EPA 910/R-03-013

United States
Environmental Protection
Agency

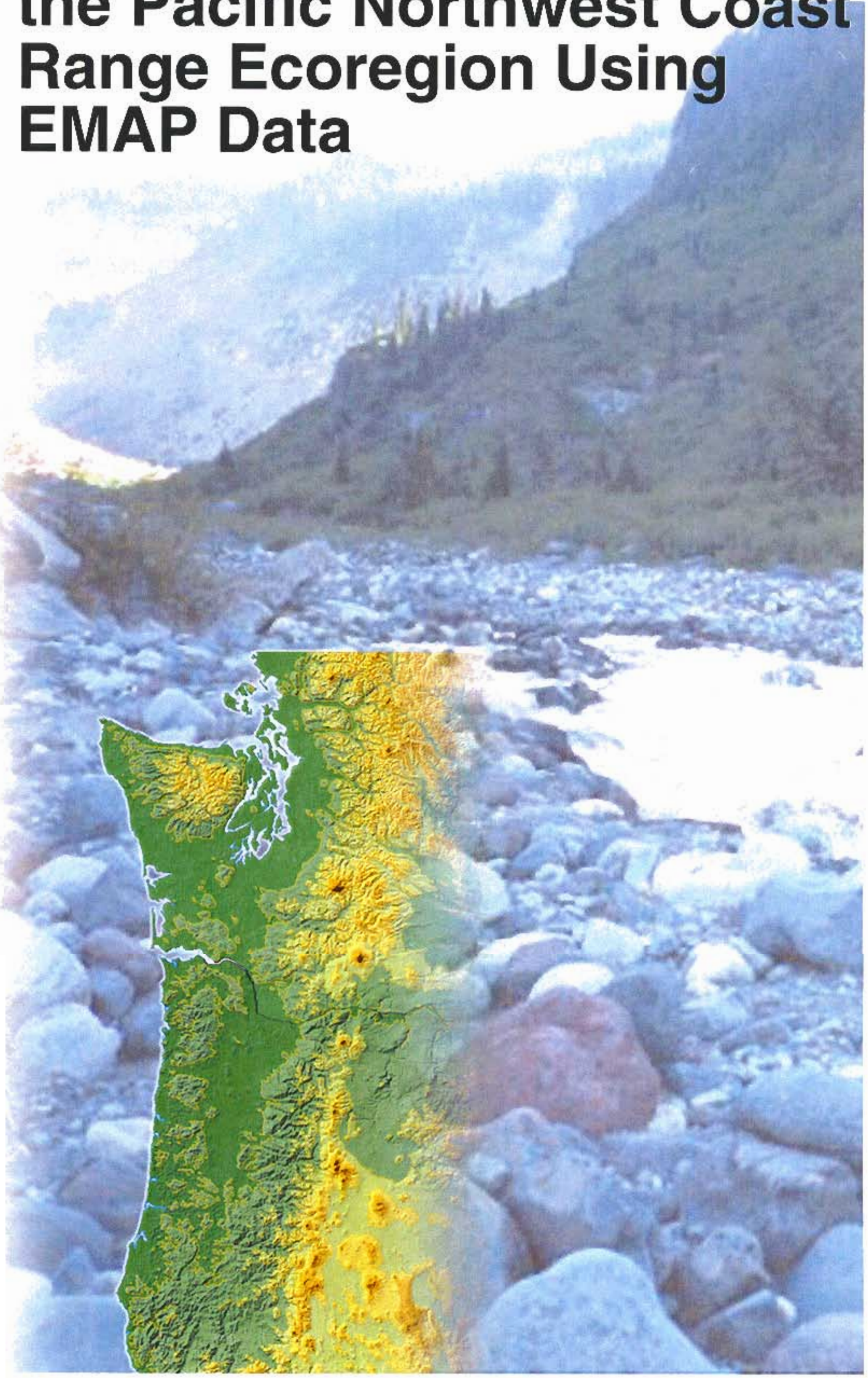
Region 10
1200 Sixth Avenue
Seattle WA 98101

Alaska
Idaho
Oregon
Washington

Office of Environmental Assessment

December 2003

Modeling Fish Distributions in the Pacific Northwest Coast Range Ecoregion Using EMAP Data



EPA 910-R-03-013
December 2003

Modeling Fish Distributions in the Pacific Northwest Coast Range Ecoregion Using EMAP Data

by

Lillian G. Herger[1], Andrew Weiss[2], Scott Augustine[1], and Gretchen Hayslip[1]

[1]U.S. Environmental Protection Agency, Region 10
Office of Environmental Assessment
1200 Sixth Avenue, Seattle, WA 98101

[2]Indus Corporation
1953 Gallows Rd. Suite 300
Vienna, VA 22182

THIS PAGE INTENTIONALLY LEFT BLANK

Note: The correct citation for this document is

Herger, L.G., Weiss, A., Augustine, S., and G. Hayslip. 2003. *Modeling Fish Distribution in the Pacific Northwest Coast Range Ecoregion Using EMAP Data*. EPA/910/R-03/000. U.S. Environmental Protection Agency, Region 10, Seattle, Washington.

Table of Contents

List of Tables

List of Figures

List of Appendices

Acknowledgements

Field data were collected by Oregon Department of Environmental Quality, Washington Department of Ecology, and Jessie Ford and Cathleen Rose of Oregon State University. U.S. EPA Offices of Research and Development provided landscape data, upstream contributing area delineation, and surface waters metric calculations. Phil Larsen helped with development of this project. Helpful comments and support provided by Phil Kaufmann, John Van Sickle, Dan Heggem. Maliha Nash and Annie Neale provided invaluable statistical advice as well as editorial comments. Valentina Haack (Indus Corporation) assisted with GIS graphics and document production. Portions of this study were completed under a contract between EPA Region 10 and Indus Corporation.

1. Introduction

1.1 Purpose of Report

EPA initiated the Environmental Monitoring and Assessment Program (EMAP) to estimate the status and trends of ecological resources and to examine associations between ecological condition and natural and anthropogenic influences. EMAP generates regional-scale assessments of ecological resource condition. These assessments describe the current geographic extent of ecological resources, what resources are degrading or improving, and how resources are responding to changing control and regulatory programs.

Information generated by EMAP can be applied to a variety of water quality management issues including: supporting States in defining beneficial uses, supporting Total Maximum Daily Load problem assessments, and augmenting existing environmental databases. The program has been highly successful in developing partnerships with state environmental quality agencies. The ultimate goal of this collaboration is for the States to integrate the EMAP design and protocols into their own water quality monitoring programs.

In support of the EMAP objectives, EPA Region 10 has conducted the following analysis of the Coast Range Ecoregion of Oregon and Washington EMAP data. Ecoregions are distinct geographic areas based on characteristics of the topography, climate, soils, geology and naturally occurring vegetation. This analysis integrates the EMAP landscape data and surface waters data to identify the key natural factors and human disturbance attributes useful for estimating fish assemblages. The purpose is to determine the extent to which natural gradients and human disturbance factors at a landscape scale account for variation in the Coast Range Ecoregion fish assemblage. The goal of this project is to contribute to the development of indicators of stress by considering metrics that can be influential at the broad scale of the ecoregion. Specific outcomes of this analysis are the following:

Develop empirical models of the relation of fish distribution to remotely sensed environmental metrics.

Compare performance of remotely sensed to reach-scale metrics for predicting fish distribution.

Determine short list of environmental metrics useful for Region 10 stream assessments.

Describe Coast Range fish assemblages. Identify distinct fish assemblages or describe species composition as transitional mixes of species along environmental continua.

1.2 Study Area

The study area is the Oregon and Washington portion of the Coast Range Ecoregion using the boundary defined in 2000 (Pater et al. 2000). The Coast Range is an extensive temperate rainforest characterized by cool summers, mild winters, and high precipitation (Franklin and Dyrness 1973). The Oregon and Washington portion of the Coast Range includes the Pacific coast mountain range and the coastal valleys and terraces (Omernik 1987). Elevation ranges 0-1900m (mean 293m) and the topography is typically steep (Figure 1). Mean annual precipitation is 240 cm (range 109 to 612 cm). The Coast Range has abundant lotic ecosystems supporting Pacific salmon species, which are biologically important as well as economically and culturally important.

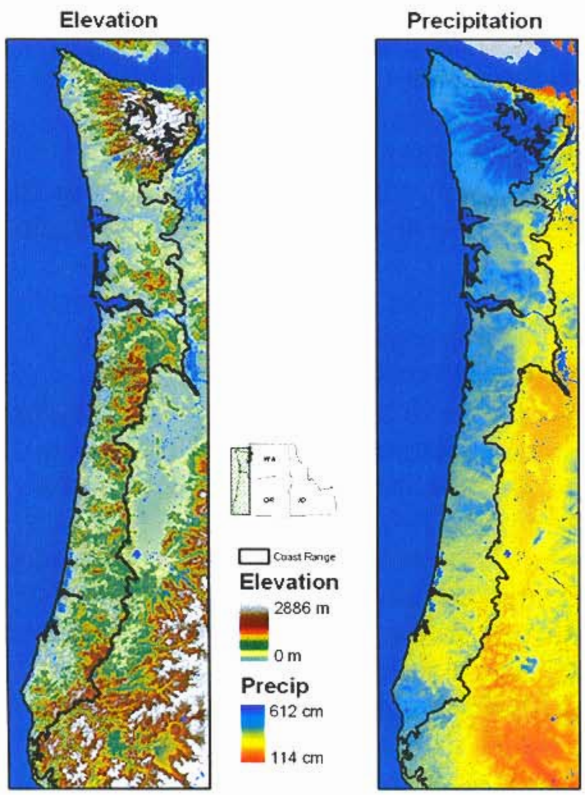


Figure1. Coast Range study area

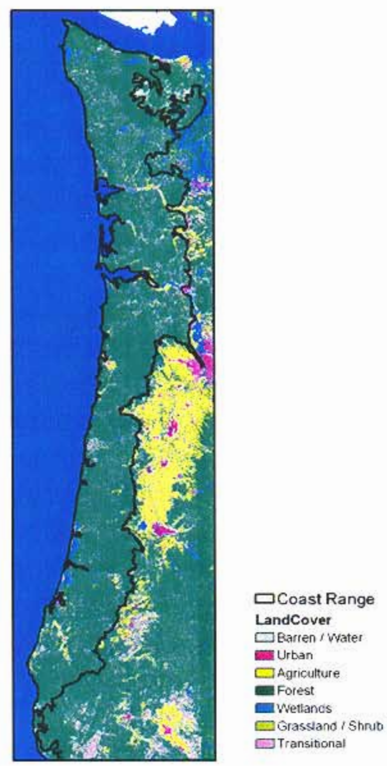


Figure 2. Human use cover classes of the Coast Range study area

The Oregon and Washington portion of the Coast Range Ecoregion contains many unique terrestrial and aquatic ecosystems. In the north, the Coast Range Ecoregion encompasses the lower elevation portions of the Olympic National Park, which has over 60 miles of undeveloped Pacific coast, (the largest section of wilderness coast in the lower 48 states) and the largest remaining old growth forests in the Pacific Northwest. The southern extent of the Ecoregion includes the dune areas of the southern Oregon coast, which is a diverse landscape of unique native plants species, wetlands and old-growth Sitka spruce forests.

The Coast Range ecoregion is dominated by forest cover (Figure 2). Much of the forested areas have been intensively logged and have been managed as Douglas fir plantations for many decades. Timber harvest is an ongoing industry in the Coast Range. Dairy cattle operations, including forage/grain cultivation and feedlots, are concentrated in larger valleys and along the coast. Human development is concentrated on land bordering rivers and ocean bays. Currently, this area is undergoing rapid human development.

2. Methods

2.1 Data Types and Sources

2.1.1Field data

Field site data were obtained from five Regional EMAP assessment datasets that were collected during 1994-1999. All sites were randomly selected based on the 1:100,000 U.S. EPA's River Reach File #3 stream basemap (National Hydrological Database). Sampled streams were wadeable with most being 1^{st}-3^{rd} Strahler stream order. Sites were sampled once during summer low flow for fish, physical habitat, and water chemistry using the EMAP Surface Waters protocols (Lazorchak et al. 1998). Length of the sample reach was 40 times the wetted-width or a minimum length of 150m for fish and physical habitat measurements. Water chemistry was collected by grab sample.

