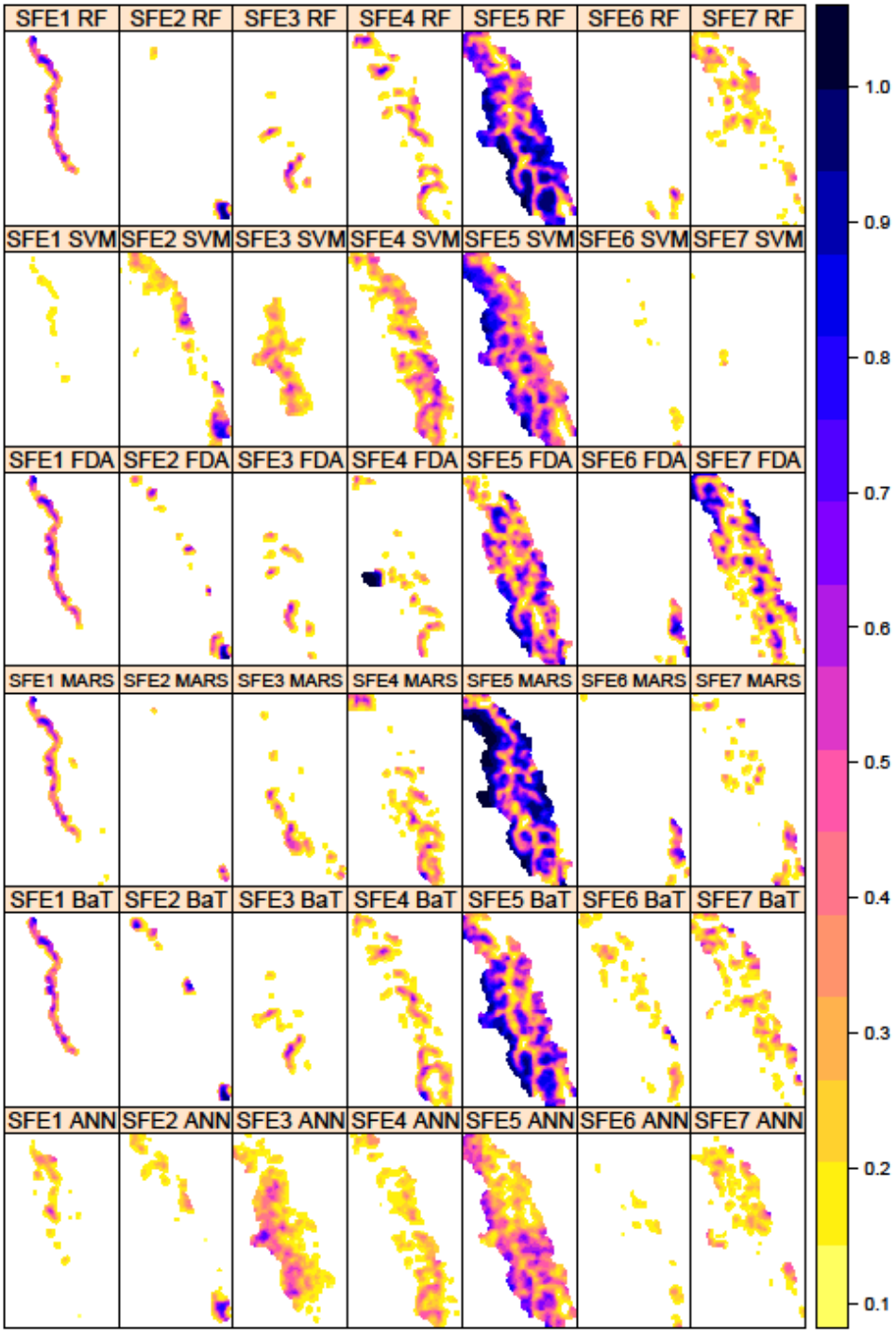


Figure 12. Relative risk maps for the 7 hydro-geomorphic classes identified in the South Fork Eel River catchment. Each row corresponds to a different classifier. The appropriate bandwidth of the smoothing kernel was determined to be ∼1000 m.

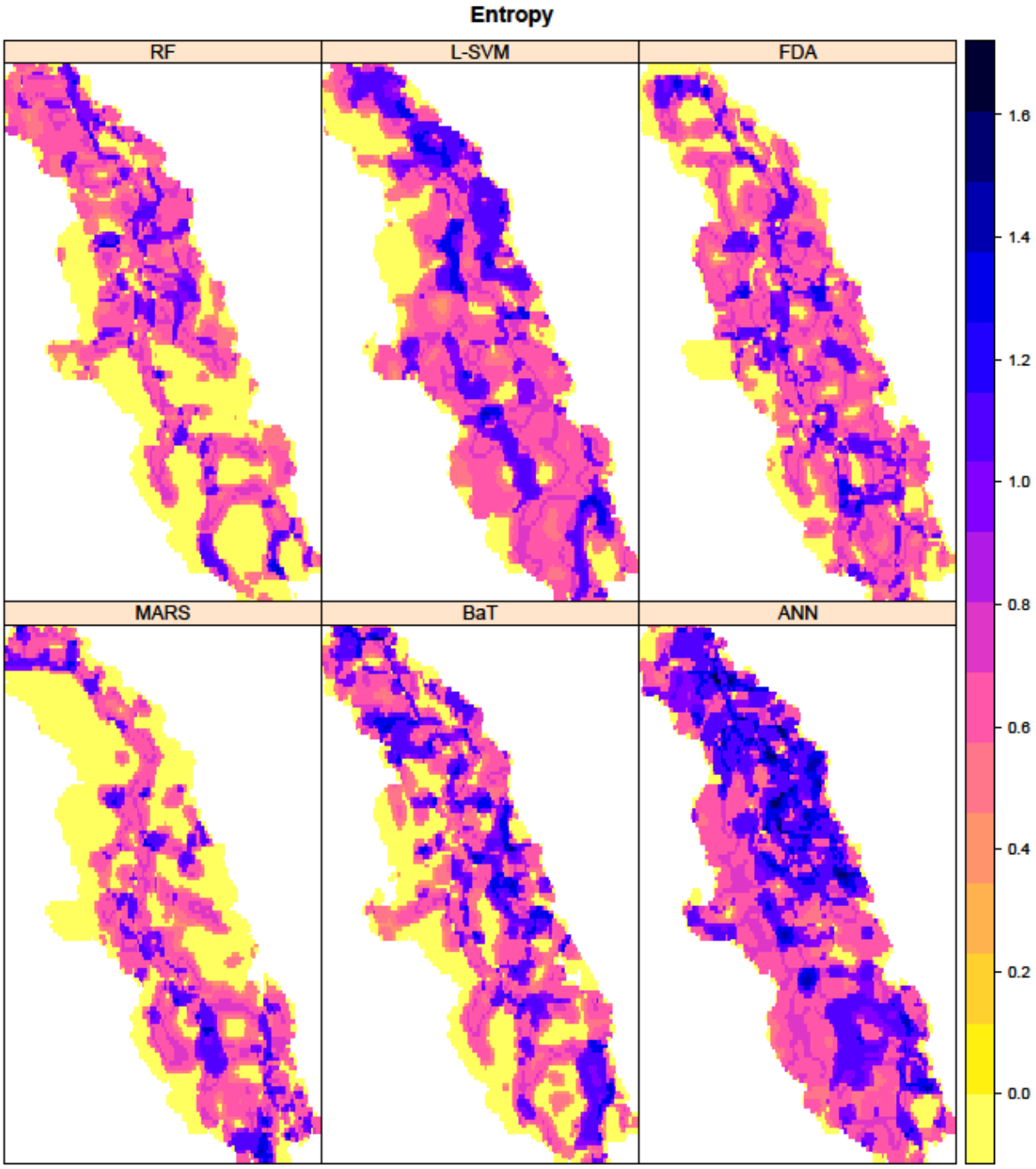


Figure 13. Map of the stability of predictions at the catchment scale from the Shannon-Weiner's entropy of the RF and L-SVM classifiers. The entropy value is computed from the relative risks maps (Fig. 12). A low entropy indicates the number of significantly probable channel types is lower and thus that 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.

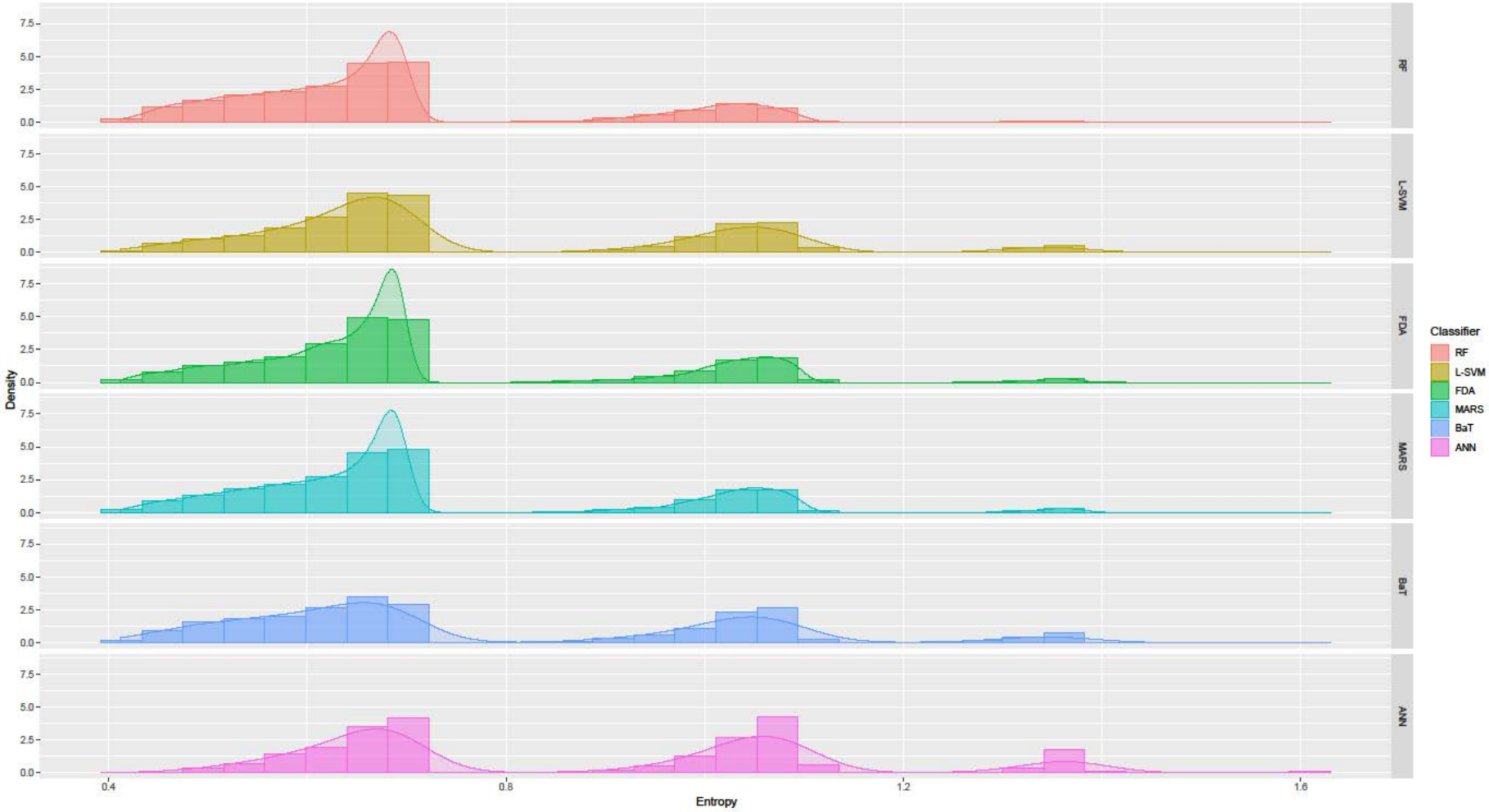


Figure 14: Histograms and densisty distributions of entropy derived from spatial statistics in the South Fork Eel River catchment for each classifiers.

