# Landscape classification of Pacific Northwest hydrologic units based on natural features and human disturbance to support salmonid research and management

Thomas R. Whittier and Alan T. Herlihy

Department of Fisheries & Wildlife, Oregon State University

and

Chris Jordan and Carol Volk

NOAA, National Marine Fisheries Service, Northwest Fisheries Science Center

June 13, 2011

Final Report for USBoR and NOAA Interagency Agreement
Reclamation IA No: 1425-06-AA-IC-4806
NWFSC IA No: 2RLEFRE

Based on work from
NOAA Contract #AB1133F10SE2464
with Oregon State University

## Executive Summary

To support salmonid research and management, we used national GIS coverages to develop two hierarchical landscape classifications of the 8,438 sixth-field hydrologic units (HUC6) in the Pacific Northwest, one based on natural features, and the other based on human disturbances. To develop the natural feature classification, we applied principal components analysis (PCA) and clustering techniques to scaled data for seven climate, land form, geology, and stream form variables. PCA showed a clear divide between Eastside and Westside landscapes. We then used a divisive clustering technique to divide the Eastside into a Mountains class and a Basins class. Thereafter, we used flexible beta clustering to develop landscape classes within each of these 3 top-level natural feature classes. The final natural feature landscape classification had 7 Westside, 8 Eastside Basins, and 9 Eastside Mountains classes. To develop the human disturbance classification, we determined proportion covered by urban land use, agricultural land use, and impervious surface, and road density in each HUC6. A flexible beta clustering of these scaled disturbance measures produced a balanced dendrogram, with the top-level division distinguishing low disturbance from high disturbance HUC6s. The final human disturbance classification had 8 classes that formed a continuum from essentially undisturbed to highly disturbed. The first principal component scores of a PCA of the four disturbance variables accounted for 65% of variability in the data, and can be used as an overall HUC6 disturbance measure. We evaluated the associations between the natural feature variables and classes, and the human disturbance variables and classes. Finally, we evaluated how well the Intensively Monitored Watersheds, an informal network of 22 salmon research and restoration projects, are distributed across the natural feature and disturbance classes.

## Introduction

Populations of the six anadromous salmon species in the Pacific Northwest (PNW) have been declining for well over a century (Nehlsen et al. 1991; Williams et al 1999; McDonald et al. 2007), due to a combination of over harvesting and widespread and often severe habitat degradation. The general habitat requirements for the various freshwater life stages (egg, larval, juvenile, returning adult) of each species are fairly well known. That knowledge has been used to develop and deploy a variety of in-stream and riparian habitat restoration actions, assumed to lead to increased salmon production, at least locally. To evaluate the effectiveness of restoration practices, a diverse set of fairly large scale projects has been established by various agencies, in which different actions (including no action) are studied in adjacent watersheds. Twenty-two of these projects across the PNW have been linked into an informal network called the Intensively Monitored Watersheds (IMW).

Due to the highly diverse character of PNW landscapes and the large geographic freshwater ranges of Pacific salmon it is not clear whether restoration practices that resulted in increased salmon production in one IMW project will necessarily be effective in other places. A landscape classification based on natural features known to be associated (positively or negatively) with salmon production could define areas of similar natural potential. Such a geographic framework could indicate areas where particular restoration actions could be expected to have similar results, as well as areas dissimilar enough to indicate less certainty about the chances of success. This framework could also be useful for evaluating whether the IMW projects are well distributed among the natural feature landscape classes or whether any “salmon landscapes” are not currently included.

There are multiple landscape classification systems based on various combinations of mapped natural features and human uses of the land that divide large geographic areas into hierarchies of ecological regions (ecoregions) (e.g., Omernik 1987; Bailey 1976). Each of these classifications was developed to support different intended applications (e.g., water quality assessment, conservation planning) often for different agencies or organizations (Loveland and Merchant 2004). Most of the widely used ecoregions systems were developed with qualitative methods to combine mapped landscape characteristics to delineate relatively homogenous regions (Omernik 2005; Loveland and Merchant 2005). In the last couple of decades, increased computing power and data storage, improved GIS software, matched with more detailed, consistently developed GIS coverages of ecological landscape data have lead to increased interest in using multivariate techniques to develop data-driven landscape classifications, assumed to be more objective (Hargrove and Hoffman 2005).

Despite the diversity of available landscape classifications, we are not aware of any developed to support salmon recovery research. To address this, we used GIS derived data and multivariate techniques to develop a classification of PNW watersheds based on the natural landscape features associated with salmonid production that should define areas of similar natural potential with regard to anadromous salmon. The resulting classification should be useful for structuring fisheries management and restoration efforts. Because humans are an integral part of the landscape, and because the kinds and intensity of anthropogenic stress are not evenly or even randomly distributed across the landscape, we also used the same approach to develop a separate human disturbance classification of PNW watersheds. We evaluated the associations between the natural feature variables and classes, and the human disturbance variables and classes. Finally, we evaluated how well the IMW research and restoration projects were distributed across the natural feature and disturbance classes.

## Methods

Operationally, we developed the natural feature classification before we began the human disturbance classification. To avoid repetition in this section we combine descriptions of both sets of analyses. We also present a subset of results in this section because we used results from early analyses to select the methods for later analyses.

### Study Area & Geographic Data

Our study covered all watersheds in the USA in the Pacific Northwest region (PNW), as defined by the USGS hydrologic unit codes (HUC) beginning with ”17": the entire Columbia River drainage within the USA, the Oregon and Washington coastal watersheds, and the Oregon interior draining watersheds. The study area included all of Washington and Idaho, most of Oregon, and portions of California, Nevada, Utah, Wyoming, and Montana. We made no attempt to delineate or remove areas inaccessible to anadromous fish, for either natural or anthropogenic reasons. The base geographic units for this study were the 8,438 sixth-field (12-digit) hydrologic units (HUC6) in the Pacific Northwest (Figure 1). In the USGS nomenclature HUC6s are subwatersheds; here we call them watersheds, recognizing that a large portion of USGS hydrologic units are not true watersheds (Omernik 2003).

We used several criteria to select landscape variables upon which to base the classifications. The natural feature landscape variables needed to be related to the ecology of the anadromous salmon life cycle. The natural feature variables should represent a variety of climate, landform and stream characteristics, while being independent of human influence. For both classifications there needed to be complete GIS coverage of the landscape attributes for study area, and each attribute needed to be expressed as a single value to represent that feature for each watershed.

For the natural feature classification, we evaluated numerous potential measures based on the criteria listed above. We evaluated the data distribution and geographic distribution characteristics of those variables, and relationships among the variables. Based on those analyses and our knowledge of issues with the underlying data, we kept seven variables, three climate, two land form, one geology, and one stream form, described below and summarized in Table 1.

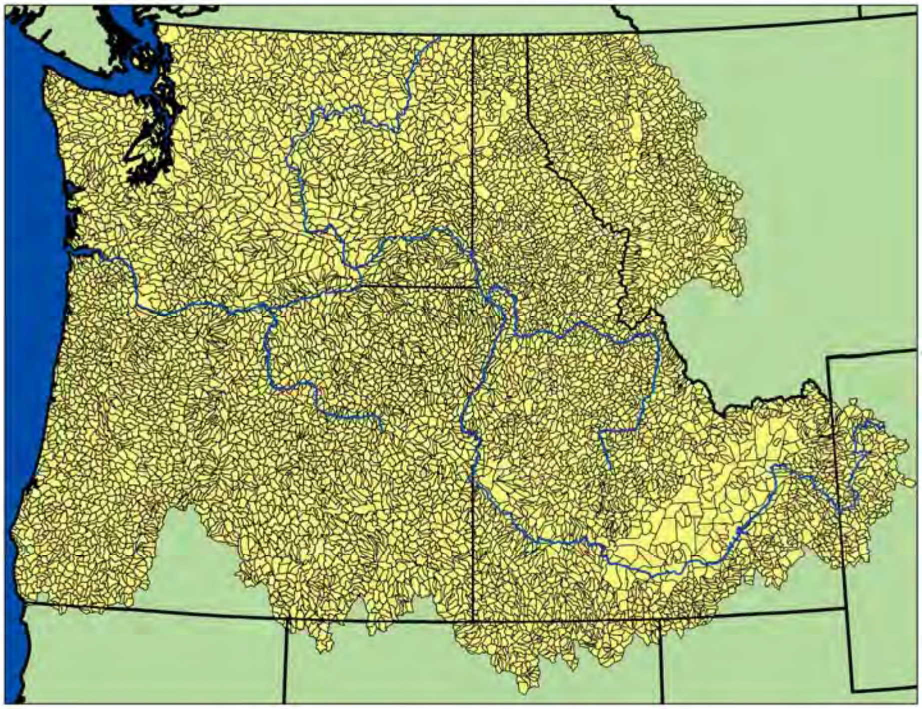


Figure 1. Map of the 8,438 sixth-field hydrologic units (HUC6) in the Pacific Northwest.

Table 1. Characteristics of the PNW $6^{th}$- field HUCs used in this study. Values are medians, interquartile ranges (IQR) and minimums and maximums. Watershed size was not used as an input variable in the classification process.

| | Median | IQR | Range |
|---|---|---|---|
| HUC6 surface area ($km^2$) | 73 | 53 - 102 | 2 - 870 |
| Elevation (m) | 1310 | 808 - 1681 | 3 - 3112 |
| Annual temperature range ($^o$C) | 35 | 31 - 37 | 16 - 45 |
| Growing degree-days | 1393 | 974 - 1817 | 19 - 3491 |
| Erodible geology (% surface area) | 12 | 0.2 - 43.6 | 0 - 100 |
| Annual precipitation (cm) | 66 | 39 - 114 | 17 - 486 |
| Watershed slope (deg) | 10.3 | 4.9 - 17.4 | 0 - 32.4 |
| Low gradient stream density (see text) | 0.0007 | 0.00002 - 0.02 | 0 - 25.8 |

Annual temperature range, annual precipitation, and growing degree day were based on 30-year mean values derived from 2 km grid PRISM data (http://www.ocs.orst.edu/prism/ for temperature range and precipitation; http://www.climatesource.com/products.html for growing degree day). Elevation and watershed slope were derived from the USGS 30 m raster digital elevation model. These five variables were each then resampled to 200 m using bilinear interpolation. The median pixel values were used to characterize the sixth-field watersheds. Erodible geology was derived from statewide bedrock and surficial geology maps. Geologic formations were classified into groups describing rock resistance to erosion. The surficial proportion of easily eroded material in each watershed was used as the erodibility metric. The low-gradient stream density was based on the 1:100,000 National Hydrology Dataset Plus and was calculated by squaring the length of channel with gradient $\leq$ 4% and then dividing by the watershed area.

For the human disturbance classification, we used the National Land Cover Database (NCLD) GIS coverages to calculate the proportion of land in each land cover type for each HUC6. We combined the appropriate specific NLCD land use class proportions to create the total proportion urban and the total proportion agricultural land uses. The NLCD has a separate impervious surface coverage which we used to develop the proportion of impervious surface in each watershed. We used the Census TIGER roads file to develop a road density measure for each watershed. We recognize that these are fairly broad human disturbance measures, and that other stressors (e.g., dams, grazing, clear cutting) likely have a more direct effect on salmon production. However, data related to each of these other stressors had problems, such as lack of consistent region-wide GIS coverages (e.g., grazing and forest practices), and unknown upstream and downstream effects (e.g., dams) that precluded their use.

To achieve approximately normal data distributions for the natural feature variables, annual precipitation values were natural log transformed, and watershed slope and low-gradient stream density values were cube root transformed. Growing degree day, annual temperature range, elevation and erodibility had approximately normal distributions and were not transformed. For the human disturbance variables, we $\log_{10}(x+1)$ transformed each measure. Despite the transformations, the high values ends of proportion urban, proportion impervious, and road density data distributions were highly skewed, such that the upper 1$^{st}$ percentile accounted for between 23% and 42% of the range of the transformed values. To partially reduce the effects of these high values we truncated the transformed values to the 99th percentile value for these three variables. All values (transformed, truncated and transformed, or untransformed as appropriate) were then scaled to a common 0 to 1 scale. These scaled values were used in all multivariate analyses.

Data Analysis

To elucidate multivariate gradients in the data, we ran separate principal components analyses (PCA) of the seven scaled natural feature landscape variables, and for the four scaled human disturbance variables, for all HUC6s. We plotted 1$^{st}$ and 2$^{nd}$ principal component scores (PC1 and PC2), and examined the amount of variability accounted for by PC1 and PC2, and the relative loadings of the input variables on those components. We examined the correlation matrix to assess potential redundancy among the input variables. The plot of PC1 and PC2 for the natural feature variables showed two distinct clouds, which generally corresponded with the west side and east side of the Cascades (Figure 2). This geographic division is commonly used in aquatic ecosystem research and management in the PNW and we eventually chose to use the sets of HUC6s as shown by the PCA, as the top level division in the natural feature classification (below). The human disturbance PCA did not show this pattern.

<u>Classification</u>

*Issues and Criteria* - Classification of multivariate data is usually accomplished by applying clustering methods. However, there are many clustering methods, and variations within those methods to choose among, as well as other choices, such as the appropriate number of classes (clusters) to be produced (Handl et al. 2005). Without clear objectives or tests, one runs the risk of selecting a classification that meets one's preconceptions. The correctness of a classification may be evaluated by assessing how well the behavior of a response variable is accounted for, or by how well novel (test) data fit into the classification. While salmon production would be an obvious response variable to test alterative classifications, those data are not available for sufficient numbers of watersheds. Likewise we do not have "extra" watersheds to test the classification(s).