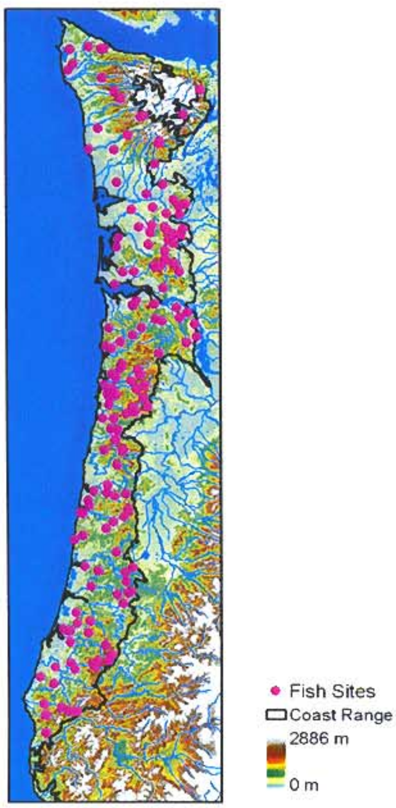


Figure 3. Location of sample sites within the Coast Range study area.

Data were available from approximately 225 sites. Many sites (47) were nested within the upstream contributing area of downstream sites, thus could not be considered independent sample points in the modeling. Also, some sites had variable levels of fish sampling. In order to have sample independence and to standardize the level of fish data among sites, only sites that met the following four criteria were included in the dataset for this analysis: 1) occurred within the 2000 Coast Range Ecoregion boundary; 2) not nested within the upstream contributing area of any downstream sample site; 3) at least 70% of the sample reach length was sampled for fish; and 4) site did not have excessive unidentified fish (>10% of individuals). The final dataset has 159 sites that meet these conditions (Figure 3).

Fish were collected by single-pass electrofishing the entire sample reach (40 times the wetted width, minimum 150m). This level of effort produces repeatable estimates of species richness (Reynolds et al. 2003). Fish were identified by species and counts. For the purpose of data compatibility among sites, two fish species categories required modification. Western brook lamprey and Pacific lamprey were only identified as Lampetra species at some sites, so these two species were merged at all other sites for data compatibility. Merging lamprey species was acceptable because it is difficult to reliably identify amnocetes (juveniles completing the freshwater phase) and the ammocoetes of these two species use similar habitats (Pers. Comm. M. Hallock, WDFW 2002). Riffle and reticulate sculpins were identified as riffle/reticulate sculpins at some sites. In Washington, separate identification of these two species has been particularly difficult and it is likely that these species hybridize where they co-occur (Pers. Comm. M. Hallock, WDFW 2002). As with the lamprey species, these two sculpin species were combined for data compatibility among sites.

2.1.2 Landscape Data

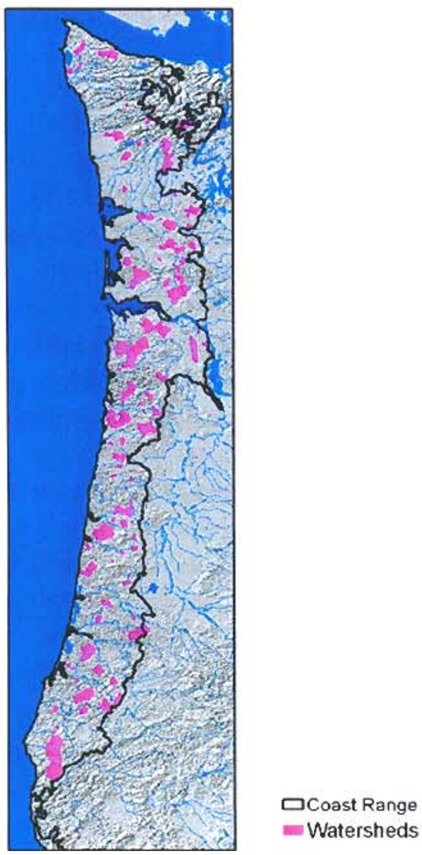


Figure 4. Upstream contributing areas associated with each point.

The upstream contributing area (hereinafter contributing area) was delineated for each sample point using 30-meter digital elevation models and ArcInfo/ArcMap GIS software (ESRI Inc. 2000). Sample contributing areas are shown in Figure 4. Digital coverages from the National Land Cover Database (NLCD), which is derived from Landsat Thematic Mapper (TM) 30-meter resolution satellite imagery, were used as the base data for land cover. Satellite imagery was collected 1991-1993. Metrics were calculated for each contributing area using the Analytical Tools Interface for Landscape Assessments (ATtILA 3.x) an ArcView Software extension (Ebert et al. 2000). Original land cover classes were re-aggregated into 20 cover classes. An additional cover class called 'forest regrowth' was generated to account for disturbance from timber harvest. The forest regrowth metric is an interpretation of the TM data that combines the barren class with recent areas of forest clearcuts and young age stands. Land cover metrics were expressed as percent of the watershed. These metrics were also calculated based on the area of watershed within 30, 60, and 120 meters of the streams. The stream network data are from U.SEPA's River Reach File 3 (National Hydrological Database).

Besides landcover and stream network GIS coverages, several other landscape data sets were used in the analysis including road and road crossing density and distribution (USGS data), and precipitation and air temperature data (Daymet data, www.daymet.org). The Topographical Position Index (TPI) compares DEM cell elevation to the mean elevation of neighboring cells providing an indication of drainage steepness. The TPI was calculated at stream buffers of 150, 300, and 2000 meters. Slope metrics were calculated using the protocol described in the ArcView manual (ESRI Inc. 2000) for the entire contributing area and within 300m of the sample site. Finally, GIS was used to calculated distance metrics that describe site sample location in terms of the length of stream distance to the divide, to higher stream orders, and to the terminal water bodies (Pacific Ocean, Puget Sound, and Columbia River). Other GIS coverages used in the analysis were elevation and road and stream density.

2.2 Data Analysis

2.2.1 Metric Reduction

An initial screening was conducted to reduce the complexity of the landscape dataset and to address collinearity (pairwise correlation) prior to multivariate analysis. The goal was to delete non-causal metrics in order to reduce the risk that non-causal metrics will be retained in the model at the expense of the causal metrics. The correlation between metrics were tested using pairwise Pearson correlation (r) to determine which metrics could act as surrogates for the redundant variables. With the number of sites used in this study, $|r| < 0.80$ was used for model development as recommended by Berry and Felman (1985). Principal components analysis was used to screen the data for important metrics. The first principal component accounts for the highest amount of variance explained by the data set and each following principal component accounts for a smaller amount. Thus, redundancy among related metrics was discerned by running PCA on combinations of related metrics. Pair wise correlation and PCA were used to identify surrogates for related metrics. Finally, The coefficient of variation was calculated to identify metrics with a relatively large spread in data values. Metrics that possibly represented a greater range of condition could be identified with this statistic. The original list of 220 landscape metrics (including metric calculations for 30, 60, 120m buffer widths). A similar process was used to identify useful habitat/chemistry metrics.

2.2.2 Ordination Techniques

The response of the entire fish assemblage in relation to environmental variables was examined using ordination techniques. Ordination extracts the major environmental gradients that are associated with the likely governing the distribution of fish species and determines the most important environmental variables (both natural and anthropogenic) forming these gradients. Ordination accounts for all species, all sites, and all environmental variables simultaneously, thus it is useful for investigating biota community structure in response to combinations of environmental variables.

The following types of ordination were used. Detrended Correspondence Analysis (DCA) provides a measure of the 'inherent order' of the species by site matrix. This is an indirect gradient analysis because the axes are implied ecological gradients and therefore are indirectly related to environmental data. The percent variance explained by the DCA analysis acts as a ceiling for the amount of variance that can be explained by the constrained ordinations. Constrained analyses, including Redundancy Analysis (RDA), and Canonical Correspondence Analysis (CCA), force the axes (or gradients) to be linear combinations of environmental variables. This generally causes the percent of the variation explained to be less than the inherent order, but allows the results to be directly related to the environmental variables. All ordination analyses were performed using the CANOCO computer program (CANOCO 1998).

2.2.3 Pair-wise Correlation and Multiple Regression Analyses

Scatter diagrams to show pair wise correlations, were used to scope the data for patterns and relations among and between landscape metrics and fish abundance, richness, and guild metrics. Multiple regression analyses with stepwise selection were used to combine landscape (remotely sensed) metrics into the best prediction of fish abundance response metrics (e.g. species richness, % tolerant species). A secondary analysis incorporated field habitat/chemistry data in an attempt to improve the model. The multiple regression models were intentionally kept simple in that metrics that provide only minor improvement in fit were omitted. A simple model has more use for future predictions, since prediction error is a function of the number of variables retained (MacNally 2000). Regression analyses were performed using S-PLUS statistical software (Insightful Corp. 2001).

2.2.4 Cluster Analysis

Cluster analysis evaluates species by site matrix with the purpose of identifying preliminary patterns in the distribution of the species among sites using fish abundance. One can determine if distinct assemblages are present or if the distribution of species is more transitional, grading gradually along a continuum of change in relative abundances. An equal emphasis approach was used where the entire assemblage is viewed objectively with no special emphasis on any particular species or family (equal weighting of all species). Analysis was conducted with S-PLUS (Insightful Corp. 2001) using the divisive hierarchical method. All sites and species were included in the analysis and data were not transformed.

3. Results

3.1 Fish Data Description

A total of 22 species were collected from the 159 sites representing eight fish families (Table 1). Reticulate/riffle sculpins, cutthroat trout, and rainbow/steelhead were the most ubiquitous species and were present in at least 50% of sites (Figure 5). These species also had the highest mean relative abundances (> 10%) across the sites (Figure 6). Ten species were considered rare, occurring at less than 5% of the sites and several of these occurred at only one or two sites. Only two alien species were captured (bluegill and brook trout) and these occurred at a total of four sites. Anadromous species, coho and Pacific lamprey, also had high occurrence across sites. Species richness ranged from 0-9 fish species (Figure 7). Eleven sites (7%) had no fish species. The relative abundance among sites of each species is illustrated in Appendix 3.

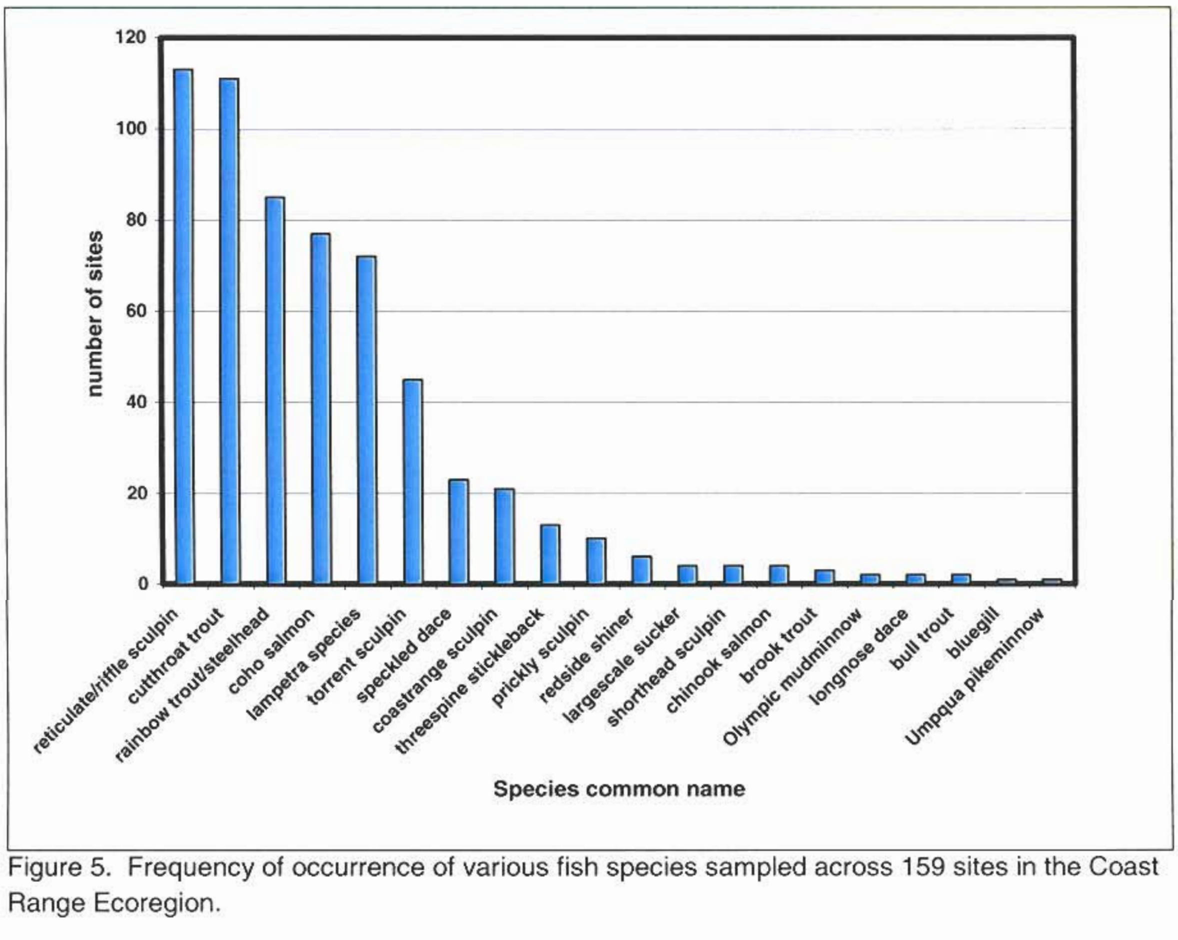


Figure 5. Frequency of occurrence of various fish species sampled across 159 sites in the Coast Range Ecoregion.