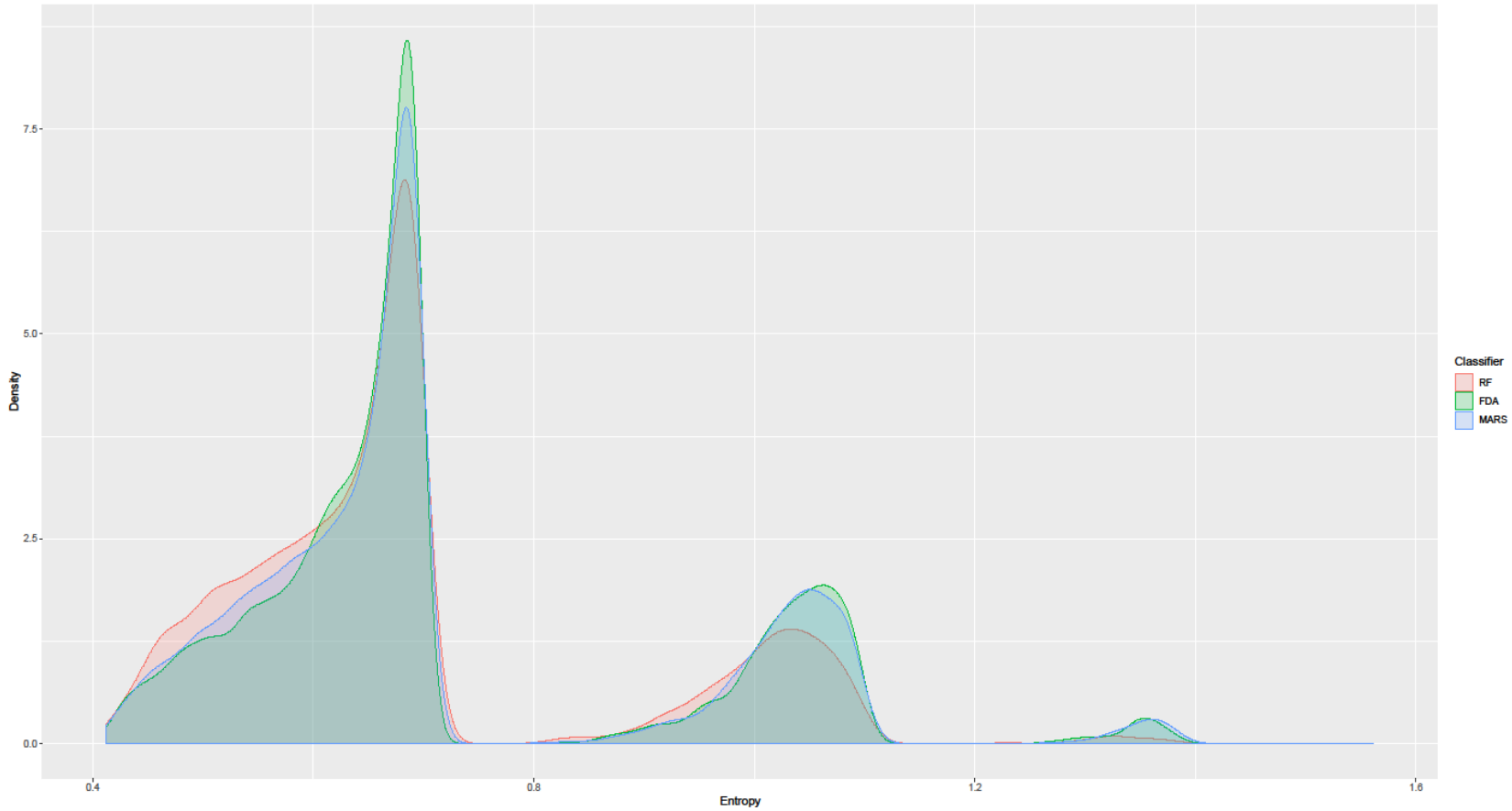


Figure 15: Stacked density distributions of entropy for the RF, FDA and MARS classifiers.

2.2.4 Random Forest Predictions Capture Large-scale and Fine-Scale Landscape Organization

The geomorphic relevance of the predictions from the RF classifier was investigated. An overview of RF predictions of channel type (Fig. 16) shows that the large-scale organization of the landscape with channel types occurring where expected e.g. SFE main stem localized in the SFE river valley, high gradient streams occuring in the most mountainous area (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. 9) and from the median cross-validation accuracy (77%).

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. 17). The predictions from the RF appear highly stable along the main stem, with more uncertainties in the steeper parts of the catchment. Interestingly, entropy does not show a clear spatial pattern. Nonetheless, a correlation analysis between the predictors in the RF classifier and the resulting entropy calculated from the posterior probabilities show that entropy correlates the most with slope (Fig. 18). This is thought to highlight the fact that slope correlates usually with a higher ability to do significant geomorphic work which translates into a higher number of possible stream channels.

The variable importance plot for the RF classifier clarifies in part the black box nature of some machine learning approaches (Fig. 19). Three variables appear significantly more important than the other predictors: drainage area, valley confinement and stream order. Apart from this 3 variables and channel slope, the most important predictors are dominated by fractal dimension predictors (Hurst coefficients) underlining their relevance. This finding is supported by similar variable importance from the Deep Artificial Neural Network.

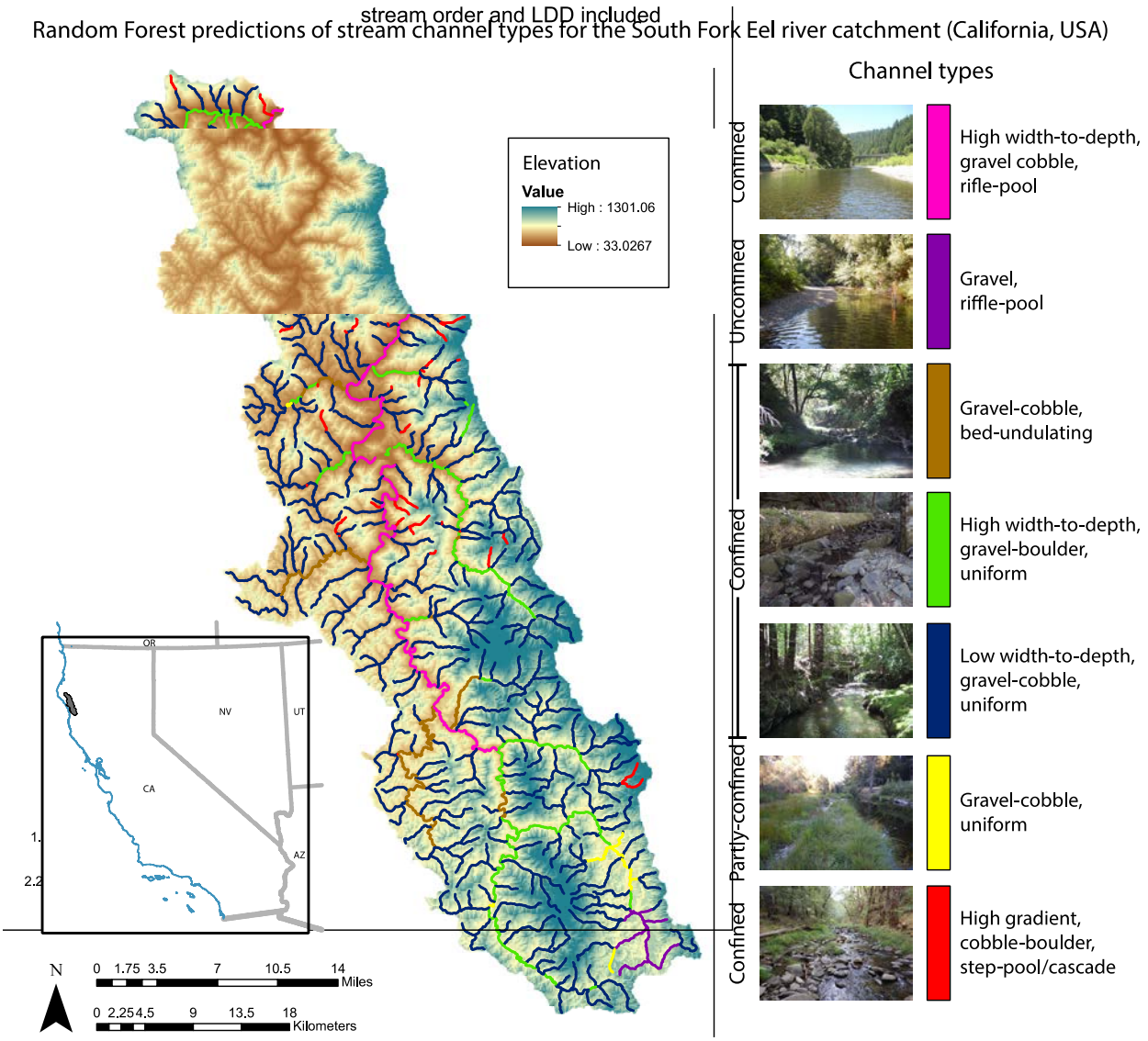


Figure 16. Map of the South Fork Eel River catchment with the spatial predictions of the 7 types of channel.