Lacking objective data-driven criteria, we established the following criteria to evaluate our choices of methods and resulting sets of classes: 1) The classification should be hierarchical. 2) The method should not produce multiple very small classes, especially at the top (coarsest) levels. Instead, the cluster tree should be at least moderately balanced. 3) Clusters should have fairly good geographic cohesion, especially at the top levels. That is, single or small groups of watersheds in one cluster should generally not be surrounded by another cluster. We assessed this by examining cluster maps for a geographic checkerboard effect, at the top levels. 4) Results should not be greatly at odds with the established understanding of freshwater ecosystem processes and salmon ecology. For example, a clustering that produced a top-level class that had watersheds at the coast and along the crest of the Cascades with the other top-level class including most watersheds in-between would be rejected. Initially, we set a minimum class size to be greater than about 50 to 70 watersheds. Smaller classes are not likely to be useful for management purposes, at this scale.

To select among clustering methods, we ran the full seven natural feature variables through several methods available in SAS and PC-ORD. We examined dendrogram structure (for hierarchical methods), compared top-level cluster membership among methods, examined boxplots of the seven input variables for each cluster, and mapped top-level clusters. For a subset of the clustering methods we repeated the above evaluation in a series of leave one out clusters with reduced variable lists. Based on these exploratory analyses, we chose to use a flexible beta method as our primary clustering method. Flexible beta is a generalized form of most other agglomerative hierarchical methods and is widely used in ecology. It takes advantage of a flexibility in the combinatorial equation that calculates cluster dissimilarity. Flexible beta allows the user to control the linkage's space-distorting properties. As beta approaches 1, it is increasingly space contracting, with chaining approaching 100%. As beta approaches zero and then becomes negative, the method ceases to be space contracting and becomes increasingly

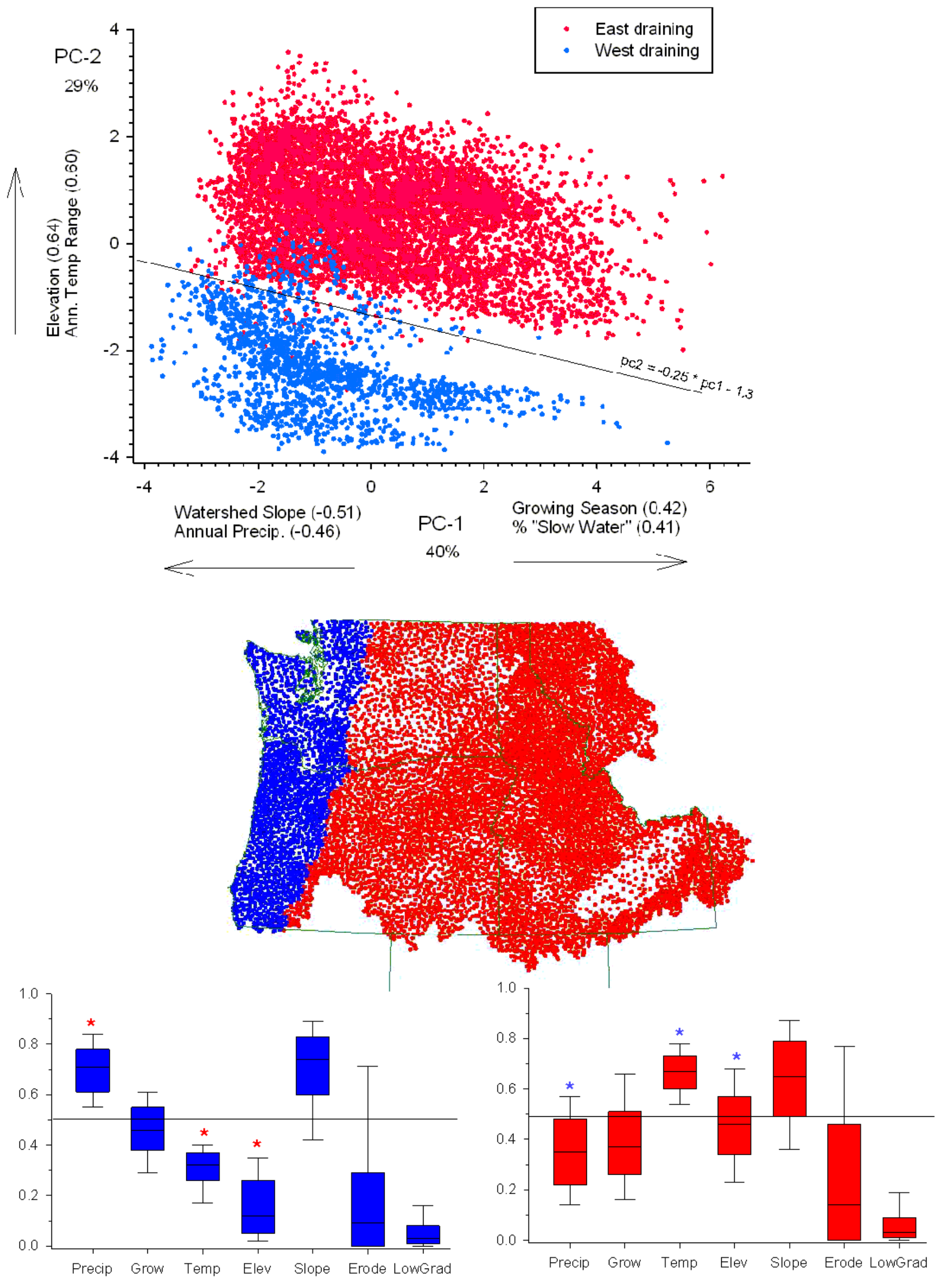



Figure 2. Plot of PCA axis 1 (PC-1) versus axis 2 (PC-2) scores for the 8,438 HUC6 in the Pacific Northwest. HUCs are color coded as either draining west or east of the Cascade Mountains crest. Box and whisker plots show scaled 0-1 scores for the natural feature variables in the PCA.

space expanding and the elements are more intensely grouped. The default beta value (-0.25), which we used, tends to produce well-balanced dendrograms.

*Establishing Top-Level Natural feature Classes* - We initially assumed that the west side versus east side divide seen in the PCA plot (Figure 2) would be in close agreement with the top-level classes of most clusterings of the data. However, none of the multiple varieties of clustering we tried produced anything close to this. Among the methods that produced fairly balanced dendrograms, the top clusters produced a mountainous class of watersheds and a low-gradient class of watersheds, in which, for example, the Willamette Valley and Puget Lowlands watersheds were in the same top cluster as the Oregon High Desert and Columbia Plateau watersheds. We attributed this result to the complexity of this large data set, in which there are multiple data structures. The clustering algorithms apparently keyed in on the land form (steep versus flat) and the erodibility structure in the data, while most fisheries ecologists would emphasize the climatic structure at the scale of the whole PNW.

Thus, we chose to use the two ordination-space clouds from the natural feature PCA as the initial (top-level) landscape classes (Figure 2). These data-derived classes met all of our criteria (above) and were distinguished primarily by precipitation, temperature range, and elevation. This step parsed the data structure complexity to the extent that flexible beta clustering of the westside class (n=1,578) yielded results that met our acceptance criteria.

However, there remained a challenging level of data structure complexity in the larger (n=6,860) eastside class. Here, the top-level flexible beta clusters produced an undesirable amount of geographic checkerboarding (i.e., there were numerous single watersheds and small groups surrounded by the other cluster). In addition, it appeared that erodibility was the key variable distinguishing the top two clusters, while to most aquatic ecologists, land form (steeper versus flatter landscapes) and climate factors should define the top-level eastside classes. A PCA of the eastside data did not produce any distinct groups of watersheds. We then applied the SAS FastClus procedure, a non-hierarchical divisive algorithm (set for two clusters) to the eastside data. This method uses an iterative nearest centroid sorting algorithm for minimizing the sum of squared distances from cluster means, beginning with random cluster seeds and at each iteration, replacing seeds with the new cluster means until no further changes occur. FastClus produced two fairly geographically cohesive clusters that divided the eastside into a mountainous (and hilly) class (4,880 watersheds) and a relatively flat landscapes (mostly deserts and basins) class (1,980 watersheds) (Figure 3). These classes showed a fairly good match with landform, when plotted over a shaded topographic-relief map (not shown). There was some degree of geographic checkerboarding at this level that reflected the distribution of basins and ranges in the northern Great Basin and occurrence of large broad valleys in some mountainous areas. These two classes were primarily distinguished by precipitation, growing season, watershed slope and erodibility, and to a lesser extent by elevation and stream gradient.

*Final classification* - Based on this work and additional clustering trials, we chose to develop natural feature classifications of watersheds within the three top-level classes described above (Westside, Eastside Mountains, Eastside Basins), by applying flexible beta clustering (beta = -0.25) to each class separately. We examined dendrogram structure and boxplots of the seven input variables, and mapped the member watersheds for the clusters at every dendrogram branching for

2 through ~ 15 clusters. Where the flexible beta distance (BD) (a measure of dissimilarity) between dendrogram branch points (off the same branch) was less than about 1, we treated the set of branches as a unit. We initially chose to prune the dendrogram above branches that produced clusters with less than about 70 watersheds, rather than using a set beta distance throughout.

To develop the human disturbance classification, we used flexible beta clustering, evaluated by mapping HUC6 locations and examining boxplots of input variables for each dendrogram branch pair for the entire PNW.

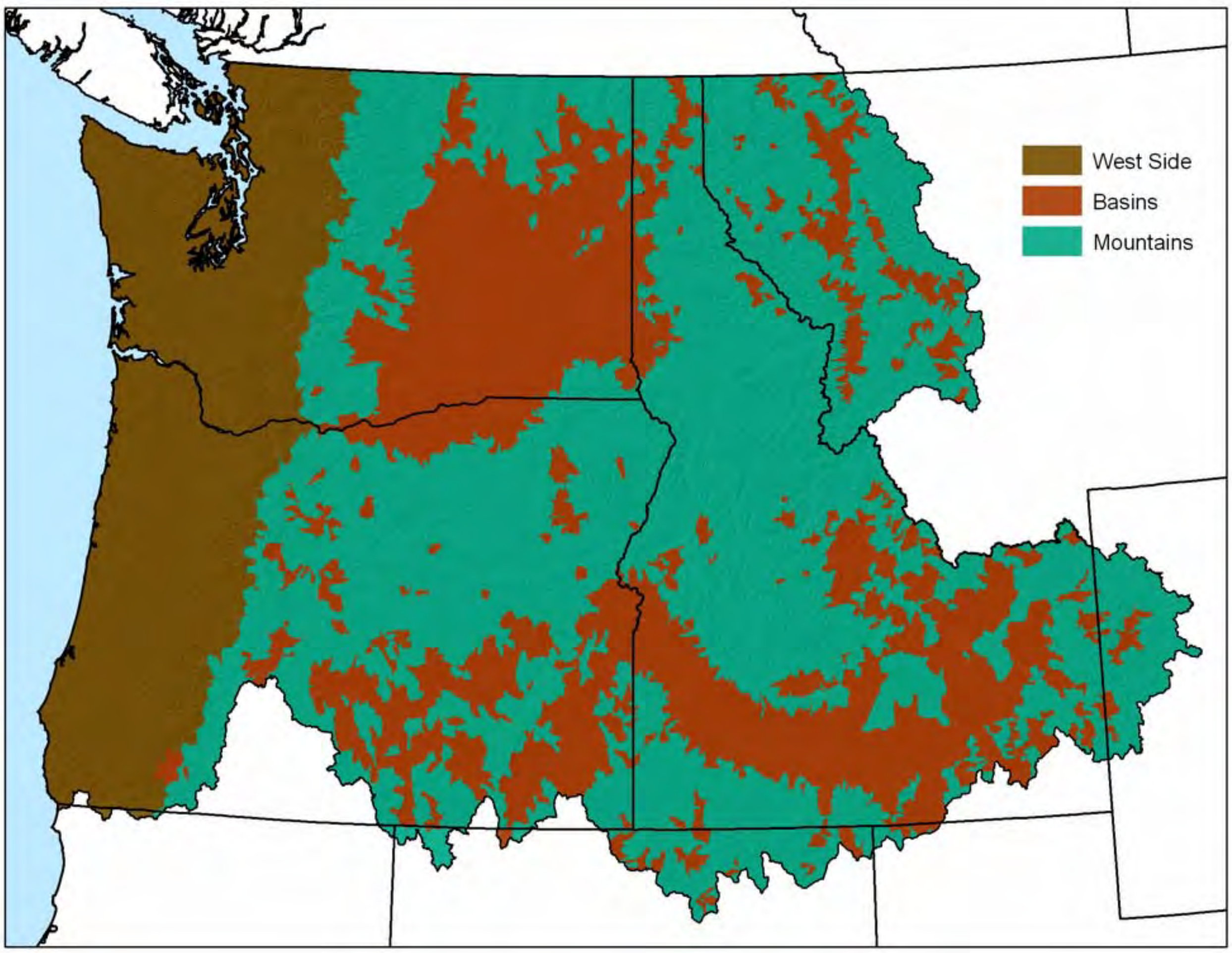



Figure 3. Map showing the location of the HUC6 among the three top level natural feature classes (Westside, Eastside Basins, and Eastside Mountains).