Table 1. Species characteristics classification for freshwater fish species identified at Coast Range ecoregion REMAP sites. Classification based on Zaroban et al. (1999).

Family/Species	Common Name	Origin[1]	Tolerance	Habitat	Thermal	Feeding
Catostomidae (suckers)						
Catostomus macrocheilus	largescale sucker	OR, WA	tolerant	benthic	cool	omnivore
Centrarchidae (sunfish)						
Lepomis macrochirus	bluegill	non-native	tolerant	water column	warm	invert/piscivore
Cottidae (sculpins)						
Cottus aleuticus	Coastrange sculpin	OR, WA	intermediate	benthic	cool	invertivore
Cottus asper	prickly sculpin	OR, WA	intermediate	benthic	cool	invert/piscivore
Cottus perplexus	reticulate sculpin	OR, WA	intermediate	benthic	cool	invertivore
Cottus gulosus	riffle sculpin	OR, WA	intermediate	benthic	cool	invertivore
Cottus confusus	shorthead sculpin	OR, WA	sensitive	benthic	cold	invertivore
Cottus rhotheus	torrent sculpin	OR, WA	intermediate	benthic	cold	invert/piscivore
Cyprinidae (minnows)						
Ptychocheilus Umpquae	Umpqua pikeminnow	OR	tolerant	water column	cool	invert/piscivore
Rhinichthys cataractae	longnose dace	OR, WA	intermediate	benthic	cool	invertivore
Rhinichthys osculus	speckled dace	OR, WA	intermediate	benthic	cool	invertivore
Richardsonius balteatus	redside shiner	OR, WA	intermediate	water column	cool	invertivore
Gasterosteidae (sticklebacks)						
Gasterosteus aculeatus	threespine stickleback	OR, WA	tolerant	hider	cool	invertivore
Petromyzontidae (lampreys)						
Lampetra tridentate	Pacific lamprey	OR, WA	intermediate	hider	cool	filter feeder
Lampetra richardsoni	western brook lamprey	OR, WA	intermediate	hider	cool	filter feeder
Salmonidae (trout and salmon)						
Oncorhynchus tshawytscha	chinook salmon	OR, WA	sensitive	water column	cold	invertivore
Oncorhynchus kisutch	coho salmon	OR, WA	sensitive	water column	cold	invertivore
Oncorhynchus clarki	cutthroat trout	OR, WA	sensitive	water column	cold	invert/piscivore
Oncorhynchus mykiss	rainbow trout	OR, WA	sensitive	hider	cold	invert/piscivore
Salvelinus fontinalis	brook trout	non-native	sensitive	hider	cold	invert/piscivore
Salvelinus confluentus	bull trout	OR, WA	sensitive	hider	cold	invert/piscivore
Umbridae						
Novumbra hubbsi	Olympic mudminnow	WA	tolerant	hider	warm	invertivore

Non-native = non-native, exotic, or introduced species OR = native to Oregon, WA = native to Washington had no fish species. The relative abundance among sites of each species is illustrated in Appendix 3.

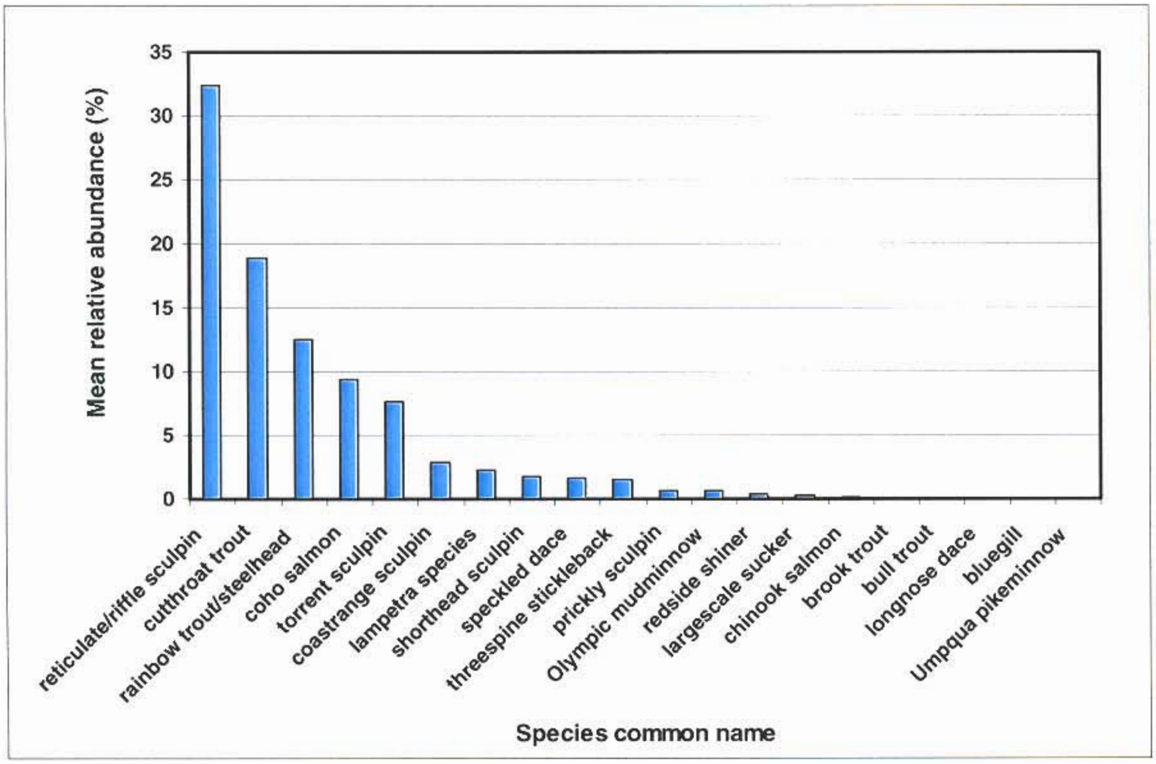
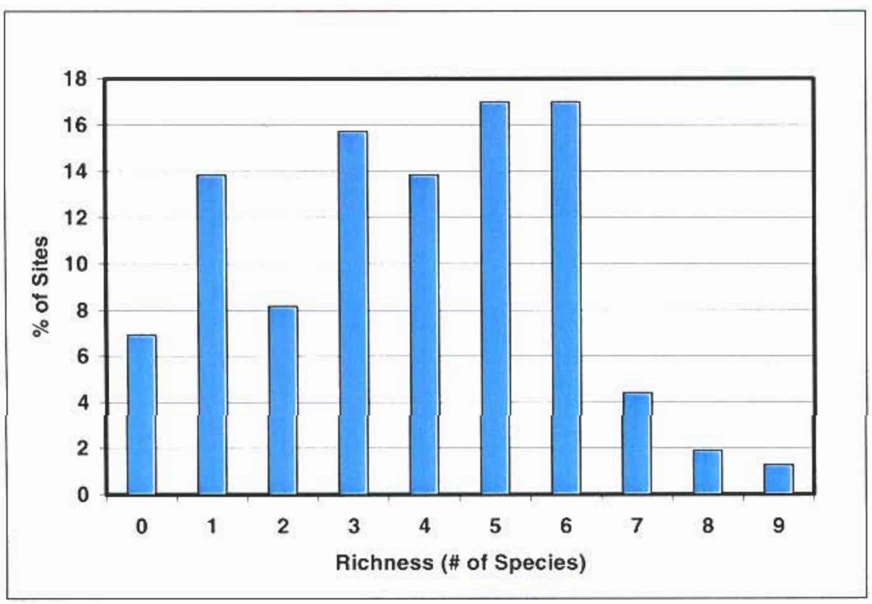


Figure 6. Mean relative abundance of fish species sampled from 159 Coast Range sites.



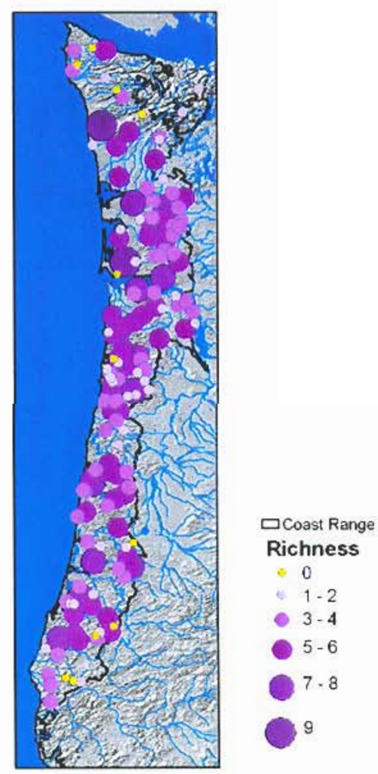


Figure 7. Histogram of fish species richness of Coast Range sample sites (n=159).

3.2 Landscape Description

The general landscape characteristics across the sample contributing areas are described by the 16 landscape metrics summarized in Table 2. There was a wide range in elevation, precipitation (Figure 1), and basin-size represented in the sample population of watersheds. Most sample stream reaches occur in upland forested watersheds where the level of human disturbance indicated by the landcover metrics was typically low (Figure 8). Forestry harvest activity was the most prevalent form of human disturbance in the sample watersheds identified from the landscape data.

Table 2. Descriptive statistics of select landscape metrics for contributing areas delineated for 159 sample points in the Coast Range Ecoregion.

Landscape metric	Units	Mean	Std. Error	Median	Std. Dev.	Range	Min.	Max.
watershed area	km^2	24.95	2.80	9.87	35.26	168.56	0.09	168.66
mean elevation	m	375.91	17.75	338.56	223.82	1183.34	55.04	1238.38
mean watershed slope	%	21.78	0.84	20.16	10.51	76.81	6.37	83.17
annual mean precipitation	cm	247.31	5.06	239.00	63.75	336.00	156.00	492.00
July mean max air temperature	c	21.11	0.11	20.74	1.37	6.01	18.72	24.73
stream distance to terminal water	km	65.11	4.20	49.95	52.96	278.16	0.58	278.75
stream density	km/km^2	0.77	0.03	0.70	0.36	2.95	0.00	2.95
road density	km/km^2	1.74	0.08	1.66	0.97	5.75	0.00	5.75
forest cover	%	90.32	0.85	92.67	10.77	67.40	32.60	100.00
Wetland cover	%	0.06	0.02	0.00	0.21	1.48	0.00	1.48
urban cover	%	0.06	0.02	0.00	0.23	1.70	0.00	1.70
agriculture cover	%	0.41	0.13	0.00	1.68	15.21	0.00	15.21
barren cover	%	2.23	0.28	0.96	3.56	27.31	0.00	27.31
transitional/forest regrowth	%	8.65	0.80	6.55	10.12	66.57	0.00	66.57
shrub cover	%	0.33	0.07	0.05	0.94	7.98	0.00	7.98
rangeland cover	%	0.08	0.06	0.00	0.71	6.65	0.00	6.65

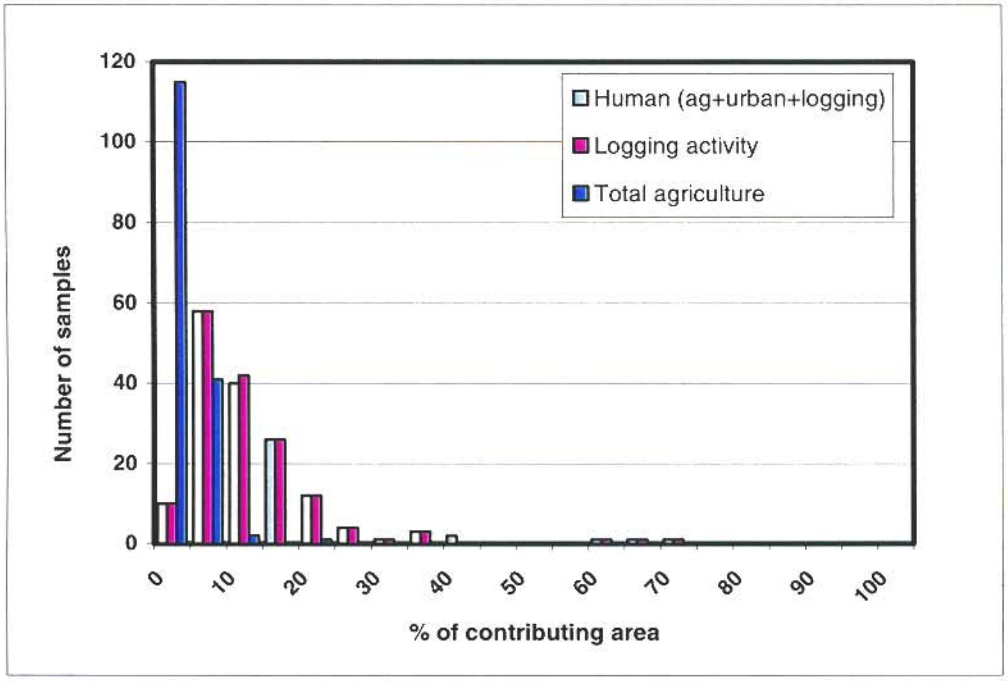


Figure 8. Histogram showing the distribution of human disturbance landscape metrics within contributing areas of the Coast Range sample sites (n=159).