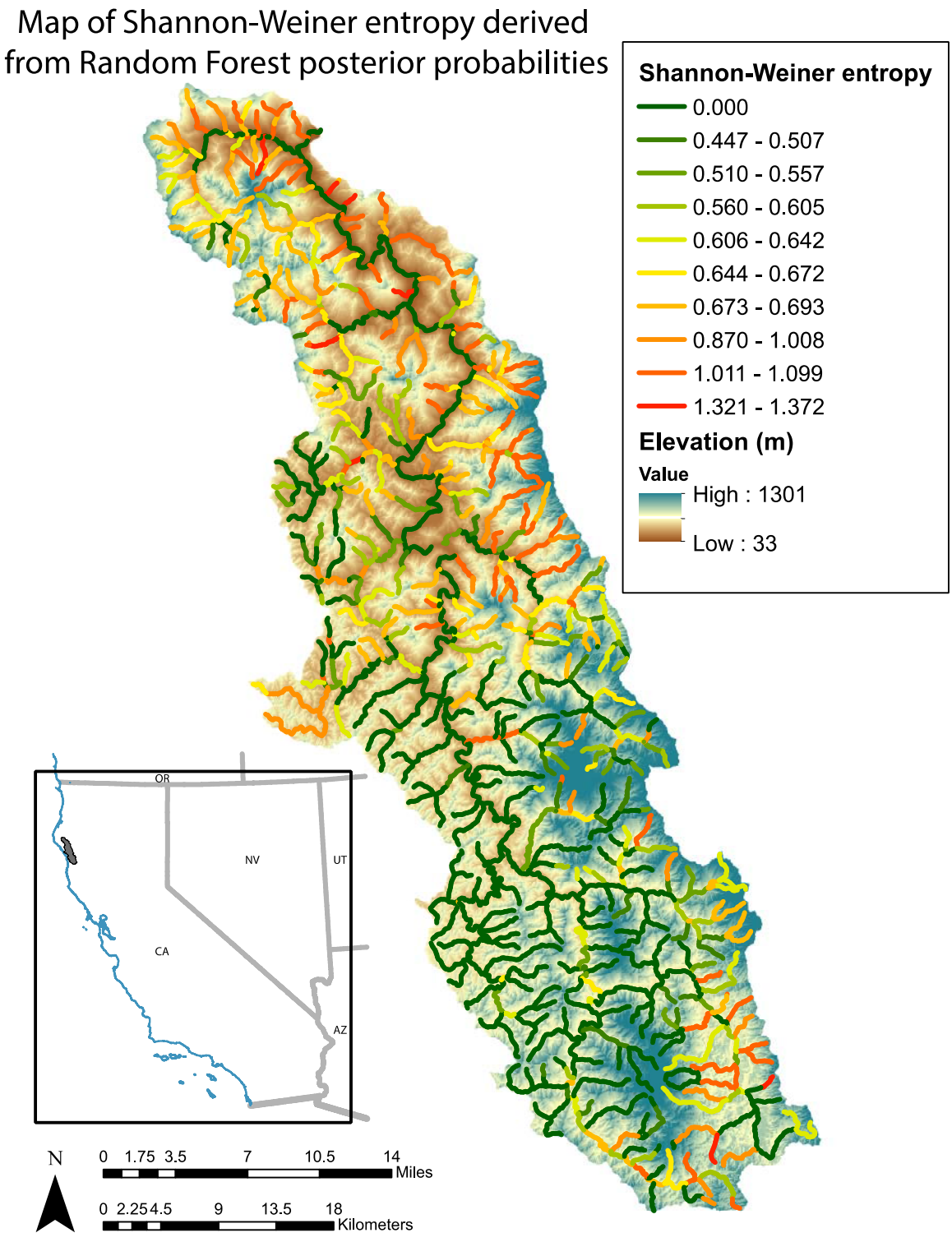


Figure 17. Map of the South Fork Eel River catchment with the entropy of each spatial predictions.

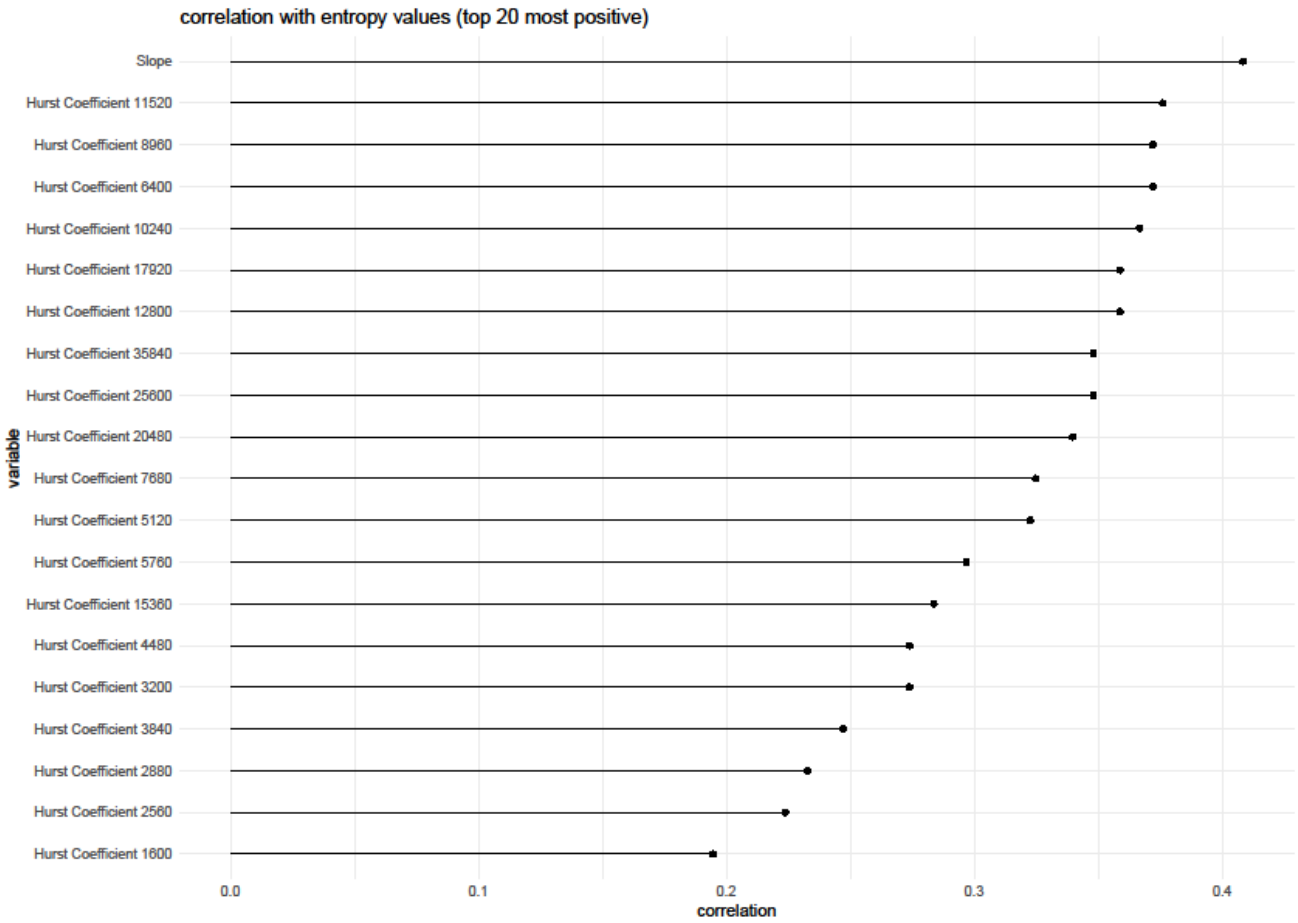


Figure 18. Correlations between predictors and entropy in the South Fork Eel River catchment.

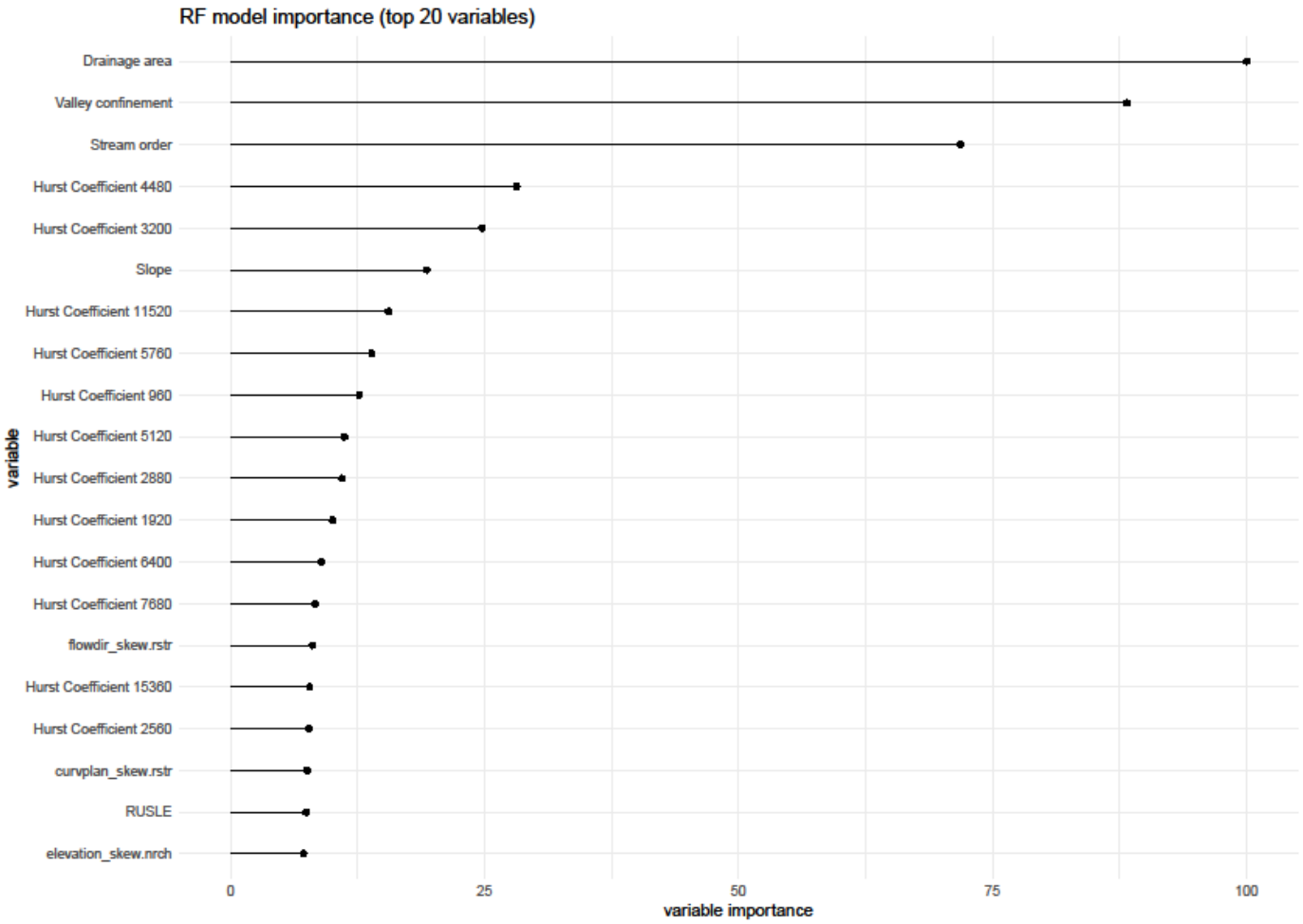


Figure 19. Variable importance of the RF model.

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.

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.

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.

Appendix 1: Stream Binning Protocol for Regional Geomorphic Classification in California

Rivers, streams, and creeks in California have a range of reach-scale geomorphic attributes, and these are often differentiated by thresholds in physical processes that yield different fluvial patterns. It is our intention to sample and characterize the diversity of streams throughout California. Because there is a very limited number of sites that can be observed during this study, much care is needed to balance site selection to not only sample widely different settings, but also obtain enough observations of each setting to have some confidence in the attributes of different channel types found in state. Consequently, this protocol lays out a mindful framework for site selection using the best available statewide information.