*Assessing the number of final classes*- In cluster analysis, there is no clear-cut empirical method to determine the best number of clusters for a classification (Handl et al. 2005). We used professional judgment, based on maps of HUC6s for each dendrogram branch pair and comparisons of the boxplots of the landscape variables, to decide at what level we ceased gaining useful distinctions for new classes. There are a variety of analyses that can be used to evaluate those decisions. For each of the natural feature classifications we calculated mean within-cluster similarities (Euclidean distance) for 2 through 50 clusters generated by three widely used hierarchical agglomerative clustering methods: flexible beta, Ward's method, and k-means clustering. Then, using the flexible beta clusters, we plotted three other statistical measures to evaluate cluster strength for 2 to 50 clusters. First, we calculated a pseudo F statistic which is the ratio of the mean sum of squares between groups to the mean sum of squares within groups. Large values of the pseudo F indicate better clusters. Second, we calculated the Anosim R which compares the within cluster dissimilarities (distance) to the between cluster dissimilarities based on ranks rather than values. Third, we calculated the KGS K statistic at each level of the tree as the mean dissimilarity across all clusters. After normalizing, the number of clusters times alpha is added. The minimum of this function over all levels (numbers of clusters) corresponds to the suggested pruning size.

<u>Associations between Natural feature and human disturbance</u>

We examined the associations between the set of natural feature input variables and classes, and the set of human disturbance input variables and classes, for each of the three top-level natural feature classes. We produced contingency tables of natural feature classes and disturbance classes at the finest classification resolution, and plotted the percent of HUC6s in each cell. We ran correlations of the natural feature variables and disturbance variables, as well as the disturbance PC1 for each of the three top-level natural feature classes. Because we expected human disturbance to be negatively associated with elevation, we made box plots of watershed elevations for each disturbance class, separately for the west side and the east side.

<u>IMW assessment</u>

We selected 22 projects that participate in the network of Intensively Monitored Watersheds. In GIS we joined the IMW location data with the HUC6s to match the natural feature classes and disturbance classes with those IMW projects. We summarized and mapped the distribution of natural feature classes in the IMW projects, and plotted the distribution of disturbance PC1 scores for the IMW HUC6s for each top-level natural feature class. We also plotted the IMW HUC6s on the natural feature PC1 by PC2 plot to show their distributions in natural feature ordination space.

## Results

Geographic characteristics

While our study watersheds are generally of similar size (Table 1), they are quite diverse with regard to other physical and climate characteristics. For example, mean elevations range from 3 to over 3100 m, mean watershed slopes range from essentially flat to over 32 degrees, and annual precipitation ranges from 17 to over 480 cm (Table 1). Five natural feature variable pairs were moderately well correlated with each other (|r| between 0.57 and 0.73) (Table 2). Proportion of highly erodible land was the least correlated with any of the other variables (|r|<0.39). The distributions of the raw values for the four human disturbance measures were highly skewed (Table 3). A high portion of watersheds had zero urban land use (40%) or agricultural land use (45%), while nearly all watersheds had some roads (96%) and impervious surface (95%). The transformed, truncated and scaled disturbance measures were all positively correlated (Table 4).

Table 2. Pearson correlation coefficients among the scaled natural feature variables for the 8,438 6$^{th}$-field HUCs in the PNW. Precip = annual precipitation; Grow = growing degree day; Temp = annual temperature range; Elev = elevation; Slope = watershed slope; Erode = proportion of watershed in highly edrodible geology; Low-Grad. = low gradient stream density.

| | Precip | Grow | Temp | Elev | Slope | Erode |
|---|---|---|---|---|---|---|
| Grow | -0.42 | | | | | |
| Temp | -0.73 | -0.02 | | | | |
| Elev | -0.15 | -0.69 | 0.59 | | | |
| Slope | 0.59 | -0.50 | -0.24 | 0.26 | | |
| Erode | -0.23 | 0.17 | 0.13 | -0.09 | -0.38 | |
| Low-Grad. | -0.34 | 0.36 | 0.13 | -0.20 | -0.57 | 0.29 |

Table 3 Raw data distribution characteristics of landscape-scale human disturbance measures.

| | % HUCs with variable = 0 | 99$^{th}$ percentile value | maximum value |
|---|---|---|---|
| % Urban | 40 | 23.5% | 80.9% |
| % Agricultural | 45 | 88.2% | 97.2% |
| % Impervious Surface | 4 | 8.5% | 47.6% |
| Road Density | 5 | 36.6 | 109 |

Table 4 Correlations (Pearson's r) among the transformed, truncated and scaled disturbance measures.

| | Urban | Agricultural | Impervious Surface |
|---|---|---|---|
| Agricultural | 0.59 | | |
| Impervious surface | 0.81 | 0.37 | |
| Road density | 0.55 | 0.30 | 0.54 |

Multivariate Gradients

In the PCA of the natural feature landscape variables, the first two principal components accounted for 69% of variability (Table 5). Watershed slope, precipitation, growing season (growing degree day) and low-gradient streams loaded onto PC1, and elevation and temperature range loaded onto PC2. The watersheds formed two fairly distinct clouds of points when plotted on PC1 and PC2 (Figure 2 top). The watersheds in the two clouds generally comprised the west side and east side of the Cascades (Figure 2 middle). This is widely recognized as the most important freshwater ecosystem divide in the PNW. These watersheds were distinguished primarily by annual precipitation, annual temperature range and elevation (Figure 2 bottom). This was not strictly a drainage split, approximately 30 eastward-draining watersheds were in the west side ordination space and approximately 50 westward-draining watersheds were in the east side ordination space (mostly in the south).

The PC1 of a PCA of the disturbance measures accounted for 65.3% of variability in the data. The four disturbance measures loaded onto PC1 fairly evenly, with eigenvectors ranging from 0.42 (agricultural land use) to 0.58 (urban land use). We judged that PC1 scores could be used as an overall landscape human disturbance measure for the watersheds.

Natural feature Classification

As described in Methods (above) our top-level natural feature classification of HUC6s divided the PNW watersheds into a west side class (n=1,578), an east side basins class (n=1,980) and an eastside mountains class (n=4,880). Here, we present the west side classification results in some detail, and the two east side classes in less detail.

*West side clusters* - The flexible beta clustering met our acceptance criteria for up to between 7 and 13 clusters; the dendrogram was reasonably well balanced, the clusters exhibited fairly good geographic cohesion (although some geographic checkerboarding occurred at the 13 clusters level, (see below). The mapped clusters matched the established understanding of landscape structure such that we could name the clusters, at least to the seven cluster level.

The top branch point (BD=38) (Figure 4) divided the watersheds into an uplands/steep class (n=1,322) versus a lowlands/flat class (n=256). Elevation, watershed slope, erodibility and low gradient stream density were markedly different between these classes (Figure 5). The second branch point (BD=18) divided the uplands/steep class into a mountains class (n=1,041) versus a foothills fringe class (n=281) that included portions of the Rouge Valley, and numerous watersheds adjacent to the coast. Again, elevation, watershed slope, erodibility and low gradient stream density distinguished these classes. The third branch point (BD=12) divided the mountains watersheds into a high Cascades/Olympics class (n=362) versus a Coast Range and low Cascades class (n=679) that differed primarily in elevation, growing season and low gradient stream density. The input variables (except for erodibility), exhibited less variability in high Cascades and Olympics class than in the Coast Range and low Cascades class.

The fourth west side branch point (BD=8) divided the Coast Range and low Cascades watersheds into a low Cascades and (generally) southern Coast Range class (n=495) versus a northern Coast Range class (n=184) that also included about 20 southern Coast Range watersheds near the ocean. These classes were distinguished primarily by precipitation, growing season, and temperature range. Two separate branches divided at BD=6. The lowlands class divided into a Willamette Valley and Puget Lowlands class (n=161) versus a low coastal and Chehalis Valley class (n=95) that included about 15 Puget Lowlands watersheds. These classes differed in precipitation, growing season, temperature range and watershed slope. We chose not divide these classes further. The other branch point at this level divided the foothills fringe class into a Puget foothills and low coastal watersheds class (n=137) versus an Oregon foothills class (n=148) that comprised the Willamette foothills, the Rouge Valley and Roseburg Valley. These classes were distinguished primarily by precipitation, growing season, and temperature range, and to a lesser extent by watershed slope. We chose not to divide these classes further.

The three mountainous westside branches divided into two, three and four clusters at BD of about 4. Although we examined these smaller classes, we chose to stop at the 7-cluster classification (Figure 4 & 5; Table 6) for the west side watersheds described above: two low-elevation relatively flat classes, and two foothills classes and three mountains classes, with between 95 and 495 watersheds.

This choice was supported by plots of mean within-cluster similarities for 2 through 50 clusters generated by flexible beta, Ward's method, and k-means clustering (Figure 6 top). For all methods, mean within-cluster similarity increased fairly quickly until about six or seven clusters, with little difference among the three methods. For the first 50 flexible beta clusters (Figure 6 bottom) the pseudo F statistic began leveling off at around 10 cluster, the Anosim R values began leveling off at around 7 clusters, and the KGS values reached a minimum in the 5 to 7 clusters range.

Table 5. Principal components analysis of seven natural feature variables for the 8,438 $6^{th}$ -field HUCs in the Pacific Northwest. Eigenvalues and proportions of variability accounted for by the first two principal components are shown, as are the PC1 and PC2 eigenvectors (loading) for each variable. Precipitation was natural log transformed, watershed slope and low-gradient stream density were cube root transformed. All variables were then scaled to a range of 0 to 1.

| | Eigenvalue | Proportion |
|---|---|---|
| PC1 | 2.82 | 0.40 |
| PC2 | 2.03 | 0.29 |
| | Loading | |
| | PC1 | PC2 |
| Climate | | |
| Annual precipitation | -0.46 | -0.34 |
| Annual temperature range | 0.23 | 0.60 |
| Growing degree day | 0.42 | -0.34 |
| Land form | | |
| Elevation | -0.18 | 0.64 |
| Watershed Slope | -0.51 | 0.03 |
| Geology | | |
| % Watershed as highly erodible | 0.30 | 0.01 |
| Stream form | | |
| Low-gradient stream density | 0.41 | -0.08 |

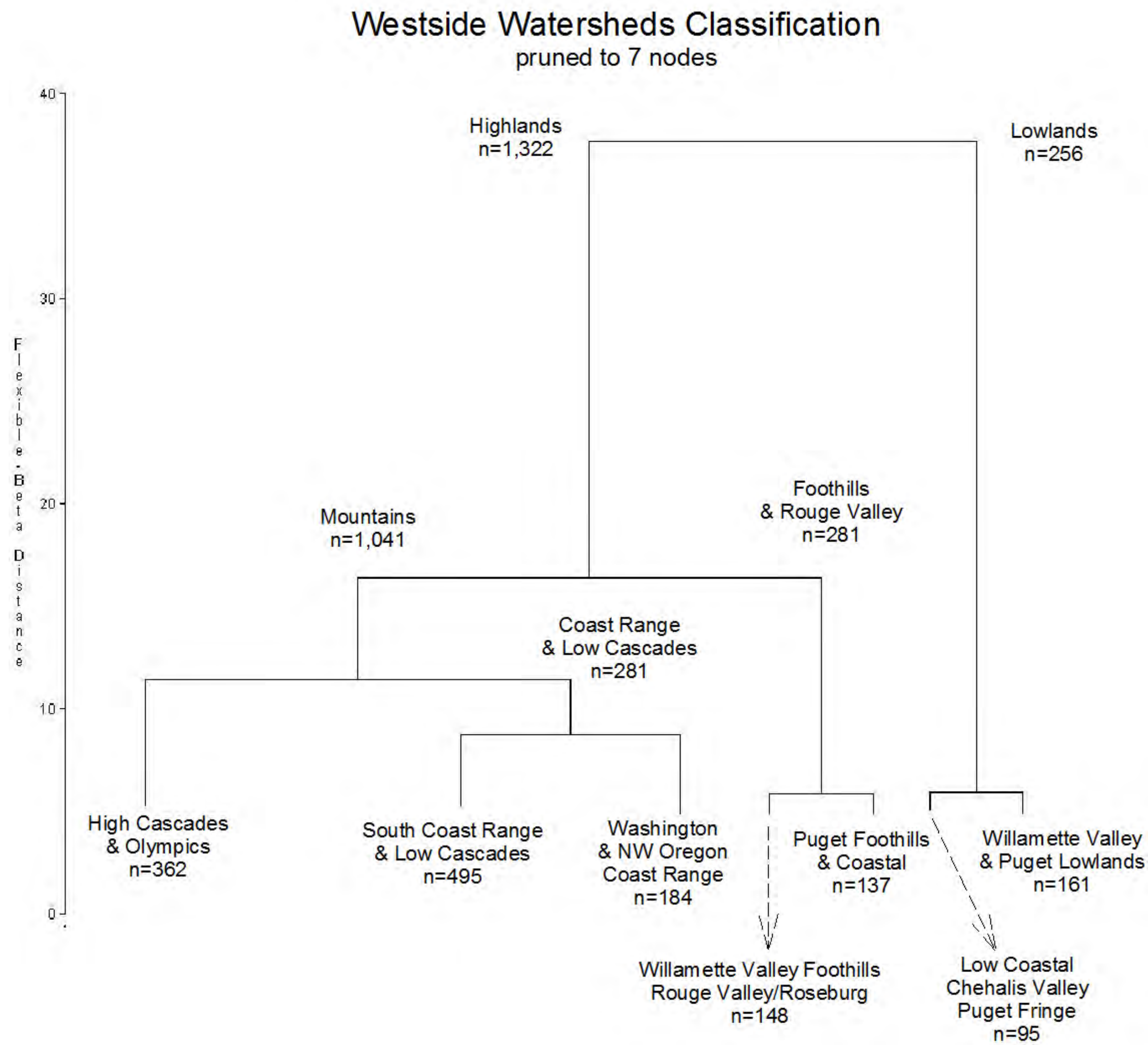



Figure 4. Natural feature cluster dendrogram for the Westside HUC6.