3.3 Reduction of Variables (Correlation and PCA)

Screening of the landscape data resulted in a list of 42 metrics for use in the multivariate analyses (Appendix 4). Land cover metrics for the entire contributing areas and for those defined within 120m of the stream were more useful than the 30 or 60 meter buffer width.

3.4 Ordination Results

The ordination analysis requires the removal of rare species and sites with no fish from the dataset (ter Braak and Similauer 1998). Species are considered rare if they occur at few sites (<5%, Gauch 1982) and if they occur at sites that are species poor (you mean low species richness) (ter Braak and Smilauer 1998, Jongman et al. 1995). Species that occurred at three or fewer sites (bluegill, Olympic mudminnow, Umpqua pikeminnow, longnose dace, and bull and brook trout) were removed. Shorthead sculpin, which occur at four sites, were also removed as these sites had very few species. The resulting data set used for all ordination analyses had 146 sites with 13 fish species. Fish data metrics were expressed as percent relative abundance to normalize among sites (ter Braak and Smilauer 1998).

Ordination results include eigenvalues, species-environmental correlation, and cumulative percent of variance explained for each of four axes. Eigenvalues measure the importance of each of the axes (between 0 and 1). The species-environmental correlation (r) measures the strength of the relation between species and environment for a particular axis. This is the correlation between the linear combinations of sample scores derived from the species data and the sample scores of the environmental variables.

3.4.1 Detrended Correspondence Analysis

The Detrended Correspondence Analysis (DCA) was used to determine the inherent order in the species by sites matrix unconstrained by environmental variables. Output from this analysis includes eigenvalues, length of gradient, and cumulative variance. Length of gradient measures how unimodal the species responses are along the ordination axis and is expressed as standard deviation units of species turnover. The length of the gradient of the first axis exceeded a gradient length of four standard deviations indicating a strong unimodal response (Table 3). The remaining axes have moderate amount of unimodality.

The first four axes of this analysis explained 43.8% of the variance. This is a relatively high value as cumulative variance (compared to that of regression of a single species with environmental variables) is typically low for species data, which is often very noisy.

Table 3. Summary statistics for DCA of species relative abundance for four axes. Sum of all unconstrained eigenvalues = 3.697.

	Axes			
	1	**2**	**3**	**4**
Eigenvalues	.674	.496	.254	.197
Length of gradient	4.343	2.640	3.566	2.555
Cum. % variance of species data	18.2	31.6	38.5	43.8

3.4.2 Canonical Correspondence Analysis

CCA were run for landscape metrics, instream metrics and both datasets combined. These analyses are summarized in Table 4. Results of the CCA are expressed as joint plots, which represent the weighted averages of species with respect to quantitative environmental variables. The joint plot axes are a linear combination of standardized environmental variables, computed by the CCA algorithm of the form: Axis1 = (-.1101 * elevmean) + (-0.1934 * ann_prec) + ... + (0.2466 * xfc_aqm). Each species' score (coordinate) on an axis is the abundance weighted average of the site scores. Each environmental variable is represented as an arrow projected onto each axis relative to its coefficient. Thus, the length of each arrow indicates its relative contribution of environmental variables to each axis. The direction and magnitude of these vectors indicate the covariation among environmental variables. Ecological gradients can be inferred, after examining the variables and coefficients for each axis. Each axis is orthogonal, (uncorrelated with the other axes).

Table 4. Summary statistics for three CCA analyses (landscape metrics, instream metrics, and landscape/instream combined) of species relative abundance for four axes. Sum of all unconstrained eigenvalues = 3.697 for landscape and combined model and 3.489 for instream-only model. Number of environmental metrics used for each model listed in heading.

	Axes			
	1	2	3	4
Landscape metrics (9 metrics)				
Eigenvalues sum= 1.115	0.399	0.318	0.123	0.112
Species-environment correlations	0.789	0.792	0.523	0.593
Cum. % variance of species data	10.8	19.4	22.7	25.7
Cum. % variance of species/envir.	35.8	64.3	75.3	85.3
Instream metrics (12 metrics)				
Eigenvalues sum= 1.162	0.385	0.302	0.189	0.108
Species-environment correlations	0.766	0.691	0.624	0.547
Cum. % variance of species data	11.0	19.7	25.1	28.2
Cum. % variance of species/envir.	33.1	59.1	75.4	84.7
Combined metrics (15 metrics)				
Eigenvalues sum= 1.568	0.42	0.378	0.255	0.18
Species-environment correlations	0.805	0.807	0.693	0.708
Cum. % variance of species data	11.3	21.6	28.5	33.3
Cum. % variance of species/envir.	26.8	50.9	67.2	78.6

Note: Site OR011 treated as supplemental due to very high NH4 value for instream-only model run.

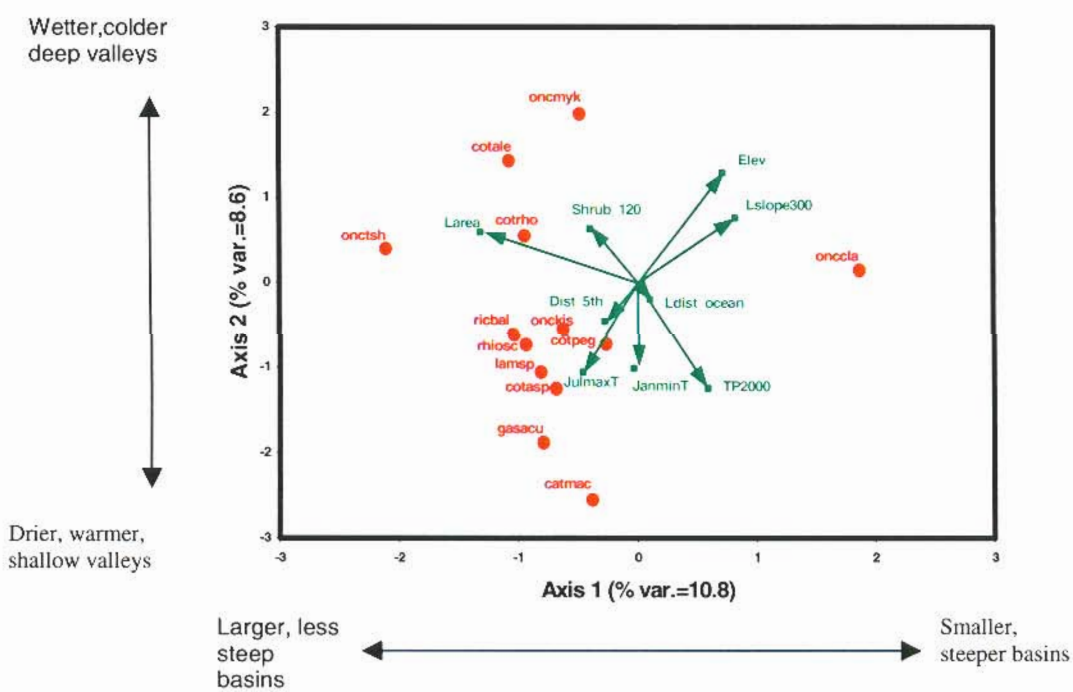


Figure 9. Biplot of axis 1 and 2 CCA results of species relative abundance and landscape metrics. Used P value cutoff of 0.05 resulting in inclusion of 9 landscape metrics. Cumulative variance of the four axis=25.7.

The landscape variables alone provided good explanation of the species matrix. The joint plot of the first two axes shows that axis one is primarily an elevation, moisture, and valley steepness gradient (Figure 9). Axis two is a hydrological gradient ranging from large wet basins to small exposed drier basins. The combined data set of landscape and habitat improves the explanation of the species by site matrix (Figure 10, Table 4). The first two axes are similar to those of the landscape-only joint plot except the combined data set shows a correlation of high and low nutrients along the first axis. Higher order axes tend to separate out a few species that are close together in lower order axes. For example, steelhead (ONCMYK) is clustered with several species on axes 1 and 2, but a graph of axes 3 and 4 shows greater separation.

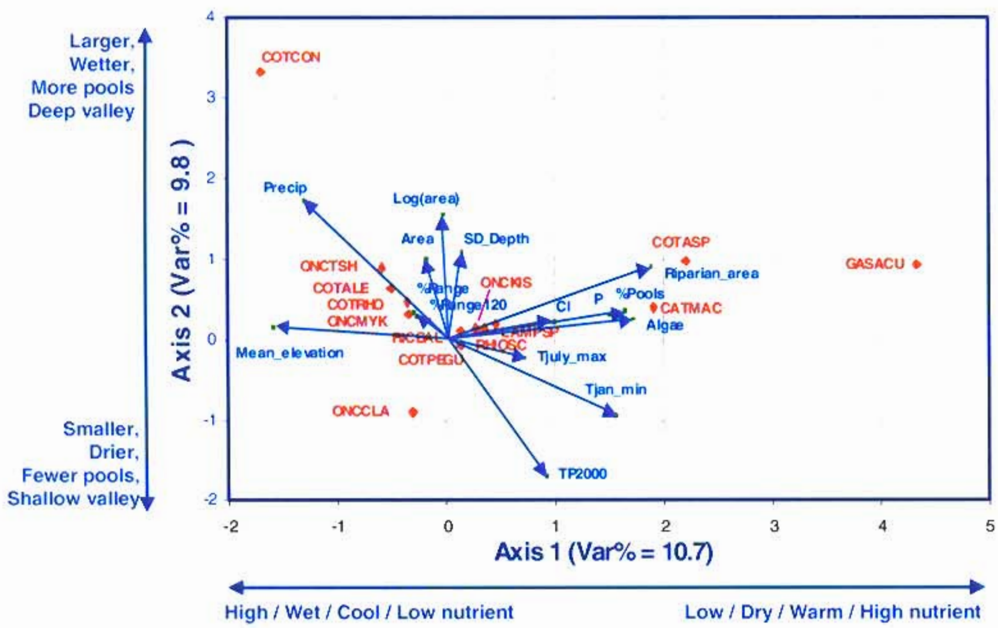


Figure 10. Biplot of axis 1 and 2 CCA results of species relative abundance and the combined data set of landscape and habitat and chemistry metrics. Used P value cutoff of 0.05 resulting in inclusion of 12 landscape metrics. Cumulative variance of the four axis=33.3.

Table 5. Summary statistics for three CCA analyses (landscape metrics, instream metrics, and landscape/instream combined) of species relative abundance for four axes. Sum of all unconstrained eigenvalues = 3.697 for landscape and combined model and 3.489 for instream-only model. Number of environmental metrics used for each model listed in heading.

	Axes			
	1	2	3	4
Landscape metrics (9 metrics)				
Eigenvalues sum= 1.115	0.399	0.318	0.123	0.112
Species-environment correlations	0.789	0.792	0.523	0.593
Cum. % variance of species data	10.8	19.4	22.7	25.7
Cum. % variance of species/envir.	35.8	64.3	75.3	85.3
Instream metrics (12 metrics)				
Eigenvalues sum= 1.162	0.385	0.302	0.189	0.108
Species-environment correlations	0.766	0.691	0.624	0.547
Cum. % variance of species data	11.0	19.7	25.1	28.2
Cum. % variance of species/envir.	33.1	59.1	75.4	84.7
Combined metrics (15 metrics)				
Eigenvalues sum= 1.568	0.42	0.378	0.255	0.18
Species-environment correlations	0.805	0.807	0.693	0.708
Cum. % variance of species data	11.3	21.6	28.5	33.3
Cum. % variance of species/envir.	26.8	50.9	67.2	78.6

Note: Site OR011 treated as supplemental due to very high NH_4 value for instream-only model run.