This protocol segregates all 200-m sections of river in California into a manageable set of broad groups differentiated by the two fundamental genetic drivers of fluvial morphology in non-glaciated landscapes: sediment supply and sediment transport capacity. Sediment supply is governed by landscape processes that erode bedrock, soil, colluvium, and alluvium. Sediment transport capacity is governed by a mix of landscape processes that convert rainfall into runoff as well as local river corridor conditions that constitute topographic steering and control hydraulics. Conveniently, both sediment supply and sediment transport capacity share some of the same governing variables, because often a variable that controls sediment production also controls water production. Examples include catchment area and local slope. There are also independent variables that make them distinct, such as rainfall impact intensity (for particle detachment from hillslopes) and valley confinement (for topographic steering of sediment transport). Some variables, such as land use, land cover, and geology may preferentially aid sediment supply more than sediment transport capacity in some settings, but not others.

To achieve the goal of sampling reach-scale geomorphic variability accounting for both sediment supply and transport capacity we have selected four specific metrics to use to stratify sites throughout regions of California. The four metrics are all achievable to extract and analyze using statewide datasets, though such datasets are coarse and uncertain. The metrics address local valley confinement, sediment supply from sheet wash and rilling, contributing area, and local stream slope. The four metrics are detailed below.

We expect this binning protocol to result in a set of survey sites with both reasonable replication and a wide range of geomorphic variability within the North, North Central, South Central, and South regions of coastal California for the 2018 Summer Field Sampling Campaign. All stratifying variables were calculated for the National Hydrography Dataset (NHDv2) streamlines used in the hydrologic classification (Lane et al., 2017). Elevation data, the 10 m digital elevation model (DEM), was retrieved from the United States Geological Survey National Elevation Dataset. This protocol was initially conducted on given NHD streamlines with variable lengths in order to ease

calculation time of valley setting and sediment supply metrics. Updated stream binning is currently underway with NHD streamlines split into 200 m segments within all four regions.

1. Calculation of Individual Parameters

Valley Confinement

Valley confinement was calculated as the distance from an NHD channel segment to a valley wall defined by 25% slope (~ 14 degree slope). This value was chosen to incorporate all potential valley floodplain slopes suggested in previous valley bottom definition literature (Gilbert et al., 2016). A sensitivity analysis of valley confinement using 5, 12.5, and 25% slopes indicated that the metric was consistent in prediction of small, medium, and large confinement distances. As 25% slope returned the greatest differences, this value was utilized to produce more distinctive bins.

The geospatial methodology, implemented using ESRI ArcGIS tools called from multiple Python scripts, included the following steps:

1. Create perpendicular cross-sections along streamline. For 200 m streams, cross-sections will be placed every 50 m resulting in four cross-sections per streamline.
2. Using a 10-m DEM, create polygons of all areas with less than 25% slope. Polygons that did not intersect with the NHD streamlines were removed.
3. Clip cross-sections with <25% slope polygons and calculate line length.
4. Calculate average distance from stream centerline to valley margin resulting in the valley confinement distance.

Valley confinement distance is not likely to be exact, but instead give a general indication of the valley setting of individual streams (Fig. 1).

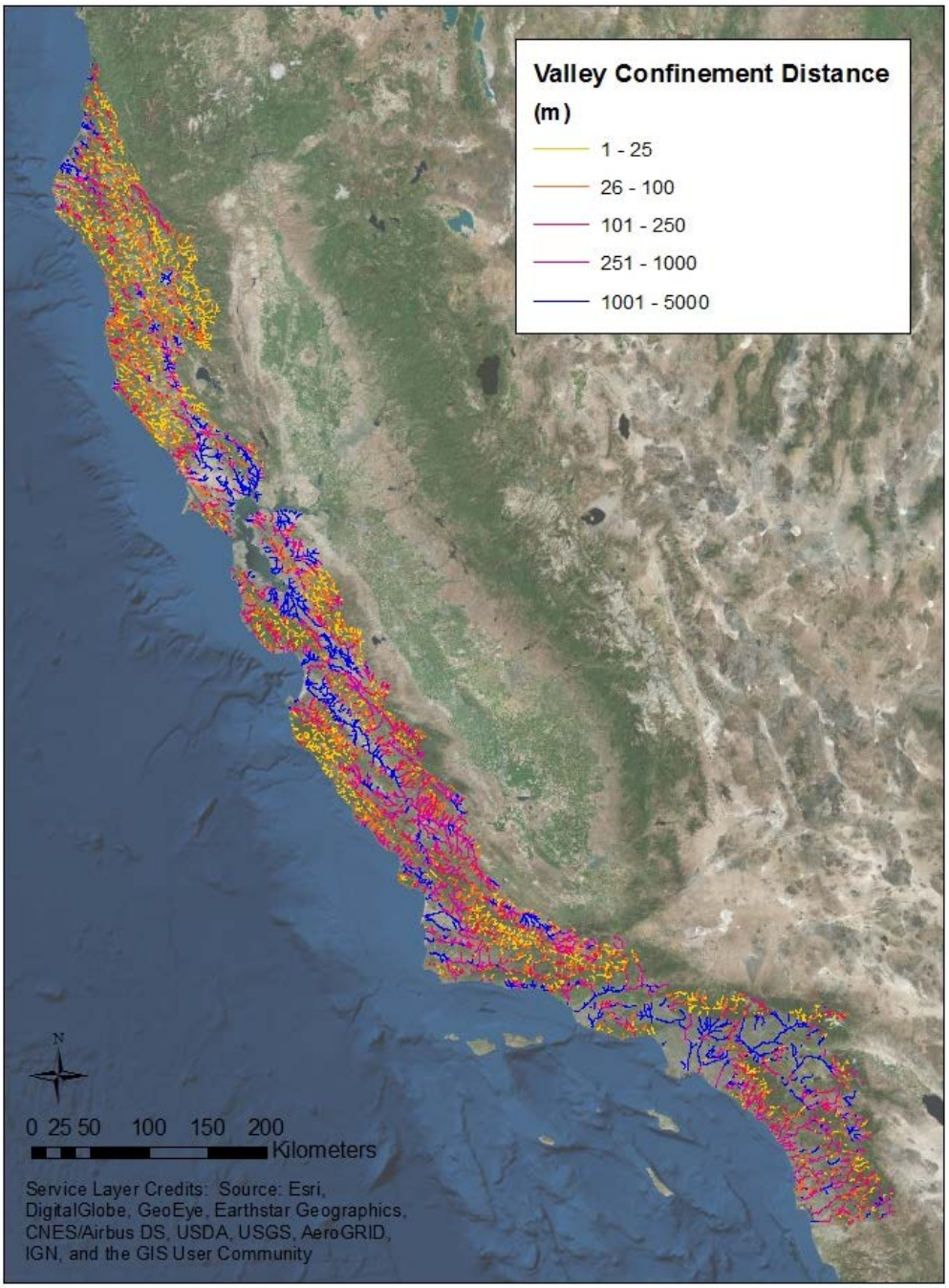


Figure 1. Distribution of valley confinement distances along the coast of California.

Sediment Supply

In order to estimate sediment supply to streams along the California Coast, an approach was implemented based on the Revised Universal Soil Loss Equation (RUSLE) (Renard et al., 1991). Although RUSLE (Eqn. 1) was developed for soil loss estimation on agricultural land, this method provided a quick and relatively easy way to estimate sediment loss over such a large area (Fig. 2).

$$A = R * K * LS * C * P \qquad \text{Eqn. 1}$$

Here A is average annual potential soil loss (tons/ac), R is a rainfall-runoff erosivity factor, K is a soil erodibility factor, L is a slope length factor, S is a slope steepness factor, C is a cover-management factor, and P is a support practice factor. The following datasets were used to estimate these factors across the state:

- R, K, and LS – maps with these factor estimates were retrieved from the California State Waterboard FTP
- C – Computed using the 2011 Stream-Catchment (StreamCat) Dataset (Hill et al., 2016) Landcover data and a table from Haan et al. (1994)
- P – Value was set equal to one

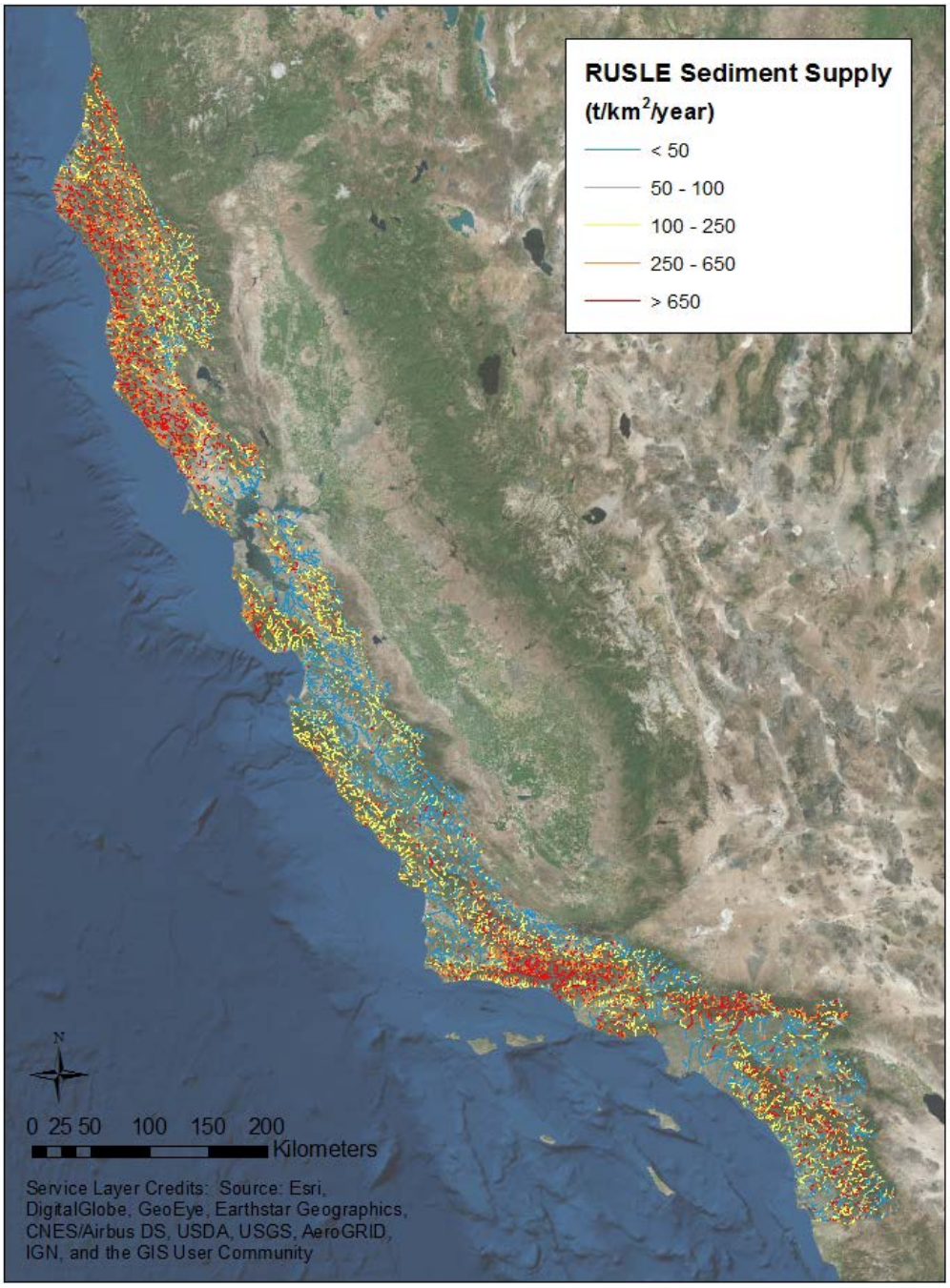


Figure 2. Estimation of annual sediment loss contributing to streams.