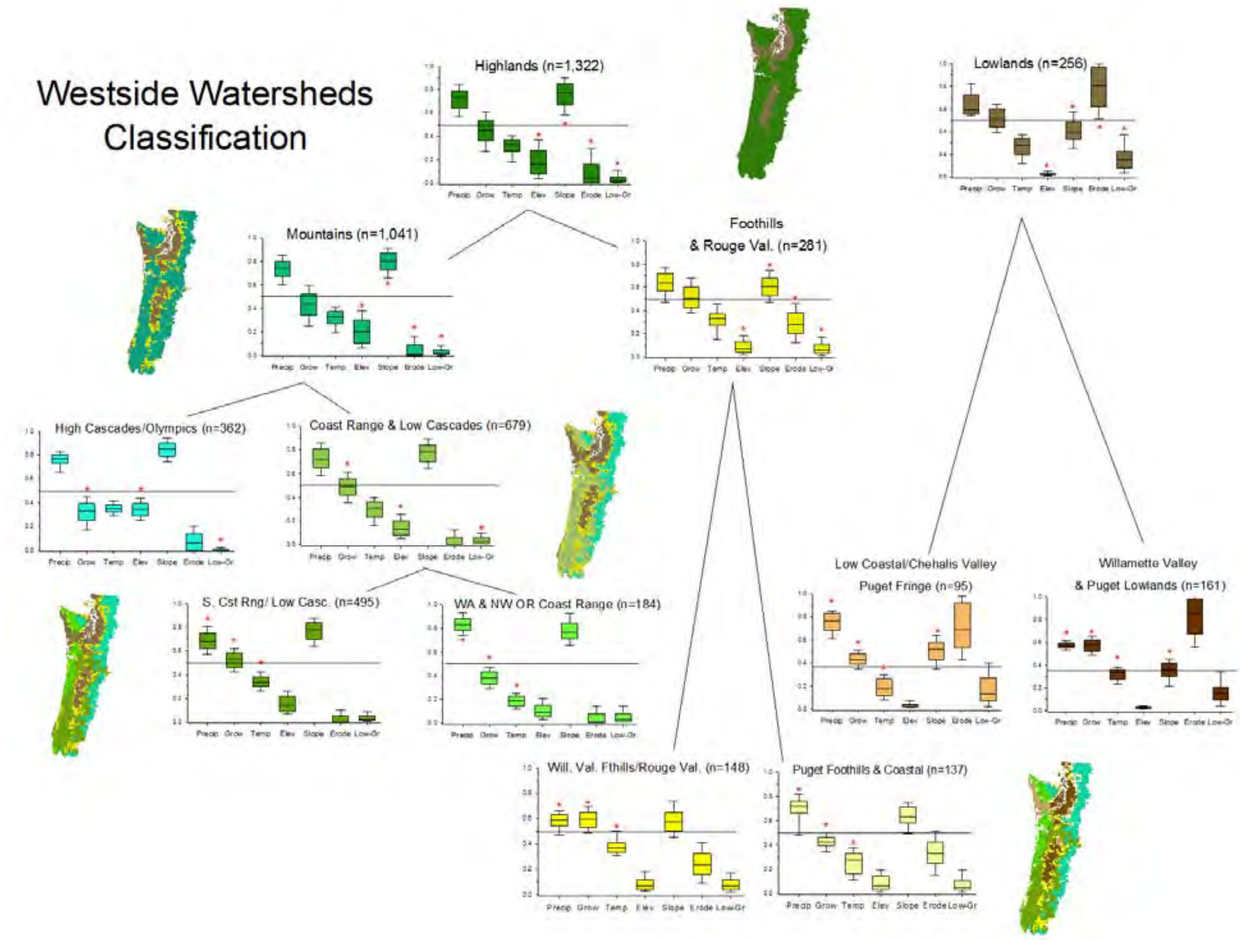



Figure 5. Geographic location and characteristics of the Westside natural feature classes. Box and whisker plots show scaled scores for the natural feature variables each class. Box colors match classes to map locations.

Table 6. Westside natural feature classes, the number of Intensively Monitored Watersheds (IMW) and the total number of HUC6 in the PNW in each class.

| Natural feature Class | IMW basins | PNW HUCs |
|---|---|---|
| I. Lowlands | 3 | 256 |
| A. Willamette Valley & Puget Lowlands | 2 | 161 |
| B. Low Coastal & Chehalis Valley | 1 | 95 |
| II. Mountains & Uplands | 29 | 1,322 |
| A. Foothills, Rouge Valley & Roseburg Valley | 1 | 281 |
| 1.Puget Foothills & Coastal | -- | 137 |
| 2.Willamette Foothills, Rouge Valley | 1 | 148 |
| B. Mountains | 28 | 1,041 |
| 1. High Cascades & High Olympics | -- | 362 |
| 2. Coast Range, Southern/Low Cascades | 28 | 679 |
| a. Southern Coast Ranges & low Cascades | 17 | 495 |
| b. Washington & NW Oregon Coast Range | 11 | 184 |

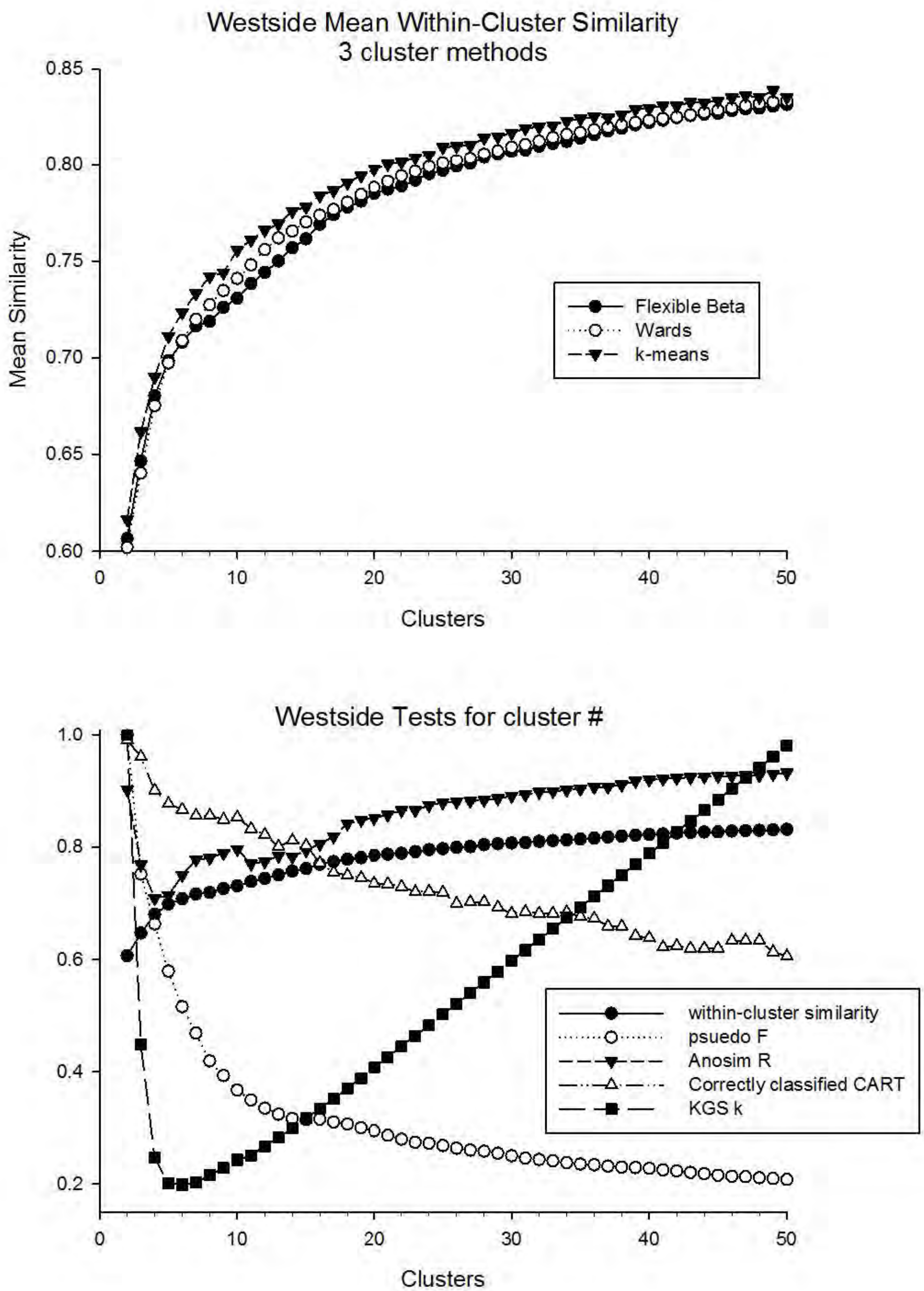



Figure 6. Changes in the statistical attributes of the Westside natural feature clustering with increasing number of clusters.

*East side clusters* - It was more difficult to find convenient names for most east side classes (both Basins and Mountains), than it was for the west side. Potential class names tended to contain multiple descriptions of key variables that were relative to the parent class (e.g., wetter, cooler). With regard to the final number of classes (clusters), for both top level east side classes, mean within-cluster similarity increased fairly quickly until about seven to ten clusters (see below).

*Eastside Basins* - The top branch divided the watersheds into a hot dry basins class (n=1,003) primarily in the Columbia Plateau and Snake River Plain ecoregions versus a more temperate basins class (n=977) primarily in the Rocky Mountains and the Idaho Batholith.

Precipitation and growing season were the major differences between these classes (Figure 7). The second branch divided the hot, dry basins class into a lower, hotter, erodible class (n=421) mostly in the Columbia Plateau and Snake River Plain versus a higher, less hot, drier class (n=582) that included the Northern Basin and Range ecoregion and foothill areas around the other class. Growing season, elevation, erodibility and precipitation distinguished these classes. The third branch divided the temperate basins watersheds into three classes: a cooler, higher class (n=321), a southern warmer, lower class (n=266) and a northern warmer, lower class (n=390). There was a mix of differences in all seven input variables among these three classes. Subsequent divisions of these five classes of Basins watersheds were into more erodible and less erodible classes of watersheds.

The hot, high basins class divided into a higher, less hot, more erodible class (n=253) and a hotter, lower, less erodible class (n=329). The hot, low, dry, erodible basin class (Columbia Plateau and western Snake River Plain) divided into a very erodible class (n=254) and a moderately erodible class (n=167). The southern warm/lower class (from the temperate basins class) divided into a low gradient streams class with highly variable erodibility (n=105) and a highly erodible class (n=161), with fewer low gradient streams. Stopping at this point produced 8 classes (Figure 8; Table 7).

The results of the assessments aimed at determining an appropriate number of classes, for the east side Basins watersheds, were similar to those on the west side. The mean within-cluster similarity plot began to level off at around 6 to 10 clusters, the pseudo F at around 8 to 10 clusters, and the Anosim R at 10 clusters. The KGS K-statistic reached a minimum at 6 to 8 clusters.

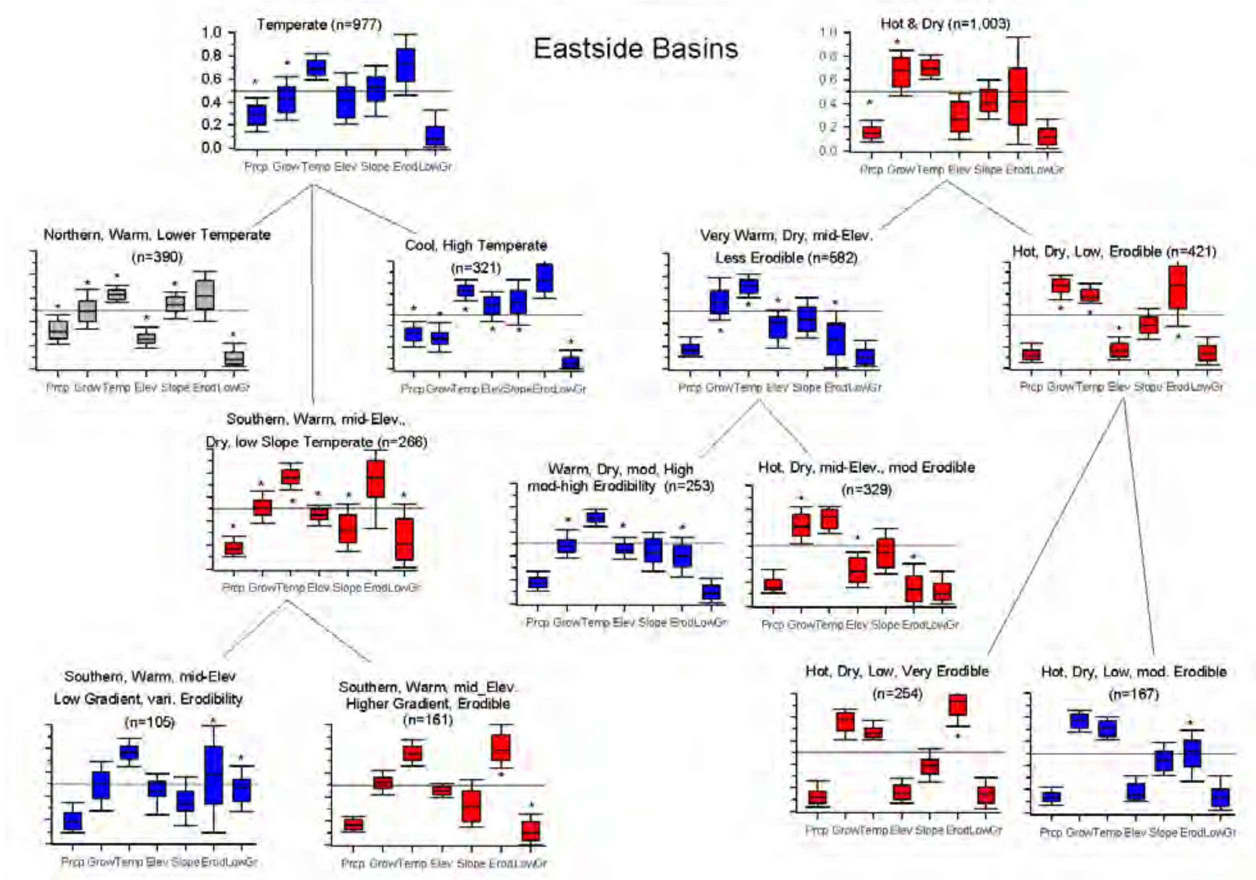



Figure 7. Characteristics of each level of the Eastside Basins natural feature classification. Box and whisker plots show the scaled 0-1 scores for each natural feature class. See Table 2 for variable definitions