The relative contribution of the landscape and instream datasets can be expressed in terms of the amount of variance explained by the unconstrained species by site matrix from the DCA analysis. The landscape data accounted for 25.7% of the cumulative variance. With a ceiling of 43.8% cumulative variance by the unconstrained analysis, this represents 59.8% variance explained by the landscape data alone. The instream variable alone also provide a good explanation fo the species matrics accounting for 28.6% of the variance. The combined dataset explains 33.3% (Figure 11, Table 4).

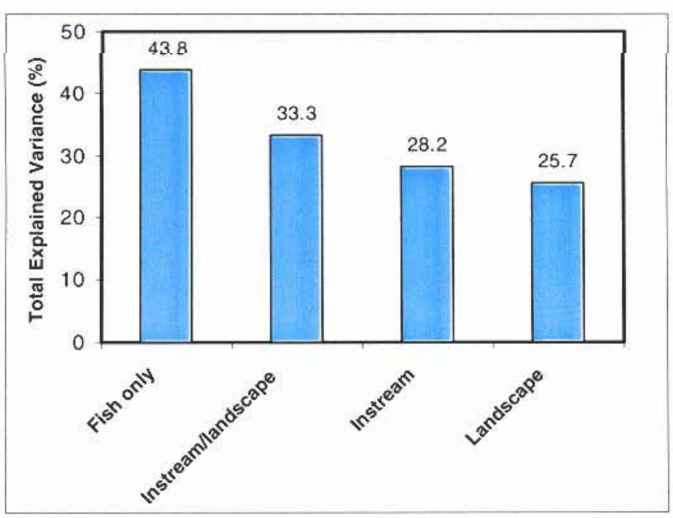


Figure 11. Summary of CCA results comparing results of analyses using only landscape, only instream, and combination of instream and landscape metrics. Fish only represents the unconstrained DCA model.

A Redundancy Analysis (RDA) was run to explore the relationship of landscape metrics to in-stream variables and to test how much explanatory power is lost by using only landscape or only in-stream variables to explain the order of the fish data. RDA is a linear response technique, so the coordinates of the response variables (in this case the in-stream measurements) are the slopes against each axis' gradient. Summary results are in Table 5 showing 32% of the variance of the instream variables can be explained by a linear relationship with landscape gradients. Statistical details regarding the ordination analysis are in Appendix 2.

Table 6. Summary statistics for RDA of landscape comparison to instream metrics. Sum of all unconstrained eigenvalues = 1.0 and sum of all canonical Eigenvalues =0.345.

	Axes			
	1	**2**	**3**	**4**
Eigenvalues :	0.240	0.045	0.026	0.012
Species-environment correlations	0.787	0.553	0.501	0.363
Cumulative percentage variance:				
of species data	24.0	28.5	31.1	32.3
of species-environment relation	69.5	82.6	90.1	93.7

3.5 Regression Analysis Results

The current status of fish abundance as predicted by remotely-sensed landscape data was modeled using step-wise multiple linear regressions. All sites (including 11 sites with no fish) except one outlier were included in the analysis (n=158). Species richness was the response variable (square-root transformed). The best fit equation (R^2 =0.657) used basin area (Larea), distance to major water (Pacific Ocean, Columbia River, or Puget Sound) (Ldist), elevation (Elev), and topographical position index (TP2000s, calculated for area within 2000m of the streams)(Table 5). The model was slightly improved with addition of slope calculated within 300m of the sample point (Slope300). Basin area has the strongest effect on species richness (R^2 =0.47) and basin area and elevation combined explained most of the variability (not transformed) (R^2 =0.60). The best regression equation using four metrics is as follows (R^2 =0.657).

Richness = -3.079 + .594*Larea + .269*Ldist – .002*TP2000S - .002*Elev

The instream data (water chemistry and habitat) were combined with the landscape data to see if the model could be improved. The resulting best model retained the same landscape metrics as in the landscape only model (see below). Two habitat metrics were added: 'coniferous riparian canopy (%) (Pcan)' and 'sand and fine sized substrate (%)' (Psafn), yielding a slight improvement (R^2 =.687). The regression model is shown below and summary results are in Table 6. Details of the regression statistics including outliers and model diagnostics are in Appendix 2.

Richness = -2.484 + .522*Larea + .291*Ldist – .001*TP2000s - .002*Elev-.580*Pcan-.004*Psafn

3.6 Classification of Fish Assemblages

As indicated by the species presence and relative abundance, there are five species that dominate the overall fish assemblage (Figures 6&7). These species tend to define the six fish assemblages present in the Coast Range based on the cluster analysis (Table 7). Although there are several distinct assemblages identified in the dataset, most of the sites are within one very large cluster. Varying abundance of reticulate/riffle sculpin combined with cutthroat or rainbow trout define the sub-groupings within this cluster. Generally, the clusters tend to be defined by abundance of species as much as by species presence due to the limited number of species.

Table 7. Summary statistics for two regression models (landscape metrics and landscape/instream metrics) with species richness (square-root transformed) as response variable (N=158).

Variable	Coefficients	Std. Error	t Stat	P-value
Landscape metrics R^2 = .657, P<.0001, intercept= -3.079				
L10_area (log basin area)	0.594	0.054	10.906	0.000
ELEVM (mean elevation)	-0.002	0.000	-8.545	0.000
Lfs_ocr (log distance to ocean/Col.R.)	0.269	0.077	3.474	0.001
TP2000S (TPI 2000m buffer)	-0.002	0.001	-2.338	0.021
Landscape/instream metrics R^2 =.687, P<.0001, intercept= –2.484				
PCAN_C (percent canopy coniferous)	-0.580	0.218	-2.664	0.009
PCT_SAFN (percent sand/fine substrate)	-0.004	0.002	-2.445	0.016
L10_area (log basin area)	0.522	0.056	9.339	0.000
ELEVM (mean elevation)	-0.002	0.000	-6.962	0.000
Lfs_ocr (log distance to ocean/Col.R.)	0.291	0.077	3.806	0.000
TP2000S (TPI 2000m buffer)	-0.001	0.001	-2.019	0.045

4. Discussion

Physiogeographic variables have the strongest correlation to species richness and a combination of these variables resulted in the best regression model predicting the current status. Because of the overwhelming effect of physiogeographic variables in the regression model, it is difficult to evaluate the effect of human disturbance on the fish distribution. One approach would be to separate the effects by building separate models or use a form of regression that lets you account for the more subtle influences (decrease effect of dominant variables).

Correlation of richness to human disturbance (both landscape and surface waters) metrics was typically very low, thus, ability to separate the natural gradients that are related to the distribution of fish assemblages from the gradients of human disturbance is limited. It is unlikely that a relationship between fish metrics and human disturbance landscape metrics can be developed with the current data set because of the low representation of human disturbance in the sample watersheds. It may be possible to do this analysis with the addition of more sample points in watersheds with a greater amount of disturbance.

The overriding influence of physical features on the fish taxonomic richness has been noted by other workers (Osborne and Wiley 1992, Allen et. al. 1999, Schlosser and Kallemeyn 2000). Couple with a weak response to human disturbance, the usefulness of this metric as an indicator of biointegrety is limited (Allen et. al. 1999). Also, peak richness has been associated with intermediate levels of disturbance (Huston 1979), particularly due to exotic species. We experimented with other response variables (e.g. % tolerant species, species diversity). Generally, these relations to landscape metrics were weak with diversity indices (i.e. Shannon-Wiener, and Simpson) having the most promise. Data sets with low species richness and low species diversity, which are characteristic of fish assemblages of the Pacific Northwest, often yield weak results in this type of analysis. Therefore, the results of our modeling were quite promising considering this limitation. The recently developed Coast Range aquatic vertebrate index of biological integrity (IBI) shows that a combination of vertebrate metrics has a clear and predictable response to human disturbance in the Coast Range (Hughes et al. accepted). This IBI will be used in future analysis of the landscape data. Another approach that may prove useful in determining importance of environmental variable would be to use presence/absence of response metrics (e.g. tolerant species) in a logistic regression analysis (Nash and Bradford 2001).

For the purpose of modeling inferences of explanatory (independent) variables to all sites over a landscape, remotely-sensed data have a couple of advantages. First, problems associated with missing data are avoided as landscape data can be generated and applied across all possible sites. Second, there are no physical limits to populating the dataset. With field data, problems with access denial or the inability to reach a site due to physical barriers (e.g. extremely far from a road, unsafe banks, waterfalls) results in entire sites that can't be sampled. Access denial is a chronic problem in EMAP site selection, resulting in a greater proportion of sites being sampled on public land. Finally, remotely sensed data is often less expensive per site.

Both data sets have a human disturbance component. The stream site data has direct human disturbance metrics based on stream reach and riparian zone observations as well as other metrics that are often related to human disturbance, such as abundance and type of fish cover, LWD presence, and fine sediment metrics. The landscape data has land-use cover types for agriculture, urban, transitional, and forest re-growth. Relations to all human disturbance metrics were weak. The human disturbance data did not have a strong relationship in the regression analysis but stronger relationships were observed with the correspondence analysis.

New metrics, waterbody distance metrics and TPI, were very significant in both types of analyses. The relation of the distance to the larger waterbody influencing fish species richness, is a well known concept in fish ecology (Osborne and Wiley 1992, Schlosser 1987) and the influence of this characteristic is often great in systems that are relatively close to coastal areas. Using GIS to automate generation of various distance metrics allowed for experimentation to determine metrics that best described species richness. Metrics that quantify distance between stream orders may be more useful in the future as the base map used in the calculations improves.

The TPI metric describes the proportion of narrow, incised drainages to broader, flatter drainages indicating the overall steepness of the terrain. This metric gives more options in how we can account for the slope. Rather that being restricted to several metrics that describe this aspect of habitat in the field data (slope, channel incision, bankfull height), TPI gives a description of a broader contributing area. Not only does this metric describe fish habitat but should prove to be a useful metric for looking at the hydrologic response to weather and snowmelt events and sediment transport due to effects of human disturbance.

5. Conclusions

-Fish assemblages in the Coast Range are structured mainly by physical and biogeographic gradients. Some residual variation may be explained by human disturbance variables.

-Large-scale landscape metrics combined with instream metrics provided the most powerful explanation of species distribution. When analyzed separately, both landscape and instream variables explained substantial proportion of assemblage composition and structure.

-Standard metrics generated by the EMAP landscape analysis are useful but additional metrics of TPI and distance measurements were substantial additions for modeling Coast Range fish distributions. These metrics will be useful in future analyses of predicting stream biota.

-Large-scale landscape metrics were useful for interpretation of stream condition. When available, it would be advantageous to use landcover data with finer spatial resolution although the gross scale NLCD did an adequate job.

-Developing the relation of human disturbance to biotic and inchannel condition will require a more complete gradient of landscape condition. This will improve our ability to relate human disturbance to stream condition so that a baseline relationship can be developed that will be useful for looking at future trends as human development increases in the Coast Range Ecoregion.

-Limiting the predictor variables to those that can be generated from universally available landscape and hydrologic data allows any site to be placed on the ordination diagram and evaluated with respect to other sites and species centroids.

References

Allen, A.P., T.R. Whitter, P.R. Kaufmann, D.P. Larsen, R.J. O'Connor, R.M. Hughes, R.S. Stemberger, S.S. Dixit, R.O. Brinhurst, A.T. Herlihy, and S.G. Paulsen. 1999. Concordance of taxonomic richness patterns across multiple assemblages in lakes of the northeastern United States. Canadian Journal of Fisheries and Aquatic Sciences 56: 739-747.

Berry, W.D. and S. Felman. 1985. Multiple Regression in Practice. Sage Publications, Beverly Hills.

CANOCO Version 4.0. 1998. Software for Canonical Community Ordination, Version 4. Centre for Biometry Wageningen, CPRO-DLO, Wageningen, The Netherlands. Distributed by Microcomputer Power. Ithaca, NY, USA.

Ebert, D., T. Wade, J. Harrison, and D. Yankee. 2000. Analytical tools interface for landscape assessments (ATtILA) User Guide. Version 1.004. Office of Research and Development, U.S. EPA, Las Vegas, NV.

ESRI. 2002. ArcInfo version 8.3. Redlands California, Environmental Systems Research Institute, Inc.