Contributing Area

Contributing area previously calculated in the StreamCat dataset was merged to NHD datasets in R using stream identification values (COMID values).

Slope

Slope was calculated from the 10 m DEM as the change in elevation along each streamline divided by the stream length.

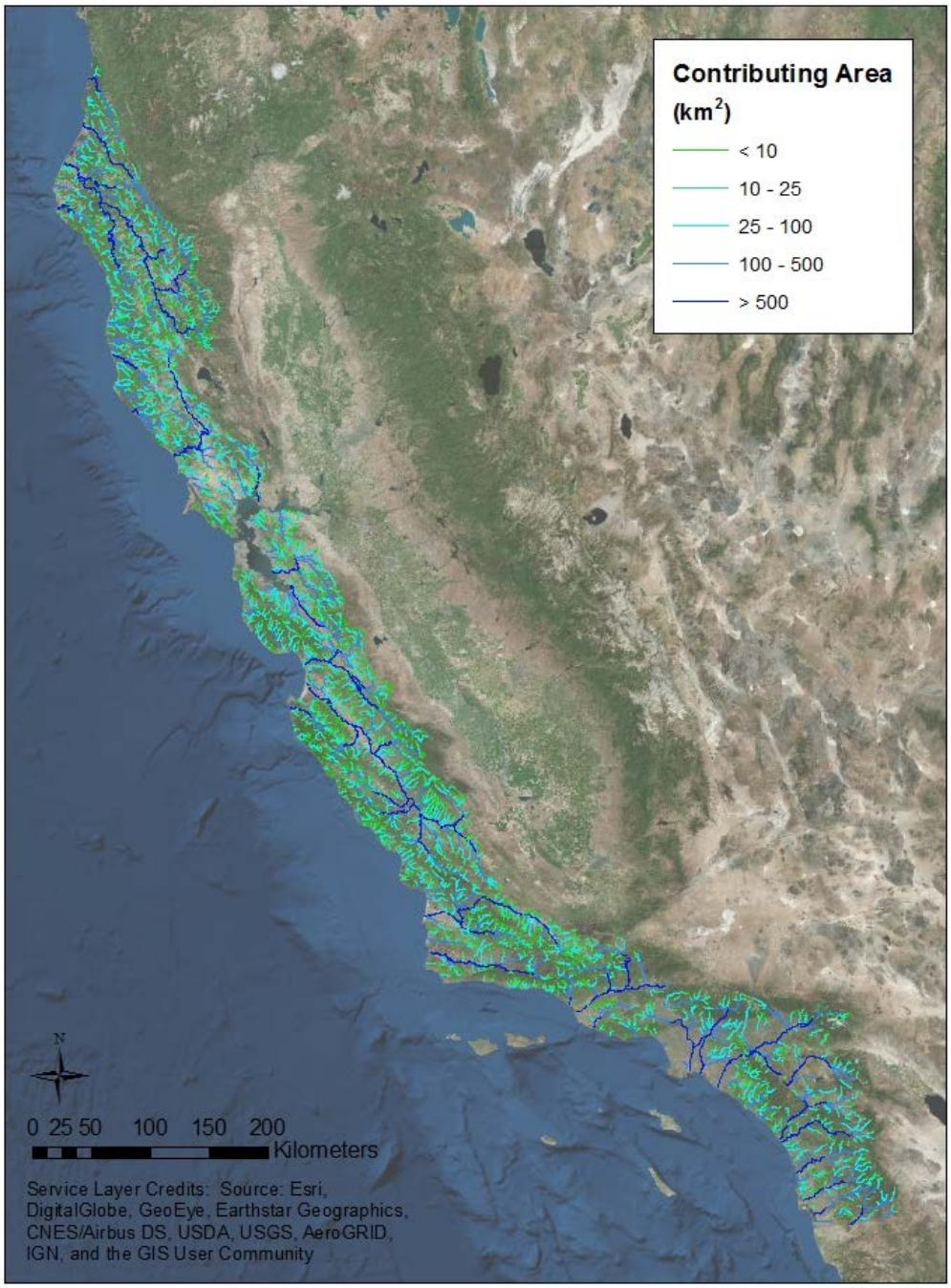


Figure 3. Total contributing area to individual stream segments.

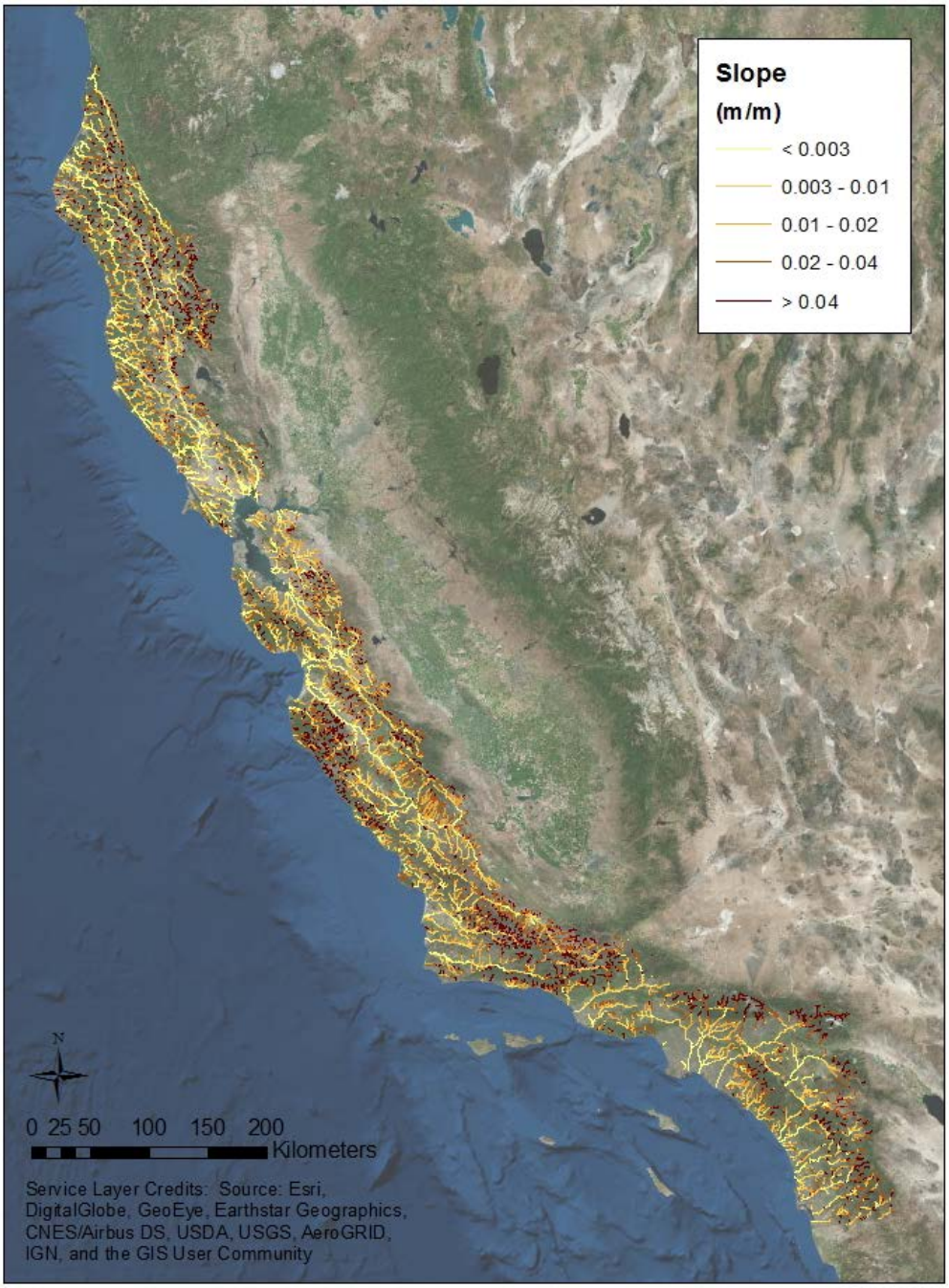


Figure 4. Slope calculated from 10 m DEMs.

2. Binning Protocol

A hierarchical approach was used to achieve a distributed set of stream segments across the four geomorphic variables. The stream network was first stratified based on valley confinement and sediment supply. Valley confinement was split into three bins based on preliminary results for the Sacramento River Basin: confined (<100 m valley confinement distance), partly-confined (>100 m and < 1000 m valley confinement distance), and unconfined (> 1000 m valley confinement distance). The log-normal distribution of sediment supply values was used to generate two bins: high sediment (> 225 t/km^2/year, or ~ 1 ton/ac/year) and low sediment (< 225 t/km^2/year).

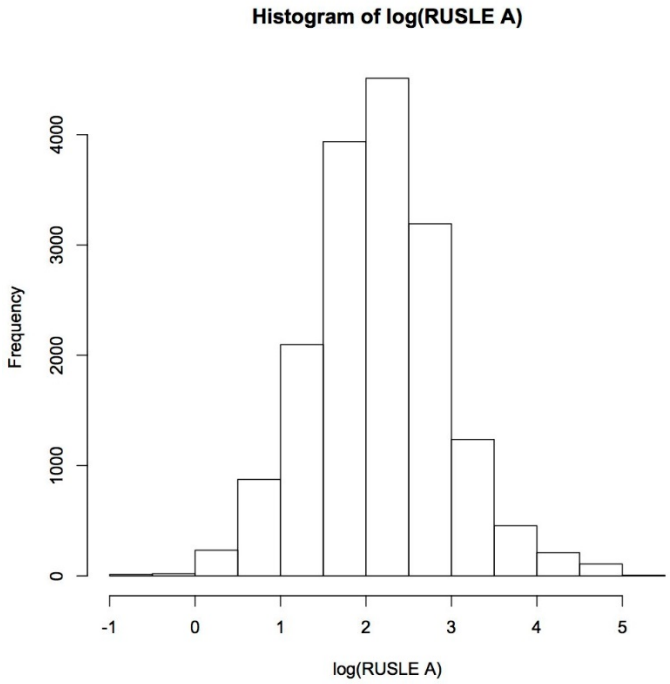


Figure 5. Distribution of the logarithm of sediment supply values.