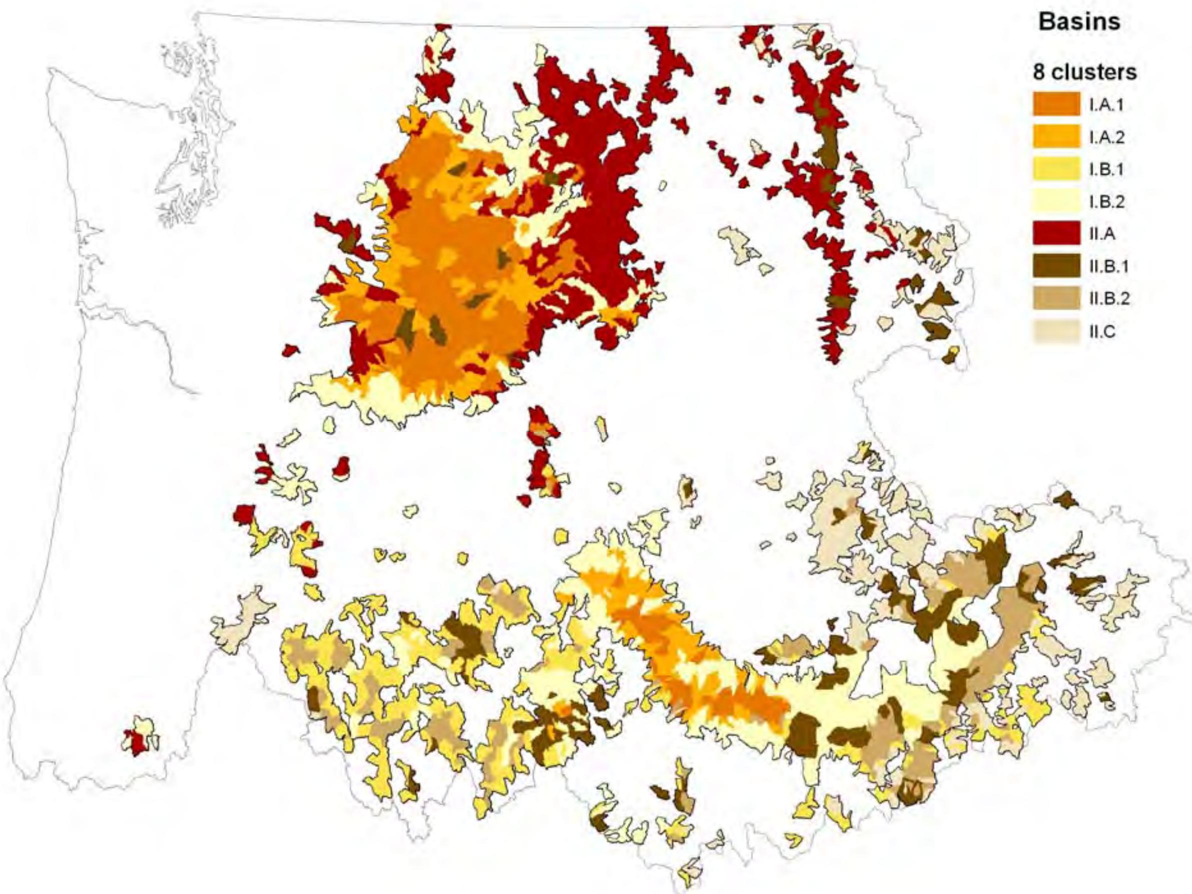



Figure 8. Location of the lowest level Eastside Basins natural feature classes. See Table 7 for class names and abbreviations.

Table 7. Eastside Basins natural feature classes, the number of Intensively Monitored Watersheds (IMW) and the total number of HUC6 in the PNW in each class.

| Landscape Class | IMW basins | PNW HUCs |
|---|---|---|
| I. Hot and Dry Basins | 1 | 1,003 |
| A. Hot, Low, Erodible Basins | -- | 421 |
| 1. Very Erodible | -- | 254 |
| 2. Moderately Erodible | -- | 167 |
| B. Mid-Elevation, very warm, Dry, mod. Erodible Basins | 1 | 582 |
| 1. Dry, very Warm, mod. High, mod. Erodible | 1 | 205 |
| 2. Fairly Hot, mod. Low, less Erodible | -- | 124 |
| II. Temperate Basins | 14 | 977 |
| A. Cool, High Basins | -- | 321 |
| B. Southern, Warm, mid-Elevation, Low Slope Basins | 1 | 266 |
| 1. Higher Gradient Streams, Very Erodible | 1 | 161 |
| 2. Lower Gradient Streams, variable Erodibility | -- | 105 |
| C. Northern, Warm, Lower Basins | 13 | 238 |

*Eastside Mountains* - This class includes 58% of watersheds in study area. The top branch divided the Eastside mountains watersheds into a wet, cool, higher elevation class (n=2,348) with variable erodibility, versus a drier, cooler, lower elevation, non-erodible class (n=2,532) (Figure 9). Geographically, the wetter, cooler mountains class tended to be at the western and eastern edges of the eastside region, with drier, warmer mountains class more centrally located. The second branch divided the cool, wet class into a relatively cooler, wetter class (n=1,385) that had steep, less erodible watersheds, versus a relatively drier, warmer class (n=963) that was less steep and more erodible. The third branch divided the warmer, drier, lower, non-erodible mountains class into a generally southern, dry, less steep, higher class (n=791) with large temperature ranges, versus a northern, wetter, lower, steeper class (n=1,741) with smaller temperature ranges. The fourth branch carved off a small (n=111) class that was very cool, high, steep and erodible, from the moderately cool, wet and erodible mountains class. This small class was mostly in the Idaho Batholith and the eastern edge of the study area. The remaining larger class (n=852) was very similar to its parent class.

The fifth branch divided the (northern) warm, steep, moderately dry mountains into a higher, cooler, steeper class (n=1,136) versus a lower, warmer, less steep class (n=605) much of which forms the outer edge of the Columbia Plateau ecoregion. The next branch divided the wet, very cool mountains into a fairly small (n=296) high and erodible class that was concentrated in the Idaho Batholith and southern portion of the Middle Rockies ecoregions. The other class (n=1,089) was very similar to its parent class, and was next divided into a drier, higher, greater temperature range, Middle Rockies class (n=445) versus a wetter, somewhat lower, Cascades, Blue Mountains, and Northern Rockies class (n=644). The moderately cool, wet and erodible mountains class divided into a wetter and lower class (n=395) and a drier, higher class (n=457) with greater temperature ranges. Geographically, these two classes generally separated into southern and northern areas, except along the Oregon Cascades. Stopping at this level produced 9 Eastside Mountains classes, with between 111 and 1,136 watersheds (Figure 10; Table 8).

The results for three of the four assessments aimed at determining an appropriate number of classes, for the Eastside Mountains watersheds, were similar to other two top level classes. The mean within-cluster similarity plot began to level off at around 8 to 10 clusters, and the Anosim R at 9 clusters. The KGS K-statistic reached a minimum at 6 to 8 clusters. The Eastside Mountains data set was apparently too large to run the pseudo-F assessment with our software.

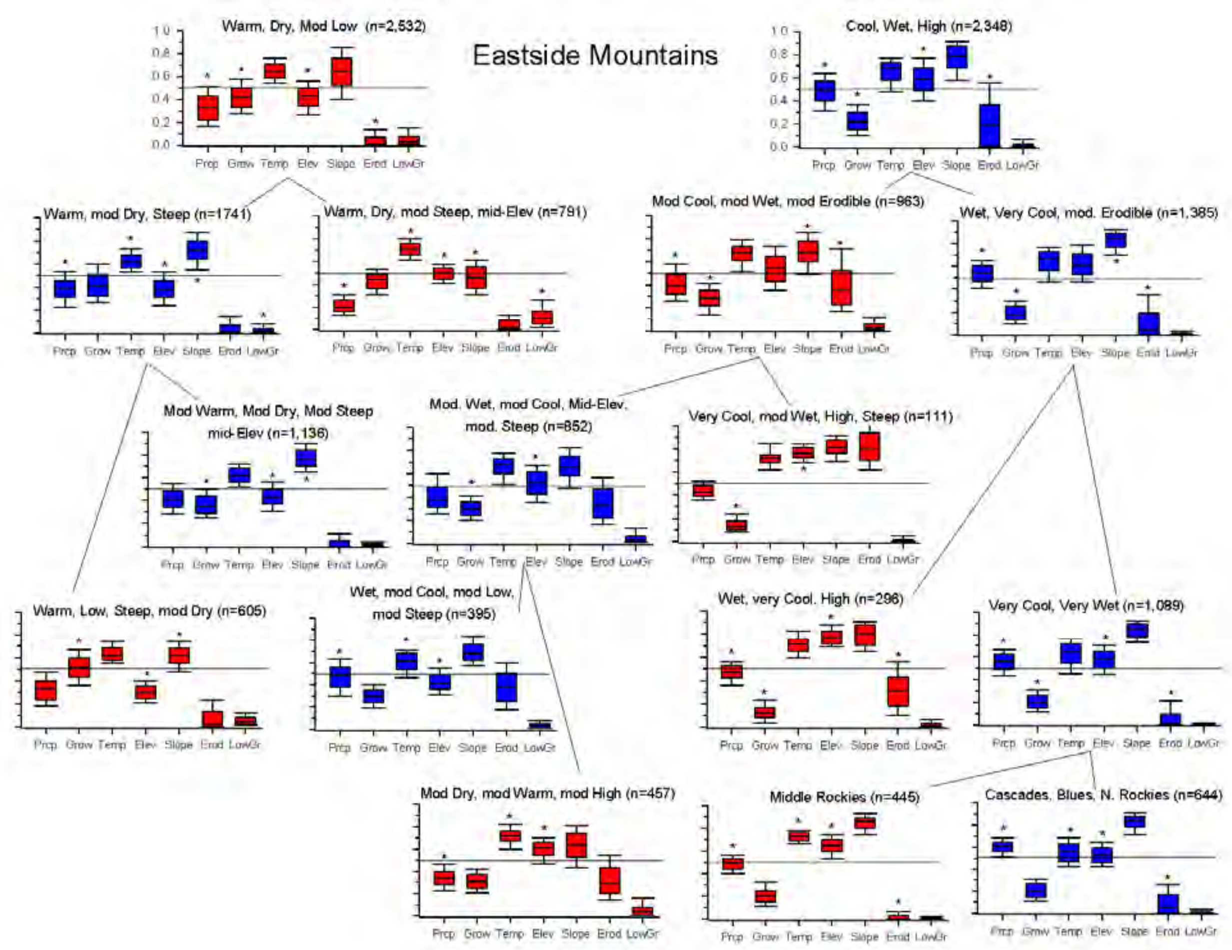



Figure 9. Characteristics of each level of the Eastside Mountains natural feature classification. Box and whisker plots show the scaled 0-1 scores for each natural feature class (see Table 2 for variable definitions).