Franklin, J.R. and C.T. Dyrness. 1973. Natural vegetation of Oregon and Washington. USDA Forest Service Gen. Tech. Rep. PNW-8, PNW Range and Experiment Station, Portland, OR.

Gauch Jr., H.G. 1982. Multivariate analysis in community ecology. Cambridge, Cambridge University Press.

Hughes, R.M., S. Howlin, and P.R. Kaufmann. *Accepted.* A biointegrity index for coldwater streams of western Oregon and Washington. Transactions of the American Fisheries Society.

Hughes, R.M., P.R. Kaufmann, A.T. Herlihy, T.M. Kincaid, L. Reynolds, and D.P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences 55:1618-1631.

Huston, M. 1979. A general hypothesis of species diversity. American Naturalist 113:81-101.

Insightful. 2001. S-PLUS 6 Professional Edition for Windows, Release 2. Insightful Corp.Seattle, Washington.

Jongman, R.H.G., C.J.F. ter Braak, and O.F.R.Van Tongeren. (eds). 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge.

Kaufmann, P.R., P. Levine, E.G. Robison, C. Seeliger, and D.V. Peck. 1999. Quantifying physical habitat in wadeable streams, EPA 620/R-99/003, U.S. Environmental Protection Agency: Washington, D.C.

Lazorchak, J.M., D.J. Klemm, and D.V. Peck (eds). 1998. Environmental monitoring and assessment program -- surface waters: field operations and methods for measuring the ecological condition of wadeable Streams. EPA/620/R-94/004F. Office of Research and Develop, U.S.EPA, Washington, D.C.

Nash, M.S. and D. F. Bradford. 2001. Parametric and nonparametric (MARS; multivariate additive regression splines) logistic regressions for prediction of dichotomous response variable with an example for presence/absence of an amphibian. EPA/600/R-01/081. U.S. Environmental Protection Agency, Washington, D.C.

Osborne, L.L. and M.J. Wiley. 1992. Influence of tributary spatial position on the structure of warmwater fish communities. Canadian Journal of Fisheries and Aquatic Sciences 49: 671-681.

MacNally, R. 2000. Regression and model building in conservation biology, biogeography, and ecology: the distinction between -and reconciliation of- "predictive" and "explanatory" models. Biodiversity and Conservation 9:655-671.

Omernik, J.M. 1987. Ecoregions of the United States: Map at a scale of 1:7,500,000. Suppl. Annals American Assoc. Geogr. 77(1).

Pater, D.E., S.A. Bryce,T.D. Thorson, J.M. Lammers, J. Kagen, C. Chappell, J.M. Omernik, S.H. Azevedo, and A.J. Woods. 2000. Ecoregions of Western Washington and Oregon (color poster with map, descriptive text, summary tables, and photographs). Denver, Colorado, U.S. Geological Service (map scale 1:1,350,000).

Reynolds, L., A.T. Herlihy, P.R. Kaufmann, S.V. Gregory, and R.M. Hughes. 2003. Electrofishing effort requirements for assessing species richness and biotic integrity in western Oregon streams. North American Journal of Fisheries Management 23:450-461.

Schlosser, I. J. 1987. A conceptual framework for fish communities in small warmwater streams (pages 17 – 26) in Community and Evolutionary Ecology of North American Stream Fishes. W. J. Matthews and D. C. Heins, (eds) Oklahoma University Press, Norman, Oklahoma.

Schlosser, I. J. and L. W. Kallemeyn. 2000. Spatial variation in fish assemblages across a beaver-influenced successional landscape. Ecology 81: 1371-1382.

ter Braak, C. J. F. and P. Smilauer. 1998. CANOCO reference manual and user's guide to Canoco for Windows. Centre for Biometry Wageningen, CPRO-DLO, Wageningen, The Netherlands.

Zaroban, D. W., M. P. Mulvey, T.T. Maret, R. M. Hughes, and G. D. Merritt. 1999. Classification of species attributes for Pacific Northwest freshwater fishes. Northwest Science 73(2): 81-93.

Appendices

Appendix 1. Habitat and chemistry metrics available for multivariate analyses.

Code	Metric	Units
DO_mgl	dissolved oxygen	mg/l
NO3_ueql	nitrate	ueq/L
TEMP_C	temperature	Celsius
NH4_ueql	ammonium	ueq/L
Ptl_ugl	total phosphorus	ug/L
XBKF_W	bankfull width-mean	m
XWIDTH	mean wetted width	m
LSUB_DMM	log10 geometric mean substrate diameter	mm
PCAN_C	fraction of reach with coniferous dominated canopy	%
PCT_BDRK	substrate bedrock	%
PCT_BIGR	Substrate ≥coarse gravel (>16mm)	%
PCT_FN	substrate fines--silt/clay/much	%
PCT_HP	substrate hardpan	%
PCT_ORG	substrate wood or organic matter	%
PCT_RC	substrate concrete	%
PCT_SA	substrate sand .6-2mm	%
PCT_SAFN	substrate sand or fines <2mm	%
PCT_SFGF	substrate ≤ fine gravel<16mm	%
W1_HAG	agricultural human disturbance	proximity wtd sum
W1_HALL	all human disturbance	proximity wtd sum
W1_HNOAG	non-agricultural human disturbance	proximity wtd sum
W1H_PIPE	pipes	proximity wtd sum
W1H_WALL	channel revetment	proximity wtd sum
XC	fraction of reach covered by canopy	%
XCL	riparian area Canopy cover >.3m DBH	%
XCMGW	riparian area covered by any woody veg	%
XCMW	riparian area covered by large woody veg	%
XFC_ALG	fish cover--filamentous algae	Areal proportion
XFC_ALL	fish cover--all types	Areal proportion
XFC_AQM	fish cover--aquatic macrophytes	Areal proportion
XFC_BIG	fish cover--LWD, RCK, UCB & HUM	Areal proportion
XFC_BRS	fish cover--brush & small debris	Areal proportion
XFC_HUM	fish cover--artificial structures	Areal proportion
XFC_LWD	fish cover--large woody debris	Areal proportion
XFC_NAT	fish cover--natural types	Areal proportion
XFC_OHV	fish cover--overhang vegetation	Areal proportion
XFC_RCK	fish cover--boulders	Areal proportion
XFC_UCB	fish cover--undercut Banks	Areal proportion
XG	riparian ground layer cover present	%
XGB	riparian ground layer barren	%
XPCM	riparian canopy & mid layer present	%
XPCMG	riparian canopy, mid, and ground layer present	%
XSLOPE	mean reach slope	%

Appendix 2. Data issues and Statistical Details

1. Data issues acknowledged prior to analysis with surface waters/fish data and landscape data.

Surface waters/fish:
--Site sample data was based on a single data collection event. Precision of the physical habitat metrics can be estimated from work by Kaufmann et al. (1999).

--Sample sites with no species or with only rare species had to be deleted from the ordination analysis. This criterion resulted in deleting some of the lowland sites (a rather rare feature of the dataset), which contained rare species, from the data set. This resulted in decreasing the length of the environmental gradient that could be modeled.

--Study sites were limited to wadeable streams so this analysis does not address entire Coast Range fish assemblage. However, the majority of Coast Range stream miles (approximately 70%) are wadeable.

--Fish species, especially migratory ones, utilize different habitats depending on life history phase. Some of species may not have been present (or catchable) at the time of sampling although a sample point may have been within a particular species' range at some times of the year.

--Some important habitat metrics could not be included due to excessive missing data (e.g. LWD abundance, pool frequency).

Landscape data:
--The transition cover type is defined as areas of sparse vegetative cover (<25%) that are dynamically changing from one land cover to another, often due to the land use activities (e.g. forest clearcuts, land temporarily cleared of vegetation) and changes due to natural processes (e.g. fire, flood). This cover type is difficult to interpret from the Landsat data so errors are likely.

--The TM data collection period pre-dates that of the surface waters data collection. However the periods are reasonably close, with Coast Range TM data collected 1991-993 and surface waters data collected 1994-1999.

--Most of the sample watersheds have low (5-10%) levels of human disturbance (Figure 8), which reflects the proportion of occurrence of cover types across the entire Oregon/Washington Coast Range Ecoregion. Although the sample is representative of the entire Ecoregion, this distribution limits our ability to assess the relation of the biota to gradients of human disturbance as cover types such as urban and agriculture are not adequately represented.

2. Missing data

The habitat/chemistry data set of 159 sites was at least partially populated for approximately 95 metrics. Of the five data sets used in the analysis, the most serious data gaps were for the Oregon Salmon plan sites (Table A2-1) where data for the thalweg measurements were not available. Also, most chemistry data were not available for the Tillamook/Kilchis sites. The statistical analyses (correlation, regression, and canonical correspondence) can only be performed on sites with complete data and sites with missing values are excluded. One option would be to delete sites that had missing data but it was more important to retain sites because of the relatively low number of species and the frequency of rare species. Metrics with many missing cells were omitted from the analysis. Some of the metrics were marginal, having relatively few missing values (~10% missing). For these metrics, missing values were replacement with the average value of all sites (Table A2-2). Extreme values (based on scatter plots and comparison to median and mean values) were omitted from the calculation of the average values for five metrics. The preferable method for selecting replacement values would be to calculate a predicted value from a regression model of that metric regressed to a surrogate variable. Unfortunately, surrogate variables, such as basin area or distance to divide, which would generate a useable model (significant level of model-F, R2, and coefficient-t values ($p \leq 0.05$) are used in the analyses so would not be independent.

Table A2-1. List of EMAP projects from which sample data was obtained for the analysis.

Year	Project Title	State	No. sample sites
94-96	Coast Range REMAP	OR, WA	72
97	Statewide wadeable sites	OR	16
98-99	Oregon Salmon Plan Sites	OR	41
98-99	Tillamook, Kilchis sites	OR	14
97	Chehalis River Basin REMAP	WA	16

Table A2-2. Replacement values used for missing data.

Metric	count	extreme values[a]	median	mean	#filled	mean w/o extremes	final fill value
DO_mgl	143	0	9.65	9.46	16	--	9.46
NO3_ueql	143	2	8.71	12.59	16	11.65	11.65
TEMP_C	143	0	13.4	13.34	16	--	13.34
NH4_ueql	144	1	1.43	2.50	15	1.62	1.62
Ptl_ugl	157	3	20.00	35.89	2	27.50	27.50
XBKF_W	158	2	8.26	9.86	1	9.34	9.34
XWIDTH	158	1	4.77	6.08	1	5.83	5.83

a=extreme values excluded from average value calculation.

3. Canonical Correlation

Deletion of outliers:
Site OR011 was found to be an outlier in the CCA runs of the instream only data due to its extreme NH4 value. Deletion of the site from the analysis would result in the species catostomus macrocheilus as being rare. Rather than deleting the site from the analysis, the site was treated as a supplemental sample. The CANOCO program (1998) allows for treating a site as supplemental where it is passive and does not influence the definition of the ordination axes.

Correlation values of metrics used in all three of the CCA analyses:

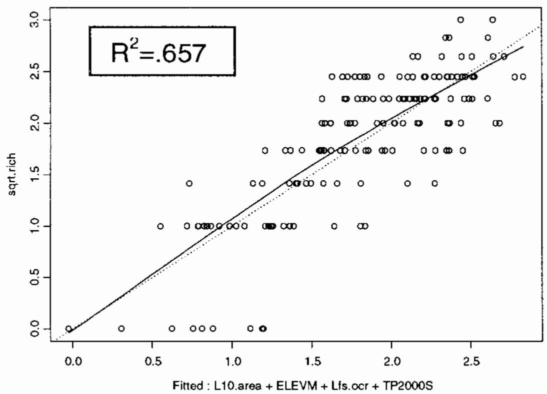
Table A2-3. Correlation matrix of metrics used in CCA.