Following the initial stratification by valley confinement distances and sediment supply, contributing area and slope bins were created within each upper-level bin (e.g. confined & high sediment) by statistical clustering using k-means algorithms in R (stats package) (Figs. 6-9). In this method, the statistical clustering is not being used to capture any sort of underlying process or connection of sites within the dataset beyond how the data fits into the Area-Slope two-dimensional space. The area-slope datasets for each upper-level bin were split into five clusters. This method allows for more control of the selection of sites by creating lower-level bins associated with distinct area-slope settings. With 60 sites per region and six upper level bins for each region, selection is limited to 10 sites per upper-level bin. With five area-slope bins, two sites can be selected from each to ensure variability in slope and area distributions. In addition, the proportion of sites in each upper-level bin is variable and differs by region as well (Fig. 10). Therefore, upper-level bins that have a low proportion of sites will be limited to five total sites

(one site per lower-level bin) and the remaining five sites will be reallocated to different bins based upon the proportion of sites in that bin.

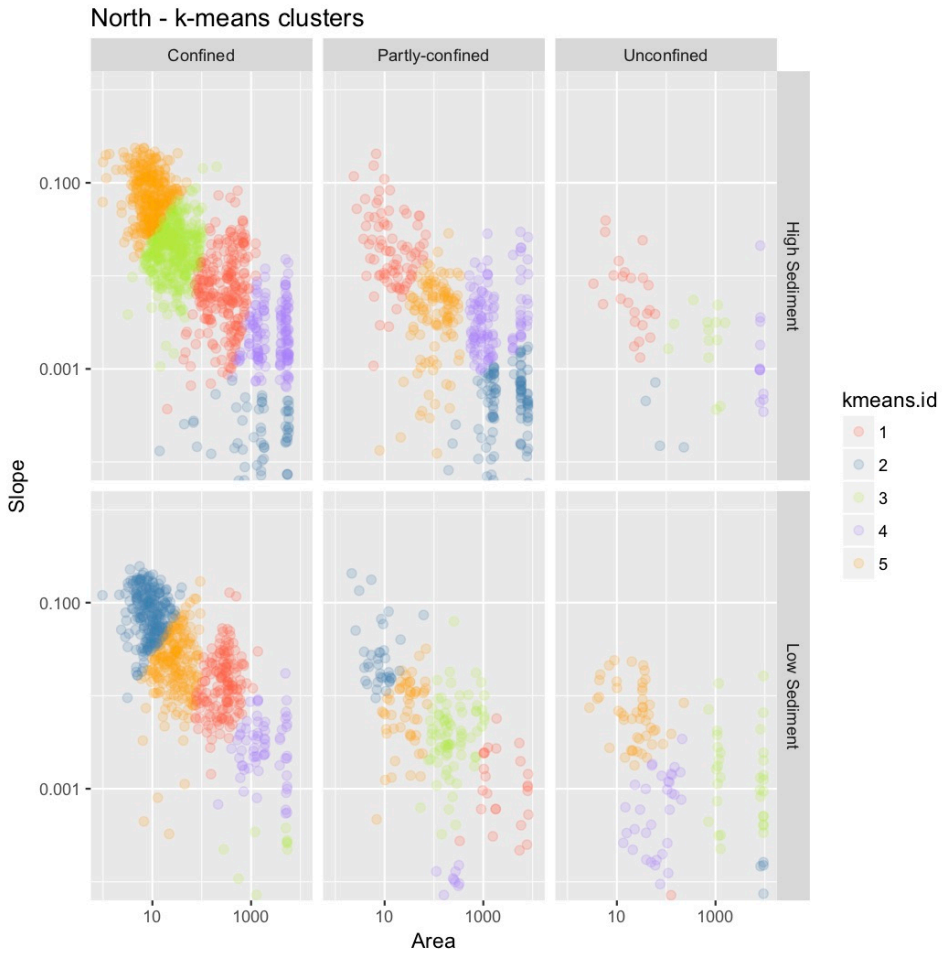


Figure 6. North Coast upper level bins with five area-slope bins.

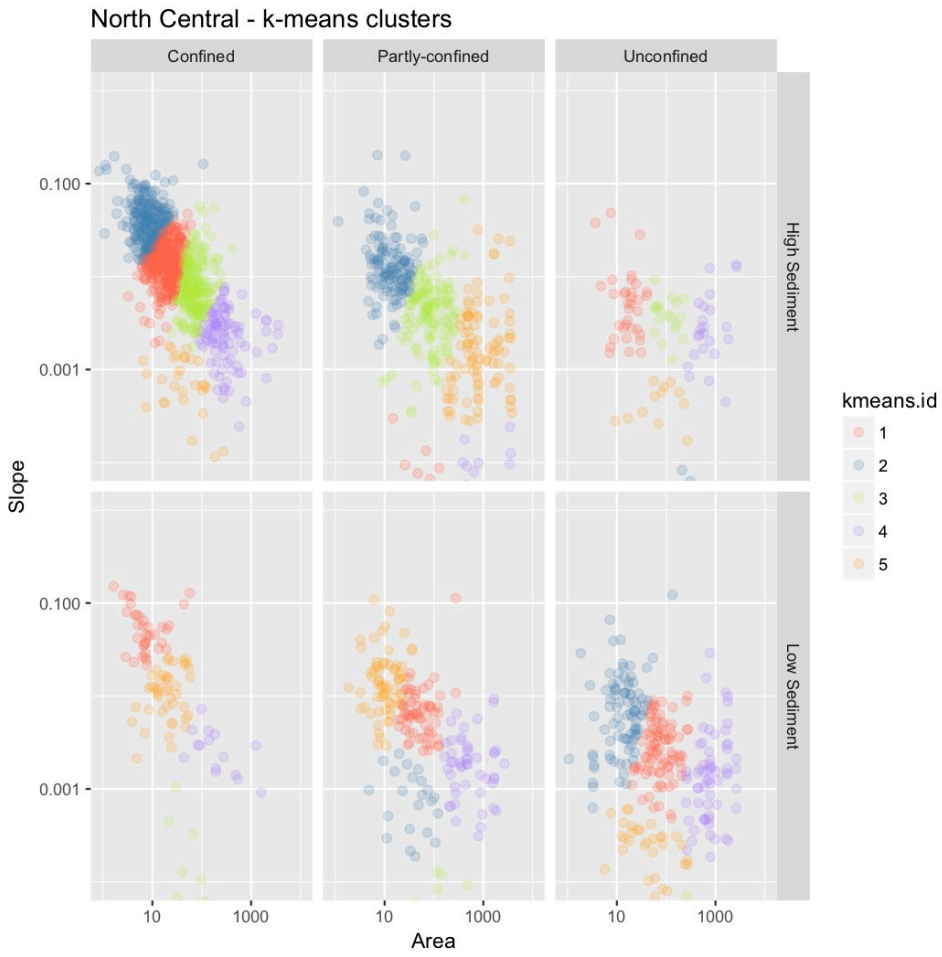


Figure 7. North Central Coast upper level bins with five area-slope bins.

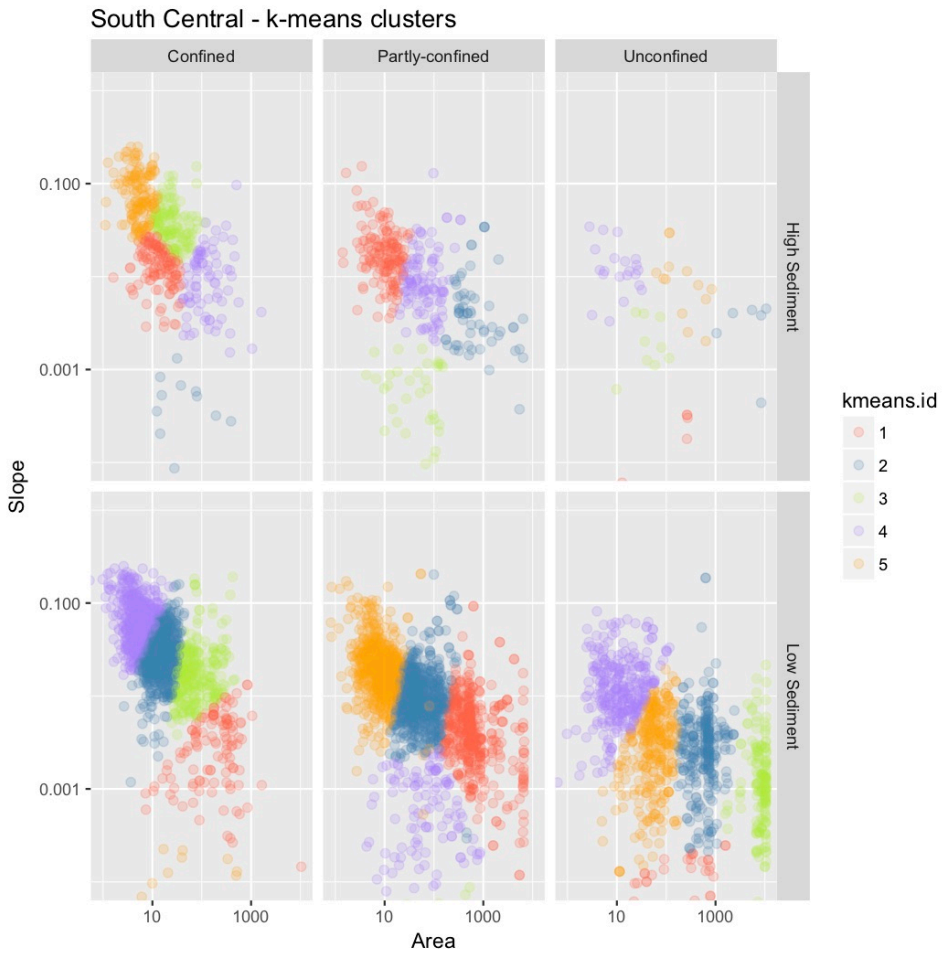


Figure 8. South Central Coast upper level bins with five area-slope bins.