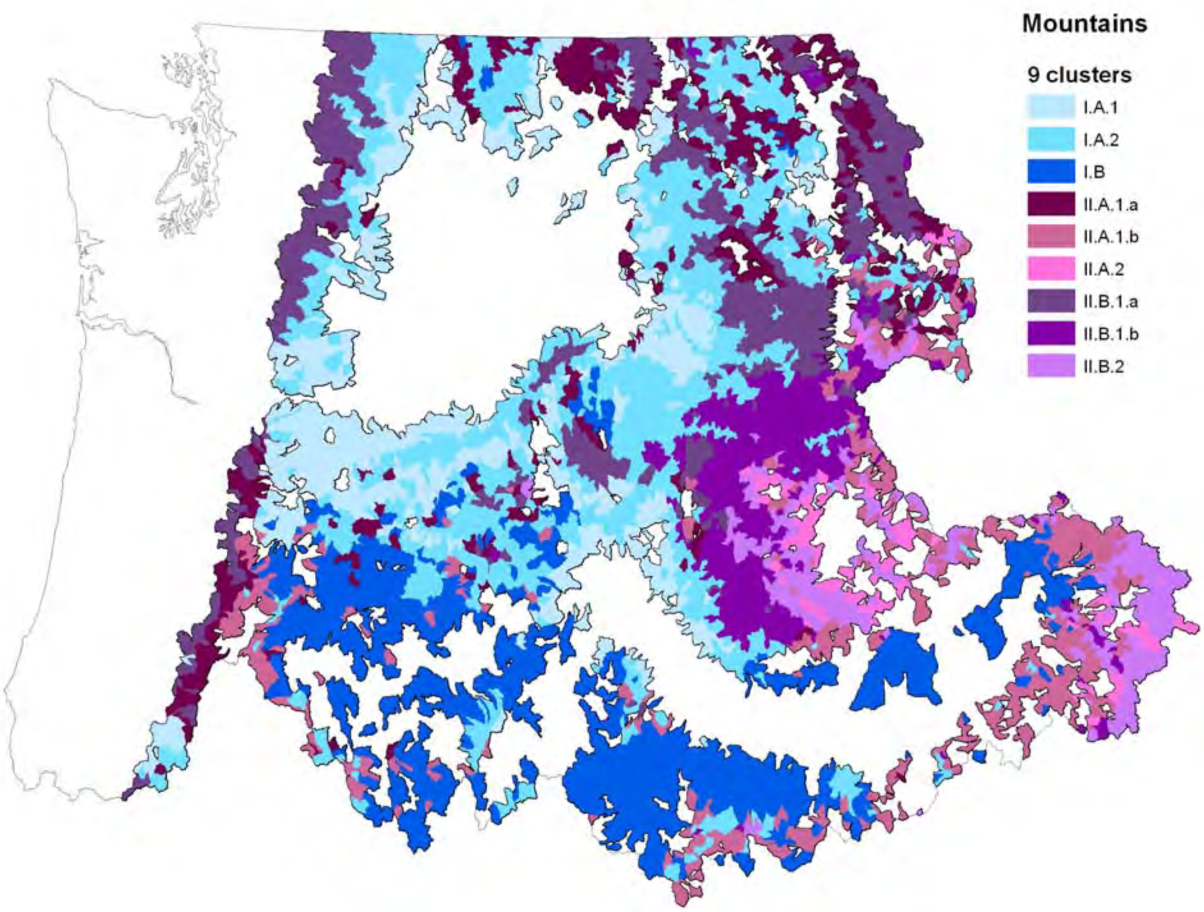



Figure 10. Location of the lowest level Eastside Mountains natural feature classes. See Table 8 for class names and abbreviations.

Table 8. Eastside Mountains natural feature classes, and the number of Intensively Monitored Watersheds (IMW) and the total number of HUC6 in the PNW in each class.

| Landscape Classes | IMW basins | PNW HUCs |
|---|---|---|
| I. Dry, Warm, moderately Low Mountains | 34 | 2,532 |
| A. Steep, moderately Dry Mountains | 32 | 1,741 |
| 1. Warm, Low, mod. Dry, mod. Steep | 8 | 605 |
| 2. Cooler, Wetter, Higher, Steep, Non-erodible | 24 | 1,136 |
| B. Dry, mid-Elevation, mod. low Gradient Mountains | 2 | 791 |
| II. Wet, Cool, High Mountains | 74 | 2,348 |
| A. Moderately Wet, mod. Cool, Erodible Mountains | 23 | 963 |
| 1. Moderately Cool, Mid-Elevation, Mod. Steep, | 18 | 852 |
| a. Wetter, Lower | 1 | 395 |
| b. Drier, Higher | 17 | 457 |
| 2. Very Cool, High, Very Steep, Very Erodible | 5 | 111 |
| B. Wet, Very Cool, Steep, mod. Erodible Mountains | 51 | 1,385 |
| 1. Very Wet, Cool, Mid-Elevation, Low Erodibility | 45 | 1,089 |
| a. Cascades, Blue Mountains, Northern Rockies | 11 | 644 |
| b. Middle Rockies | 34 | 445 |
| 2. High, Wet, very Cool, mod. Erodible | 6 | 296 |

Human Disturbance Classification

The flexible beta cluster analysis of the four scaled disturbance measures produced a balanced dendrogram that divided (BD ~105) the PNW watersheds into a generally low-disturbance class (n=4,794) and a generally high-disturbance class (n=3,644) (Figure 11). In the low-disturbance class urban, agriculture and impervious surface were quite low, but the median (scaled) value for road density was ~0.5 (out of 1). The map of the low disturbance watersheds is similar to a map of public lands in the PNW (Figure 12). All disturbance measures were distinctly higher in high-disturbance class.

The low-disturbance class divided (BD ~45) into an essentially undisturbed class (n=1,021) and a larger (n=3,773) generally low-disturbance class with distribution of disturbance measures similar to its parent class. The undisturbed class had very low values for urban, agriculture and impervious, as well as the lowest values for roads of any of the final classes. This was a fairly distinct class at this level of the dendrogram, in that it did not divide again until BD~10. The second level low-disturbance class divided at BD~24 into a very low disturbance but roaded class (n=2,736) and a moderately low disturbance class (n=1,037). The very low disturbance class had urban, agriculture and impervious values very similar to those in the undisturbed class, but had a median value for roads ~0.55. This class next divided at BD~20 into clusters differentiated only by road density. Most of the watersheds in lower road density cluster at this division were outside the area of anadromy, thus we chose not to use this division in the final set of human disturbance classes.

The top-level high-disturbance class divided, at BD~55, into a mostly westside class (n=1,790) and mostly eastside class (n=1,854). The westside disturbed class had distinctly higher levels of urban, roads, and impervious surface. The eastside disturbed class tended to have somewhat higher levels agricultural land use. At BD~30 the eastside disturbed class divided into a highly agricultural class (n=682) with relatively high urban land use, and a somewhat less disturbed mixed land use class (n=1,172) with partially overlapping agricultural and urban land use coverage. Road density and impervious surface were not different between these two classes.

The westside disturbed class divided at BD~42 into a mostly western low mountains class (n=997) and a mostly large valleys class (n=793). The large valleys class had distinctly higher urban, agriculture and impervious surface values. The western low mountains disturbed class had higher values for urban than for agricultural land use, while those two measures were similar to each other in the big valleys disturbed class. Road densities were not different between these classes. The big valleys class divided at BD~25 into a highly disturbed class (n=183) and a somewhat less disturbed mixed urban and agricultural class (n=610). The highly disturbed class HUC6s were located in and around urban centers with very high values for urban, roads and impervious surface, and highly variable agricultural land use. The mixed urban / agriculture big valleys disturbance class generally had higher agricultural land use than did the more urban highly disturbed class.

The eight final human disturbance classes could be subjectively ordered from least-disturbed to most-disturbed (Figure 13 top). Boxplots of disturbance PC1 scores for each disturbance class generally support this ordering. Two pairs of disturbance classes had very similar ranges of PC1 scores (Figure 13 bottom), which suggests that while two classes can differ in the key sources of disturbance, they may have similar overall levels of disturbance.

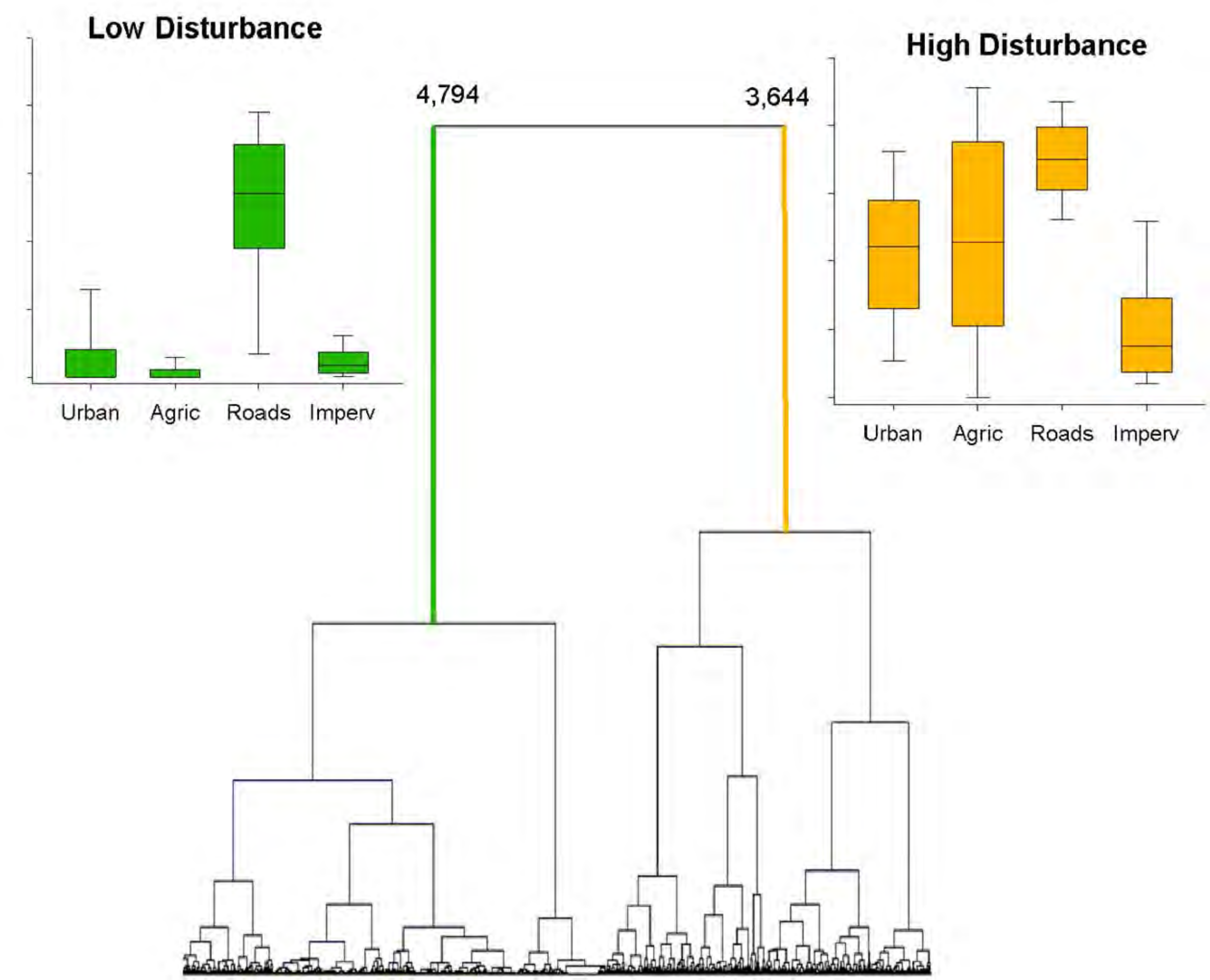
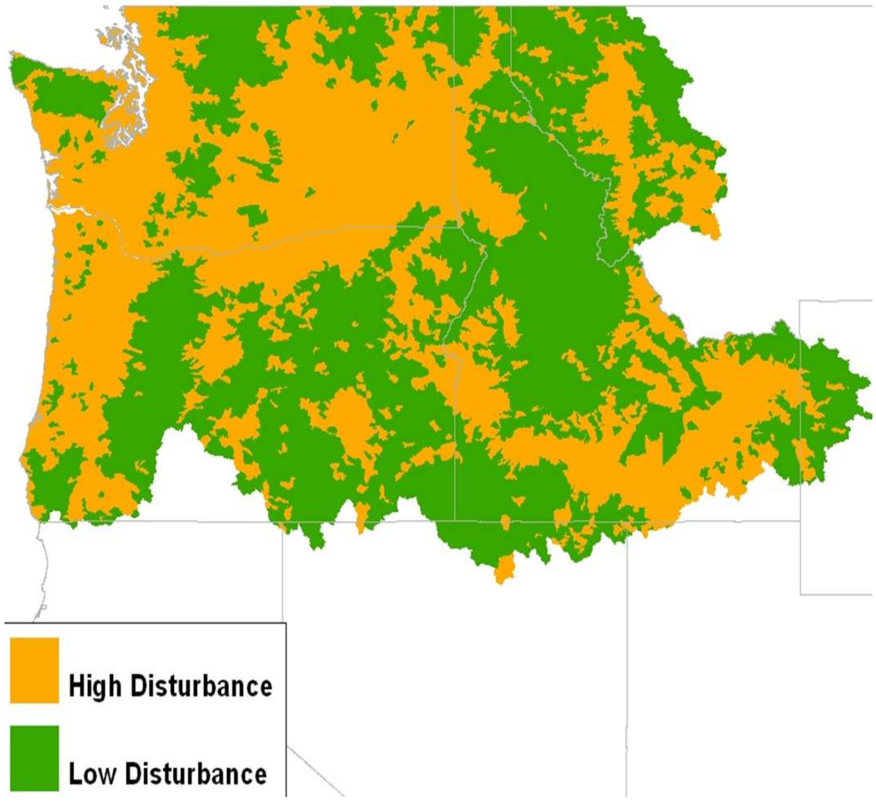



Figure 11. Cluster dendrogram of the 6th field HUCs based on four human disturbance variables. Box and whisker plots show scaled disturbance scores for the high and low disturbance class.