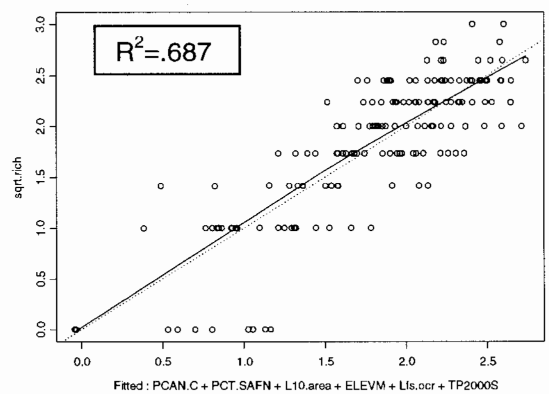
	DO	*ELEVm*	*FDXSLOPE*	*FS_S5*	*janmin*	*julmax*	*L_area*	*ln_fcRCK*	*ln_LWD*	*lnbkfl*	*lnxcmw*	*lnXPCM*
DO	1.00											
ELEVm	0.24	1.00										
FDXSLOPE	0.12	0.39	1.00									
FS_S5	0.21	0.07	0.00	1.00								
janmin	-0.31	-0.69	-0.22	-0.05	1.00							
julmax	-0.14	-0.37	-0.12	0.11	0.01	1.00						
L_area	0.11	0.16	-0.45	-0.02	-0.16	-0.07	1.00					
ln_fcRCK	0.23	0.56	0.20	0.16	-0.31	-0.32	0.31	1.00				
ln_LWD	-0.19	0.04	0.21	-0.06	-0.03	0.02	-0.28	-0.04	1.00			
lnbkfl	0.17	0.34	-0.22	-0.02	-0.27	-0.29	0.85	0.44	-0.18	1.00		
lnxcmw	0.06	0.21	0.13	0.08	-0.23	0.00	-0.01	0.28	0.46	0.08	1.00	
lnXPCM	0.03	0.19	0.10	0.06	-0.21	0.07	0.06	0.25	0.37	0.11	0.91	1.00
log_fs.ocr	0.13	0.23	0.04	0.20	-0.23	0.30	-0.01	0.03	-0.06	-0.04	0.02	0.10
logslp300	0.13	0.39	0.33	0.09	-0.04	-0.20	-0.21	0.52	0.07	-0.03	0.30	0.27
LSUB_DMM	0.34	0.53	0.08	0.01	-0.44	-0.31	0.42	0.66	-0.11	0.53	0.37	0.35
NH4B	-0.25	-0.18	-0.10	-0.11	0.19	0.09	0.09	-0.06	0.00	0.02	-0.07	-0.01
Ptl	-0.31	-0.23	-0.13	0.00	0.09	0.17	-0.07	-0.25	-0.08	-0.14	-0.21	-0.11
RSHRB120	0.03	0.37	-0.02	0.00	-0.37	-0.02	0.15	0.08	0.05	0.12	0.04	0.03
TP2000S	-0.18	-0.59	0.04	-0.08	0.38	0.32	-0.52	-0.51	0.22	-0.62	-0.05	-0.07
W1_HAG	-0.20	-0.31	-0.22	-0.16	0.19	0.15	0.11	-0.33	-0.20	-0.08	-0.44	-0.38
XFC_ALG	-0.01	-0.08	-0.15	-0.15	0.08	0.09	0.18	-0.15	-0.26	0.14	-0.17	-0.15
XFC_AQM	-0.07	-0.23	-0.12	0.04	0.33	-0.05	-0.11	-0.24	-0.17	-0.18	-0.33	-0.20
XFC_BRS	-0.23	-0.20	0.15	0.14	0.28	0.05	-0.45	-0.18	0.43	-0.43	0.09	0.05

	log_fs.ocr	*logslp300*	*LSUB_DMM*	*NH4B*	*Ptl*	*RSHRB120*	*TP2000S*	*W1_HAG*	*XFC_ALG*	*XFC_AQM*	*XFC_BRS*
DO											
ELEVm											
FDXSLOPE											
FS_S5											
janmin											
julmax											
L_area											
ln_fcRCK											
ln_LWD											
lnbkfl											
lnxcmw											
lnXPCM											
log_fs.ocr	1.00										
logslp300	-0.01	1.00									
LSUB_DMM	0.04	0.35	1.00								
NH4	-0.21	-0.09	-0.21	1.00							
Ptl	0.05	-0.25	-0.36	0.09	1.00						
RSHRB120	0.01	-0.08	0.09	-0.10	-0.07	1.00					
TP2000S	-0.06	-0.22	-0.46	0.08	0.17	-0.29	1.00				
W1_HAG	0.05	-0.52	-0.32	0.24	0.31	-0.01	0.17	1.00			
XFC_ALG	0.09	-0.08	-0.04	-0.02	-0.02	-0.02	-0.04	0.08	1.00		
XFC_AQM	-0.12	-0.15	-0.41	0.20	0.31	-0.01	0.14	0.12	0.02	1.00	
XFC_BRS	0.00	0.06	-0.42	0.13	0.11	-0.08	0.35	-0.05	-0.13	0.35	1.00

4. Regression Review of Regression Model

The following graphs show the fit of the points to the regression lines for the two models. Dashed line is regression line and solid line is smoothed line showing general trend in the data.





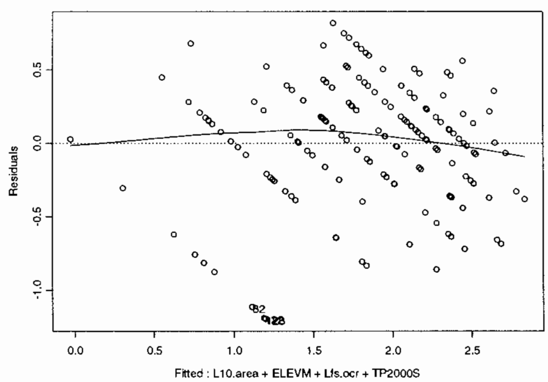
Deletion of outliers for regression model:
One outlier, fishless site WA004, was deleted from the regression analysis. The site would be expected to be fish-bearing based on the elevation, slope, and basin area. However, it is a wetland-type site with very small stream size (1.2m wetted width), 100% fine-sized substrate, and 100% pool type habitat.

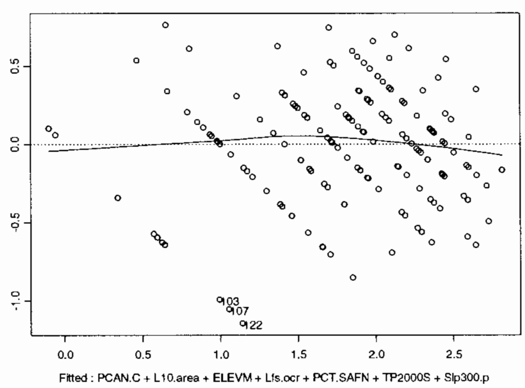
Regression model collinearity:
Collinearity among variables included in both regression models was low, with Pearson r values ranging from 0.41 to -.55.

	PCAN_C	PCT_SAFN	L10_area	ELEVM	Lfs_ocr	TP2000S
PCAN_C	1.000					
Pct_SAFN	-.062	1.000				
L10_area	-0.285	-0.353	1.000			
ELEVM	0.409	-0.547	0.065	1.000		
Lfs_ocr	0.070	0.025	-0.034	0.268	1.000	
TP2000S	0.040	0.448	-0.547	-0.483	-0.029	1.000

Cook's distance:
The graph of the Cook's distance indicated several sites that may be heavily influential on the model. Three sites had Cook's values that were ≥ .10 in the combined landscape/instream model. The models were run without these points (WA001 and WA836 for landscape model and without WA001, WA836, and WA026 for the landscape/instream model). The removal of these sites did not substantially improve the models (R^2=.667 for landscape and R^2=.707 for landscape/instream model).

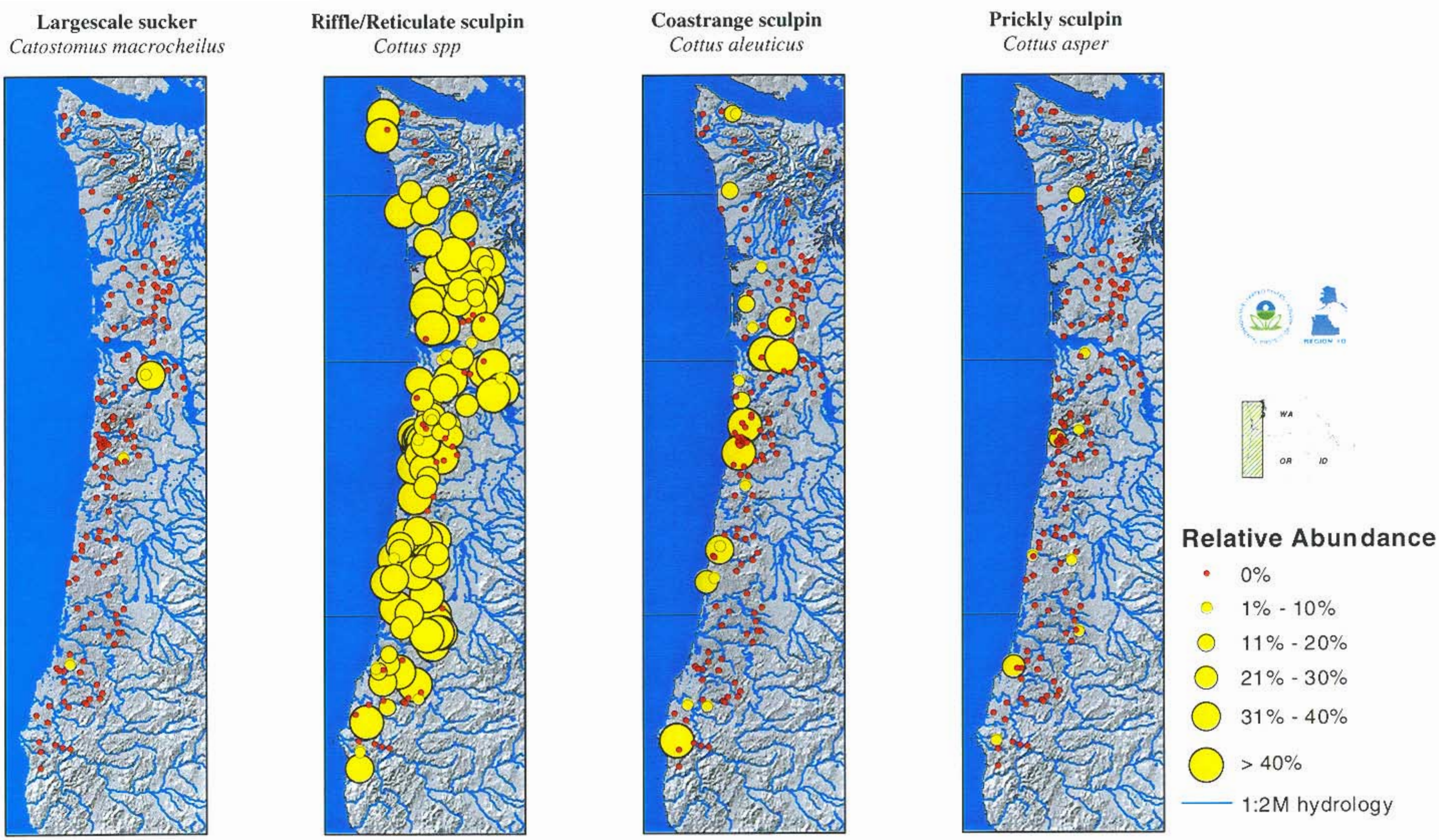
Residuals:
Graphing the residuals on the fitted line showed a good distribution of residuals both above and below the line. Residuals were well aligned on the line of quantiles of standard normal with some deviation on the lower tail of the graph. Analysis of residuals can highlight sites that contain fewer species than expected from their physical characteristics, which might reflect anthropogenic affects, barriers, or historical biogeographic affects. Sites that are richer than expected may reflect exotic species or deliberately stocked and managed sites. A plot of residuals from the basin and elevation regression showed no geographic pattern.