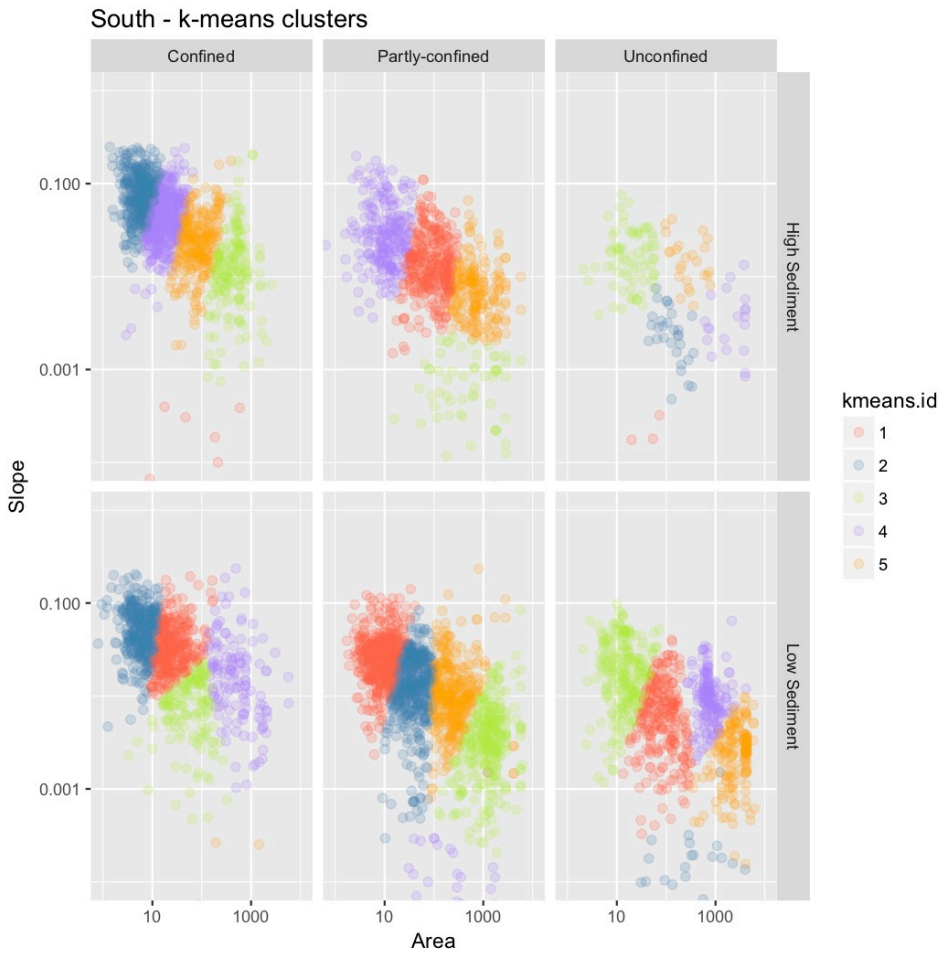


Figure 9. South Coast upper level bins with five area-slope bins.

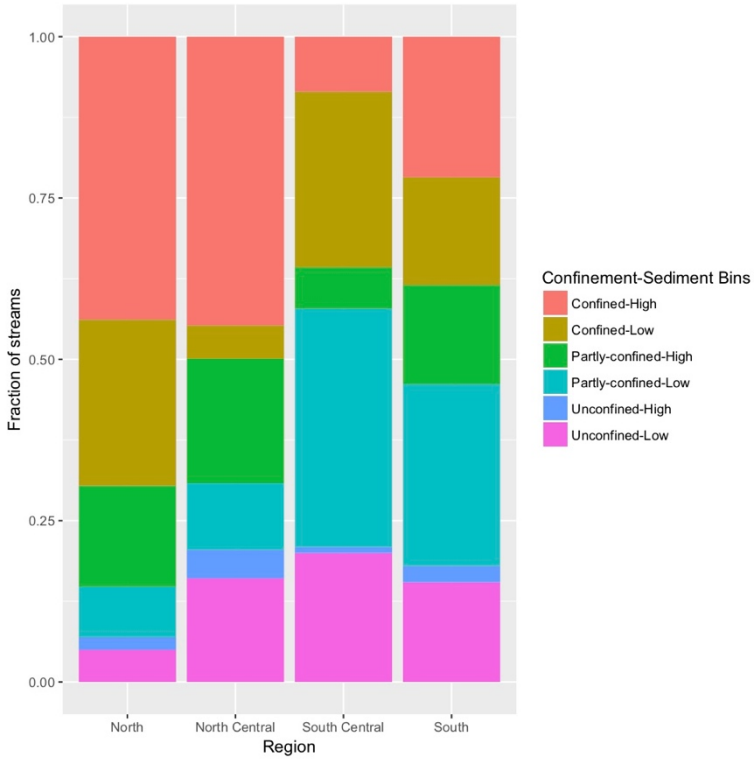


Figure 10. Proportion of sites within North, North Central, South Central, and South regions.

Following this two-tier stratification, regional streamlines will be classified by bin and site selection can proceed based upon specific region needs with a specified allotment of sites per bin based on the proportional analysis. In other words, several possible survey sites will be provided in each bin. Figures 11 through 14 provide an example of streamlines in each region with specific bins. Shapefiles will be provided with appropriate bin numbers.

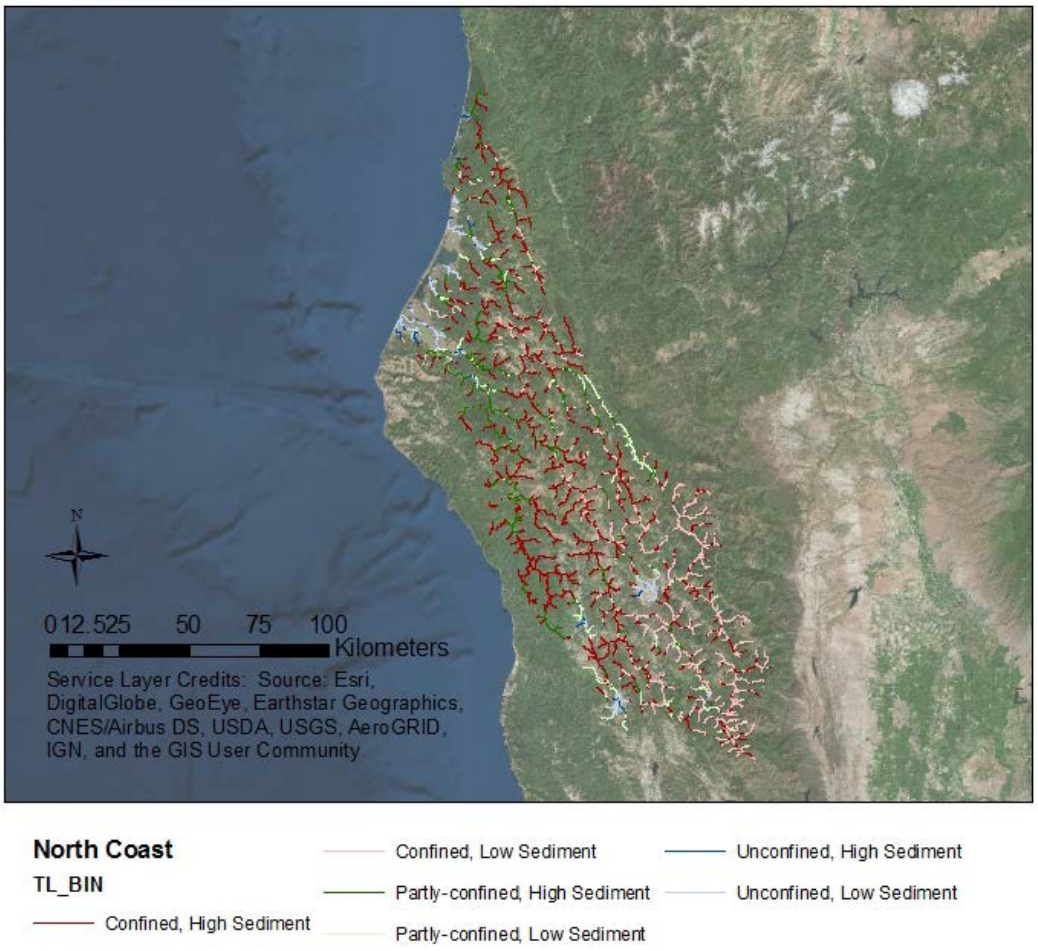


Figure 11. Distribution of streams in the North Coast region based upon upper and lower level bins.

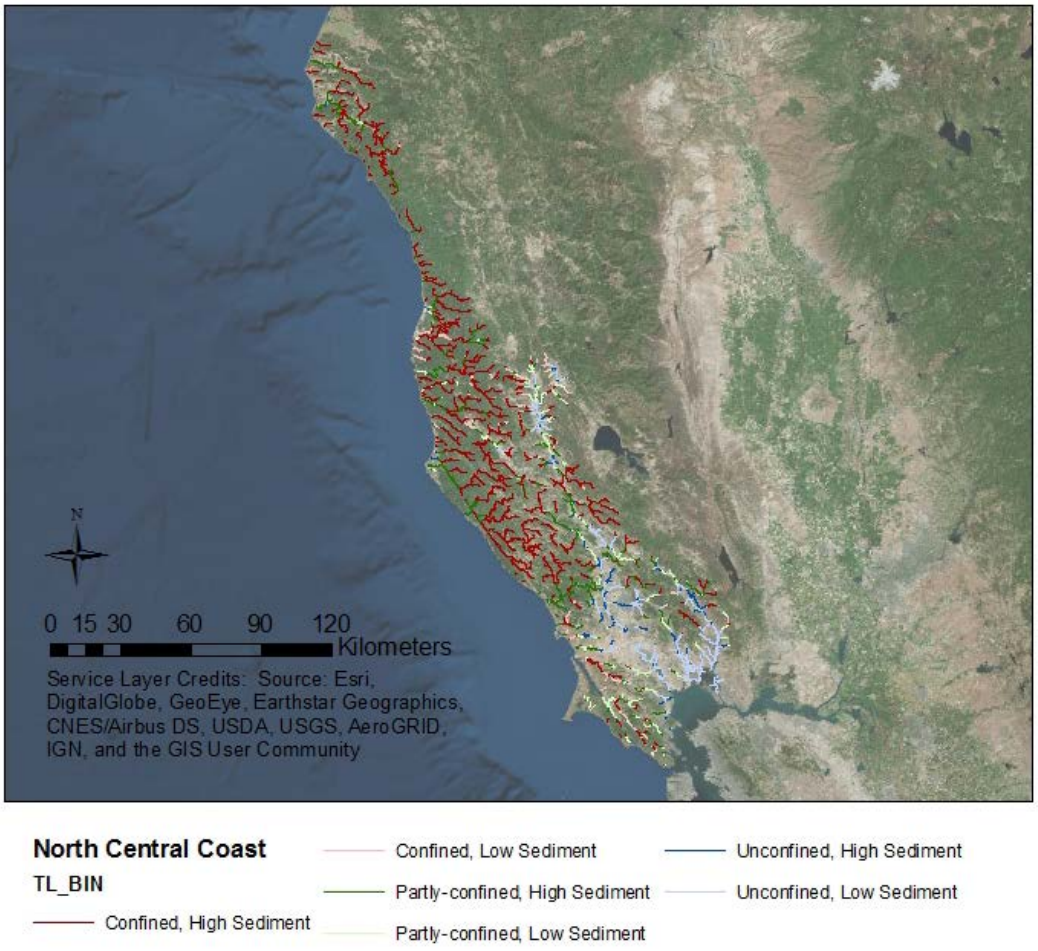


Figure 12. Distribution of streams in the North Central Coast region based upon upper and lower level bins.