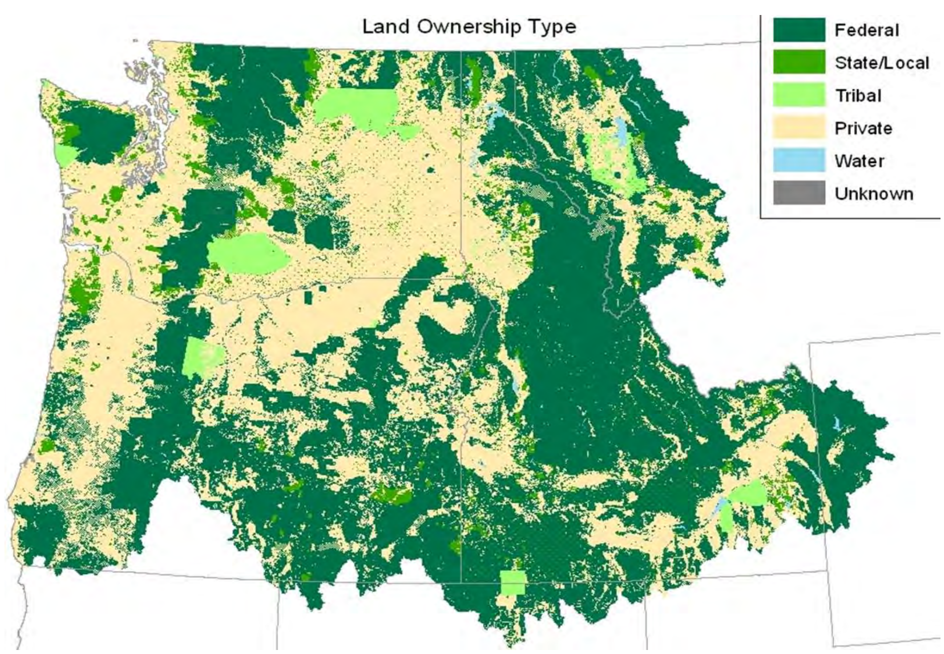




Figure 12. Map showing location of high and low human disturbance classes (top) and land ownership (bottom).

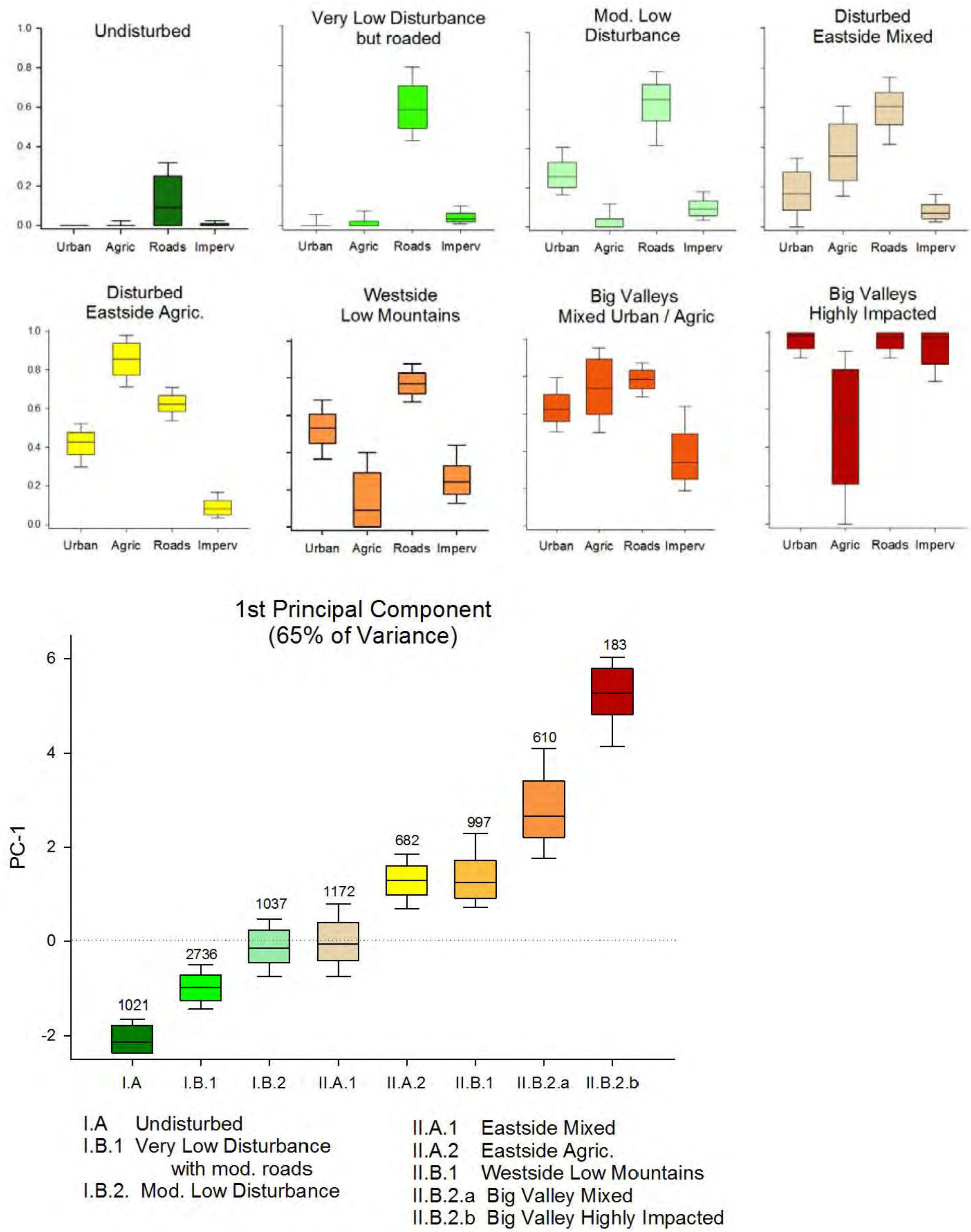



Figure 13. Scaled human disturbance scores (0-1) for each of the lowest level human disturbance classes in the Pacific Northwest (top). Gradient in overall disturbance PCA axis 1 (PC1) score for each of the disturbance classes (bottom).

Associations between natural features and human disturbance

As expected, less disturbed watersheds were at higher elevations than more disturbed watersheds (Figure 14). Likewise, the watersheds in each natural feature landscape class were not evenly distributed among the disturbance classes (Figure 15). In the westside natural feature class, nearly all of the watersheds in the two lowlands and the two foothills classes were in the two most-disturbed human disturbance classes, while very few of the watersheds in the three mountains classes were in those disturbance classes. Interestingly, most westside mountains watersheds were in the intermediate disturbance classes and very few were in the least disturbed class. In contrast, a large majority of the eastside mountains watersheds were in two least disturbed classes. Most of the few eastside mountains watersheds in more disturbed classes were in the lower, drier natural feature classes. The eastside basins watersheds tended to be mostly in the intermediate disturbance classes with very few at the extremes. The eastside basins watersheds were less concentrated in particular disturbance classes than were watersheds in the other two top-level natural feature classes.

Correlations between the seven natural features input variables and the four human disturbance variables were strongest in the westside and weakest in the eastside basins (Table 9). Annual temperature range was only weakly correlated ($r<|0.30|$) with any of the disturbance measures in all three top-level natural feature classes, as were annual precipitation, watershed erodibility and low gradient stream density in the two eastside classes. Among the other natural features, elevation was most strongly and consistently correlated with disturbance measures in all three natural feature classes. Unexpectedly, agricultural land use was negatively correlated with annual precipitation in all three natural feature classes.

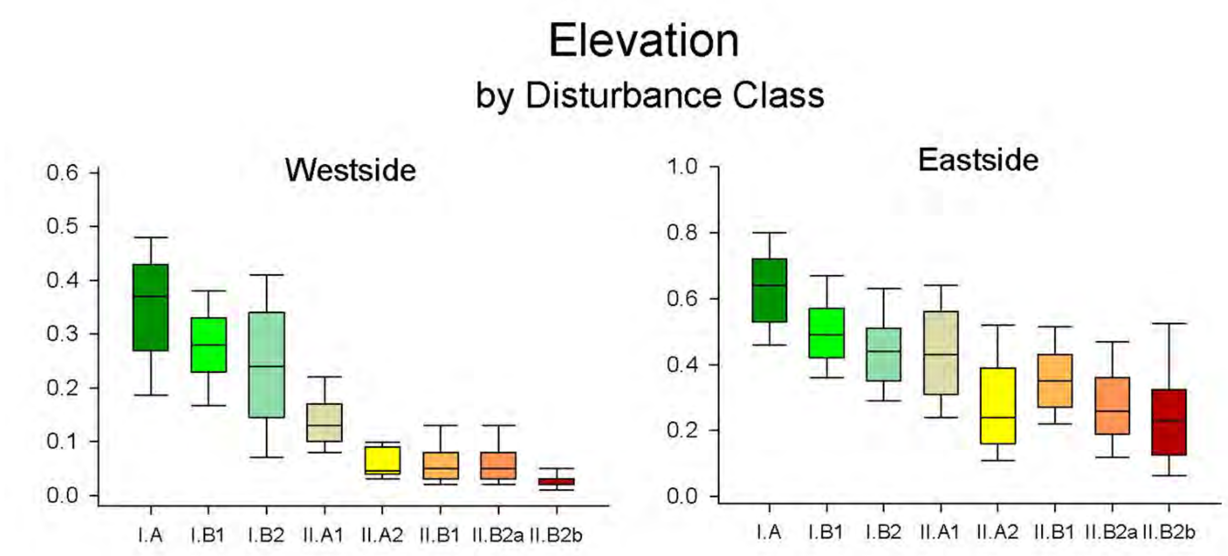



Figure 14. Box and whisker plot of scaled (0-1) elevation for each disturbance class.

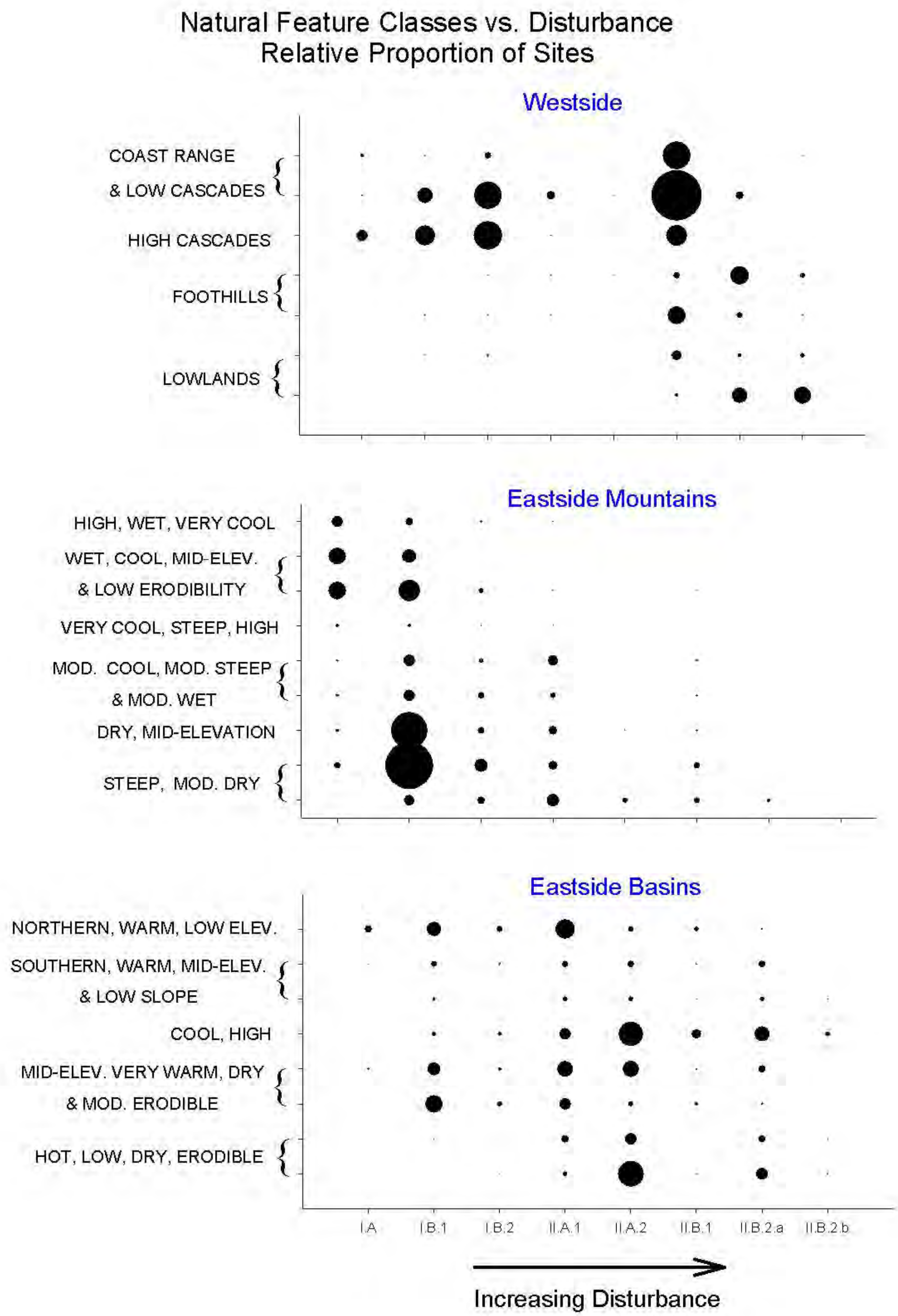



Figure 15. Cross tabulation showing relative proportion of 6th field HUCs between natural feature classes (y-axis) and disturbance classes (x-axis). Disturbance classes are ordered by level of disturbance (see Fig. 13). Diameter of the circle is proportional to number of HUCs in that cell.