Appendix 3

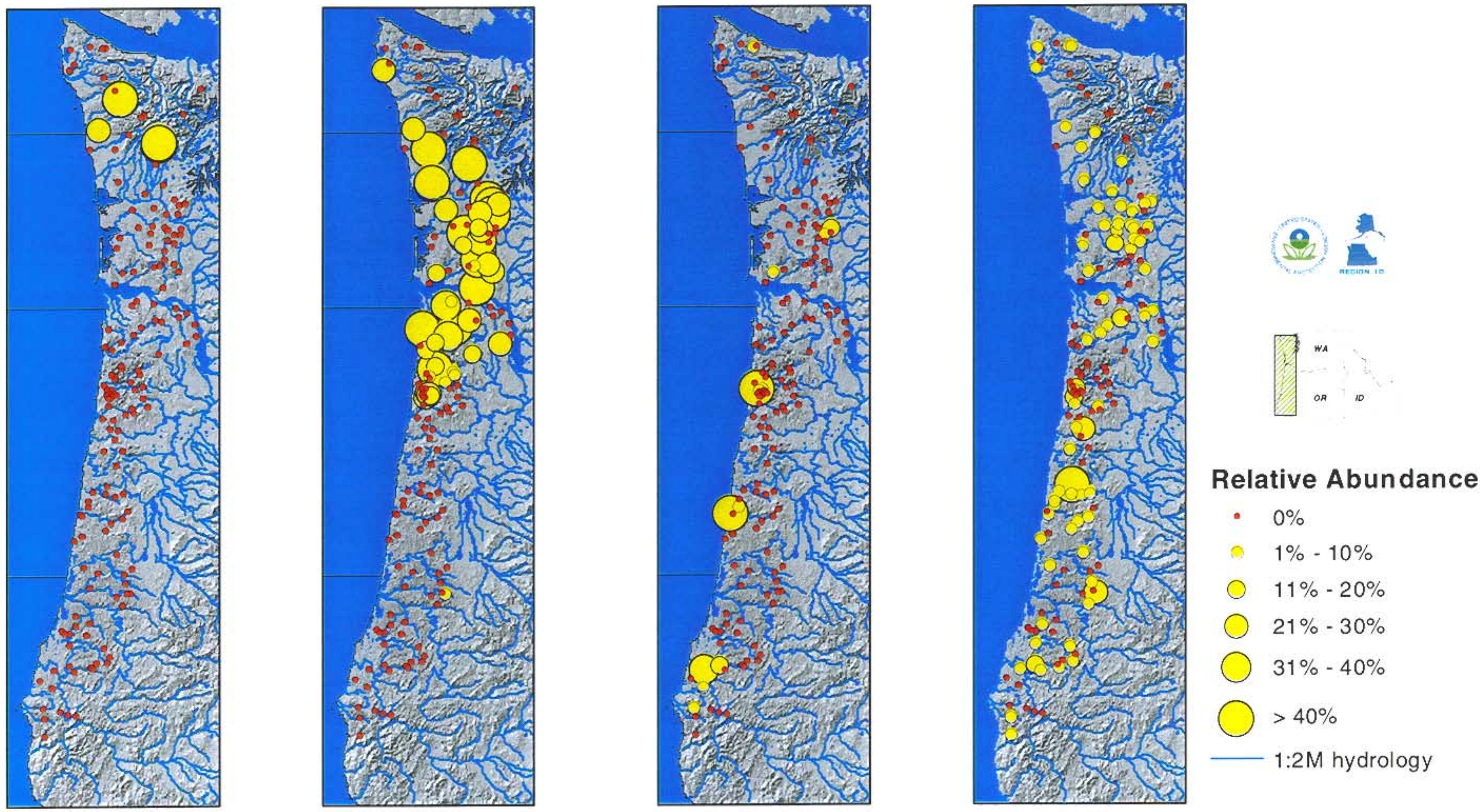
Fish species relative abundance across sample sites.



Appendix 3

Fish species relative abundance across sample sites.

Shorthead sculpin *Cottus confusus* | **Torrent sculpin** *Cottus rhotheus* | **Threespine stickleback** *Gasterosteus aculeatus* | **Lamprey** *Lampetra spp*



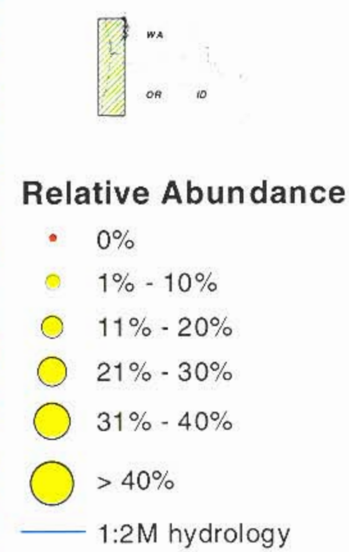
Appendix 3
Fish species relative abundance across sample sites.

Bluegill
Lepomis macrochirus

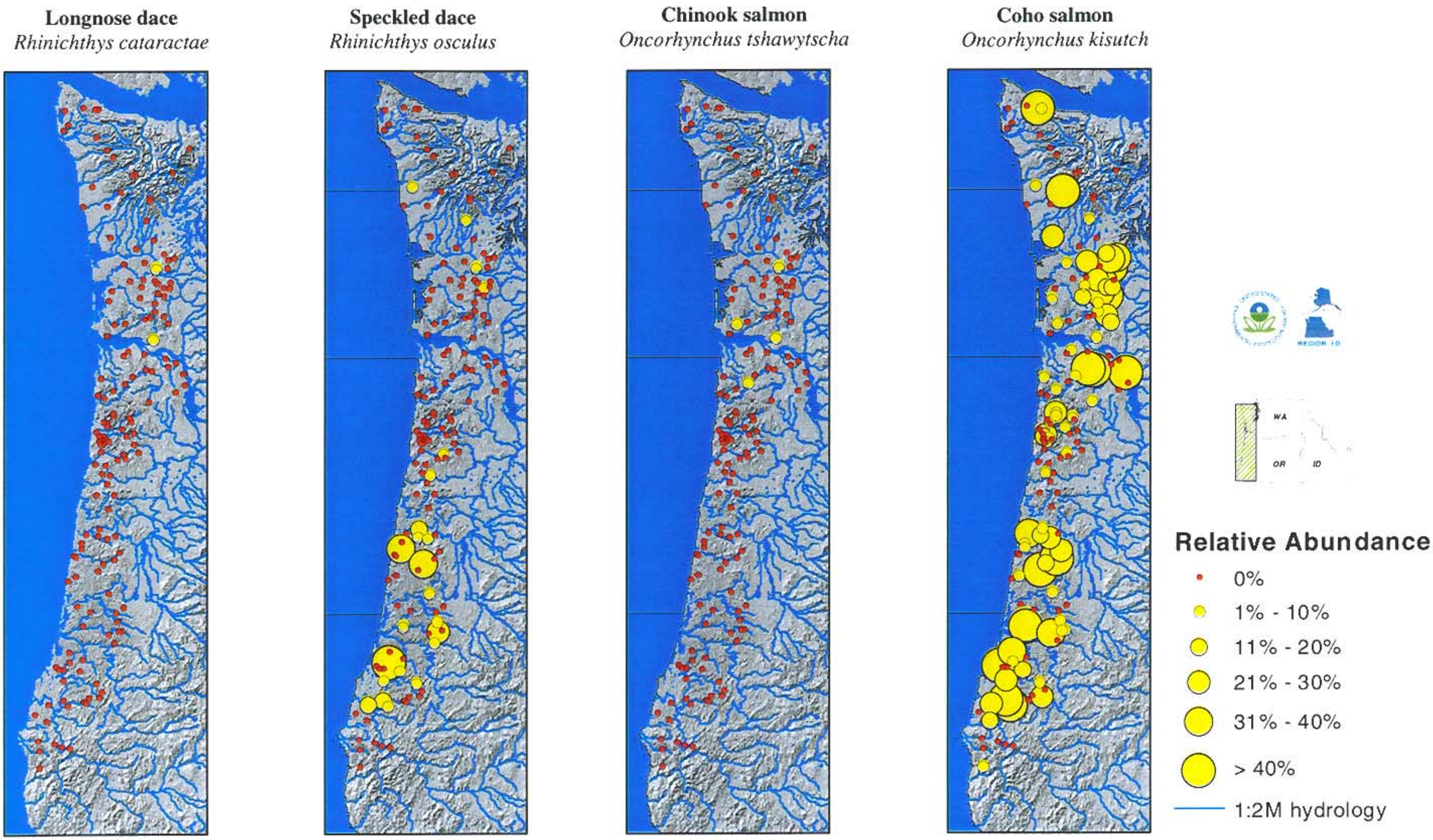
Umpqua pikeminnow
Ptychocheilus umpquae

Olympic mudminnow
Novumbra hubbsi

Redside shiner
Richardsonius balteatus



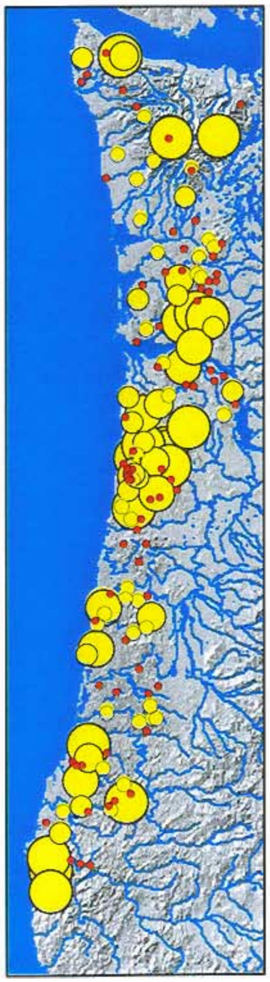
Appendix 3
Fish species relative abundance across sample sites.



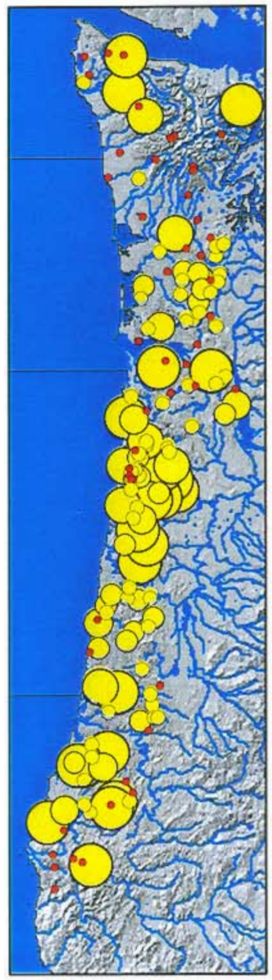
Appendix 3

Fish species relative abundance across sample sites.

Rainbow trout/steelhead
Oncorhynchus mykiss

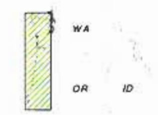


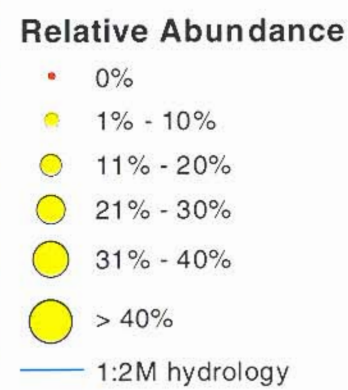
Cuthroat trout
Oncorhynchus clarki



Bull trout
Salvelinus confluentus

Brook trout
Salvelinus fontinalis





Appendix 4. Metrics used in preliminary landscape multivariate analyses.

Code	Metric	Units
basin_AREA	watershed area	m^2
SLPMEAN	Mean watershed slope	%
Slope300	Mean slope under streams +/- 300m of sample point	%
Strahler_gis	Strahler order of sample stream reach	
ELEVMEAN	Mean elevation	m
STRMDENS	Length of stream feeding into point/basin area	km/km^2
TP300	medium scale topo position (calc from area within 300m of streams)	
TP150S	fine scale topo position (calc from area within 150m of streams)	
TP2000S	Broad scale topo position (calc from area within 2000m of streams)	
ANN_PRECIP_MEAN	Mean of annual precipitation values within the watershed	cm/yr
TJAN_MEAN_MIN	Mean maximum temperatures in January	°C
TJULY_MEAN_MAX	Mean maximum temperatures in July	°C
PFOR	percent forest	%
PWETL	percent wetland	%
PURB	percent urban	%
PAGT	percent agriculture--total of pasture and row crops	%
PBAR	percent barren	%
PUSER	percent area in transitional (#33) + forest regrowth (#44)	%
PSHRB	percent shrub	%
PNG	percent rangelands	%
U_INDEX	percent of watershed area sum urban, ag, mining, regrowth and transitional	%
USERSL3	percent forest regrowth on slopes	%
AGTSL3	percent agriculture on slopes	%
H	Cover type diversity--shannon weiner index	
RURB120	percent urban within 120m riparian buffer	%
RFOR120	percent forest within 120m riparian buffer	%
RUSER120	percent transitional and forest regrowth within 120m riparian buffer	%
RHUM120	percent total human disturbance classes within 120m riparian buffer	%
RAGT120	percent total agriculture within 120m riparian buffer	%
RSHRB120	percent shrub class within 120m riparian buffer	%
RNG120	percent rangeland within 120m riparian buffer	%
RWETL120	percent wetland within 120m riparian buffer	%
P_load	phosphorous load based on weighting of different land use types	kg/ha/year
N_load	nitrogen load based on weighting of different land use types	kg/ha/year
RDDENS	total length of roads per unit basin area	km/km^2
STXRD	# of road stream crossings per km of streams in watershed	#/km
FS_OCR_D	stream distance to ocean, Puget Sound, or Columbia River	m
FS_S4	distance to nearest 4th order stream	m
FS_S5	distance to nearest 5th order stream	m
distnxtrdr_km	distance to the next highest stream order	km
StrmLn_dvde_KM	longest upstream stream length	km
PT_Dvde_KM	maximum distance to divide	km