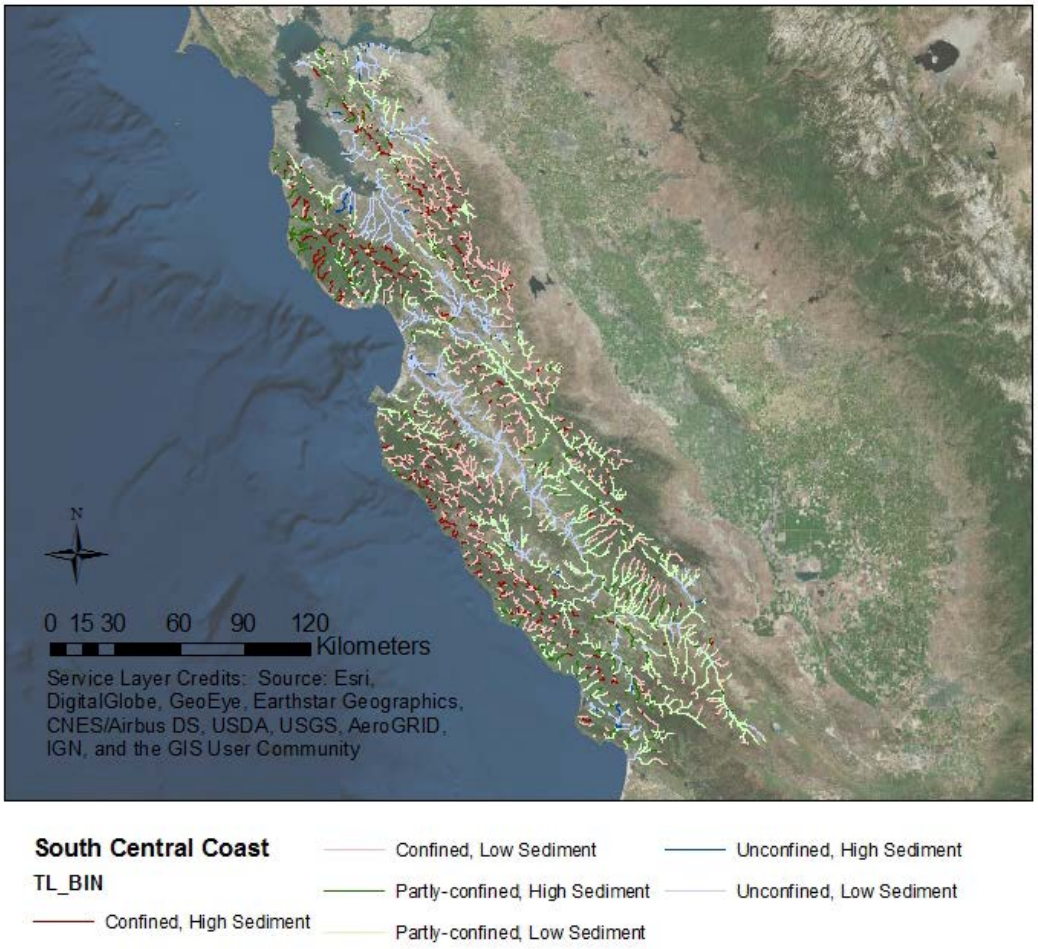


Figure 13. Distribution of streams in the South Central Coast region based upon upper and lower level bins.

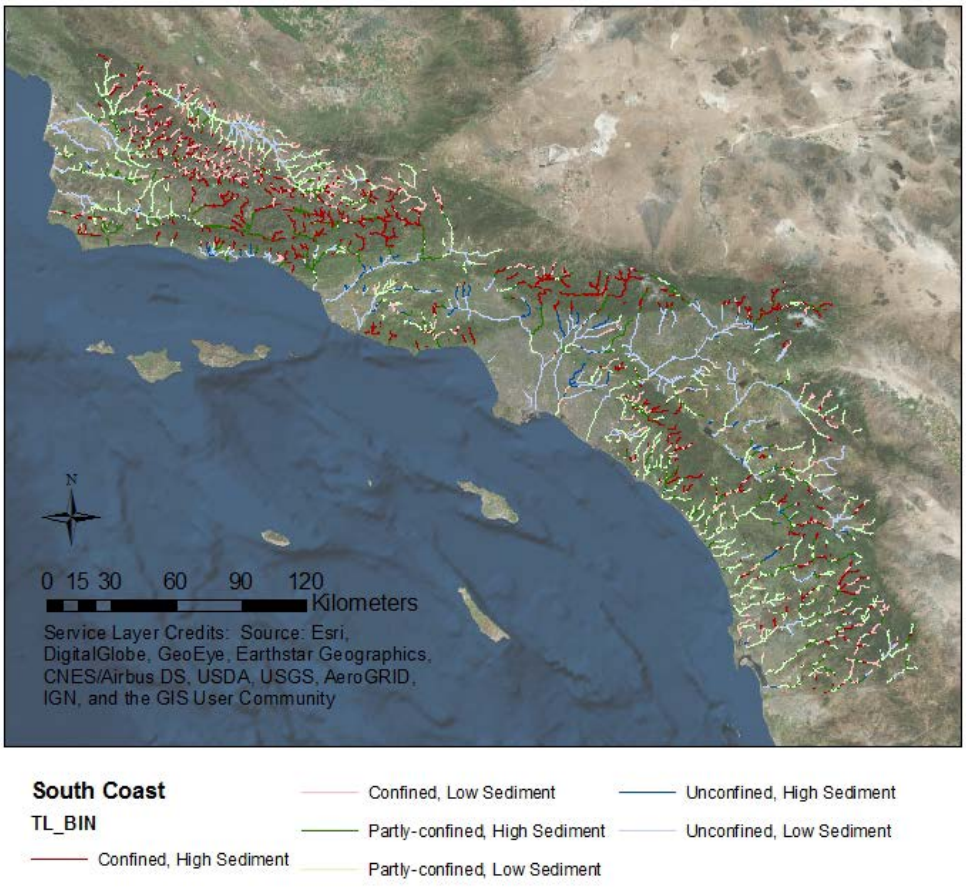


Figure 14. Distribution of streams in the South Coast region based upon upper and lower level bins.

References:

Gilbert, J.T., Macfarlane, W.W., Wheaton, J.M., 2016. The Valley Bottom Extraction Tool (V-BET): A GIS tool for delineating valley bottoms across entire drainage networks. Computers & Geosciences 97, 1–14. https://doi.org/10.1016/j.cageo.2016.07.014

Haan, C.T., Barfield, B.J., Hayes, J.C., 1994. Design hydrology and sedimentology for small catchments. Elsevier.

Hill, R.A., Weber, M.H., Leibowitz, S.G., Olsen, A.R., Thornbrugh, D.J., 2016. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. J Am Water Resour Assoc 52, 120–128. https://doi.org/10.1111/1752-1688.12372

Lane, B.A., Dahlke, H.E., Pasternack, G.B., Sandoval-Solis, S., 2017. Revealing the Diversity of Natural Hydrologic Regimes in California with Relevance for Environmental Flows Applications. JAWRA Journal of the American Water Resources Association 53, 411–430. https://doi.org/10.1111/1752-1688.12504

Renard, K.G., Foster, G.R., Weesies, G.A., Porter, J.P., 1991. RUSLE: Revised universal soil loss equation. Journal of Soil and Water Conservation 46, 30–33.

Appendix 2: Pictorial illustration of the workflow used to derive the fractal dimension

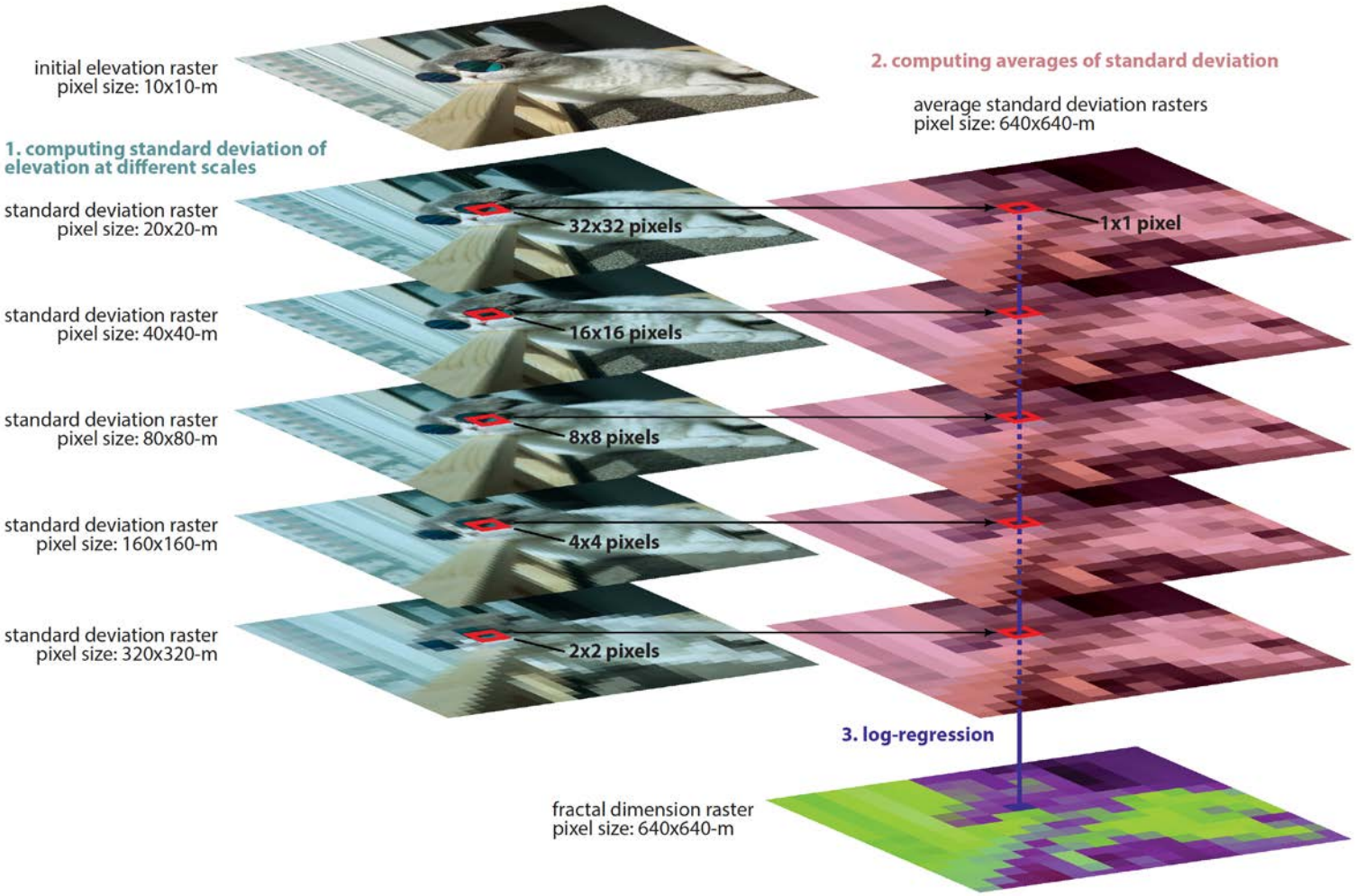


Figure 1: Pictorial illustration of the workflow used to derive the fractal dimension.