Table 9. Associations between disturbance measures and natural features for the primary natural feature classes: Westside, Eastside Mountains and Eastside Basins. PC-1 is the 1st principal component of a PNW-wide PCA of the 4 disturbance measures used in disturbance clustering and could be considered as an overall landscape disturbance measure. Correlations >|0.50| are in **bold**, correlations <|0.30| are not shown, ns=not significant

| | PC-1 | Urban | Agric | Roads | Imperv |
|---|---|---|---|---|---|
| Westside | | | | | |
| Annual Precipitation | -0.48 | -0.31 | **-0.56** | -0.37 | -0.42 |
| Degree Days | 0.45 | 0.35 | **0.50** | 0.41 | 0.32 |
| Ann. Temp. Range | | | | | |
| Elevation | **-0.70** | **-0.67** | **-0.53** | **-0.60** | **-0.56** |
| Watershed Slope | **-0.77** | **-0.61** | **-0.70** | **-0.51** | **-0.73** |
| Watershed Erodibility | **0.59** | 0.48 | 0.49 | | **0.63** |
| Low Gradient Streams | 0.49 | 0.43 | 0.46 | | 0.46 |
| Eastside Mountains | | | | | |
| Annual Precipitation | | | | | |
| Degree Days | 0.41 | 0.30 | 0.33 | 0.37 | |
| Ann. Temp. Range | | | | | |
| Elevation | **-0.53** | -0.40 | -0.30 | **-0.50** | -0.37 |
| Watershed Slope | -0.33 | | | -0.34 | |
| Watershed Erodibility | ns | | | | ns |
| Low Gradient Streams | | | | | |
| Eastside Basins | | | | | |
| Annual Precipitation | | | | ns | ns |
| Degree Days | 0.30 | | 0.42 | | |
| Ann. Temp. Range | | | | | |
| Elevation | -0.49 | -0.46 | **-0.54** | -0.38 | |
| Watershed Slope | | | | | |
| Watershed Erodibility | | | | ns | |
| Low Gradient Streams | | | | | ns |

Intensively Monitored Watersheds

A majority of Westside (91%) and Eastside Mountains (86%) IMW HUC6s had natural feature PC1 scores <0, compared with 55% of all PNW HUC6s (Figure 16). Only 13% of Eastside Basins IMW HUC6s had natural feature PC1 scores <0. The Eastside Mountains IMWs HUC6s appeared to be well distributed along the Eastside portion of the natural feature PC2, while 14 of the 15 Eastside Basins IMW HUC6s had PC2 scores >1. The distributions of the human disturbance PC1 scores (Figure 17) for the Eastside Mountains and Basins IMW HUC6s matched those of all HUC6s for their respective areas, while on the range of Westside IMW HUC6 disturbance PC1 scores was compressed around the regional median (i.e., few low or high scores).

The thirteen west side IMW projects (Table 4) were located in 32 Westside HUC6s, in five of the seven Westside landscape classes. The 88% of western IMW watersheds that were in the Coast Range and Southern/Low Cascades split fairly evenly between the Southern Coast Range/Low Cascades, and the Washington and NW Oregon Coast Ranges classes. There were no IMW watersheds in the High Cascades/High Olympics or the Willamette Foothills/Rouge Valley classes. Four of the western IWM projects spanned two landscape classes.

The nine Eastside IMW projects were located in 108 Eastside Mountains HUC6s and 15 Eastside Basins HUC6s (all part of the Lemhi Project). . All but 2 of 15 Eastside Basins IMW watersheds were in the Northern, Warmer, Lower Temperate Basins class (Table 5). The 108 Eastside Mountains IMW watersheds occurred in all 9 Mountains classes (Tables 6) and were primarily in the top level class of Wet, Cool High Mountains (n=74). Of these, about 2/3 were in the second-level class of Wet, Very Cool, less Erodible Mountains, primarily in the Very Wet, Mid-Elevation, Low Erodibility third-level class (Table 6). In the Dry, Warm, Moderately Low Mountains top-level class, about 70% of IMW watersheds were in the landscape class of more moderate conditions (i.e., Cooler, Less Dry, less Steep, non-Erodible watersheds).

All Eastside IMW projects included between two and four landscape classes. Five of these projects had watersheds belonging to both of the top level Mountains classes. This diversity of landscape classes within projects implies that was a tendency to place the projects in transitional areas. This also suggests that caution may be needed in order to extend IMW project results to other areas, particularly to areas more centrally located within Eastside landscape classes (i.e., non-transitional areas).

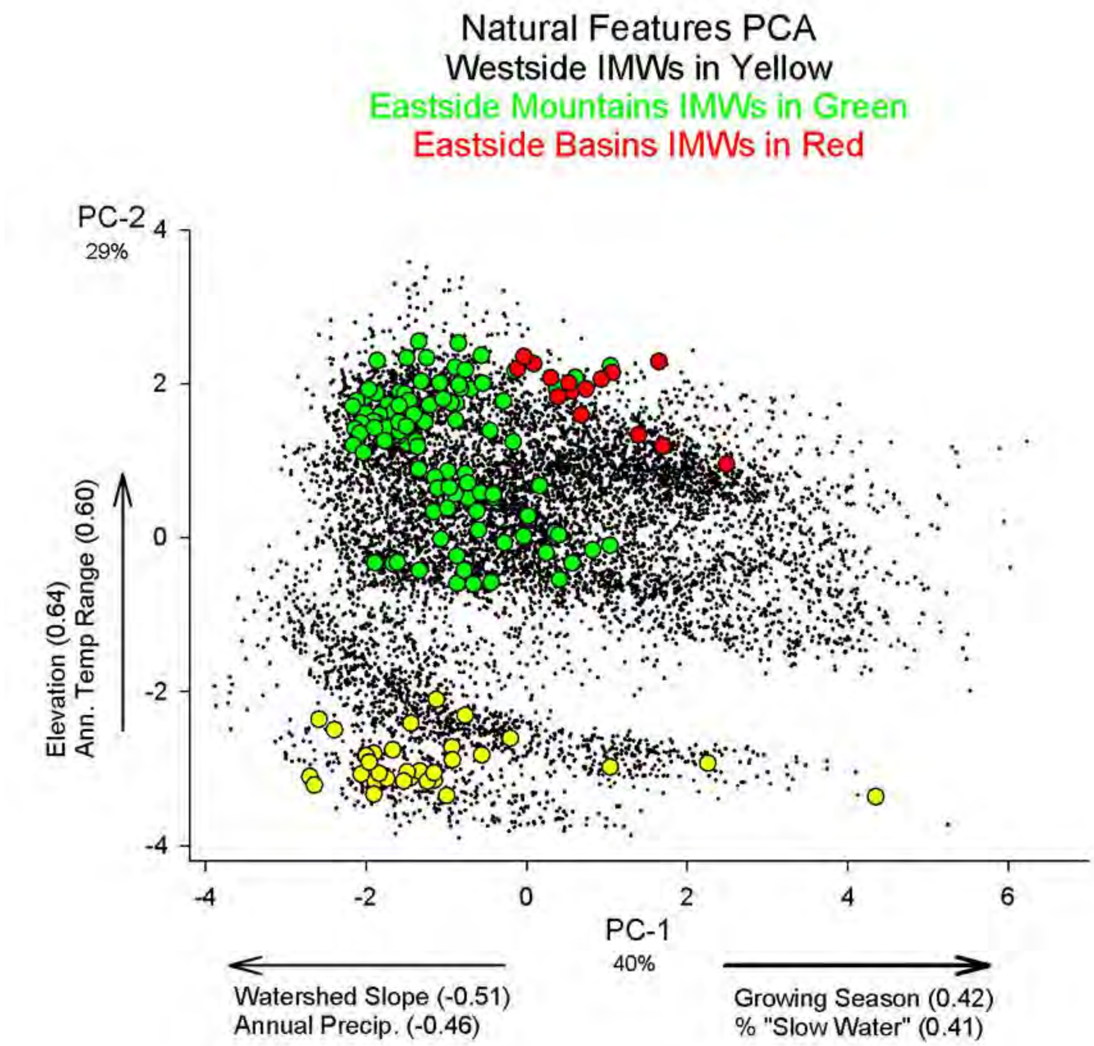



Figure 16. Natural feature variable PCA axis 2 (PC-2) versus axis 1 (PC-1) score plot (same plot as figure 2) showing the location of intensively monitored watersheds (IMW) as colored dots. PC-1 explained 40% of the variance and PC-2 explained 29% of the variance. Numbers in parentheses after the variable names indicate variable loadings on the PCA axes.

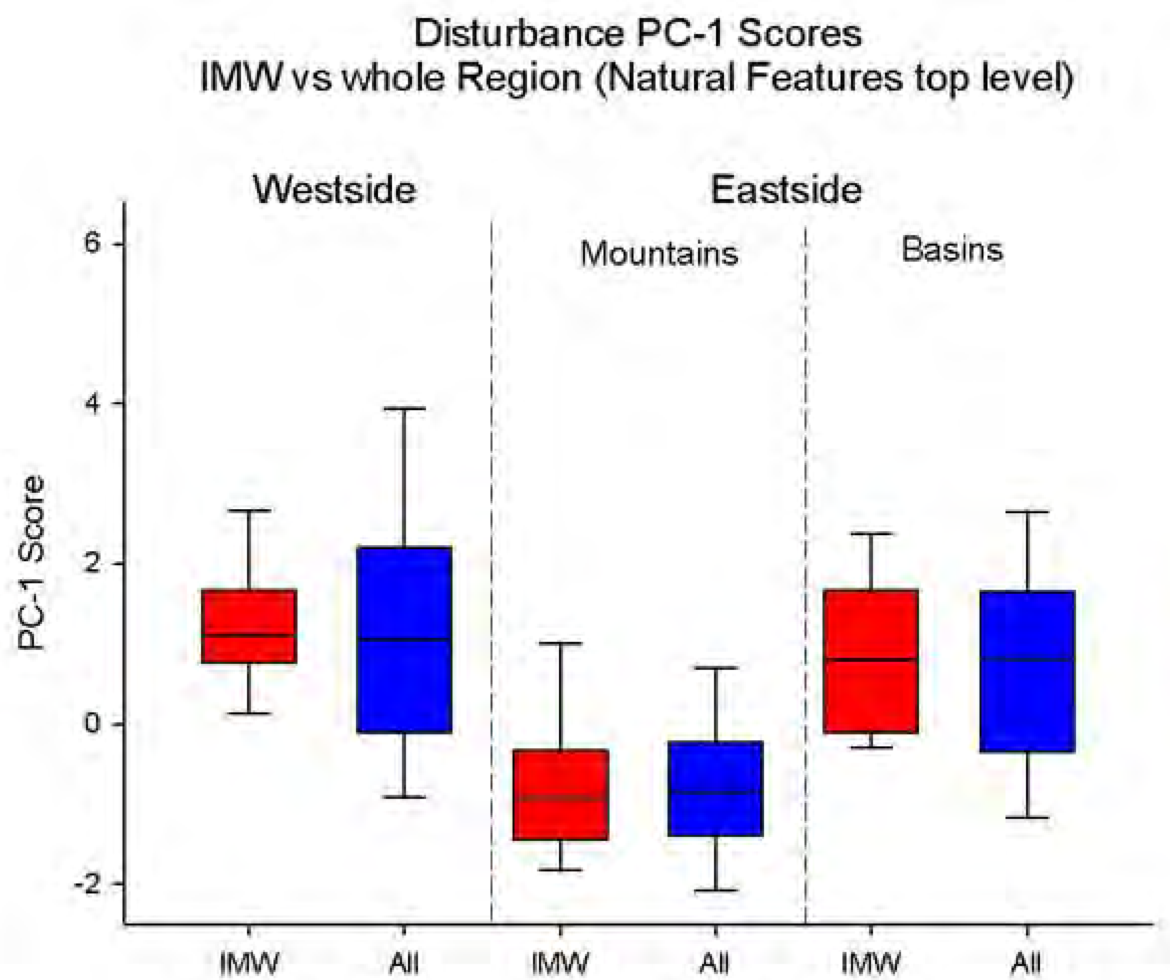



Figure 17. Box and whisker plot comparing disturbance variable PCA Axis 1 (PC-1) scores for intensively monitored watersheds (IMW) versus all 6$^{th}$-field HUCs by top level natural feature class.

## Literature Cited

Bailey, R.G. 1976. Ecoregions of the United States (1:7,500,000-scale map). Intermountain Region, USDA Forest Service, Ogden, Utah.

Fausch, K.D., C.E. Torgersen, C.V. Baxter, and H.W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. BioScience 52:483-498.

Handl, J., J. Knowles and D.B. Kell. 2005. Computational cluster validation in post-genomic data analysis. Bioinformatics 21:3201-3212.

Hargrove, W.W. and F.M. Hoffman. 2005. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environmental Management 34(Suppl.1): S39-S60.

Loveland, T.R. and J.M. Merchant. 2005. Ecoregions and ecoregionalization: geographic and ecological perspectives. Environmental Management. 34(Suppl 1): S1-S13.

McDonald, L.L., and 15 others. 2007. Research, monitoring, and evaluation of fish and wildlife restoration projects in the Columbia River basin. Fisheries 32:582-590.

Nehlsen, W., J.E. Williams, and J.A. Lichatowich. 1991. Pacific salmon at the crossroads: stocks at risk from California, Oregon, Idaho and Washington. Fisheries 16(2):4-21/

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American Geographers. 77:118-125.

Omernik, J.M. 2003. The misuse of hydrologic unit maps for extrapolation, reporting, and ecosystem management. Journal of the American Water Resources Association. 39:563-573.

Omernik, J.M. 2005. Perspectives on the nature and definition of ecological regions. Environmental Management. 34(Suppl 1):S27-S38.

Williams, R.N., and 12 others. 1999. Scientific issues in the restoration of salmonid fishes in the Columbia River. Fisheries 24(3):10-19.

## Acknowledgements

This research was funded by the National Oceanic and Atmospheric Administration through Contract #AB1133F10SE2464 with Oregon State University. We thank Jeannie Sifneos for help with cluster analysis and dendrogram pruning and the NOAA Science Center staff for GIS and initial data acquisition and analyses.